Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse - Jean Jaurès

Approche intégrative du rythme de la parole en L2: Enjeux pour les apprenants et impact de l'enseignement de la prosodie en français L2

Thèse présentée et soutenue, le 18 octobre 2024 par

Lucie DROUILLET

École doctorale
CLESCO - Comportement, Langage, Éducation, Socialisation, Cognition

Spécialité
Sciences du langage

Unité de recherche
LNPL - Laboratoire de NeuroPsychoLinguistique

Thèse dirigée par
Corine ASTESANO et Charlotte ALAZARD-GUIU

Composition du jury

Mme Elisabeth DELAIS-ROUSSARIE, Présidente, Université de Nantes
Mme Pilar PRIETO-VIVES, Rapporteure, Universitat Pompeu Fabra
M. Fabian SANTIAGO VARGAS, Examinateur, Université Paris 8
M. Paolo MAIRANO, Examinateur, Université de Lille
Mme Corine ASTESANO, Directrice de thèse, Université Toulouse - Jean Jaurès
Mme Charlotte ALAZARD-GUIU, Co-directrice de thèse, Université Toulouse - Jean Jaurès

AN INTEGRATIVE APPROACH OF L2 SPEECH RHYTHM: LEARNERS' CHALLENGES AND IMPACT OF PROSODIC INSTRUCTION IN L2 FRENCH

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to address my deepest gratitude to my supervisors: Corine Astésano and Charlotte Alazard.

Corine, c'est vraiment grâce à toi que je suis là aujourd'hui. Tu m'as fait découvrir le monde merveilleux de la prosodie et depuis je suis "hooked"! Tu m'as accompagnée depuis le Master dans mes premiers pas de chercheuse, tu m'as donné confiance en moi et m'a soutenu et je t'en suis profondément reconnaissante. Tu fais partie des rencontres les plus influentes de ma vie. Merci pour tout.

Charlotte, nous avons fait un sacré chemin aussi depuis que tu as été jury de ma soutenance de Master! Je suis extremement reconnaissante de l'accompagnement et du soutien que tu m'as apporté pendant ces 4 années. Ayant eu l'honneur d'être ta toute première doctorante, je peux affirmer que ton encadrement est de la meilleure qualité! Merci pour ta confiance et ta douceur. J'espère sincèrement que nos projets de collaboration aboutiront car c'est un véritable plaisir de travailler avec toi. Merci pour ton amitié et ton écoute aussi. Coeur de beurre ;)

I would like to thank my jury committee: Elisabeth Delais-Roussarie, Pilar Prieto, Fabian Santiago, and Paolo Mairano. Thank you for accepting my invitation, I am sincerely honoured (and a little terrified) that you accepted to review my work. I sincerely admire your work and even though we have only briefly crossed path, you all strike me as wonderful human beings. I am excited to see you all at my defense!

Je voudrais remercier chaleureusement les âmes charitables qui m'ont généreusement apporté leur aide sur des points techniques : Robin Vaysse, Noémie Te Rietmolen, Fanny Guitar-Ivent, Saulo Santos, Romain Mendes. Un grand merci d'avoir répondu à mes appels de détresse!

Un très grand merci aussi à Laure Fesquet et Sébastien Palusci de m'avoir ouvert les portes de leur fantastique plateforme Fonetix.
Un très grand merci aussi à Florence Baills d'avoir partagé avec moi une partie de son materiel expérimental.

Since we are in the experimental topic, HUGE thanks to all my participants without whom I wouldn't have much to talk about in this dissertation.

À mes incroyables, formidables, fantastiques co-doctorantes (et Sangho!) sans qui ces 4 années auraient été... vachement plus nulles. À Noémie, Sangho, Mireille, Regina et Jasmin. Le thé vanille aura pour toujours une saveur "Bureau E111" ! Merci pour l'ambiance de folie, les memes hilarants, la solidarité sans failles, et les discussions qui n'avaient rien à voir avec le boulot... Quel bonheur d'avoir partagé mon quotidien avec vous! Promis je resterai aussi longtemps que possible à vos côtés.
À Iona et Cécile pour les serrages de coudes réguliers, vos encouragements inconditionnels, et votre génialitude innée.
À Anne-Flore, Nour, Katia, Layana, Victoria, Salma, Emma, Chloé. Vraiment je me sens privilégiée d'avoir évolué dans une si belle cohorte!

Un grand merci au LNPL de m'avoir accueillie, et à Cyrille pour son énergie et son engagement qui donne au labo son atmosphère positive et conviviale. Merci aussi pour tes encouragements Cyrille, et aussi ceux de Mélanie, Chris, Inès, Barbara, Anna, Simona.

Enfin je remercie de tout mon coeur mes proches qui m'ont encouragé et soutenu super fort pendant cette aventure, même s'ils n'ont pas toujours compris de quoi il s'agissait...

Ma famille d'amour : Maman, Papa, ma sis Maud, et le ptit chou Oboè. Merci pour tout l'amour et toute la force que vous me donnez. J'ai hâte d'être plus disponible pour vous désormais!

À Jérôme, merci d'avoir été et d'être toujours là. Ta patience et ta douceur me surprendront toujours. Merci de n'avoir jamais douté de moi, de m'avoir supporté dans les pires moments, et de toujours savoir comment me réconforter. You're the best!

À mes ami.es du coeur: mes Bobbe Maurane Quinn sisters pour briller toujours plus fort, mes Chicas d'amour, my darling Chiara (mais tellement merci pour les tips de rédaction!!!), ma Catcat, Antonin, Groro, Arnaud. Merci de rendre ma vie si fun et d'avoir été présent.es tout au long de ce périple. On va aller boire tellement de coups...

Finally, I just want to say that I never thought research and academia was for me (I am still not sure though!). I never thought a Ph.D was in the cards, I thought a Ph.D is something other people do but not me, "I'm not that kind of people", and yet here I am. It just amazes me how getting old actually means constantly proving yourself wrong.

FOREWORDS

Be aware that this dissertation is written in UK English.

The pronouns "they-their-them" are used as singular gender-neutral pronouns, except when the subject is plural.

Because I changed my last name in the middle of my Ph.D. years, two publications are under my former last name: Judkins, and the two most recent ones are under my current name Drouillet.

Related publications from the main author are provided in Appendix 7 (p. 347).

INTRODUCTION

It is stating the obvious that the world we live in today is multilingual. With the development of travel opportunities, the internet, social media, and access to cultural content from all over the world, languages are all around us. My first language crush was for English, then came Japanese, and Italian. But my feelings towards any language are first and foremost connected to how they sound. I fall in love to their melody, their rhythm, their music. To me, learning a language means being able to change my "talking music".

Unsurprisingly, this intuitive interest towards languages led me to become a teacher of French as a foreign language (FLE[1]). Through this experience, I realised that I was not at all equipped to teach the aspects of language that thrill me the most. Then started my journey in the world of Linguistics and research. I discovered that the melody and rhythm I love so much are called prosody, and that there is much more information, descriptions, theories, and models about it that I could have ever imagined. So I dived right into it, and the rest is history.

The work presented in this dissertation reflects my interest for the study and measure of prosodic aspects, cross-language comparison, and the transfer of research findings to teaching methods.

The first chapter introduces our subject of study: speech rhythm. In the context of music, there is a common idea of what rhythm is. We might struggle to explain it as it is very much something that we feel rather than something we intellectualise, but we can collectively agree that it refers to the under-lying beats, the patterns they form, and maybe the tempo too. But what is rhythm in speech? How is speech rhythmical? What and where are the beats?

Paradoxically, speech rhythm is at the same time the most fundamental element that structures spoken language, but it is also the most difficult to grasp. Several approaches will be presented, to which correspond distinct acoustic correlates of speech rhythm. Because these contrasting views are in fact complementary, we will propose an integrative approach that considers them all

[1] From French: Français Langue Etrangère

as different levels of analysis that intertwine and create, altogether rhythm in speech.

Chapter II focuses on the specificities of non-native (L2 hereafter) speech. Speaking in L2 engages cognitive and motor processes that are not as automatic as speaking in our first language (L1), leading to difficulties. In addition, transfer from the L1 linguistic system and universal acquisition processes interweave and impact the production of L2 speech. Models of L2 acquisition will be presented as well as empirical studies that support their theoretical assumptions. The literature review will include L2 studies of speech rhythm from the different approaches presented in Chapter I, and will also address the relation between L1 and L2 speech patterns within-subject.

In Chapter III, we shift focus from the production of L2 speech to the perception of thereof. The definition of what constitute a foreign accent will be interrogated, and its impact on the perception of L2 speech by native listeners will be explored. In fact, native listeners are often called upon in L2 pronunciation studies to make the link between the performance of the L2 speaker, and how it is perceived in terms of intelligibility.

We will then turn to the L2 speaker's perception and listening abilities towards the target language as spoken by natives. Indeed, the L2 experience implies listening and understanding as well as speaking. The first step into accessing meaning is the ability to segment the continuous stream of speech into individual words. This seamless mechanism in L1 becomes an arduous challenge in L2.

The fourth Chapter reviews methods and techniques used to teach L2 pronunciation. Over the last century, pedagogical approaches have shifted from explicit instruction of L2 sounds and imitation exercises to more holistic communicative goals. Some methods include the use of gestures, props, and even music in order to help learners perceive and acquire the L2 sounds and prosody. A literature review of studies testing and comparing the effectiveness of different pronunciation instruction methods will be presented.

Finally, the last two chapters of this dissertation will present the experimental design and results of the study we conducted. In L2 French, teaching prosody is not a common practice. However, the preceding chapters highlight the crucial role of prosody in L2 pronunciation and listening abilities. Therefore, the study questions the impact of a specific prosody training on the performances of L1 English learners of L2 French. By comparing this training with general speaking and listening activities commonly used in L2 French classes, we are looking to see if a more direct, explicit and multimodal approach of L2 French prosody is more effective than what is commonly done in L2 French classes. The effect of both training will be assessed on acoustic measures of speech rhythm, following our integrative approach of it.

In addition, we will assess the progress of the learners after training through perceptual measures of comprehensibility (ease of understanding) and accentedness (degree of foreign accent) attributed by native French listeners. This will allow to evaluate if the changes measured in the speech rhythm of participants are relevant to native listeners' perception.

Lastly, the listening abilities of the participants will be evaluated to see if a prosody training can also help learners segment the speech of native French speakers.

The methodological choices made to build the experimental design put an emphasis on the ecological validity of the outcomes. By means of a delayed posttest, and the analysis of spontaneous speech samples, we ensure that the changes from before to after training are not limited to the classroom setting, but rather transfer onto natural speech and are still visible one week after the end of the training. Furthermore, speech samples in L1 will also be analysed. The differences between the L1 and L2 production will help interpret the L2 results.

The implications of the findings in terms of pedagogical practice, methodological decisions, and level of analysis of speech rhythm will be discussed.

CHAPTER I - RHYTHM IN SPEECH

INTRODUCTION

In this first chapter, we start by contextualising the study of language in its spoken form. An overview of the historical evolution of the study of spoken language within the field of linguistics is provided. We will see how the specificities of oral language as opposed to written language came in focus later than its counterpart. We then turn to a specific aspect intrinsic to spoken langue: prosody. Through its definition and the presentation of its functions, we grasp the transversal nature of speech prosody, as it impacts several linguistic levels such as phonological, syntactic, and pragmatic dimensions. We discuss the essential role of prosody in language acquisition, its universal aspects as well as language-specific ones. All of which goes towards building the argument that the study of prosody in L2 speech is of importance if we are to understand L2 learning mechanisms.

The main focus of the work presented in this dissertation is more specifically one aspect of prosody: speech rhythm. Before defining it in detail, we will first discuss the difficulty of defining rhythm itself. Adopting lenses from different disciplines, we will gradually narrow down the main principles that allow to characterise a phenomenon as rhythmical, principles that also apply to speech rhythm. We will briefly discuss speech rhythm in relation to motor rhythm and brain rhythm. This introduction to rhythm as an essential aspect of human life and cognition supports the relevance of its study.

Only then will we dive into the core of this work by introducing the different approaches and acoustic correlates of speech rhythm. Four different perspectives are introduced, namely: the phonological view, the prosodic view, temporal variables, and lastly fluency. We then present an argument in favour of the integration of all the aforementioned perspectives into a global approach of speech rhythm. This integrative view constitutes the theoretical ground for the study presented in Chapter V and VI.

Finally, we will close this chapter with a presentation of the differences between French and English across the different dimensions of speech rhythm previously defined.

1. SPEECH AND ORALITY

At the root of language is orality. From a chronological perspective, written language came second after spoken language, and some languages remained exclusively oral. Therefore, written language can be viewed as a "translation", a reflection whose function is to render the primary source that constitutes spoken language, or speech. However, the written form of a language differs from its spoken counterpart. Paradoxically, the study of language has extensively been focused on language's written form, which, from this view, could be considered a secondary source (Chafe & Tannen, 1987).

The study of language in its oral form is fairly recent as it really took off towards the end of the 20th century. Gadet (1996) writes that the scientific interest for spoken language did not manifest earlier because of the ideology of the norm, which confuses the necessary process of standardisation with the allegiance to a language model that would be homogenous, and mostly close to written language. The author also explains that despite the affirmation of the primacy of oral over written language in Linguistics, it would seem that linguists do not fully take into consideration the implications of the difference between oral and written language, and as a whole, do not consider oral for what it is, to the point of studying it as such.

The study of large speech corpora (in both English and French) led to the description of oral forms and to re-think the dichotomy between written and oral language which, rather than being considered as two distinct systems, should be approached as two varieties or two dialects of the same language (Blanche-Benveniste & Bilger, 1999; Chafe & Tannen, 1987).

Besides the pioneering work on oral language discourse analysis and syntax in the 1990's (Blanche-Benveniste, 1997; Gadet, 1989), the field of experimental and acoustic phonetics has been providing descriptions of the sounds of language. Notably, the work of Marey and later Rousselot on articulatory phonetics significantly advanced the understanding of physiological mechanisms involved in speech production, at the beginning of the 20th century (Rosset et al., 2010). As technology evolved, possibilities to study the acoustics of speech widened. Parameters of duration, voice quality, pitch movements, and intensity allow us to describe not only the realisation of phones, but also the prosodic marking at play in a string of speech.

Oral language has not been given the same attention or value in Linguistics as written language for a collection of reasons, but one of them is certainly its intrinsic volatility. As opposed to written language, oral language is not fixed nor stable. Rather, it is subject to variation stemming from a plurality of factors: time, geographic situation, social background and status, communication context, idiosyncratic characteristics of the speaker (Foulkes, 2020; Labov, 1981; Reynolds, 1994). The combination of all these sources of variation leads to the polymorphic nature of oral language, both within and between speakers. To the point that one can wonder how we - language users - manage to understand each other despite such inconsistency.

Another challenge raised when studying oral language relates to the inadequacy of the analysis framework and units used for the study of written language. In the oral language, there is no capital letter and period to frame a sentence, there is no space between words, and there is no indentation and line break to signal a paragraph. The organisation and structure diverge from that of written language, and units' boundaries are signalled acoustically. In addition, because of its spontaneous nature, oral language also involves markers of online speech production processes such as hesitations and repetitions.

Lastly, oral language and communication are multimodal in nature. Aside from the sounds coming out of the mouth, facial mimics and gesture systematically accompany speech, and play an important role in language processing (Cosnier & Vaysse, 1997; Kendon, 1972; McNeill, 2005).

In sum, speech itself is a multi-layered construct where the verbal material (phonemes, words and syntactic organisation) intertwines with the prosodic system (rhythm, accentuation, intonation), all of which is supported and enhanced by facial cues and gestures.

Our work focuses on speech rhythm, which is one aspect of speech prosody. In order to provide some context, the following section defines and presents speech prosody, and its functions.

2. SPEECH PROSODY

Speech prosody is often referred to as the music of speech. The comparison is indeed tempting since, like in music, speech prosody includes melodic features (voice frequency - f0 - variations), and rhythm (variation of timing and intensity). Historically, the term prosody comes from ancient Greek *prôsoidia* when it used to refer to the distinctive melodic accentuation (similar to tones in tone-languages) that characterised that language at that point in time (Di Cristo, 2013). Progressively, the meaning of the term shifted to refer to the *metric*, especially in the field of literature and poetry, or intonation in linguistics. Nowadays, the term covers both. From a linguistic point of view, prosody is defined as the ensemble that constitute the intonation, accentual system, and rhythm in speech. These elements are also referred to as suprasegmentals, since they occur at levels beyond the segmental chain (sequences of consonants and vowels).

Each language possesses its own unique prosodic system, that is, patterns and rules that fundamentally structure its spoken form. A language prosody, along with its phonemic inventory, constitutes the core of its identity. When hearing a foreign language, most people are able to hypothesise what language it might be, based primarily on "how it sounds" i.e., its phonetic and prosodic characteristics.

Prosodic features are the first aspects of language we acquire as infants. In the gradual development of language in our first language (L1), prosody comes first. As early as 6 months of life, babies already babble following the prosodic patterns of their L1 (Fitzpatrick, 2002). The accentual system (alternation of strong and weak syllables) sets up as early as 13 to 16 months (Konopczynski, 1990), and for what concerns perception, infants are able to distinguish their L1 from other languages pretty much from birth (Mehler et al., 1988; Nazzi et al., 2000). This early sensitivity and tuning to L1 prosody show how it constitutes the entry key to language. Essentially, prosody functions as the back-bone of language, the first foundation layer onto which the following ones (syntax, lexicon etc.) will build.

The rhythmical aspects specifically instantiate the structure onto which speech is organised (speech rhythm is defined in more detail in section 3., p. 24). The accentuation system plays a role in the framing of rhythmic units such as the foot or the prosodic word, and rhythm determines the rules of alternating weak and strong beats. This function is crucial for the segmentation of speech - the first stage for access to meaning (Cutler, 2012).

Intonation is actualised by f0 variations, creating melodic contours anchored onto accented syllables. As such, accentuation prevails intonation (Astésano, 2001). Intonation serves focus marking, sentence modality (declarative/interrogative), pragmatics and communicative intention.

Prosody also bears interactional functions such as speaking turn regulation, and informs on the emotional state of the speaker, as well as idiosyncratic characteristics (age, region of origin...).

In sum, prosody assumes a plurality of functions from linguistic (structural organisation) to extra-linguistic (speaker's characteristics). Intonation, accentuation and rhythm are observable in the speech signal through the acoustic parameters of duration, intensity, and f0. All three aspects are not categorical but rather, they interact and are intertwined in co-dependent relationships.

Languages share certain universal aspects of prosody such as its boundary marking function and pragmatic functions (Lewandowska-Tomaszczyk, 1996; Vogel, 2009), while other aspects distinguish between them, e.g., accentual vs syllabic rhythm (see section 3.3., p. 29), or tonal vs non-tonal languages.

Prosody transcends all language dimensions (lexicon, syntax, pragmatics) and levels of constituency, from the syllable to the utterance. As such, it is a fundamental block of language acquisition and is specific to each language.

Our work being essentially focused on rhythmic aspects, intonation will not be discussed here. The following section provides an introduction to the concept of rhythm, first in a general sense, then specifically in speech.

3. SPEECH RHYTHM

3.1. (TRYING TO) DEFINE RHYTHM

Out of all aspects of speech prosody presented above, rhythm is the least clearly defined. In fact, the definitions of rhythm as a general concept are abundant to say the least. As You (1994) puts it:

> "There is no unified theory of rhythm just as there is no single conception of time, body or life. Intrinsic difficulties aside, the difference in the conception of rhythm reflects the different views of life and the world." (p. 361).

Whether in science, the humanities, arts, philosophy etc.; authors from all fields have written about and try to define rhythm relatively to their prism. One thing that everyone agrees on is the ubiquitous quality of rhythm:

> "Natural phenomena very generally, if not universally, take a rhythmic form. There is a period recurrence of a certain phenomenon, sometimes accompanied by others, going on continuously in all that pertains nature." (Bolton, 1894, p. 146)

> "The most characteristic principle of vital activity is rhythm. All life is rhythmic." (Langer, 1953, p. 126)

> "Rhythm is a universal scheme of existence." (Dewey & Simon, 1989, p. 154)

Whether it is the alternation of daylight and night throughout the day or the moon phases, the physiological rhythms of the breath and the heartbeat, or the rhythm we hear or dance to in music, rhythm is at the heart of all human activity and behaviour.

Yet, however obvious rhythm might be, there is no generally accepted and precise definition of rhythm (Fraisse, 1982).

Concepts that are regularly mentioned in definitions include the alternation of contrasting elements, the grouping and structuration of these elements, a notion of periodicity, recurrence or regularity. In his definition, Fraisse (1990, cited in

Missire, 2007) puts forward the perceived character of rhythm and the notion of structure (which we will come back to in regards to speech rhythm specifically):

> "Caractère perceptif de stimulations successives lié à leur organisation en des ensembles structurés. Cette structuration se produit sur la base de différences de durée, d'intensités, ou d'intervalles entre les stimulations, ces différences pouvant être subjectives ou objectives." (p. 396)
> [*Perceptive nature of successive stimulations associated to their organization into structured ensembles. This structuration occurs on the basis of duration, intensity, or intervals differences between the stimulations - these differences being potentially subjective or objective.*][2]

Sauvannet (2000) collects and gives a remarkable inventory of rhythm definitions, and summarises the core principles in his own. Is considered rhythmical:

> "[...] tout phénomène perçu, subi ou agi, auquel on peut attribuer au moins deux des critères suivants : structure, périodicité, mouvement" (p. 195).
> [*all perceived, endured or acted phenomenon to which can be attributed at least two of the following criteria: structure, periodicity, movement*]

In these two definitions above, the *subjectivity* of Fraisse rejoins the *action* of Sauvannet in that, as humans, our perception of a rhythm is not passive but rather, we tend to group stimuli together into patterns automatically. Around the turn of the 20th century, several studies in experimental psychology investigated the perception of pattern and grouping of auditory stimuli (Bolton, 1894; Woodrow, 1909).They found that in a sequence of sounds alternating in intensity, the more intense stimulus tended to be perceived as beginning the sequence; whereas in a sequence of sounds alternating in duration, the longer stimulus was perceived as ending the sequence. From these findings, emerged the basis of the Iambic-Trochaic law (Hayes, 1995) which is called for in early description of the metrical theory to account for the possible structures of feet in languages (see section 3.3.2., p. 34 on metrical theory).

[2] All translations from French to English throughout this dissertation are ours.

In sum, rather than the resulted organisation of successive phenomenon, rhythm is the *active* structuring principle creating temporal organisation (Astésano, 2001).

The most universal and conscious way of experiencing rhythm is of course through music. Interestingly, definitions of rhythm from the perspective of music theory might as well be interchangeable with those pertaining to rhythm in speech:

> "Rhythm may be defined as the way in which one or more unaccented beats are grouped in relation to an accented one." (Cooper et al., 1963, p. 8)

The following section elaborates on what constitutes rhythm in speech.

3.2. SUPRAMODAL RHYTHM: LANGUAGE, MOTRICITY, AND NEURAL ACTIVITY

Although language(s) and speech are intuitively perceived as rhythmical, and rhythm is always mentioned as a component of speech prosody, there is no consensus in the literature regarding what elements of speech bear the rhythmic structure, and what the rhythm units are.

Following the three essential criteria mentioned in the definition of Sauvannet (2000), we can say that *structure* emerges in speech through the grouping and hierarchical organisation of the syllables into feet or prosodic words, rhythmic groups (or accentual phrase), intonational phrase and so on. *Movement* is created by the alternation of strong and weak syllables (accented vs unaccented), continuous sound and pauses, as well as contrasting melodic contours. Finally, the recurrence of prominences, at all levels of the hierarchy, creates patterns and *periodicity*.

As was emphasised above, rhythm is present in all aspects of human life and activity and as such, speech rhythm is not an independent module, but rather interacts and is co-dependent with motor rhythm and brain rhythm (Astésano, 2022).

3.2.1. Speech rhythm and motricity

There is no doubt about the engagement of motricity in speech. For one thing, producing speech necessarily involves the action of several muscles and organs in the body. From pushing air through the pharynx which activates the vibration of the vocal folds, to the movements of the jaw, tongue and lips involved in the articulation of phones, speaking means motor activation.

But this activation is not limited to the mouth, co-speech gestures and facial mimics have been described and shown to be linked to several linguistic levels, and most notably, to speech rhythmical structure. Efron (1941) and later Kendon (1972) had already shown the relationship between the speech structure and the occurrence of gestures in terms of temporality. Since then, several authors have provided detailed classifications of co-speech gestures (Cosnier & Vaysse, 1997; McNeill, 1992, 2005). These have paved the way for further investigation of the co-occurrence of prosodic marking and gestures (among others Jannedy & Mendoza-Denton, 2005; Loehr, 2007; Rohrer et al., 2019; Shattuck-Hufnagel & Ren, 2018).

Co-speech body-movements constitute strong evidence in favour of the Embodied Cognition perspective, which posits that cognition is deeply rooted in the body, through constant interactions with the sensory-motor system and the environment (amongst others Barsalou, 2008; Shapiro, 2019; cited in Baills, 2022). According to this view, the connection between mind and body goes both ways, therefore the engagement of the motor and sensory systems is crucial to the maintenance and development of our cognitive capacities.

In fact, both foreign language teachers and speech therapists tend to intuitively use activities based on motricity and rhythm. Speech therapists report the observation of a strong relationship between speech, rhythm, and movements in the remediation of speech pathologies (Daigmorte et al., 2022). In foreign language classes, using hand gestures seems fairly common, especially for teaching pronunciation (Tellier, 2006, 2008). One of the methods that advocates for it is the Verbo-Tonal Method (Guberina, 1956, 1975) for phonetic correction whose tenet is to approach the pronunciation of difficult sounds through a first phase of rhythmic exercises, usually involving movements (such as tapping or walking), in order to set the rhythmical premises of the target language (see Chapter IV, p. 143).

Clearly, speech rhythm and motor rhythm are inter-connected, and this interaction is already exploited in the context of speech rehabilitation, and (not as commonly) L2 pronunciation teaching. The latter will be further developed in Chapter IV (p. 133) of this dissertation, which elaborates on L2 pronunciation teaching practices.

3.2.2. Speech rhythm and brain rhythm

Research on speech processing in cognitive science and neuroscience has highlighted the relationship between the rhythm of neural activity and the perception of sensory stimuli. Regarding the processing of spoken language Peelle & Davis (2012) explain that on the listener's side, the perception and comprehension of the speech signal relies on their ability to integrate and anticipate acoustic information in real-time, as they unfold in a linear temporal fashion.

Furthermore, the rate of information delivery being entirely up to the speaker, the listener is forced to tune to the speaker's rhythm in order to successfully follow along. In essence, predictability is the key to efficient processing.

The pre-requisite to predictably is regularity. Not in the strict sense of a metronome for instance, but in the sense of a perceptible recurrence of events in the stimulus. In the speech signal, amplitude modulation at the syllabic level while not perfectly regular, is not random and has a perceptible recurrence. Consequently, such temporal information in speech enables listeners to anticipate and make predictions about the incoming signal and this way optimise comprehension.

Now addressing brain rhythm: neural activity is oscillatory in nature, it alternates between phases of high and low excitability of neuronal populations (for detail see Lakatos et al., 2005; cited in Peelle & Davis, 2012). Research has shown that the efficiency of information processing varies according to the oscillatory phase at the moment of reception. A stimulus is processed more efficiently when it aligns with a high excitability phase of oscillation. Further, if sensory information follows a predictable temporal pattern, processing efficiency can be enhanced by adjusting the timing of ongoing neural oscillations to match the phase of the stimuli (Busch et al., 2009; Romei et al., 2010; cited in Peelle & Davis, 2012).

Thus, returning to speech processing and comprehension, through the perception of the recurring amplitude modulations of syllables in the speech input, the auditor's neural oscillations automatically synchronise to the signal in a process called *entrainment*, such that relevant information (e.g., accented syllables) coincides with phases of high excitability, optimising processing.

In the terms of the Dynamic Attending Theory (DAT, Large & Jones, 1999), internal oscillations (also called attentional rhythms) generate expectations enabling the prediction of future events (top-down process). On the other hand, external rhythmic stimuli can entrain internal rhythms (bottom-up process). The DAT propose to explain the entrainment mechanism as a result of an attraction force between the internal and external rhythms, itself inducing a neural resonance resulting in the phase synchronisation of the internal rhythm with the external one (Snyder & Jones, 1999 cited in Gindre, 2024). Moreover, adaptative processes ensure that the timing of internal oscillations can continuously be adjusted to re-align with the timing of the sensory input.

Turning to the processing of L2 speech, it has been shown that phase-synchronisation is reduced in the case of L2 speech perception, in comparison to native speech perception (Pérez et al., 2015). While this area of research lies outside our current focus, investigating neural synchronisation to speech rhythm in L2 in relation to intelligibility and the development of listening skills could help advance our understanding of the role of speech rhythm in L2 speech processing.

3.3. SPEECH RHYTHM CORRELATES: SEVERAL APPROACHES

The boiling question in the study of speech rhythm is: through which correlates, which acoustic features can we observe and quantify it. In sum, what do we measure?

The most widely used acceptation of the term *rhythm metrics* refers to a set of measures developed under the framework of the Rhythm Class Hypothesis (Abercrombie, 1967/2019; Pike, 1945). However, these represent only one face of the dice, and other speech-rhythm related measures are used outside of this framework. Such measures are simply not identified as measures of rhythm, even though in essence, they are. In this section, we present the so-called rhythm metrics, and the different "anonymous" correlates to speech rhythm that have emerged from various theoretical viewpoints.

3.3.1. The phonological view of rhythm & rhythm metrics

Pike (1945) and later Abercrombie (1967/2019) looked at speech rhythm across languages and classified them into rhythm classes based on the unit that exhibited isochronous recurrence. So called syllable-timed languages (Romance languages) present a syllabic rhythm where syllables are all of (near) equal length. This category is opposed to stress-timed languages (Germanic languages) where all between-stress intervals (feet) are isochronous, thus inducing an accentual rhythm. A third category was also proposed to account for languages relying on the mora as their rhythm unit (such as Japanese). Therefore, a mora-timed language present isochronous successive morae (Bloch, 1950; Han, 1962).

These categories were viewed as completely distinct, languages belonged to either one or the other. Several studies empirically disproved the isochrony principle in languages from all three categories (among others Bolinger, 1965; Dauer, 1983; Wenk & Wioland, 1982). However, Lehiste (1977) investigated and demonstrated the reality of isochrony as a perceptual phenomenon. The study concerned the English language only, and used a perceptual judgement task of synthetised speech composed of intervals of varying length. The study's conclusion was two-fold:

> "First of all, many actual differences in the duration of interstress intervals may be below the perceptual threshold. Second, listeners tend to impose a rhythmic structure on stretches of sounds and thus subjectively to perceive isochrony even in sequences where the durational differences should be above the perceptual threshold." (p. 259)

The search for acoustic features that could account for rhythmic differences between languages then shifted from the duration of the rhythm units - syllable, foot, mora - to their internal structure. Amongst others, Bertinetto (1977) and then Dauer (1983) argued that the distinctive traits of the syllable-timed vs stress-timed languages lied at the phonological level, in the differences in syllable complexity, the degree of vowel reduction, and the characterisation of accents (distribution and duration). Furthermore, rather than a strict separation of the two categories, she proposed to consider languages rhythm to belong to a continuum (see Figure 1 below) between a syllable-timed-like structure, which prototypically involves syllables of simple structure (mainly CV) and no vowel reduction (as in French), and stress-timed-like structures where syllables are more complex (more vowel and consonant clusters), and important vowel-reduction (as in English).

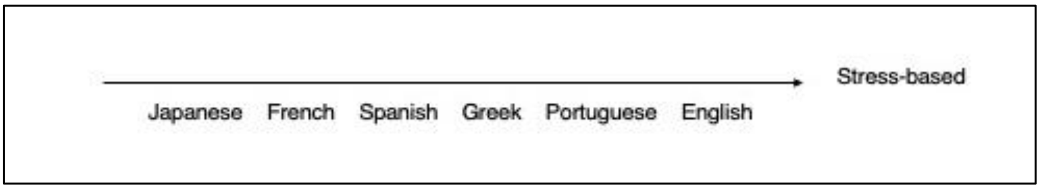


Figure 1- Dauer's comparison of languages over her "more or less stress-based rhythm" continuum (p.60).

Following Dauer's work, Ramus, Nespor & Mehler (1999) developed three quantitative measures accounting for syllable complexity and vowel reduction. The standard deviation of consonantal interval duration (ΔC) was thought as an indicator of syllable complexity, while the same measure applied to vocalic intervals (ΔV) supposedly indicated the degree of vowel reduction of unstressed syllables. Finally, the proportion of time taken by vowel intervals (%V) was supposed to capture both the complexity and reduction degree. Following these assumptions, a stress-time language should be characterised by a high degree of syllable complexity and vowel reduction should present high values of ΔC and ΔV, and a low %V caused by vowel reduction and a higher proportion of consonants.

The authors compared these three measures obtained for eight languages. Based on the results, they observed that languages were grouped in a way that reflected the initial stress vs syllable vs mora time classification. English, Dutch and Polish on one side (stress-time), French, Spanish, Italian and Catalan on another (syllable-time), and Japanese by itself (mora-time). Thus, their conclusion corroborated the phonological grounding of rhythmic differences between languages.

Low & Grabe (1995), Low, Grabe & Nolan (2000) and Grabe & Low (2002) also developed and tested measures of durational variability of vocalic and consonantal intervals, however their approach included the temporal succession of the intervals. The Pairwise Variability Index (PVI) is a measure of the difference in duration between each pair of successive intervals in a string of speech, compiled into a mean. However, studies have shown that the PVI on vocalic intervals (PVI-V) is influenced by the speech rate (Grabe & Low, 2002; Ramus, 2002), therefore the raw version of the PVI (rPVI) is used on consonantal intervals while a normalised version of the index, where the durational difference between two intervals is divided by the mean duration of the pair, is used on vocalic intervals (nPVI).

The authors applied these measures to 18 languages and found that supposed stress-timed languages (English, German, Dutch and Thai) presented indeed high values of nPVI, while supposed syllable-timed languages (French, Spanish, Tamil, and Singapore English) presented lower nPVI. This measure also accounted for "mixed" languages such as Polish and Catalan. rPVI-C however was a weaker indicator.

Dellwo et al. (2003) also proposed speech rate-normalised version of ΔC and ΔV by divided them by the mean of all intervals: VarcoC and VarcoV.
All of the above-described measures are commonly referred to as rhythm metrics and are summarised in Table 1 below.

	Metrics name	Description	Level	Reference
Raw metrics	Δ (delta)	Standard deviation of duration of interval (ΔV for vocalic, ΔC for consonantal, ΔS for syllable)	Global	Ramus, Nespor, & Mehler, 1999
	rPVI	Raw pairwise variability index for intervals (rPVI-V for vocalic, rPVI-C for consonantal, rPVI-S for syllable)	Local	Grabe & Low, 2002
Normalised metrics	Varco	Coefficient of variability in duration of intervals (VarcoV = ΔV/meanV, VarcoC = ΔC/meanC, VarcoS = ΔS/meanS)	Global	Dellwo et al., 2006
	nPVI	Normalised pairwise variability index for intervals (nPVI-V for vocalic, nPVI-C for consonantal, nPVI-S for syllable)	Local	Grabe & Low, 2002
	%V	Percentage of vocalic intervals	Global	Ramus, Nespor, & Mehler, 1999

Table 1- Summary of rhythm metrics (adapted from Ordin & Polyanskaya, 2015)

These measures have since been criticised and several authors have brought evidence to their limitations. Most notably, Arvaniti (2012) experimentally tested these measures and showed that they were as much influenced by inter-individual variation, speech style, and syllabic structure complexity within a language, as by inter-language variation.

However, in regards to syllable complexity, Prieto, del Mar Vanrell, Astruc, et al. (2012) conducted a study in which they controlled this variable in the speech material analysed, and still found that nPVI-V, ΔV and VarcoV discriminated between English, and Spanish and Catalan. Furthermore, they found that the English vs Spanish/Catalan distinction also appeared in the way these languages mark prosodic heads and pre-final lengthening (in terms of duration). This suggests that these metrics do encapsulate languages differences at the prosodic level (at least those analysed there) since they discriminate between languages even when phonotactic properties are controlled for.

In fact, results found by Li & Post (2014) in a study combining rhythm metrics and accentual and final lengthening corroborated these findings.

Nevertheless, rhythm metrics cannot be taken as rendering a full account of speech rhythm, and should be used in full awareness of their sensitivity to intra-individual variation and speech style. In addition, these metrics being exclusively based on the duration parameter, the role of f0 in prominence marking and speech rhythmicity in general is entirely overlooked. To address this, Fuchs (2014) actually proposed a version of the PVI that includes f0 values: nPVI-V(dur*f0). The argument is that f0 plays a role in the perception of interval duration and that this parameter needs to be included in order to account for the multi-dimensional nature of rhythm in speech.

Besides f0, **rhythm metrics fail to capture the organisation of prominences and the hierarchy of nestled constituents (Cummins, 2002). In that sense, rhythm metrics are confined to what we will call the micro-level of speech rhythm** (see section 3.4., p. 44). The next section presents a radically different approach based on prosodic theories and focused on prominences distribution and constituents marking.

3.3.2. The prosodic view & the Metrical Theory

In the metrical approach to prosody (Halle & Vergnaud, 1987; Liberman & Prince, 1977), rhythm refers to the temporal organisation of weak and strong syllables into hierarchical structures. Stress is considered as the linguistic manifestation of the speech rhythmical structure (Hayes, 1995). Thus, in this approach, the accentuation system is truly at the core of speech rhythm.

The theory introduces the notion of *relative prominence* which is in direct relation to the constituent's position in the hierarchy. The weight of a prominence is only defined in relation to what surrounds it. Each constituent (foot, accentual phrase etc.) is marked by a prominent syllable constituting the head of that constituent. The hierarchical structure can be represented in a metrical tree or grid (see Figure 2 below). According to this theory:

> "The perceived 'stressing' of an utterance [...] reflects the combined influence of a constituent-structure pattern and its grid alignment."
> (Liberman & Prince, 1977, p. 249)

The fundamental rhythm unit in this approach is the metrical foot, comparable to the meter in poetry or the bar in music. A foot is necessarily composed of a prominence (the head) and of a variable number of weaker elements (unaccented or less accented syllables). The possible structures of the foot i.e. - the position of the head relatively to the weaker elements - differ across languages and its inventory is defined by the Iambic-Trochaic Law (Hayes, 1995).

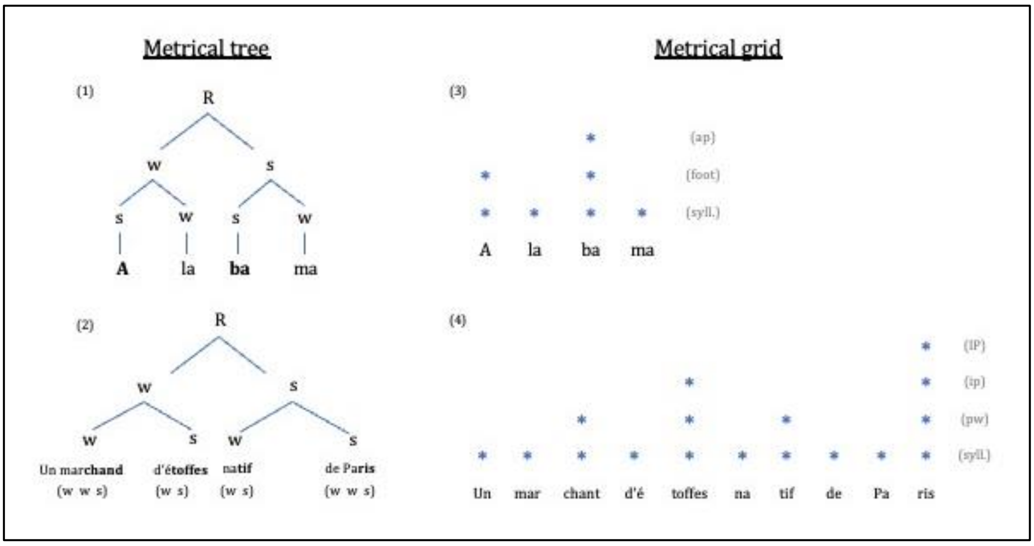


Figure 2 - Metrical tree and metrical grid representation for an English word in (1) and (3) and a French utterance in (2) and (4) (adapted from Di Cristo, 2013).
R = root, s = strong, w = weak. The rightmost columns in (3) and (4) indicate the constituent levels: ap = accentual phrase, syll. = syllable, pw = prosodic word, ip = intermediate phrase, IP = intonational phrase.

Figure 2 illustrates the notion of relative prominence, in A-la-ba-ma (example (1) and (3)), "A" and "ba" are prominent in relation to "la" and "ma" respectively, at the level of the foot. However, at the accentual phrase level (which in this example corresponds to the word however it is not always the case) "ba" is more prominent (primary stress) relatively to "A" (secondary stress).

In French, the notion of relative prominence is illustrated within an intonational phrase in Figure 2 (example (2) and (4)). However, it also occurs at the level of the accentual phrase or even of the prosodic word (a content word with its clitics, Selkirk, 1996/2014). An accent in final position of a constituent is obligatory, therefore the final accent in French is considered as primary accent. Yet, constituents can also be accented on the initial syllable of a content word, such as in ***Président.*** The initial accent is considered secondary relatively to the primary final accent, meaning that French accentuation is dominantly right-headed[3]. The possibility of a constituent receiving an initial and a final accent is referred to as the *bipolarity principle*[4] (Delais-Roussarie & Di Cristo, 2021; Di Cristo, 2013).

The metrical tree, however informative of the prominence hierarchy, does not include the temporal dimension necessary to represent the eurhythmic principle, i.e., the tendency of languages to favour a balanced alternation of strong and weak beats. This translates to a preference for the avoidance of *stress-clash,* when two successive syllables bear primary stress, and *lapse*, a succession of several unaccented syllables. The metrical grid representations in (3) and (4) make this alternation quite clear.

As can be seen though, this alternation is not a strict one (one strong syllable for one weak syllable). In French it is common to find two or three unaccented syllables between prominences, and the secondary initial accent plays an essential role in maintaining eurythmy (Di Cristo, 2013).

The metrical theory was originally developed to describe the rhythmic structure of English, where the foot is of trochaic nature, meaning it is left-headed and marked by an increased intensity. The metrical theory has also been at the base of descriptions of prosodic systems of other languages than English. As mentioned above, Delais-Roussarie & Di Cristo (2021) have proposed a description of French accentuation based on these principles. In French, the foot is of iambic nature, meaning right-headed and marked by an increased duration. The differences between the English and the French prosodic system are presented in detail later in this Chapter (section 4., p. 46).

In sum, through the metrical theory lens, **speech rhythm is the fundamental structure that emerges from the language's metrical rules, through the combination and subordination of constituents, instantiated by the relative prominence of their heads**. The acoustic correlates are therefore

[3] In French: le principe de dominance à droite (Delais-Roussarie & Di Cristo, 2021).
[4] In French: le principe de bipolarité (Delais-Roussarie & Di Cristo, 2021).

these prominences, physically manifested by an increase in pitch, intensity and/or duration.

This view of speech rhythm based on the accentuation system constitutes an intermediate level - the meso-level - of analysis between the micro-level described in the previous section and the next two angles developed here: temporal variables, and fluency.

3.3.3. Temporal variables

The term *temporal variables* refers to variables related to timing in speech, mainly duration and frequency of silent pauses and utterance length and rate. As Goldman-Eisler (1968) explains in her pioneer book *Psycholinguistics: Experiments in Spontaneous Speech,* a number of authors turned their interest from studying language on the basis of written corpora (whether originally written or transcriptions of oral language) with a focus on language's norms and the description of linguistic categories, to studying "the live product" (p. 8), that is speech in action.

This shift came largely from a psycholinguistic perspective, with the aim of studying - through the act of language - the underlying psychological processes involved in speech production. In this approach, physiological and cognitive factors are taken into account, and the flow of natural speech is viewed as an "indicator of the speaker's generative activity" (p.9). As such, the primary object of these studies is spontaneous speech.

Goldman-Eisler's work helped to recognise that temporal variables and mainly pauses and hesitation are de facto part of the act of speaking and worth studying. She concluded from one of her first study on the matter:

> "The actual proportion of pausing time in utterances, while subject to considerable variation, was high enough to justify the conclusion that pausing is as much part of speech as vocal utterance." (p. 31)

Interest in this area of research grew and hesitation markers such as filled pauses[5], repetition and false starts were also included in so-called temporal

[5] We use the term filled pause as it is the one used in the literature we are referring to here. However, in our own experimental studies we have preferred the term voiced pause for its added precision. "filled" does not necessarily indicate that the pause is filled with voicing, it

variables. After several studies focused on temporal variables within a given language (amongst others Duez, 1976 for French; Henderson et al., 1966 for English), authors started to extend the field to cross-language studies.

Grosjean (1972) proposed a systematised description of temporal variables which was then picked up and adopted by a number of authors in the field. Speed measures such as speech or articulation rate, length of uninterrupted speech run, and length of silent pauses constitute primary variables as they are necessarily present in spoken language, whatever the speech style. Secondary variables include hesitation markers (filled pauses, repetition etc.) which do not necessarily occur in speech, especially so in read or very fluent speech (see Table 2 below).

Primary variables	Secondary variables
- Speech rate	- filled pauses
- Phonation-Time Ratio	- drawls
- Articulation Rate	- repetition
- Length of silent pauses	- false starts
- Length of runs	

Table 2 - Primary and secondary temporal variables according to Grosjean, 1980, pp. 40–42

Within-language studies informed on speech style differences (e.g., cartoon description vs interviews, Grosjean & Deschamps, 1973), and normal vs pathological speech (Quinting, 1971/2019); whereas cross-language temporal variables studies allowed to discriminate language-specific aspects from shared ones.

One of the first and most famous study comparing two languages in terms of temporal variables is that of Grosjean & Deschamps (1975) who compared radio interviews from native English and French speakers. They found similarities in rate measures, silent pauses length, and frequency of hesitations. However, they found speech runs and filled pauses to be longer in French. We found similar tendencies in run length and voiced pauses duration in a previous study (Judkins et al., 2022; this study is discussed in Chapter II, p. 92). This indicates a difference between the two languages at a level we propose to call **the macro-level of rhythm, in that it pertains to the grouping (speech runs) and pausing organisation**.

might as well be a sigh, a tongue click, a throat-clearing noise... Voiced pause refer exclusively to a "uhm"-like hesitation marker.

While studies concerned with temporal variables rarely claim to belong under the speech rhythm umbrella, by looking at speech surface structure they de facto are. Where metrical theory is concerned with the abstract metrical structure underlying speech temporal organisation, temporal variables are in essence on the concrete production level of speech rhythm (as opposed to meter), following Di Cristo (2013)'s dichotomy.

The term temporal variables is now either used when referring to the psycholinguistics framework of studying processes at play in speech production, mainly in L1; or interchangeably with the term fluency measures. While fluency studies can be approached in connection with underlying cognitive processes (Segalowitz, 2010), fluency measures are mostly associated to L2 language proficiency (e.g., Saito et al., 2018; Tavakoli et al., 2020). Therefore, speech performance is viewed as window into the degree of mastery or ease of the processes involved, rather than focusing on understanding such processes.

For the sake of clarity and in accordance with the integrative approach of speech rhythm we propose later in this chapter (p. 44), we follow Grosjean (1972) in the distinction between primary and secondary temporal variable. **Primary temporal variables, such as length of run and silent pauses, are an essential part of speech, and participate in its rhythmical structuration, at what we call the macro-level. Secondary temporal variables such as filled pauses and all kinds of voiced hesitation markers evidently participate in the perception of speech rhythm but are not structural in nature. We will consider them as fluency measures. We make one exception regarding the articulation rate which we categorise into fluency measures, contrary to Grosjean's categorisation of speed measures into primary variables.**

3.3.4. Fluency

Even though temporal variables described above are now largely included in and referred to as fluency measures, the two terms - while overlapping - do not cover the same grounds. **Temporal variables are strictly of acoustic nature, conversely, fluency encompasses a larger spectrum of linguistic aspects such as segmental, semantic, syntactic, and pragmatic accuracy, however its definitions are plural.**

One of the oldest and most referred to definition of fluency comes from Fillmore (1979) who broke down the concept of fluency into four dimensions:

1. The capacity to unravel speech in a fluid manner, with minimal pauses and without stopping the stream of thoughts expressed.

2. The semantic coherence and density of the content.

3. The appropriateness of the content in relation to the context of communication (i.e., social and pragmatic skills).

4. The creativity displayed in the way of expressing oneself (use of original phrasing, vocabulary, humour, metaphors etc.). (p. 93)

In the context of L2 acquisition, this definition of fluency was related to the language competence in L1. Being fluent in L2 meant expressing oneself in the most natural way, displaying a competence similar to that in L1 (Brumfit, 1984).
This definition makes fluency a synonym of overall oral proficiency which authors in the field later called "broad fluency" (Chambers, 1997; Koponen & Riggenbach, 2000; Lennon, 1990), measurable through perceptive judgements from native listeners.

In the narrow sense, fluency corresponds to the objective and measurable aspects of speech delivery such as speed and flow (Lennon, 1990), in alignment with the primary and secondary temporal variables described by Grosjean (1972) previously mentioned in section 3.3.3. (p. 37).

Definitions of fluency emerged from psycholinguistics, as it is seen as the reflection of underlying cognitive processes involved in speech planning and production. In models of cognitive processing, the notion of procedural automaticity is central. The development of a skill (here speaking in L2) is understood as a progression from the conscious application of declarative knowledge - which require attention and conscious effort - at the early stages, to the sub-conscious and automatic retrieval and implementation of ready-to-use sequences - which does not require attention nor effort (Anderson, 2014; Schmidt, 1992).

Stemming from this approach, Segalowitz (2010, 2016) proposes a model that comprises three "types" of fluency: *cognitive fluency*, *utterance fluency* and *perceived fluency*. Cognitive fluency relates to speech planning processes such as conceptualisation and formulation (as described in Levelt's model, presented in the next Chapter), utterance fluency corresponds to what is actually produced by the speaker and can be measured objectively through different types of correlates (presented below), and perceived fluency concerns the appreciation of the listener in terms of ease of following the speaker's speech.

Most L2 fluency studies focus on measures of utterance fluency, often in relation to perceived fluency by mean of listeners' judgement. Utterance fluency measures are plentiful, Skehan (2003) and Tavakoli & Skehan (2005) propose to categorise them into three groups: speed measures, reflecting the continuity and flow of speech, i.e., speech or articulation rate; breakdown measures, which have to do with interruptions in the speech flow, i.e., silent pauses; and repair measures which correspond to disfluencies, i.e., voiced pauses, false starts (see Table 3 for a comprehensive list and description of fluency measures). This categorisation has since been widely adopted and these measures have been the most commonly used in L2 fluency studies.

Fluency measures have also been distinguished on the basis of their *pure* vs *composite* nature. A pure measure captures only one of the three categories defined above, whereas a composite measure captures several. For example, the articulation rate (number of syllables divided by the utterance time excluding pauses) indicate solely the speed of syllable articulation. In contrast, the speech rate (number of syllables divided by the utterance time including pauses) captures both the speed and breakdown aspects. In the following table, we indicate pure measures in blue and composite measures in green.

CATEGORY	DEFINITION
Base measures	▪ **Performance time**: total duration of time allocated to the task ▪ **Phonation time**: time between the first and last phoneme produced ▪ **Total number of syllables or words** per minute or performance
Speed measures	▪ **Speech rate**: total number of syllables divided by the phonation time including pauses (expressed in syll./s. or syll./min.) ▪ **Articulation rate**: total number of syllables divided by the phonation time excluding pauses (expressed in syll./s. or syll./min.) ▪ **Mean syllable length**: phonation time excluding pauses divided by total number of syllable (reverse of articulation rate) ▪ **Pruned syllables per second**: speech rate excluding all types of disfluencies ▪ **Mean Length of Run** (MLR): mean number of syllables between 2 silent pauses ▪ **Phonation run**: uninterrupted phonation time between 2 pauses
Breakdown measures	▪ **Phonation/Time ratio**: proportion of time spent speaking over total performance time ▪ **Mean duration, quantity, distribution of silent pauses** ▪ **Mean duration, quantity, distribution of voiced pauses** ▪ **Mean duration, quantity, distribution of lengthenings** (Quantity is expressed as the total number divided by the total phonation time, distribution relates to the location within or outside a clause)
Repair measures	▪ **Mean quantity of all repair** (false start, repetition, self-correction): total number divided by the phonation time ▪ **Mean quantity of each kind of repair** separately: total number divided by the phonation time
Prosodic measure	▪ **Pace**: number of stressed words per minute ▪ **Space**: proportion of stressed words to total number of words ▪ **Prominence and pitch**: number of prominent syllables per tone unit

Table 3 - Summary of fluency measures, compiled and adapted from Tavakoli and Wright (2020), Derwing et al. (2009), Kormos (2006).
Pure measures are indicated in blue, composite measures are indicated in green.
syll. = syllable; min. = minute

In this table, authors have also added less common measures such as what they have called "Base measures" and "Prosodic measures". We also added lengthenings, what Grosjean calls drawls (duration extension of a segment part of a word) but it is generally not included in repair or breakdown measures (Tavakoli & Wright, 2020). However, lengthenings have been studied in L1 fluency studies and have been described as having similar functions than voiced pauses, i.e., marking hesitation, signalling a repair, and/or holding a speaking turn - unless it assumes a semantic and/or pragmatic function such as in an enumeration, or to create a cliff-hanger effect (Di Cristo, 2016; Johnsen & Avanzi, 2020).

Measures tend to differ across studies. Notably, the chosen silent pauses threshold has varied between 100ms and 400ms (Tavakoli & Wright, 2020), however most recent studies settle on 250ms based on de Jong & Bosker (2013) who demonstrated its relevancy. The definition of a "run" in the MLR measure has been based on the syntactic structure (clause + dependents, called T-Unit such as in Lennon, 1990), later Foster et al. (2000) proposed the Analysis of Speech Unit (ASU) which also takes into account semantic and prosodic aspects, and has been used in several studies since. Alternatively, the run is defined as an uninterrupted stretch of speech between 2 silent pauses. However, some authors have also defined it as stretches of speech without any hesitation or agrammatical pauses (Baker-Smemoe et al., 2014). Finally, some authors categorise filled pauses as breakdown measure and others as repair phenomenon. While it does not change the measure per se, it changes its interpretation. As far as we are concerned, voiced pauses are considered as disfluencies.

As opposed to the three previously presented dimensions of speech rhythm (the micro, meso, and macro levels), disfluencies cannot be pinned down to a structuration level because they are not structural in nature but rather, they are by-products of speech production. They are therefore transversal, as they intervene at all levels of structuration and constituency. A repetition can occur at the syllable level, a false start at the word or group of word level, and voiced pauses can occur within or at the intonational phrase's boundaries.

In the past 40 years, L2 fluency studies have multiplied, with the aim of better defining this concept and its measures, its relation to L2 proficiency, L2 oral competence assessment, and its effect on L2 speech perception by native listeners. However, to our knowledge, **very few of these studies address the relationship between fluency and speech rhythm**.

In the next section, we draw relations between the aforementioned levels of speech rhythm and propose an approach that integrates them all.

3.4. PROPOSITION FOR AN INTEGRATIVE APPROACH TO SPEECH RHYTHM

The previous section gave an overview of different dimensions of speech rhythm stemming from diverse theoretical standpoints. While the so-called rhythm metrics and the descriptions of accentuation systems of languages emanating from the metrical theory claim their rhythmical nature, temporal variables and fluency measures - however relevant - have not been considered through this lens.

We advance the argument that all these aspects are not mutually exclusive but rather, should all be considered as speech rhythm correlates. As such, we propose that speech rhythm should be understood as a multifaceted construct, combining parameters belonging to different domains and level of analysis that interlace and complement one another. Figure 3 below illustrates this proposition.

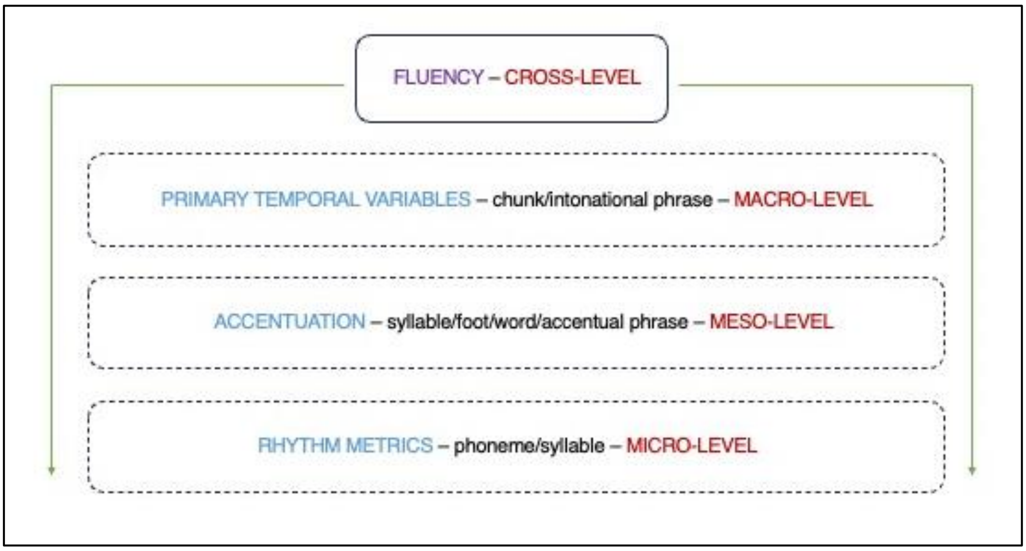

FLUENCY – CROSS-LEVEL
PRIMARY TEMPORAL VARIABLES – chunk/intonational phrase – MACRO-LEVEL
ACCENTUATION – syllable/foot/word/accentual phrase – MESO-LEVEL
RHYTHM METRICS – phoneme/syllable – MICRO-LEVEL

Figure 3 - Proposition of an integrative approach to speech rhythm.

Rhythm metrics such as %V, ΔC, (n)PVI actualised at the phoneme and syllable level concern the micro-level of rhythm. Accentuation rules of alternating weak and strong syllables actualised at the foot, word, accentual phrase and intonational phrase levels concern the meso-level of rhythm. Primary temporal variables such as length of runs and silent pauses instantiate the chunk and intonational phrase levels, and concern the macro-level of rhythm. Fluency measures (disfluencies frequencies and length, speed of delivery) pertain to all levels.

The three structuration levels (micro, meso, macro) are represented inside dotted lines as they are not exclusive categories. Manifestations of rhythm at each level necessarily crosses over to the others - just like rhythm metrics can capture prosodic features from the meso-level. Fluency however, is a phenomenon that transcends all levels and is not structural in nature. Yet, fluency participates in the rhythm of speech. One could argue that disfluencies mostly interfere with rhythmic patterns of each level. Yet, in L1, voiced pauses can also have a regularity that participates in the perception of patterns.

In the past, we ran a study solely on the macro-level of rhythm (Judkins et al., 2022; see Chapter II, section 2.4., p. 92), but in the study presented in Chapter V and VI of this document, aspects from each levels presented above are taken into account.

To provide a full picture, we would have to integrate intonation. We deliberately left it aside in this presentation, but we acknowledge that melodic contours also mark rhythmicity, maybe more so at the macro-level. We aim to integrate intonation in future research.

In the rest of this document, the term speech rhythm will refer to this conception, i.e., the combination of the micro, meso, and macro levels, as well as fluency, unless specified otherwise.

The next section presents the specificities of the French and English languages in regards to all levels of speech rhythm.

4. FRENCH & ENGLISH RHYTHMS

Our research focus has been on L1-French learners of L2-English and L1-English learners of L2-French. These two languages have notoriously been opposed and cited as prototypical examples of syllable-timed and stress-timed languages in the classification of languages based on the isochrony principle (see section 3.3.1., p. 30). Even though the original claim of this theory has since been disproved (strict rhythm categories and isochrony of the rhythm units), English and French have indeed very different rhythmic patterns, whether at the micro, meso or macro levels we have just identified, and even disfluencies follow different patterns.

The measures developed by Ramus et al. (1999) and Grabe and Low (2002) allowed to quantify the difference in syllabic structure complexity and vowel reduction degree between the two languages. Ramus et al. (1999) operationalised the degree of vocalic reduction as the standard deviation of the duration of vocalic intervals (ΔV). The presence of reduced vowels in unstressed syllables should involve a high degree of variability in syllable length since the difference between stressed and unstressed syllables would be enhanced by the reduction phenomenon. As for the syllable structure complexity, authors propose it would be reflected by the standard deviation of consonantal intervals (ΔC), as languages with complex syllable structure (categorised as stressed-timed) allow for both complex and simple consonantal clusters, raising the durational variability of consonantal intervals. The vowel proportion (%V) was said to indicate both vowel reduction and syllable complexity. According to the original rhythm classes (Pike, 1940; Abercrombie, 1967/2019), English being a stress-timed language should then display higher ΔV and ΔC than French, while French, a syllable-timed language, should display a higher %V than English. Their results, shown in Figure 4 and 5, confirm the expected tendencies.

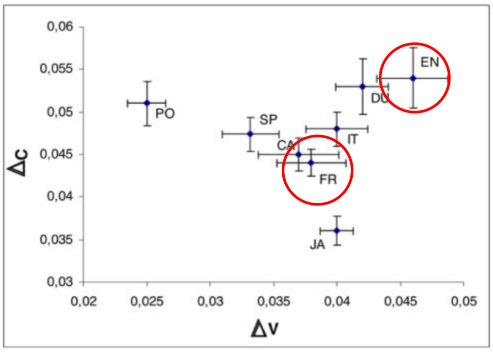


Figure 4 - Standard deviation of consonantal (y) and vocalic (x) intervals in 8 languages including French and English calculated on 5 read sentences by 4 speakers of each language. From Ramus et al. (1999, p. 273)

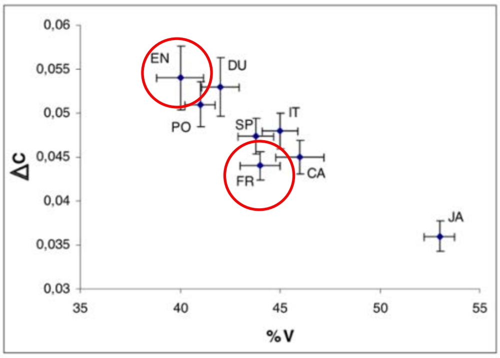


Figure 5 - Standard deviation of consonantal intervals (y) and vocalic proportion (x) in 8 languages including French and English calculated on 5 read sentences by 4 speakers of each language. From Ramus et al. (1999, p. 273)

In Figure 4, we can see that English presents higher scores than French in both measures. Therefore, English has a greater range of structure complexity as shown by a high variability in consonantal intervals' durational variability, and a higher degree of vocalic intervals' durational variability reflecting the alternation between short and reduced unstressed vowels and full stressed ones. Figure 5 shows that French presents a higher proportion of time dedicated to vowels which reflects the absence of the vowel reduction phenomenon as well as a simpler structure involving less consonant clusters than English.

Grabe & Low (2002) also found an important difference between French and English in terms of durational variability of consonantal and vocalic intervals using the Pairwise Variability Index. As shown in Figure 6, once again French presents a smaller degree of durational variability than English for consonantal and vocalic intervals.

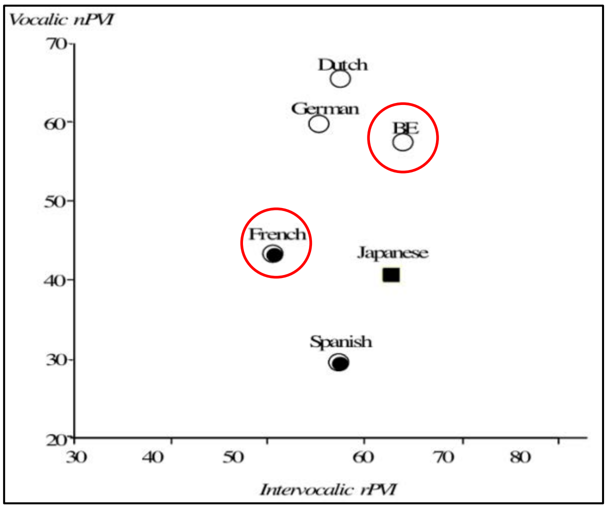
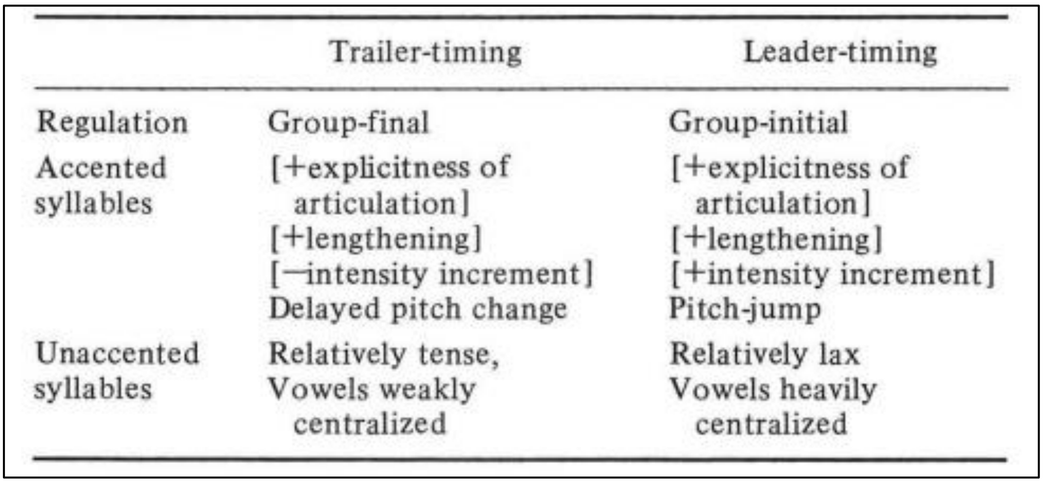


Figure 6 - Normalised Pairwise Variability Index of vocalic intervals (y) and Pairwise Variability Index of consonantal intervals (x) in 6 languages including French and English calculated on the "North Wind and the Sun" text read by 1 speaker per language. From Grabe & Low (2002, p. 4/16). BE= British English.

In contrast to Ramus et al.'s measures, the PVI and nPVI (its normalised version) give an account of the durational variability between two successive intervals, making this measure more robust to speech rate variations.

Overall, Ramus et al. and Grabe & Low's results attest of the difference between French and English in terms of their micro structure. French is more stable in terms of durational contrast, and English shows a greater range of intervals' duration.

Wenk & Wioland (1982) and Wenk (1985), in reaction to the isochrony principle and the categories of syllable and stress-timed propose a rhythmic typology based on the position of the accented syllable in the rhythmic group. Stemming from studies in experimental psychology demonstrating the different perceptions of grouping according to the type of prominence, i.e., intensity perceived as starting the group vs duration perceived as ending the group (Allen, 1975; Fraisse, 1974; Woodrow, 1951), Wenk proposes that French, with its duration-based prominences, can be described as a trailer-timed language, where the accented syllable is positioned in final position of the rhythmic group. In contrast, English marks prominence with an increase in intensity which makes it a leader-timed language where the accented syllable is positioned at the beginning of the rhythmic group. Table 4 below presents the characteristics of accented and unaccent syllables in both categories.



	Trailer-timing	Leader-timing
Regulation	Group-final	Group-initial
Accented syllables	[+explicitness of articulation] [+lengthening] [−intensity increment] Delayed pitch change	[+explicitness of articulation] [+lengthening] [+intensity increment] Pitch-jump
Unaccented syllables	Relatively tense, Vowels weakly centralized	Relatively lax Vowels heavily centralized

Table 4 - Phonetic characteristics of trailer-timing (such as French) and leader-timing (such as English) rhythmic patterns. From Wenk & Wioland (1982, p. 204).

These authors assume that:

> "Rhythmic patterns in speech, to the extent that they correspond to muscular events, involve successive phases of tension and relaxation" (p.204).

Figure 7 below illustrate what Wenk & Wioland called the "rhythm curve".

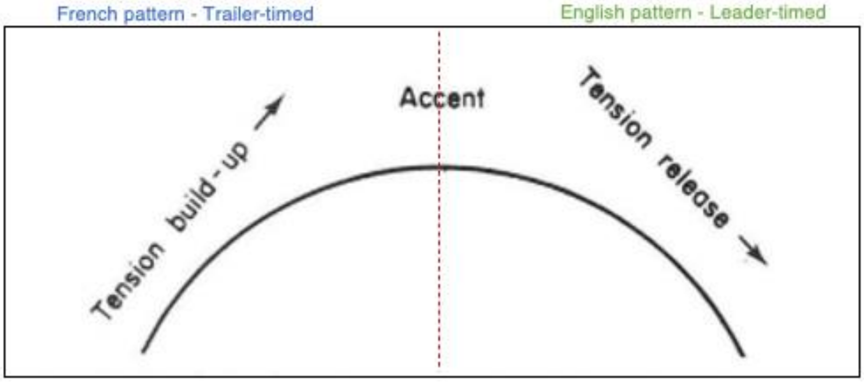

French pattern - Trailer-timed
English pattern - Leader-timed
Accent
Tension build-up
Tension release

Figure 7 - Illustration of the "rhythm curve".
Adapted from Wenk & Wioland (1982, p. 205).

In both English and French (and all other languages for that matter), accented syllables correspond to a peak in articulatory energy. In French, unaccented syllables precede the accented syllable and are realised as muscular tension is building, hence the fact that unaccented syllables are relatively tense and not centralised. Contrastively, in English, unaccented syllables follow the accented one and are therefore realised as muscular tension releases, leading to a high degree of centralisation.

Delattre (1966) also compared French and English (amongst other languages) and noted the difference in terms of accent realisation, with English using primarily intensity and French lengthening. Delattre's work also showed that the duration ratio of accented syllables to unaccented syllable differs in the two languages. Logically, since French using duration as a primary cue to prominence, its ratio is higher (1.78) than that of English (1.6). That is, the difference in duration between accented and unaccented syllable is greater in French than in English.

English and French have also been found to present contrastive patterns in terms of temporal variables. Grosjean & Deschamps (1972, 1975) showed that productions in French tend to yield longer runs (uninterrupted speech) than in English, and that pauses in English tend to be shorter than in French. This tendency was recently confirmed by one of our studies (Judkins et al., 2022).

This divergent organisation of alternation between speech runs and pauses also involves that a greater proportion of silent pauses is dedicated to breathing in French, as compared to English (Grosjean & Deschamps, 1972; Judkins et al. 2022). Grosjean (1980) also observed a contrast in the location of pauses. Where in English it is common to pause within a verbal phrase, it is much rarer in French. Lastly, even though Grosjean (1980) found similar quantities of disfluencies in both languages, he noted that French displayed a somewhat equal number of draws (lengthening of the final syllable of a word) and filled pauses, whereas English presented quite a lot more filled pauses than drawls.

The literature provides quite a large array of the rhythmical differences between English and French at all levels of analysis. Table 5 below gives a summary of the rhythmic characteristics of both languages.

	FRENCH	ENGLISH
Micro-level	▪ low durational variability ▪ simple interval structure ▪ no vowel reduction	▪ high durational variability ▪ complex interval structure ▪ vowel reduction
Meso-level	▪ trailer-timed ▪ accent in final position ▪ accent = duration increase ▪ high duration ratio accented to unaccented syll.	▪ leader-timed ▪ accent in initial position ▪ accent = intensity increase ▪ moderate duration ratio accented to unaccented syll.
Macro-level	▪ Longer runs ▪ Longer silent pauses ▪ Greater proportion of breathing in silent pauses ▪ Silent pauses outside verbal phrases ▪ Similar speech rate	▪ Shorter runs ▪ Shorter silent pauses ▪ Smaller proportion of breathing in silent pauses ▪ Silent pauses within and outside verbal phrases ▪ Similar speech rate
Disfluencies	▪ Near equal proportion of drawls to filled pauses	▪ More filled pauses than drawls

Table 5 - Summary of rhythmical characteristics of French and English by level of analysis (syll. = syllable).

We can assume that these differences will have an impact when learning one of these two languages as an L2 from one of these two languages as an L1. This guided our choice of including L1 data in the study of L2 speech rhythm.

The acquisition of L2 rhythm, transfer phenomena between L1 and L2, and overall effects of the L1 rhythmic typology on the acquisition of the L2 are topics that are developed in the next chapter.

CHAPTER SUMMARY

In this opening chapter, we contextualised and defined our main object of study: speech rhythm. We saw how rhythm is at the heart of all human activity and behaviour, and amongst all definitions of rhythm presented, we note the one provided by Sauvannet (2000) who proposes that the essential criteria pertaining to a rhythmical phenomenon are structure, periodicity, and movement.

In speech, structure emerges through the grouping and hierarchical organisation of the syllables into larger constituents. Movement is created by the alternation of strong and weak syllables (accented vs unaccented), continuous sound and pauses, and contrasting melodic contours. Finally, the recurrence of prominences, at all levels of the hierarchy, creates patterns and periodicity.

We discussed the relationship between speech rhythm and motor rhythm, and highlighted the fact that the interaction between the two has long been exploited in the context of speech rehabilitation, and L2 teaching - albeit for the most part intuitively (see Chapter IV, section 2.2.1., p. 147). We also briefly talked about the role of neural entrainment to speech rhythm in language perception and processing. Investigation of such phenomenon in L2 development could help advance our understanding of the role of speech rhythm in L2 speech processing.

After this broad introduction, we turned to the presentation of the acoustic correlates to speech rhythm across four different theoretical views. From the phonological view, so-called rhythm metrics focus on quantitative measures of the proportion of vowels and consonants which reflect the syllable structure complexity of a language, and measures of durational variability of intervals (vocalic, consonantal, syllabic) that capture the degree of vowel reduction. Because these measures leave aside higher-level prosodic aspects, they are confined to a **micro-level** of analysis of speech rhythm.

Conversely, the prosodic approach based on the metrical theory considers that speech rhythm emerges from the language's metrical rules, through the combination and subordination of constituents, instantiated by the relative prominence of their heads. The acoustic correlates are therefore these prominences, physically manifested by an increase in pitch, intensity and/or duration. This view of speech rhythm based on the accentuation system constitutes an intermediate level of analysis: **the meso-level**.

A third level of analysis concern temporal variables that pertains to the chunking and pausing patterns. Durations of speech runs and silent pauses in between inform on the **macro-level** of speech rhythm.

Finally, fluency measures such as the distribution and number of disfluencies, and measures of speed of delivery also belong under the speech rhythm umbrella. However, such measures cannot be pinned down to a structuration level because they are not structural in nature but rather, they are by-products of speech production. They are therefore transversal, as they intervene at all levels of structuration and constituency.

We believe that all these different windows into speech rhythm are not mutually exclusive, but rather exert an influence on one another and should all be considered for an integrative approach of speech rhythm. This is the view we adopt in this dissertation and as such, **the term speech rhythm will from now on refer to this conception, i.e., the combination of the micro, meso, and macro levels, as well as fluency**, unless specified otherwise.

Lastly, because our work focuses on speakers and learners of French and English, we compared these two languages in terms of their rhythmical structure, across all levels of analysis. Table 5 gives a summary of the differences found.

The next chapter turns to the specificities of L2 speech and the acquisition of speech rhythm.

CHAPTER II - L2 SPEECH PRODUCTION

INTRODUCTION

The previous chapter gave an overview of the different definitions of rhythm in speech and presented correlates to rhythm in the speech signal. The last part pointed out the rhythmical specificities of the French and the English language, our focus being on these two.

This second chapter turns to the specificities of L2 speech production. We start by the presentation of Levelt's (1989) language production model, which later was adapted to account for the production of L2 speech specifically (de Bot, 1992; Bock & Levelt, 1994; Kormos, 2006). These theoretical contributions help to understand the direct link between speech rhythm and the underlying cognitive processes involved in speech production.

We then give an overview of important theories on the acquisition of L2 features. From the Contrastive Analysis Hypothesis (Lado, 1957) to the L2 Intonation Learning Theory (Mennen, 2015), models recognise the effects of L1 transfer, and/or universal L2 acquisition processes.

The second part of the chapter presents a literature review of experimental studies investigating the acquisition of speech rhythm in L2. Most studies offer an interpretation of their findings in relation to L2 acquisition theories presented in the preceding section. The review is organised into sub-sections corresponding to the focus of the studies on either the micro-level, the meso-level, or the macro-level and fluency aspects.

Finally, the chapter closes on a section which discusses the relation between L1 and L2 macro-level and fluency patterns. A study on L1 and L2 English and French we conducted and published is presented (Judkins et al., 2022).

1. SPEAKING IN AN L2

We can all agree that speaking in an L2 is different than speaking in our first language (L1). We can observe this through introspection, when we are confronted to speaking an L2, while travelling in a foreign country or in any situation that requires an interaction with someone who doesn't speak our L1. What happens then? How could we characterise that difference?

From my point of view, the L2 speaking experience entails consciously taking the time to prepare a sentence, then articulate it to the best of our capacity/willingness, sometimes stumbling on sounds or words. But despite all this effort, our interlocutor often responds in English (when they have the capacity to do so!). Most likely to ease our pain, because it took them just a second to identify that we are not a native speaker of that language. When the situation is reversed, in an interaction with someone unknown, we are able to identify very quickly if that person is speaking our L1 as an L2, and even sometimes we can make a guess at what their L1 is.

From these very common and intuitive observations, we can already characterise L2 speech as a) being less automatic than our L1 (hence the conscious preparation phase), b) containing identifiable traces of another language (usually the L1 but it can also be another L2), and c) easily identifiable by native listeners. Unsurprisingly, these three aspects are the main foci in L2 speech descriptions and research.

In this section, we will start by presenting speech production models, then move to the question of the influence of the L1 in the acquisition of L2 phonology.

1.1. LANGUAGE PRODUCTION MODELS

1.1.1. Speech production in L1

To understand and study L2 speech production and development, we must first take a look at the processes involved in speech production in general, therefore turn to cognitive psychology. The reference is Levelt's speech production model (1999; 1989) illustrated in Figure 8 below.

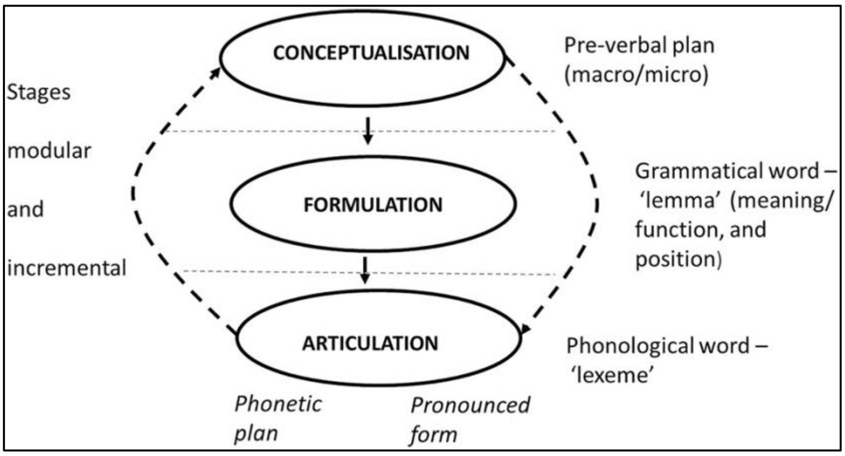


Figure 8 - Levelt's speech production model. From Takavoli & Wright, 2020, adapted from Levelt, 1989, p. 9.

The first component is the conceptualisation, where the speaker plans the general content of what they intend to say and the overall form it will take (e.g., statement vs question). Planning the content of the message is referred to as macro-planning and is understood as being language-independent, whereas planification of the form is referred to as micro-planning and seen as language-dependent as it is connected to the encoding of semantic and pragmatic information.

Both macro and micro planning processes generate a pre-verbal message which then progresses into the formulation phase. Formulation corresponds to the encoding of appropriate grammatical, lexical and phonological forms. The message then progresses to the articulation stage, where it takes its phonetic form and speech is uttered. A fourth monitoring component (not represented in Figure 8) serves as an online checking tool which allows correction at each of the three stages.

The model postulates that while the conceptualisation phase requires a certain level of consciousness and attention, the formulation and articulation are automatic processes, making it possible for the three stages to work in parallel, with little cognitive effort. This high level of efficiency is usually what characterises L1 speech and highly proficient L2 speech. The result is speech that is fast and smooth i.e., fluent.

In the case of a not-so-proficient L2 speaker, each stage will occur one after the other and require much more attention and effort, thus hindering speech fluency. The link between speech fluency and the degree of automaticity of Levelt's language production components constitutes the base of L2 fluency definition and framework (Kormos, 2006; Segalowitz 2010; Tavakoli & Wright, 2020), and will be further discussed in the coming sections (1.1.2., p. 60, and 2.3., p. 83).

Although shortcomings of Levelt's model have been raised (Pawley & Syder, 1983 cited in Kormos, 2006), and other models have been proposed (Dell, 1986; Laver, 1980; Nooteboom, 1980), it remained the reference when researchers started to develop L2 speech production models.

1.1.2. Speech production in L2

A bilingual speech production model was proposed by de Bot (1992), Bock & Levelt (1994) and enriched by Kormos (2006). This latter version is illustrated in Figure 9 below. In a bilingual individual, the question relates to the shared (L1 + L2) vs duplicated quality of each component of the original model.

It is postulated that the macro level of planning in the conceptualisation of the message is language-independent. L1 and L2 concepts are therefore stored together, and concepts can be shared between the two languages, or separate (when a concept exists in one word in a language but does not in the other for instance). In the early stages of L2 acquisition (low proficiency), L2 concepts are aligned onto L1 concepts and become independent as proficiency develops (Kormos, 2006). The micro level of planning requires the selection of a language so that the pre-verbal message contains a language cue.

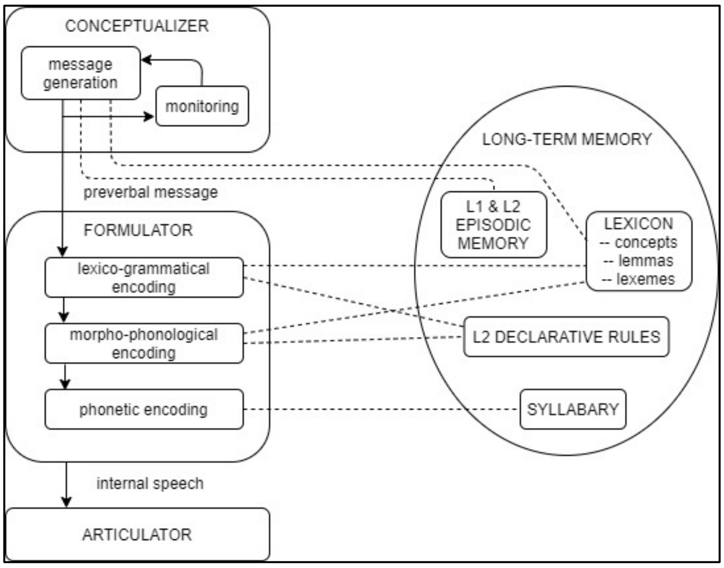


Figure 9 - Bilingual speech production model. From Kormos, 2006, p. 168.

Formulation is de facto language specific since it requires syntactic, lexical, and phonological shaping. However, L1 and L2 syntactic and morphological forms (lemmas) as well as words (lexemes) are thought to be stored in a common mental lexicon. The selection of the correct form results from the competition between activated forms in both L1 and L2 (amongst others Costa et al., 2000), and the selection of the language is made in accordance with the pre-verbal language cue generated at the previous conceptualisation stage.

Phonological encoding involves the retrieval of the selected language's phonemes which are stored in a common L1-L2 network (Poulisse, 2000). At the beginning stages of L2 learning, similar L1-L2 phonemes might mistakenly be associated to a single representation (this assimilation phenomenon described by Flege (1987, 1995) - amongst others - is developed in section 1.2., p. 62 which presents speech acquisition models). Whereas in L1, suprasegmental rules (syllabification, accentuation patterns, intonation) are applied automatically, L2 learners - depending on their proficiency level - might need to consciously retrieve L2 phonological rules from their declarative memory. If L2 phonological rules are not yet available to the learner, they will rely on their L1's (L1 to L2 phonological transfer phenomena are developed in the next section).

Lastly articulation relies on the syllabary (Levelt, 1999) which stores syllable articulation programs for both L1 and L2. As for phonological rules, it is postulated that at low proficiency levels, speakers will rely on automatised L1 syllable sequences, and develop L2 specific programs as they gain proficiency.

Disfluencies might arise at each stage of the process, as the monitoring component detects errors or divergences from the speaker's original intention. Problems at the conceptualisation phase might be related to an inadequacy between the intended concept and its expression in L2, which is a common issue when attempting to translate a thought in L1 to an utterance in L2. The formulation stage in L2 might require a lot of attention and conscious retrieval of the different linguistic aspects involved, which might slow down the production process and generate ruptures and disfluencies in the signal, and the same goes for the articulatory phase where L1 habits tend to die hard.

According to Kormos' view, prosodic encoding intervenes also at each major stages of the process. In the micro-planning phase, which is related to the sentence modality, it seems fair to assume that the prosodic shape associated to the given modality is also selected. In the formulator, accurate metrical patterns and intonational contours must be selected in relation to the phrase structure and the semantic and pragmatic aspects of the message.

This L2 speech production model helps us understand the direct link between fluency and the underlying cognitive processes involved in speech production. Since the beginning of the 2000's, L2 fluency studies have blossomed and still heavily rely on this cognitive-based approach (Segalowitz, 2010; Tavakoli & Wright, 2020). We will come back to L2 fluency and its measures in the second part of this chapter (section 2.3., p. 83).

1.2. L2 ACQUISITION MODELS

1.2.1. The Contrastive Analysis Hypothesis

The influence of the L1 on the L2, theorised as the construct of transfer, has been at the base of L2 acquisition theories, and especially so in the acquisition of L2 phonology where transfer is most prevalent.

The Contrastive Analysis Hypothesis (CAH) was developed based on the early work of several authors from structural linguistics and behaviourism (Fries, 1945; Lado, 1957; Trubetskoy, 1939; Weinreich, 1953) who all acknowledged the undeniable influence of the L1 on the nature of errors occurring in L2. Trubetzkoy

(1939) proposed the - now widespread - concept of phonological filter, according to which the L2 learner perceives L2 sounds through the filter of their L1 phonological system, thus inducing interpretation and production errors when L2 sounds differ or lack an equivalent in the L1.

Lado's work (1957) and the Contrastive Analysis Hypothesis in general postulates that the analysis of the differences between two linguistic systems - especially phonetic features - allows to predict learners' difficulties. Similarities between the two linguistic systems will lead to positive transfer - i.e., transfer resulting in a successful interpretation or production of a target sound - whereas differences will lead to negative transfer (errors). However, the CAH does not allow to predict the degree of difficulty or the specific areas that will be more difficult than others beyond the broad principle according to which whatever is different between L1 and L2 will be difficult (Major & Kim, 1999).

Another strong criticism against the CAH is that transfer is considered the (only) explanation for all errors in L2 productions. Selinker (1972), amongst other authors, argued that L2 learners' interlanguage is influenced and shaped by many factors, transfer being one of them, yet certainly not the only one. New L2 acquisition models were proposed, in which in addition to transfer, non-transfer processes such as developmental factors similar to those in L1 acquisition and universal acquisition processes were put forward as alternative explanations to errors in L2 (Major, 2008).

1.2.2. The Markedness Differential Hypothesis

Eckman's Markedness Differential Hypothesis (MDH) model (1977) still relies on the comparison of the L1 and L2 linguistic systems in order to predict L2 leaners' difficulties, but introduces the notion of typological markedness. Any language feature is more or less marked, relatively to other traits. A language trait A is more marked than a language trait B if the presence of A implies the presence of B but not the opposite. That is, B can exist by itself, whereas A exists only if B exists too, A cannot exist by itself.

A common example is that of voiceless obstruents. Some languages have in their phonemic inventory only voiceless obstruents (e.g., Korean). Other languages have both voiceless and voiced obstruents (e.g., French). However, there is no language with only voiced obstruents. Therefore, if a phonemic inventory includes voiced obstruents, it necessarily includes voiceless ones. But a phonemic inventory can include only voiceless obstruents and no voiced ones. Conclusion: voiced obstruents are more marked than voiceless ones.

According to the MDH, the more marked the feature, the more difficult it is to acquire, whether in an L1 or L2 context. As such, markedness is considered as a universal phenomenon that predicts how difficult a language trait will be to acquire for the learner. Predictions under the MDH have been supported by empirical studies (e.g., Major & Faudree, 1996; Rasier & Hiligsmann, 2007) and present an interesting alternative to explain L2 productions that would not fit a transfer hypothesis.

In fact, a markedness scale of sentence prosody has been proposed by Zerbian (2015), based on and extended from Rasier & Hiligsmann (2007)'s work. The latter postulated that pragmatically determined sentence prosody is more marked than structurally determined sentence prosody. Zerbian added that within the category of pragmatic sentence prosody, the marking of given information is more marked than prosodic focus marking.

Both the CAH and MDH are applicable to all aspects of language learning, but later on models on the acquisition of L2 phonology specifically were developed.

1.2.3. Perception-based models

Examples of influential models are the Native Language Magnet Theory (NLMT; Iverson & Kuhl, 1995), the Speech Learning Model (SLM; Flege, 1995), and the Perceptual Assimilation Model for L2 speech perception (PAM-L2; Best, 1995). Despite some differences, all three theories rely on two major principles to account for learning difficulties: the fact that L2 sounds are perceived through the L1 phonological filter (Trubetskoy, 1939), and the degree of similarity of L2 sounds with those of the L1.

The main postulate is that L2 sounds that are different but resemble L1 sounds will be difficult to discriminate and consequently to produce because they will tend to be (wrongly) assigned to a pre-existing L1 sound category. This phenomenon is called *magnet effect* in the NLMT, *equivalence classification* in the SLM, and *assimilation* in the PAM but the same principle stands. Accordingly, it is possible to predict the segments with which learners will struggle the most based on their degree of similarity with segments in their L1.

Figure 10 below presents a schematic representation of the different types of assimilations described in the PAM (Best, 1995; Best et al., 2007). In the *two-category assimilation,* L2 sounds are categorised into two L1 categories, leading to excellent discrimination of the sound pair. The *single-category assimilation* means that two L2 contrastive sounds are categorised in the same L1 category as they equally differ from an ideal L1 sound, which makes discrimination difficult. The *category goodness difference assimilation* is similar but is this case, one of the two L2 sounds is perceived as a good exemplar of the L1 sound, while the other is perceived as a bad exemplar. This might lead to moderate to good discrimination. The *uncategorised-categorised assimilation* leads to the assimilation of one L2 sound into an L1 category while the other is not assimilated, in this case, discrimination should be facilitated. And finally, the *uncategorised-uncategorised* assimilation corresponds to two L2 sounds that are both not assimilated to any L1 category. In this case, discrimination can range from poor to very good.

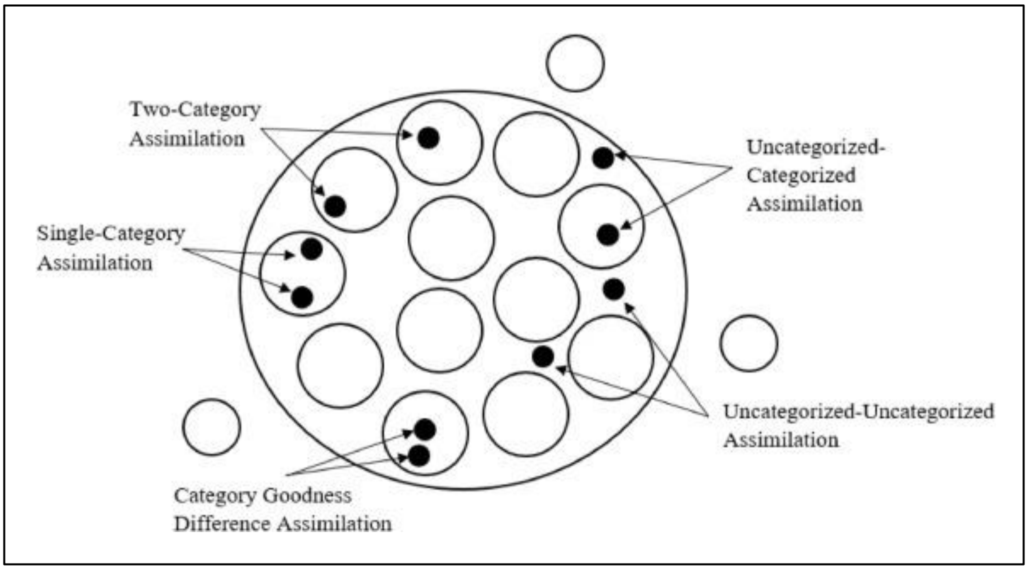

Two-Category Assimilation
Uncategorized-Categorized Assimilation
Single-Category Assimilation
Uncategorized-Uncategorized Assimilation
Category Goodness Difference Assimilation

Figure 10 - Assimilation types proposed by PAM (from Best, 2014).

While these models have been proposed to account for learners' development of L2 individual segments, there has been a few attempts to apply them to the acquisition of certain prosodic features. Notably, So (2010, 2012) and So & Best (2010, 2011, 2014) have extended the PAM-L2 into the PAM-S, a version that accounts for the categorisation of non-native suprasegmental contrasts, especially in tone vs non-tone language pairs. For example, So & Best (2014) tested L1 English and French speakers' perception and categorisation of Mandarin tones.

They found that both groups assimilated Mandarin lexical tones into their L1 categories of sentence modality (statement, exclamation etc.).

Despite the growing interest in L2 prosody, a majority of research has focused on the acquisition of segments only. As Sönning (2023) points out in his recent review of L2 phonology acquisition models, theories on phonological acquisition have been predominantly concerned with individual segments, and a lot less has been developed to explain the development of suprasegmental aspects, let alone rhythm.

1.2.4. The Ontogeny Phylogeny Model & Linguistic Theory of L2 Phonological Development

Sönning, however, presents a study on speech rhythm in which hypotheses from two acquisition models are tested against data from L1 German learners of L2 English. First a hypothesis is formulated in accordance with Major's (2001) Ontogeny Phylogeny Model (OPM). The OPM combines L1 transfer, markedness, and L2 similarity to account for the learners' progression in L2, and assumes that the role of the L1, L2, and universals weigh differently throughout stages of acquisition. Chronologically, the role of the L1 is supposed to follow a downward slope whereas the L2 an upward one, while the role of universals is supposed to first increase then decrease, forming a hill shape. All of the above being nuanced by the degree of markedness and similarity of the language features in focus. Following these assumptions, the author expects to see a U-shaped curve when observing durational variability measures (nPVI, VarcoV, %V) in the L2 English speech of L1 German speakers, across proficiency levels ranging from low to high.

Additionally, Sönning also tests James's (1988) Linguistic Theory of L2 Phonological Development (LTD). The LTD distinguishes the lexical, prosodic and rhythmic levels of phonological representation and focuses on their interactions. The prosodic level is represented similarly as in the metrical theory, with levels of constituency ranging from the syllable to the sentence, and an alternation of strong and weak elements at each level. The rhythmic level is represented as a hierarchical organisation of three types of elements: unstressed elements preceding a stressed one (proclitics), stressed elements (heads), and unstressed elements following the head (enclitics). To each of these corresponds a tempo: faster tempo for proclitics, and a slower one for heads and enclitics (because of accentual and final lengthening).

James posits a universal bottom-up acquisition of L2 rhythm from the lexical level, to the prosodic and finally the rhythmic level, where the acquisition of higher-level structure rests on the lower-level ones. The LTD also predicts that at the prosodic level, the marking of strong elements precedes that of weak ones, and similarly at the rhythmic level, heads and enclitics properties are acquired before proclitics'.

In sum, the LTD postulates that the acquisition of rhythm follows a universal progression: an increase in durational variability between each element (strong/weak - proclitics/heads/enclitics). Sönning looked at the realisation of strong and weak syllables in four level of constituency (a simplified version of the original) and following the LTD, predicted that accurate realisation of prominence should develop in a bottom-up fashion (from lower-level to higher-level as a function of proficiency).

Sönning's results did not follow his predictions. Measures of nPVI and VarcoV showed an upward trajectory instead of the expected U-shape suggested by the OPM. Regarding the realisation of stress/unstress at different level of constituency, results were again - for the most part - inconsistent with the LTD predictions. Consequently, these results disprove the propositions of both the OPM and the LTD, but more empirical testing of these models are needed before drawing any firm conclusion on their relevance.

1.2.5. Archibald's model of word stress acquisition

Of the few theories focusing on the acquisition of prosodic properties, Archibald (1993, 1994) conceives the acquisition of word stress (in L2 English) as a combination of universal principles, language specific parameters, and L1 transfer. The Universal Grammar the author refers to is based on the foundations of metrical phonology that dictates the possible stress patterns in languages. The precise settings such as size of the metrical feet and head location (right or left) are language specific. He argues, providing empirical evidence from L1 Hungarian, Polish and Spanish learners, that learners' L2 speech respect the principles of Universal Grammar, show a resetting of stress assignment rules to adopt the L2 stress pattern (in the case of success), and also includes stress placement errors due to the transfer of the L1 settings. However, in a later study (Archibald, 1997) involving learners from so-called non-accentual languages (Chinese and Japanese), L1 transfer appeared in the form of how word stress was treated. Instead of using stress assignment rules, learners treated stress as a lexical phenomenon, therefore memorising stress location in each word (just like tone in Chinese and pitch accent in Japanese).

One of the limitations of Archibald's work lies in the fact that it has been solely centred on the acquisition of English stress. Moreover, in our view his propositions do not seem to enable precise predictions regarding the acquisition of stress, and does not provide any elements regarding the acquisition of a non-stressed language.

1.2.6. The L2 Intonation Learning Theory

One of the most recent contributions to L2 prosody acquisition models is the L2 Intonation Learning Theory (LILt) from Mennen (2015). Grounded on the autosegmental-metrical framework (AM, Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988), LILt posits that in order to generate predictions about the learners' areas of difficulty in L2 intonation, it is crucial to identify precisely cross-language similarities and or dissimilarities. To that effect, it is essential to distinguish the phonological and the phonetic levels, as prescribed by the AM and advocated in the author's previous work (Mennen, 1999, 2004, 2007; cited in Mennen, 2015). However, LILt goes further and specifies four dimensions to take into account in order to characterise elements of a language's intonation system, and operate cross-linguistic comparisons:

1. The *systemic dimension* - organisation, combination rules, and inventory of categorical phonological elements.
2. The *realisational dimension* - phonetic implementation of the elements (alignment and shape).
3. The *semantic dimension* - elements' semantic function(s).
4. The *frequency dimension* - usage frequency of the elements.

The theory builds on language-specific intonational phonology descriptions provided by Ladd (1996). L2 intonation literature support the fact that learners' use of tones can deviate from that of native speakers in all four dimensions, and in most cases such deviations are the result of an L1 transfer (see Mennen, 2015 for a review). From these previous studies and what perception-based learning models on segmentals (SLM & PAM) have revealed, Mennen forms several assumptions (to be tested) regarding the acquisition of L2 intonation.

Concerning the L1 and L2 systems, she posits that deviations in L2 intonation production are perception-based (phonological filter) and similarity leads to assimilation. Secondly, similarity/dissimilarity can arise in more than just the systemic dimension, and the influence of context has to be taken into account. And lastly, L1 and L2 categories share a common phonological space which leads to L1-L2 interactions and potential merging. Studies have shown that productions can fall in between the L1 and the L2 values (for pitch range for example) in both the L1 and L2 of a speaker (De Leeuw et al., 2012; Mennen et al., 2014).

Turning to individual characteristics, Mennen assumes that the age of first contact and start of learning of the L2 influences the successful acquisition of L2 intonation, yet she acknowledges that this influence might differ across the four dimensions. In addition, learners have the ability to get closer to native targets as they gain experience in the L2, but their progression is not parallel across all four dimensions. Studies indicate that phonetic realisation and the semantic dimension are more problematic than the systemic dimension (for instance Jun & Oh, 2000, and recently Sánchez-Alvarado, 2022). Therefore, the four dimensions represent varying degree of difficulty for learners.

Finally, the LILt acknowledges that intonation is connected to other prosodic and segmental aspects and that its acquisition also depends on the development of these other speech features. In this first version, the theory gives a major role to L1 transfer in the acquisition of L2 intonation, especially so in the initial stages of development, and does not make any prediction regarding universal processes.

Despite the multiplication of L2 phonological acquisition models over the years, research in this area still needs to be developed in order to build empirical evidence to support them. What clearly emerges is that **the acquisition of L2 phonology is not a monochrome process. While transfer from L1 is undeniable, universal processes are also at play and the effect of different factors intertwine in a dynamic fashion. L1-L2 similarity is also a crucial factor that most models rely on to determine areas of difficulty.** However, defining similarity degrees is not always straight forward, especially when looking at prosodic features, as opposed to segmentals (Mennen, 2015).

Some of the theories presented above are compatible with the study of prosodic features and there is an exciting growing body of research around the acquisition of suprasegmentals (e.g., Li & Post, 2014; Rasier & Hiligsmann, 2007; Sánchez-Alvarado, 2022). However, **to this day, a comprehensive description of the acquisition of L2 speech rhythm that would include micro, meso, macro levels and fluency, have yet to be developed.** The following section of this Chapter presents a review of studies focusing on the acquisition of (several aspects of) speech rhythm with the underlying question of transfer vs universal processes' role.

2. *THE ACQUISITION L2 SPEECH RHYTHM*

L2 speech rhythm, or rhythmical aspects of the learners' interlanguage, has been an object of study in the field of L2 acquisition. However, as seen in the first chapter, speech rhythm can be approached through different lenses (rhythm metrics, metrical theory, temporal variables, fluency).

Following the rhythm metrics frenzy - which were primarily used to discriminate languages in their native form - L2 acquisition researchers quickly started to apply these metrics to L2 speech. The research goal of such studies was mainly to observe the influence of the learners' L1 rhythm on their speech rhythm in L2, especially in the case where the L1 and L2 belonged to different rhythm categories (typically, going from a syllable-timed language to a stress-timed one or the opposite, such as in Ordin & Polianskaya, 2015). Some studies have also focused on comparing the different rhythm metrics to try and find the most suited one to capture the specificities of L2 speech rhythm in a given language (amongst others Yazawa & Kondo, 2022).

Shifting perspective, a number of studies have also looked at L2 speech rhythm in a more qualitative fashion through the observation of stress realisation both in terms of location and phonetic realisation accuracy (e.g., Rasier & Hiligsmann, 2007). The question of the influence of the L1 onto the L2 realisations vs the existence of L2 universals underlies this type of work.

Finally, a large number of L2 acquisition studies are concerned with fluency. Fluency studies generally aim at defining and describing the relationship between L2 utterance fluency, perceived fluency and proficiency level. Interest for L2 fluency has grown exponentially in the last 40 years, in connection with the need to define and assess L2 oral competency. Temporal variables as defined in the previous chapter (chunk and pausing patterns) have been absorbed in fluency measures and are now commonly referred as such.

This section presents a literature review of L2 speech rhythm studies from all aforementioned perspectives.

2.1. L2 SPEECH RHYTHM AT THE MICRO-LEVEL

Rhythm metrics have been applied to L2 speech to account for L1 vs. L2 differences, and investigate L2 rhythm acquisition stages, yet the results are mixed.

Gut (2003) investigates the acquisition of the German rhythm by learners from three different L1: Polish, Italian and Mandarin. Their productions (reading + re-telling tasks) are compared to that of native German speakers. In this study, the focus is solely on contrasts between adjacent non-reduced and reduced syllables occurring in a specific German suffix.

The author does not actually use common rhythm metrics (see Table 1, p. 33) but calculates a "Syllable Ratio" which consists in dividing the duration of a non-reduced syllable by the duration of the following reduced syllable for each non-reduced/reduced pair. Then the results of all pairs are added together and divided by the number of pairs, so as to obtain an average. This measure differs slightly from the nPVI since it focuses solely on the difference between pairs of unreduced and reduced syllables, as opposed to considering all pairs of adjacent syllables. In doing so, this measure captures the realisation of stressed syllables in comparison to unstressed ones (we use a similar ratio in our study, see Chapter V, p. 202).

Results show that in the reading task, all three groups of L2 speakers present significantly lower ratios (that is, a smaller durational difference between reduced and non-reduces syllables) than L1 German speakers. Interestingly, the learners' L1 does not seem to have an impact as they all have very similar results. In the retelling task, only the L1 Mandarin group differs significantly from the L1 German group, and the results are a bit more contrasted between groups.

Carter (2005), using sociolinguistic interviews, investigates the difference between the nPVI of L1 English, L2 English spoken by bilingual Mexican immigrants residing in the US, and L1 Spanish (spoken by Mexicans residing in Mexico). He finds that the nPVI-V values of the L2 English group sit right between the low values of the L1 Spanish group and the high ones of the L1 English group. Results are explained as a persistent transfer from L1 Spanish where vowel reduction is rare, and point to the relevance of the nPVI-V for measuring interlanguage varieties between rhythmically contrasting languages.

Lin & Wang (2007) used %V and ΔC to compare L1 English, L1 Mandarin, and L2 English spoken by L1 Mandarin speakers (upper intermediate proficiency level). They used read sentences as well as conversation recordings. Results showed that L1 Mandarin differed significantly from L1 English, L1 Mandarin showing syllable-time-like scores and English stress-time-like scores as always. The L2 English group differed significantly from the L1 Mandarin group but not from the L1 English group, except for %V in the reading condition. This indicate that these L2 speakers are closer to an L1 English realisation than an L1 Mandarin. Authors also found that scores for both measures were higher in the conversation task, for all three language groups, once again pointing to the sensitivity of these measures to the task type.

In an influential paper on the topic, White & Mattys (2007) combined an analysis of the distinctive power of rhythm metrics between several L1, then between several L1 and L2, with both rhythmically similar and contrasting languages. They applied the whole battery of rhythm metrics (%V, ΔV, ΔC, VarcoV, VarcoC, nPVI-V, rPVI-C) to L1 English, Dutch (both stress-timed), French and Spanish (both syllable-timed) and found that %V, VarcoV and nPVI-V best accounted for the stress vs syllable time distinction, as well as differences between languages within these categories. In L2, they compared L1 English-L2 Dutch with L1 Ducth-L2 English, and L1 Englilsh-L2 Spanish with L1 Spanish-L2 English. There is no precise information on the proficiency level of the L2 speakers but according to the authors they all present a noticeable non-native accent. Authors found that different measures reflected the difference between L1 and L2 in the English-Spanish pair and in the English-Dutch pair. VarcoV, nPVI-V and %V allowed to discriminate between L1 and L2 versions of English and Spanish. However, none of the vocalic interval variability measures reflected the distinction in the English-Dutch pair. An effect of the L1 was observable on %V though.

A number of studies used the rhythm metrics not only to look at the distinction between languages' L1 and L2 version, but also to question their relevance in the distinction of L2 proficiency level.

Guilbaud (2002) measured a variability index (similar to nPVI) on syllables in the speech of L1 English learners of L2 French of low-intermediate level (n=3) and upper-intermediate level (n=3). The scores did not differ significantly between the two L2 French proficiency levels, and actually less proficient learners had a slightly lower score (closer to that of native French) than more proficient ones. However, it should be noted that the difference in proficiency level between the two groups was actually quite small and levels were slightly overlapping.

Ordin et al. (2011), in a study involving a much larger number of speakers than the usual 5 or 6, apply all rhythm metrics (%V, Δ, Varco, rPVI, nPVI) on vocalic, consonantal, and syllable intervals to the speech of 51 L1 German learners of L2 English at 3 distinct proficiency levels: lower-intermediate (n=12), upper-intermediate (n=9), and advanced (n=22). They found that Δ and rPVI measures did not significantly differ between groups (most likely because of their sensitivity to speech rate variation), however all other measures did (Varco and nPVI measures increased along with proficiency level). Furthermore, by mean of a discriminant analysis, they found that metrics calculated on syllable durations yielded a much clearer distinction between proficiency levels than vocalic and consonantal intervals metrics.

Stockmal et al. (2005) investigated the rhythm of L1 Russian speakers (a stress-time language) in L2 Latvian (a syllable-time language), in comparison with L1 Latvian speakers. They applied all rhythm metrics as did White & Mattys (2007), on two read sentences per speaker. They compared scores obtained for L1 Latvian (n=10), L1 Russian (n=1), low-proficiency L2 Latvian (n=5), and high-proficiency L2 Latvian (n=5). All metrics were successful in distinguishing between all groups (however %V was very similar for all groups). But while in most studies, L2 scores fall between the speakers' L1 scores and the target language natives' scores, Stockmal et al. found that their low-proficiency group presented variability scores that were outside of the L1 Russian - L1 Latvian range, especially for the variability measures (ΔC, ΔV, rPVI-C and nPVI-V).

The authors explain these results by the relationship between speech rate and consonantal measures, arguing that low-proficient learners' slow speech rate and co-articulation issues lead to more consonantal interval variability. As for the vowels, they assume that the high variability is due to a lack of mastery of the Latvian quantity system and a tendency to lengthen vowels in stressed syllables.

However, Ordin & Polyanskaya (2015) propose to explain the results of Stockmal et al. (2005) study in light of universal rhythm acquisition processes. From several studies on the development of L1 speech rhythm in children (Bunta

& Ingram, 2007; Grabe, Gut, Post et al., 1999; Ordin & Polyanskaya, 2014; Payne et al. 2012; cited in Ordin & Polyanskaya, 2015), they observe that no matter the language and its rhythm tendency (syllable vs stress), durational variability systematically increases with age (all together these studies looked at the development of English, German, Spanish, Catalan and French).

Since syllable-time rhythm is associated by a rather low durational variability and stress-time rhythm by a high one, Ordin & Polyanskaya advance that L1 speech rhythm acquisition follow a syllable-time to stress-time progression regardless of the target language.

They extend this trend to L2 rhythm acquisition and cite Stockmal et al.'s study as an example: there is no increase in vocalic interval variability between low to high proficiency L2 speakers because the target language (Latvian) is more syllable-time-like than stress-time-like. This means that going from a stress-timed L1 to a syllable-timed L2, low vocalic variability should show in the first stages of acquisition as a consequence of the universal syllable-time to stress-time progression, and sort of stagnate as proficiency increases since the target language's rhythm presents low vocalic variability.

The absence of increase in vocalic variability between low and more proficient level in Stockmal's study sure goes towards confirming Ordin & Polyanskaya's theory, yet what they fail to mention is that the vocalic variability of the low-level group is actually higher in their L2 than in their L1 (Russian, which is supposed to be stress-time therefore have a high nPVI-V). According to Ordin & Polyanskaya's theory, there should be a dramatic drop in vocalic variability from the L1 Russian to the L2 (whatever it might be), therefore that part of Stockmal's result contradicts the theory they advance.

Nonetheless, Ordin & Polyanskaya have conducted studies looking at the rhythmic progression of learners of English from various L1 (German, Ordin et al., 2011; Italian and Punjabi, Ordin & Polyanskaya, 2014; German and French, Ordin & Polyanskaya, 2015) and found similar progression for all groups from low vocalic or syllabic variability in lower proficiency levels to an increase as learners gain proficiency.

Li & Post (2014) also found results that support the idea of that universal progression tendency in the L2 acquisition of speech rhythm. They compared L1 Mandarin and L1 German learners of English L2 at B1 (low-intermediate) and C1 (advanced)[6] proficiency levels. They applied all normalised rhythm metrics to the speech samples of 10 speakers of L1 Mandarin, L1 German and L1 English, and 5 speakers of L2 English for each L1 and proficiency condition. As can be seen in Figure 11, the progression of the two groups of learners in terms of vocalic durational variability follow the same tendency.

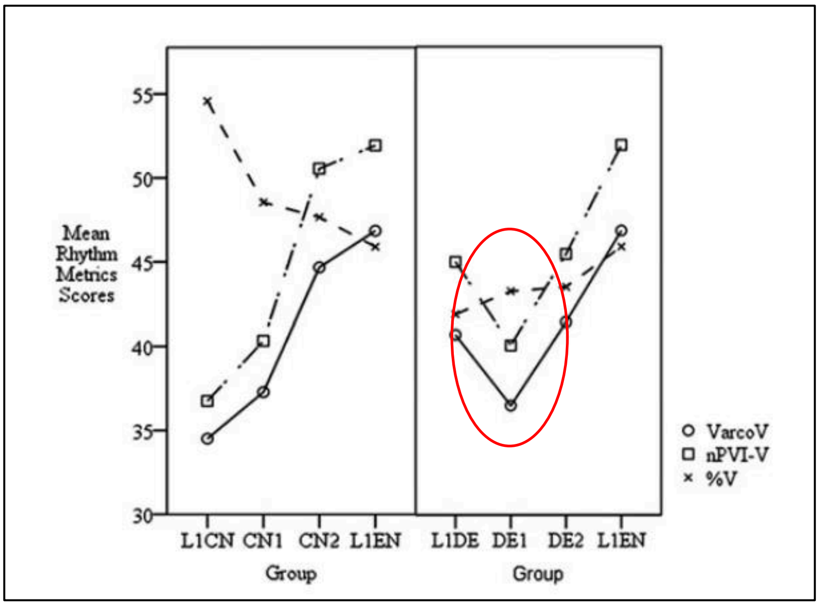


Figure 11 - Vocalic proportion and durational variability metrics per language and proficiency level groups. CN = L1 Mandarin, DE = L1 German, EN = L1 English, 1 = B1 proficiency level, 2 = C1 proficiency level. From Li & Post, 2014, p. 244.

What is striking in this Figure is the drop in the variability scores in the L1 German group between their productions in L1 and B1 level L2 English (circled in red). However, results on the vocalic proportion (%V) in the two L1 groups follow opposite trajectories. The authors conclude that vocalic durational variability seem to develop in a similar fashion (from low to high) beyond L1 typological differences (in line with Ordin & Polyanskaya's theory). Conversely, the vocalic proportion seems to be largely influenced by the L1 which supports a transfer hypothesis.

Giving a comprehensive review on studies using rhythm metrics in L2 speech is beyond the scope of this dissertation. **The inconsistency of the results across studies can be explained by the differences pertaining to the studies' design, such as the number of speakers, the elicitation method, the volume of the corpus, the segmentation into intervals of different nature, and the**

[6] Levels of the CEFR (2020)

languages compared (Gut, 2012). However, when controlling these factors, **rhythm metrics remain relevant to study certain aspects of L2 speech rhythm development. Normalised metrics seem more robust as they eliminate the risk of an influence of the speech rate, and the nPVI has been shown to be especially suited to distinguish between proficiency levels** (Ordin et al., 2011; Li & Post, 2014). For this reason, we have chosen to include a measure of nPVI in the study presented in Chapter V & VI.

As for **the interpretation of metrics' results relatively to L2 speech rhythm acquisition, several studies point towards a universal process regarding durational variability of vocalic and syllabic intervals in the shape of an upward trajectory, consistent with findings from L1 rhythm acquisition studies. However, such studies are too often focused on the acquisition of a stress-time-like language, and only a few of them include speakers at early stages of L2 acquisition. To support the universality of this developmental process, evidence is lacking from design involving the acquisition of a syllable-time-like language, and L2 speakers at an elementary proficiency level.** The study presented in Chapter V & VI of this dissertation - while not solely focused on rhythm metrics - partly addresses this gap.

As exposed in the previous chapter, rhythm metrics alone cannot be taken as representative of speech rhythm in its entirety, rather they constitute one side of the prism. The following section turns to the prosodic side of the prism, what we have called the meso-level of speech rhythm.

2.2. L2 SPEECH RHYTHM AT THE MESO-LEVEL

Studies concerned with L2 speech rhythm from a prosodic perspective have focused mostly on the realisation and distribution of accented syllables and the marking of prosodic boundaries such as utterance final lengthening. As highlighted in the previous chapter, metrical patterns that dictate accents' possible location and acoustic realisation are language-specific, and therefore constitute a major step in L2 acquisition.

Difficulties in the acquisition of L2 prosody can concern different dimensions: the positioning of prominences, pauses, and intonation contours; their phonetic realisation; and the interpretation of linguistic or paralinguistic meanings of such cues (Rasier & Hiligsmann, 2007).

2.2.1. Accent assignment and distribution in L2

Several studies on the acquisition of pitch accent in L2 (mainly English and Dutch as target languages) have shown that learners tend to over-produce (in quantity) pitch accents in L2. This tendency seems to be common to learners of different L1 background: German (Grosser, 1993; Wieden, 1993), Polish and Hungarian (Archibald, 1997), French (Hiligsmann & Rasier, 2002; Rasier, 2003), Spanish (Backman, 1979; Verdugo, 2003).

Some of these studies reveal that the overuse of pitch accent seems to be characteristic of early stages of L2 proficiency (Grosser, 1993; Wieden, 1993) while others have shown the persistence of this phenomenon in advanced L2 learners (Hiligsman & Rasier, 2002; Rasier, 2003). This tendency could be explained by the fact that learners have trouble distinguishing old and new information in an utterance and therefore tend to mark every word (Rasier & Hiligsmann, 2007).

Other authors make the observation that learners tend over-articulate when speaking in L2, defaulting to a syllable-time-like pattern (Barry, 2007). Barry shows that speakers from stress-time languages (Russian, English) and supposed syllable-time language (Korean) all show a reduction of the accented-unaccented ratio in L2 German, deviating from their native pattern. The same phenomenon was found by Gut (2003) and Benkewitz (2003; cited in Barry, 2007). It seems to us that "over-articulation" or "pitch accent over-use" are just two ways of naming the same phenomenon, which nevertheless has been demonstrated by several authors.

However, other studies on the topic obtain contrastive results showing that learners are able to accurately place pitch accent in L2 English (Barlow, 1998; with speakers from L1 Spanish, Italian and Chinese of varying L2 proficiency levels), and that rather than the assignment (presence vs absence of pitch accents), it is the quality, i.e., the accuracy of the phonetic realisation of the accent, that distinguishes between L2 proficiency levels (Frost & O'donnell, 2018). As usual, a vast majority of the literature on the topic concerns the acquisition of the L2 English stress system.

Rasier & Hiligsmann (2007)'s study is one of the rare that looks at the acquisition of L2 French accent pattern. The study focuses on accents at the utterance level and includes 20 L1 Dutch learners of French L2 and 20 French learners of Dutch L2, all at advanced proficiency level. The analysed speech material is elicited from a picture description task, yielding noun phrases in which information status varies (new, given or contrastive). Authors describe Dutch accentuation as being *plastic*, i.e., accent distribution in utterances is heavily influenced by information status (as in other Germanic languages). Conversely French (and other Romance languages) has a *non-plastic* accentuation that is essentially of structural function, and is fixed final[7] (for descriptions of *plastic* vs *non-plastic* languages see Ladd, 1996; Vallduví, 1990). The study compares the acquisition of accents distribution from one language to the other, and authors formulate their hypothesis under Eckman (1977)'s Markedness Differential Hypothesis (MDH, see section 1.2.2., p. 63).

The L1 French and Dutch data collected by the authors confirms that pragmatic constraints (in the form of information status) influence accent distribution in Dutch (words that are not in focus are deaccented), but not in French. Following descriptions of other languages in that regard, Rasier & Hiligsmann advance that pitch accents may be structural only (as in Spanish and Italian), primarily structural and secondarily pragmatic (as in French), or primarily pragmatic and secondarily structural (as in Dutch). However, there seem to be no language in which accents obey to pragmatic rules only. Therefore, under the MDH, pragmatically motivated accents are more marked and consequently difficult to acquire than structural accents.

Results of this study show that indeed, L1 Dutch learners of L2 French produce more L1 French like patterns than L1 French learners of L2 Dutch do. Authors conclude that their result support the MDH. However, because of the speech material used (noun + adjective phrases), the L1 French pattern corresponds to an accentual bridge where both content words were accentuated (e.g., "une **é**toile **jaune**", p. 57), whereas in L1 Dutch the word that was not in focus was deaccented. In L2, both L1 Dutch and L1 French speakers produced the sentences with accents on both words. Authors interpret the results as L1 French speakers transferring their L1 pattern onto L2 Dutch because of their difficulty in acquiring the more marked Dutch pattern. L1 Dutch speakers successfully adopt

[7] This study focuses on the primary accent of French which indeed is final, however we have seen that French also has initial accents (see Chapter I, section 3.3.2., p.34)

the L1 French pattern because it is easier to acquire a less marked pattern (the L1 French pattern) than a more marked one (the L1 Dutch). However, how can we be sure that what is observed in L2 - i.e., accents on all content words - is not the result of the supposed universal tendency towards over-articulation or over-use of pitch accent? From our point of view, the speech material used in this study does not allow to make any firm conclusion regarding the origin of the difference observed between the two L1 groups in L2.

2.2.2. Acoustic realisation of prominences in L2

In a more recent study, van Maastricht et al. (2019) investigated L2 speech rhythm and the interaction between prominence (accentual and boundary marking) and syllable structure. The study crossed data from L1 Spanish, L1 Dutch, L2 Spanish by Dutch speakers, and L2 Dutch by Spanish speakers. Spanish has been described as presenting a majority of simple syllable structures (CV or CVC), while Dutch shows more complexity with syllables that can reach up to 7 segments. Regarding prominence marking, Spanish employs little accentual and final lengthening while Ducth employs them both extensively (Delattre, 1966; Cambier-Langeveld & Turk, 1999; Prieto et al. 2012; cited in van Maastricht et al., 2019).

The authors predict L2 rhythm acquisition under the MDH, and Ordin & Polyanskaya (2015)'s deduction from L1 rhythm acquisition that a syllable-time pattern is less marked than a stress-time pattern because the latter does not exist without the presence of the former. Following the same reasoning, complex syllable structures are also assumed to be more marked than simple ones, and high lengthening ratio are more marked than smaller ratios. Therefore, Dutch is more marked in all these areas than Spanish, implicating that L1 Dutch learners of L2 Spanish should be more successful in reaching the L2 rhythm target than L1 Spanish learners of L2 Dutch.

The study compared L1 Spanish (5 speakers), L1 Dutch (5 speakers), L2 Spanish-L1 Dutch (30 speakers), L2 Dutch-L1 Spanish (30 speakers), each divided into groups of 5 speakers across CEFR levels (A1 to C2). Stimuli were 30 sentences (5 of only CV syllable structure, 5 CVC structure, 20 of mixed structures) for participants to read. Measures of syllable duration according to prominence status (unaccented, accented, nuclear accented) and phrasal position (non-final, ip-final, IP-final) were extracted.

Results show that in L2 Spanish, L1 Dutch proficiency levels groups all progress towards the native target from higher values of syllable duration to lower ones, reaching native-like durations in all prominence and phrasal position conditions for the most advanced learners. Conversely, in L2 Dutch, L1 Spanish speakers present a non-linear progression across proficiency levels, and the most proficient groups (C1 and C2 level) do not get as close to native-like values. Authors conclude that these results support their hypothesis of a greater difficulty for the L1 Spanish to adapt to Dutch rhythm because of its more marked nature (as did Rasier & Hiligsmann, 2007; and Ordin & Polyanskaya, 2015).

However, syllable structure influences the results, especially so in the L1 Spanish-L2 Dutch groups. In fact, none of the proficiency groups differ from native Dutch speakers when syllable structure is controlled (CV only). Authors explain this phenomenon as being similar to that in L1 rhythm development, where difficulties in the articulation of complex consonantal clusters alter the production of the rhythm target (as in Ordin & Polyanskaya, 2014; Payne et al., 2012).

Ueyama (2003, 2016) looks more closely at 2 parameters of accent realisation: f0 and duration, in L1 and L2 English, Japanese, and Italian. From her first study (2003), she showed that prosodic transfer varies according to acoustic parameters, and that the difference in transferred patterns can be associated with contrastive phonological status of the acoustic parameter observed.

The 2016 study looks at f0 and duration of word accents in L1 Italian-L2 Japanese speakers. L1 Japanese word accent is actualised using f0 but not duration, whereas in L1 Italian and L1 English both f0 and duration are used. L1 Japanese, L1 Italian, L1 Italian-L2 Japanese at beginner level, and L1 Italian-L2 Japanese at intermediate level are compared (with between two and four speakers per group). Participants read carrier sentences containing three target minimal pairs of homophonous words differing only in word accent. In L1 Japanese a high pitch is produced on accented syllables with no difference in duration between accented and unaccented syllables. In L1 Italian, accented syllables receive a low pitch (contrasting with a preceding rise) and longer duration than unaccented syllables.

Results show that L2 Japanese groups of both proficiency level successfully assign higher pitch to accented syllables in L2 Japanese. This suggests the absence of a negative transfer from L1 Italian to L2 Japanese. However, the pitch rise is realised inaccurately in terms of ratio with either too much or not enough f0 contrast between accented and unaccented syllables (the direction of the difference does not correlate with the proficiency levels). However, both L2 groups show a durational contrast between unaccented and accented syllables, which indicates the presence of a negative transfer from the L1. Yet, the durational ratio in L2 is reduced in comparison with L1 Italian. As observed by Frost & O'Donnell (2018), Ueyama's results show that L2 learners have more difficulty in the phonetic realisation of accents than at the phonological level (presence or absence of accent). In addition, her work shows no clear influence of the proficiency level on accent realisation, although the limited number of speakers per group prevents from generalising.

Similar observations were made by Trofimovitch & Baker (2006), who looked at the influence of L2 experience (length of residence in the target language country) on L2 suprasegmental production. Native Koreans learners of L2 English separated into 3 groups according to length of residence (3months, 3 years, 10 years; 10 speakers in each) were compared with L1 English speakers, on 6 sentences elicited with a delayed repetition task. Unstressed to stressed syllable duration ratios were calculated. L1 English ratio is of .5 on average, whereas L1 Korean ratio is closer to 1 (i.e., no durational difference between unstressed and stressed syllables). The two less experienced groups presented higher ratios (but below .6) and differed significantly from the English natives, while the more experienced group did not. This result suggests that the acquisition of the English stress pattern increases as a function of length of residence.

However, authors found contrasting results on the alignment of pitch peak. In L1 English, maximum pitch value in an intonational phrase necessarily falls on a stressed syllable, aligns with its onset, and carries pragmatic weigh (indicating the most important word). Conversely in Korean, pitch peak marks the boundary of an accentual phrase and falls on the offset of the syllable in the last word of an accentual phrase. Results in L2 English showed that all L2 groups were aligning pitch peak inaccurately and groups did not differ significantly, suggesting that experience does not have any effect on peak-alignment accuracy.

As in Ueyama's studies (2003, 2016), these results show that different rhythm correlates evolve in different ways through the L2 acquisition process, and the phonetic level seem to be more problematic than the phonological one.

In addition to studying L2 rhythm through rhythm metrics (see previous section, p. 72), Li & Post (2014) also included in their study measures of duration of accented and phrase-final syllables. The study included L1 German, Mandarin, and English data, as well as L2 English data from the L1 German and Mandarin speakers, both divided into two proficiency level group: low-intermediate and advanced. In L1 German and English, accented syllables are lengthened, though in smaller proportion in German (Delattre, 1966). Conversely, L1 Mandarin does not have accentual lengthening. All three languages present phrase-final lengthening, English and German have similar lengthening ratio, while Mandarin's is distinctively smaller.

Results in L2 English show similar progressions for both L1 groups in terms of accentual and phrase-final lengthening. Low-intermediate learners lengthened accented and phrase-final syllables less so than advanced learners, who themselves presented smaller durational increase than native English speakers (although the difference was not significant between L2 advanced and native English speakers).

The similarity between both L1 groups suggests universal mechanisms. The low-intermediate L1 Germans, who have accentual lengthening in their mother tongue - albeit in smaller proportion than L1 English - present similar accentual lengthening in L2 English than in their L1, which suggests a transfer from L1 German. Yet, low-intermediate L1 Mandarin speakers, who do not have accentual lengthening in their L1, show a similar lengthening ratio in L2 English than low-intermediate L1 Germans. This observation leads to consider the hypothesis of a universal process where learners - no matter their L1 - apply a default lengthening ratio to accented syllables in the early stages of L2 acquisition.

Results obtained on these prosodic features tally with some rhythm metrics, suggesting that certain metrics (%V mainly) do capture prosodic features.

Research still needs to be developed in the area of prosodic rhythm acquisition in order to distinguish universal processes from L1 transfers. In addition, studies involving the acquisition of other L2s than English, and spontaneous speech samples still lack.

In the following section, we present studies focused on temporal variables (macro level) and fluency measures in L2.

2.3. MACRO-LEVEL & FLUENCY IN L2

As highlighted in Chapter I (sections 3.3.4., p. 39), fluency studies emerged largely from a psycholinguistic perspective, where it is seen as the expression of speech planning and production procedural knowledge and automaticity (Goldman-Eisler, 1968). Within this framework, fluency is mostly understood in its narrow sense i.e., aspects of speech production that relate to fluidity, smoothness and speed (Lennon, 1990). Most studies on L2 fluency in the past 40 years have focused on the relationship between L2 utterance and perceived fluency, and L2 fluency and proficiency level. Authors have compared fluency correlates with the aim to isolate the most reliable predictors for proficiency, accentedness, comprehensibility and intelligibility; with potential direct application in L2 assessment of oral competency. Given the tremendous amount of research related to L2 fluency, this section presents a selection of the most influential and recent work on the topic.

2.3.1. L2 utterance fluency and perceived fluency

A large number of studies on L2 fluency have focused on the relationship between utterance fluency and perceived fluency, with the aim of discriminating which speech features would best predict fluency ratings, and how these features develop with L2 experience.

One of the most commonly cited early(-ish) study on L2 fluency development is the one reported by Lennon (1990). The author was interested in finding *objective* correlates of fluency in the speech of L2 English speakers, as indicators of what listeners base their fluency judgements on. In other words, Lennon was not interested in the underlying processes causing such and such feature in speech production, but rather sought to find those features that correlate with perceived fluency.

The author recorded four L1 German-L2 English learners at advanced proficiency level at the beginning and at the end of a six-month stay in England. The pre and post-test consisted of a story retelling task based on prompt images. Nine native English speakers and English as a Foreign Language (EFL) teachers rated the fluency of the L2 speakers on the pre and post-test recording. For each speaker, a majority of the raters assigned higher values for the post-test recordings than for the pre-test recordings, suggesting that all speakers had improved their fluency.

He then examined 12 potential fluency indicators belonging to all three categories of speed, breakdown and repair, in the speech of the L2 speakers. Results showed that speakers increased their pruned speech rate, produced less filled pauses and repetitions but not fewer self-corrections, made shorter silent pauses, and longer uninterrupted runs between pauses. The main findings are that the most significant indicators of improved fluency are the increase of the pruned speech rate and of the run length (uninterrupted speech between two pauses), and the decrease of the number of filled pauses, repetitions and silent pause time. Self-correction however does not come out as a good indicator of fluency. Lennon also observed that the between-speaker speech rate variation was in fact related to the pause time variation rather than the speed of articulation, hence the usefulness of extracting the articulation rate in order to isolate the speed of articulation proper.

Even if the small sample size limits the generalisation of these findings, several later studies (partially) support them (Ejzenberg, 2000; Freed, 1995, 2000; Riggenbach, 1991; Towell et al., 1996).

In another influential study, Kormos & Dénes (2004) also investigated the correlation between several fluency measures and fluency ratings of native and non-natives speakers, teachers of the target language. They compared two groups of L1 Hungarian learners of English, 8 speakers at an advanced proficiency level, and 8 at a low-intermediate proficiency level. They were recorded on a narrative task based on cartoon prompts. Three L1 Hungarian listeners, teachers of EFL and three native speakers of English also EFL teachers were asked to rate the fluency of each L2 speaker on a scale from 1 to 5. Recordings were transcribed and 10 measures of utterance fluency were extracted: speech rate, articulation rate, phonation/time ratio, mean length of runs (number of syllables between pauses > 250ms), number and length of silent pauses > 200ms per minute, number of filled pauses per minute, number of disfluencies per minute (repetition, restart and repair), pace (number of stressed words per minute), and space (proportion of stress words over the word total).

Results showed that the best predictors of the fluency ratings by both groups of listeners were the speech rate (higher for the advanced group), the mean length of run (longer for the advanced group), and the pace (higher count for the advanced group). In contrast with Lennon's findings, the number of filled pauses and the number of disfluencies (although these included self-correction) did not correlate with the perceived fluency scores. The authors also found that pace, speech rate, phonation/time ratio, and mean length of run (MLR) and pause were strongly correlated. Number of filled and silent pause and disfluency also clustered together but were not as strongly correlated with fluency ratings.

The findings of this study indicated that both native and non-native EFL teachers relied on the same pool of speech features to rate the fluency of L2 speakers. It also introduced pace as a strong correlate of perceived fluency. However, one can wonder to what extent this correlation can hold in L2 languages other than English.

A recent meta-analysis conducted by Suzuki, Kormos, & Uchibara (2021) provides an insightful account of 22 fluency studies carried out between 1994 and 2020. Unsurprisingly, 15 of them concerned L2 English, while three were on L2 Dutch, two on L2 French, one on L2 Spanish, and one on L2 Japanese. This goes to show that a large majority of data available on L2 fluency concerns L2 English. Therefore, findings should be interpreted carefully as they cannot be generalised to L2 acquisition, but solely to the acquisition of L2 English. This motivated the choice of conducting our study (see Chapter V & VI) on the acquisition of L2 French.

Susuki, Kormos, & Uchibara's analysis reports that previous studies have shown that perceived fluency is primarily associated with speed delivery and pause pattern (Saito et al., 2018; Suzuki & Kormos 2020), the contribution of disfluencies however differs across studies.

Several studies have shown that perceived fluency is most strongly associated with speed measures, less so with breakdown measures (Bosker et al. 2013, Kormos & Dénes 2004), conversely, other studies have shown the opposite with breakdown measures more strongly correlated with perceived fluency than speed measures, especially when studies considered pause location relatively to clauses (Cucchiarini, Strik & Boves, 2002; Suzuki & Kormos, 2020). Such studies showed that mid-clause pauses quantity were strong predictors of perceived fluency. It also seems that silent pauses measures are better indicators then filled pauses (Bosker et al. 2013; Cucchiarini, Strik & Boves, 2002; Suzuki & Kormos 2020). Stronger correlations are found with repairs when they are observed separately rather than all together.

Composite measures such as speech rate and MLR strongly correlate with perceived fluency because they actually capture multiple utterance fluency phenomena, however they do not enable to finely analyse which specific feature might play the biggest role.

Also, the extent to which utterance fluency measures are correlated with fluency ratings varies across studies, most likely because of methodology differences such as languages, speech task, sample size, and listeners characteristics and rating methods. The meta-analysis looked at aggregated effect sizes to observe the overall impact of fluency measures on fluency ratings. All measures were significantly associated with perceived fluency; MLR and speech rate (both composite measures) showed strong effects, articulation rate a little less, then pause frequency, pause duration, and finally disfluency rate.

Medium to strong effect sizes were observed in L2 Dutch, English and French suggesting that the relationship between utterance (through the selected measures) and perceived fluency is fairly stable across L2s. Japanese actually had the strongest effect size, maybe because of its mora-timed nature which implies less rhythmic variation, and therefore maybe less effects of suprasegmentals on fluency ratings which, in turn, rely even more on fluency.

Finally, stronger effect sizes were found in controlled task vs spontaneous speech, most likely because in controlled speech, there is no linguistic variation that could influence ratings.

2.3.2. L2 French fluency studies

Since our work is mainly concerned with the acquisition of L2 French, let us zoom-in onto the findings specific to the L2-French studies included in that meta-analysis.

Préfontaine, Kormos, & Johnson (2016) recruited 40 adult learners of L2 French participating in a 5-week immersion program at a university in Quebec. Participants were all native speakers of English (Canadian, American, and British), at beginning, intermediate, and advanced proficiency levels. They were recorded on three narration tasks elicited from different prompts (random pictures, retelling a text, cartoon strip). 11 native French teachers of French rated the recordings and assigned a proficiency level (one of the six CEFR levels), and rated their speed and breakdown fluency separately, on 6-point scales. Articulation rate, mean length of run, quantity and average duration of silent pauses >250ms were extracted.

The authors found that mean length of run and articulation rate strongly influenced both fluency and proficiency ratings, supporting previous studies' findings (such as Towell, Hawkins, & Bazergui, 1996 on L2 French). However, while we have seen that the length of silent pauses tends to have a weaker impact on fluency ratings, this study reports a strong effect of this variable. Moreover, they found that the longer the pauses, the higher the fluency score. This makes sense in light of findings regarding pause pattern in L1 French (Grosjean & Deschamps, 1975; Judkins et al., 2022), which showed that French speakers make less but longer pauses than L1 English speakers. Finally, the quantity of pauses was found to be the weakest predictor of fluency ratings. However, the authors found that their results slightly varied across tasks.

In another study conducted in Quebec, Trofimovich, Kennedy, & Blanchet (2017)[8] investigated the progression of 30 L2 French learners (intermediate proficiency level), before and after a 12-week speaking and listening course. Learners came from a diversity of L1 (Mandarin, Russian, Farsi, Spanish, Portuguese, Cantonese, Korean, Malay, and Romanian). They were recorded on a reading task, and a picture-based speaking task. The authors found that fluency ratings correlated once again with the mean length of run, but only in the reading task. In the picture-based speaking task, the number of hesitations (i.e., all pauses - no duration threshold specified) was correlated with fluency ratings while the mean length of run was not, which contradicts previous studies' results (most notably, Préfontaine et al., 2016). However, the mean length of run was measured as the number of syllables between silent or filled pauses > 400ms which differs from Préfontaine et al.'s study where a 250ms threshold was applied.
Interestingly, the authors also included a measure of intonation errors (intonation contours judged "unnatural" by expert raters) which they found to correlate with fluency ratings in both tasks. This suggests that prosodic features contribute to the listeners' impression of fluency.

All in all, **substantial empirical evidence supports the correlation between utterance fluency and perceived fluency. However, the impact of each individual measure is subject to variation across language pairs, tasks, and methodological choices** (thresholds for pauses, listeners' characteristics etc.). Finally, studies have shown that listeners' perception of fluency can be influenced by other factors than temporal variables, such as intonation errors (Trofimovich et al., 2017).

[8] It should be noted that the aforementioned L2-French studies concern Canadian French. Yet, it has been shown that Canadian and European French differ with regards to segmental and suprasegmental aspects (Guilbaud, 2002; Bissonnette, 2003). Therefore, findings from these studies might not be directly comparable to that of European French studies such as ours (presented in Chapter V & VI).

2.3.3. L2 fluency and proficiency level

While a large majority of L2 fluency studies have focused on the link between utterance fluency and perceived fluency, fewer have directly addressed the relationship between utterance fluency measures and L2 overall proficiency. The link between L2 fluency and oral proficiency has been demonstrated by studies looking at both utterance and perceived fluency. Since oral proficiency is exclusively assessed by humans (although these studies have paved the way for automatic measures of fluency for assessment), and even though other factors than temporal features influence the raters' judgements (accuracy, syntax), utterance fluency plays a major role.

However overall proficiency is usually assessed through standardised tests such as IELTS and TOEFL for English and DALF or DELF for French, and obviously includes all linguistic competence (vocabulary, grammar etc.). Because utterance fluency reflects the degree of automaticity in language processing and fluidity of access to language structures and forms (Segalowitz, 2010), it is fair to assume that the more fluent the speech, the more proficient the speaker. In fact, Segalowitz & Freed (2004) and de Jong, Steinel, et al. (2013) support the hypothesis of a link between cognitive and utterance fluency.

In their study, Baker-Smemoe et al. (2014) collected 126 samples from an oral proficiency interview in an official language test (ACTFL). They belonged to native English speakers learners of French, German, Japanese, Russian, and Arabic. Samples were first rated by ACTFL certified raters, and assigned a proficiency level. 80 of the 126 interviews were pre and posttests, 6 months at least apart, from students who either went abroad of followed a language course. Levels of proficiency in the 126 samples varied from novice mid to superior, with a majority concentrated between intermediate-mid and advanced-mid. Samples analysed for fluency measures were 20 seconds long, taken after three minutes of interview.

Results showed that speech rate, mean length of run (MLR), number of runs, number of pauses, and length of pause significantly differed between all proficiency level groups, regardless of the L2. However, the effect of the L2 on utterance fluency measures was tested separately and showed that the MLR, the number of runs, and the number of hesitations significantly differed in L2 French, German, and Arabic, while the proficiency level as a covariate was not significant. It was found that L2 German speakers produced significantly less hesitations than the two other L2 groups, and that L2 Arabic speakers made significantly less and shorter runs than L2 German and L2 French speakers.

This study's findings show that **the composite measures of speech rate and MLR, as well as the number of pauses are relevant indicators of overall proficiency**, especially at higher proficiency levels (advanced and superior). This study also highlighted the fact that **fluency measures differed according to the L2**. The authors explained these differences by the varying degree of learning difficulty of the languages involved (for English native learners), as determined by their degree of similarity (for native English speaker, German is easier to acquire than French, than Arabic). Therefore, findings regarding **fluency development in a pair of language cannot be generalised to other language pairs**.

Similarly, Saito et al. (2018) found articulation rate to be the best predictor of proficiency, followed by pause frequency. In addition, listeners' reliance on repairs (repetition and self-correction) to rate fluency were found inconsistent. This study included 90 L1 Japanese learners of L2 English at three proficiency levels as indicated by their length of residence (LOR) in Canada (10 inexperienced: 0 years; 40 experienced: 0.1 to 5 years; 40 attainers: 6 to 18 years). The speakers were recorded on a picture-description task and rated for perceived fluency by 10 English native listeners on a 9-point scale. It was found that proficiency as a function of LOR correlated with perceived fluency scores.

In L2 Dutch, de Jong & Bosker (2013) found that the number of pauses >250ms correlated with proficiency levels, as indicated by a vocabulary knowledge score. However, the mean length of pauses was not related to proficiency, while it was to fluency ratings. The speech samples came from 27 speakers of L1 English and 24 of L1 Turkish, recorded on eight spontaneous speech tasks (formal/informal, persuasive/descriptive, simple/complex).

Recently, Tavakoli, Nakatsuhara, & Hunter (2020) conducted a study to investigate how much the different aspects of fluency (speed, breakdown, repair, composite) could differentiate between CEFR levels, once again in L2 English. The authors used samples from a standardised, computer-based speaking test (Aptis Speaking test from the British Council) in which test-takers answer questions spontaneously. 32 participants were selected based on their test ratings provided by expert, so as to obtain groups of eight participants from four CEFR levels (from

A2 to C1). Results showed that articulation rate, speech rate, and mean length of run distinguished between A2 and B1, and between these levels respectively and the higher levels. However, B2 and C1 did not differ, which suggested that these speed and composite measures are subject to a ceiling effect. Additionally, the length of silent pauses distinguished A2 from higher levels but no difference was found between these higher levels. A linear relationship between fluency and proficiency was found for this measure: silent pauses length decreased from A2 to C1. However, the length of filled pauses did not distinguish between any levels, which is consistent with the results found by Segalowitz et al. 2017 (cited in Tavakoli et al., 2020).

As for pause frequency, the number of silent pauses distinguished between A2-B1 and B2-C1, and only for pauses in mid-clause position. The number of filled pauses distinguished between A2 and other levels, and surprisingly, was greater in higher proficiency levels (B2 and C1). In a similar fashion, repairs (false start and reformulation) distinguished A2 and B1 level groups, and the latter, in fact, used more repairs than all other levels. B2 and C1 levels used more repairs than the A2 group, but less than the B1.

Essentially, this study once again supports the fact that **speed and composite measures are strong predictors of L2 proficiency** as they tend to increase in a linear fashion from lower to higher levels. Similarly, silent pauses length progressively decreases as proficiency increase, and longer pauses are characteristic of the A2 level. However, filled pause and repairs are more frequently used by B1 level speakers and, to a lower extent, B2 and C1; whereas they are almost absent in A2 speakers. This suggests that the use of such strategies develops with proficiency in a non-linear fashion.

Overall, all measures discussed above are reliable to distinguish the elementary level (A2) from the others, but their distinctive power is more moderate in regards to higher proficiency levels (especially between B2 and C1 where composite and speed measure seem to reach a ceiling).

However, an important limitation of Tavakoli et al.'s study is the absence of data on the speakers' L1. As a matter of fact, it has been shown that L2 fluency patterns tend to be influenced by L1 fluency patterns. Therefore, it is impossible to know if the tendencies observed by the authors are solely the result of L2 behaviour, or if they can be explained by a transfer of L1 patterns. The next section provides a discussion on the relationship between L1 and L2 fluency.

2.4. THE RELATIONSHIP BETWEEN L1 AND L2 FLUENCY

L2 fluency studies, such as those presented in the previous section, have for the most part considered L2 fluency separately from L1 speech production. However, just like for any aspect of L2 speech, the question of the role of L1 transfer and universal processes still stands. Although studies on the relationship between L1 and L2 fluency have emerged mostly in the last 15 years, the topic is not exactly new. Raupach (1980; cited in Tavakoli & Wright, 2020) already highlighted the fact that patterns found in L2 and considered as disfluent were, in fact, also found in L1 speech; namely filled and unfilled pauses and repairs.

Ultimately, fluency patterns - whether in L1 or L2 - are not only shaped by cognitive fluency, but also by the individual's performance "style"; as such, fluency can be seen as an individual trait (Raupach, 1980; Segalowitz, 2010; Towell et al., 1996).

For instance, if we pay attention to this in our daily L1 interactions, we will certainly notice that some people tend to use a lot filled pauses while others barely do; some use repetitions extensively while others make long silent pauses etc. For the sake of the example, let's say that the person who uses repetitions in L1 also does in L2. By looking only at their L2 productions, we might conclude that this person is very disfluent, however as soon as we take into account the fact that this pattern is also present in their L1 speech, we must adapt our interpretation.

This section presents findings from empirical studies on the relationship between L1 and L2 fluency, and how it can be nuanced by other factors such as the proficiency level, or the language pair.

2.4.1. Moderating role of proficiency and language pair

Derwing et al., (2009) conducted a longitudinal study of two immigrant populations in Canada over a period of two years. 16 native speakers of Slavic languages (Russian and Ukrainian) and 16 native speakers of Mandarin were recorded on a cartoon-based narrative task in their L1 and in L2 English four months after their arrival in Canada, then 10 months later, and finally a year later.

Several fluency measures were extracted from the recordings, and native listeners rated their fluency. Results showed that there was indeed a correlation between L1 and L2 speech rate, pruned speech rate and number of silent pauses (>400ms), especially at early stages of exposure. A stronger correlation was observed in the Slavic group, presumably because of the closer structural properties of their L1 with English. Towell & Dewaele (2005) had also found a stronger correlation between L1 and L2 speech rate in their learners of L2 French before a study-abroad period; the L1-L2 correlation weakened when testing the participants after they had returned.

These findings suggest that the relationship between L1 and L2 fluency evolves with L2 proficiency development, and can differ depending on the language pair. However, Derwing et al.'s study did not include measures of repairs, nor data from high L2 proficiency levels.

Partly addressing these limitations, De Jong et al., (2015) examined the relationship between L1 and L2 fluency using the same speech data as in De Jong & Bosker (2013; see previous section). The speakers were L1-English and L1-Turkish at intermediate and advanced levels (B1 to B2) in L2 Dutch. The authors found positive significant correlations between all L1 and L2 fluency measures.

Speed was measured by the mean syllable duration (reversed articulation rate), breakdown measures included silent pause duration and number, and filled pause number, repair measures included number of repetitions and number of corrections. Contrary to Derwing et al. (2009) and Towell & Dewaele (2005)'s results, in this study, the smallest correlation coefficient was that of the speed measure (0.37), whereas breakdown and repairs all had higher coefficients (from 0.6 to 0.76). Although, this difference might be explained by the use of a composite measure (speech rate) in Derwing et al. and Towell & Dewaele, vs a pure measure (reversed articulation rate) in the currently discussed study. Differences between L1 Turkish and L1 English also emerged: syllables were longer in L1 English, L1 Turkish produced fewer silent pauses and fewer repetitions. However, between groups differences disappeared in L2 Dutch, suggesting that these L1 patterns did not transfer over to the L2.

The originality of this study lies in the fact that the authors tested the predictive power of L2 fluency measures on proficiency levels in their *raw* form (no transformation after extraction, like in all studies), and also in a *corrected* form, where L2 fluency measures were shaved off of L1 behaviour by partialing out the L1 variance from the L2 measures. Results showed that for the measure of syllable duration, the corrected version better predicted the proficiency level than the raw version. However, for all other measures, both raw and corrected measures equally predicted the proficiency level. Yet, the authors advocate for the use of corrected measures for more precise models of prediction of L2 proficiency from fluency measures. Finally, it was found that the typological difference in L1 (Turkish vs English) did not impact the relationship between L1 and L2 fluency, meaning that the results might be generalised to other language pairs.

Contrastively, Huensch & Tracy-Ventura (2017) found that variation in L2 fluency was influenced by L1 fluency behaviour, cross-linguistic differences, and proficiency level. In their study, L1 English learners of Spanish (n = 24) and French (n = 25) recorded a picture-based narration task before (T1) and after (T2) a 5-month stay in a country speaking the target language. Proficiency was measured by means of an Elicited Imitation Task (EIT) at T1 and T2, and a significant improvement was found for both L2 Spanish and L2 French groups.

First, the authors looked at fluency measures in each language in L1. They found that L1 English, Spanish and French differed significantly in mean syllable duration, number of filled pauses, number of repetitions, number of corrections, but no difference was found on measures of silent pauses (both quantity and duration), which is in contradiction with De Jong et al. (2015), and Peltonen (2018).

Unsurprisingly, English had longer syllables than French, and French than Spanish. This reflects differences in syllable structures and supports other studies findings regarding speech rate differences across languages (de Jong et al., 2015; Roach 1998; cited in Huensch & Ventura, 2017). However, contrary to Grosjean & Deschamps (1975), no difference was found in pause frequency between L1 English and L1 French, although the task used differed. Concerning repairs, L1 Spanish produced less filled pauses and repetitions than English, and English used more repetitions than Spanish and French. It was found that the target language variable had a significant impact on these same fluency measures in L2 at T2.

Correlation tests between L1 and L2 fluency measures revealed that in the L2 Spanish group, syllable duration, silent pause number and duration were significantly correlated at both T1 and T2. In the French group, syllable duration and number of silent pauses also correlated at T1, and at T2 five more measures were found to correlate significantly. This suggests that the relationship between L1 and L2 fluency is influenced by both the L2 (French vs Spanish), and the proficiency level (since there was improvement between T1 and T2). These results are also in partial contradiction with those of De Jong et al. (2015) who found significant correlations for all measures.

On the whole, this study showed that the weight of each factor changed over time. L1 fluency had a significant effect on all L2 fluency measures at T2, as opposed to only a couple (mean syllable duration and number of silent pauses) at T1. L2 target language had a significant effect at T2 on measures for which native speakers were shown to be different. However, proficiency was only significant on mean syllable duration at T1.

The effect of proficiency on the relationship between L1 and L2 fluency was the focus of the very recent study conducted by Suzuki & Kormos (2024). In contrast with previous studies which operationalised proficiency as either longitudinal change (Derwing et al. 2009, Huensch & Tracy Ventura ,2017), or vocabulary size (De Jong et al., 2015), Suzuki & Kormos used a fine-grained measurement of L2 proficiency. Participants were tested both in the dimensions of linguistic knowledge (vocabulary, syntax structures), and processing speed (lexical retrieval, syntax encoding, articulatory speed), which together define cognitive fluency.

The study involved 104 L1 Japanese learners of L2 English between B1 and C1 CEFR level as indicated by the university they were recruited at. They were recorded on argumentative tasks where they gave their opinion on diverse societal topics in L1 and L2. Measures of speed, breakdown, and repair fluency were extracted.

Results showed that all L1 and L2 fluency measures were significantly associated. L1 fluency had a strong predictive power (as indicated by effect sizes) over L2 fluency for mean length of run, filled pause ratio, and repetition ratio; while for articulation rate and other types of repairs the effect of L1 was negligible. This contrasts with Huensch & Ventura's results were L1-L2 association was especially strong on speed and breakdown measures (silent pauses). The authors argue that the typological difference between Japanese and English might be in part responsible.

Regarding the role of proficiency as operationalised as linguistic knowledge scores and processing speed scores, the authors found that both significantly affected the L1-L2 fluency relationship but only on articulation rate, speech rate, and mean length of run. Interestingly, the higher the linguistic knowledge score was, the greater the articulation rate in L2, and the more it was dissociated from the articulation rate in L1. Conversely, the L1-L2 association was enhanced as a function of processing speed scores, showcasing the link between speed of processing and speech delivery.

2.4.2. L1 English and French macro-level patterns

In a previous study (Judkins et al., 2022)[9], we investigated the difference between L1 English and L1 French macro-level (i.e., Inter Pausal Units - IPU, and pauses) and fluency patterns, and their effect on L2 spontaneous speech.

The speech samples used for this study were selected from a bilingual corpus created originally for research on bilingual vocal register variations (corpus B-FREN3, Drouillet et al., 2023). Spontaneous speech samples collected from a conversation task belonged to 6 L1-English-L2-French speakers and 6 L1-French-L2-English speakers. Their proficiency level ranged from A2 to B2 from the CEFR. Participants were recorded both in their L1 and their L2 in similar experimental conditions. **The idea was to be able to conduct a within-speaker comparison in order to gage the difference between L1 and L2 productions, but also a between-group comparison (L1 English vs L1 French groups) to look at the potential difference between the two L1, and see how this difference could nuance the L2 productions.**

120-second excerpts taken from the middle part of each of the 24 recordings (12 participants x 2 languages) were annotated. Measures extracted included the length and quantity of IPU, respiratory IPU-external pauses >250ms, and IPU-internal voiced pauses (which also included lengthenings).

Results revealed that the L1-L2 comparison (within-speakers) was consistent with previous studies that found more and shorter IPU and voiced pauses (Fauth & Trouvain, 2018; Kaglik, 2009), as explained by a reduced cognitive fluency in L2. However, the between-group comparison confirmed that L1 English and L1 French differ in regards to those measures, and between-group differences were also found in L2.

In L1, it was found that L1 English produces more and shorter IPU, longer respiratory external pauses, and less and longer voiced pauses than L1 French. For the most part, these results are consistent with Grosjean & Deschamps (1975). In L2, while the IPU quantity and length follow similar tendencies across groups, the

[9] As a reminder, Judkins is my former last name therefore I (Lucie Drouillet) am the first author of this publication.

L1 English group produced shorter external pauses while the L1 French group produced longer ones. Regarding voiced pauses, the L1 English group greatly increases the quantity of voiced pauses without much change in duration, while the L1 French group shows the opposite with a small difference in quantity but a significant increase in duration.

On the whole, **this study showed that French and English differ in macro-level and fluency patterns, and that these L1 habits seem to affect outcomes in L2.** However, one important limitation of this study is that the effect of the speech rate was not taken into account.

Working on a bi-directional corpus such as B-FREN3 allows for within-subject and between-group comparison which enables the observation of L2 patterns in light of L1 habits and L2 targets.

2.4.3. Concluding remarks

To conclude this part, we will draw upon the conclusions raised in Gao & Sun (2024)'s meta-analysis of L1-L2 fluency studies. The authors collected results from 16 L1-L2 fluency studies between 2007 and 2022. **It emerges clearly that L2 fluency cannot be taken as an L2 only phenomenon, but rather, it is strongly influenced by L1 fluency patterns. The quantity and duration of between-speech-unit pauses in L2 is the feature most correlated to L1 patterns. Overall, breakdown measures are strongly correlated to L1, speed measures moderately so, and repair measures weakly.** Gao & Sun also concluded that L1-L2 relationship tend to be stronger in later stages of proficiency (Huesch &Tracy-Ventura, 2017; Gagné, French, & Hummel, 2022); however, we have seen that Suzuki & Kormos (2024), Derwing et al. (2009), and Towell & Dewaele (2005) obtain contradictory results.

Gao & Sun also looked at the moderating effects of different methodological choices and concluded that **the L1–L2 fluency relationship appeared to be more pronounced when participants learned the target language in a study-abroad context as opposed to in-class instruction, when an "open" task was used rather than a closed one (e.g., free speech vs reading), and when different prompts were used in the L1 and L2 tasks** (rather than the same elicitation topic for instance).

Even though the effect of individual fluency traits present in the L1 is more and more included in L2 fluency studies, we have noticed that **the variability between two test times in L1 is never looked at in relation to the same variability in L2**. L1 fluency is systematically taken as a set of measures at a single point in time, however just like any other aspect of speech, it is highly likely to be subject to a certain amount of intra-subject variability from one time to another. So, in addition to the role L1 plays in L2 patterns, when employing a multiple test-time design, we could obtain even more accurate measures that are also shaved off of the L1 between-time variability. Even though we were not able to perform such transformation on our data, we raise this question by looking at T1-T2 variability in L1 in comparison to that in L2 (see Chapter VI, section 1.3., p. 248).

CHAPTER SUMMARY

We started this chapter by presenting language production models in L1 and L2. We saw that each stage of the process (conceptualisation, formulation, articulation) is prone to difficulties for the L2 speaker. Disfluencies might arise as the monitoring component detects errors or divergences from the speaker's original intention. Problems at the conceptualisation phase might be related to an inadequacy between the intended concept and its expression in L2. The formulation stage in L2 might require a lot of attention and conscious retrieval of the different linguistic aspects involved, which might slow down the production process and generate ruptures and disfluencies in the signal, and the same goes for the articulatory phase where L1 habits remain tenacious. According to Kormos' (2006) view, prosodic encoding also intervenes at each major stages of the process.

The overview of different L2 acquisition models showed that authors, for the most part, consider the role of both universal acquisition processes, and L1 transfer. Clearly, the acquisition of L2 phonology is not a monochrome process, and the effect of different factors intertwine in a dynamic fashion. L1-L2 similarity is also a crucial aspect that most models rely on to determine areas of difficulty. These models help to build hypotheses regarding the origin of L2 speech phenomenon, and there is an exciting and growing body of research around the acquisition of suprasegmentals specifically (e.g., Li & Post, 2014; Rasier & Hiligsmann, 2007; Sánchez-Alvarado, 2022). However so far, studies have focused on one, sometimes two levels of analysis, but the acquisition of speech rhythm has yet to be discussed in a comprehensive manner that would include micro, meso, macro levels and fluency together.

The second half of this chapter was dedicated to a literature review of experimental studies focused on L2 speech rhythm. From research on micro-level rhythm, we noted that normalised metrics seem to be more robust as they eliminate the risk of an influence of the speech rate. Specifically, the nPVI has been shown to be especially suited to distinguish between proficiency levels (Ordin et al., 2011; Li & Post, 2014). For this reason, we have chosen to include a measure of nPVI in the study presented in Chapter V & VI. In addition, several studies point towards a universal process regarding durational variability of vocalic and syllabic intervals in the shape of an upward trajectory, consistent with findings from L1 rhythm acquisition studies. However, such studies are too often focused on the acquisition of a stress-time-like language, and only a few of them include speakers at early stages of L2 acquisition. Therefore, evidence is lacking from designs involving the acquisition of a syllable-time-like language, and L2 speakers at an elementary proficiency level. The study presented in Chapter V & VI of this dissertation - while not solely focused on rhythm metrics - partly addresses this gap.

Research focused on meso-level rhythm highlight what looks like a universal phenomenon in the tendency to over-distribute pitch accents in L2, along with an over-articulation phenomenon which reduces the durational variability between accented and non-accented syllables (Barry, 2007; Li & Post, 2014; Rasier & Hiligsmann, 2007; Verdugo, 2003). It seems like the fine details of the acoustic realisation of prominences is particularly problematic, and dependent on proficiency levels (Frost & O'donnell, 2018; Ueyama, 2003, 2016).

Turning to macro-level rhythm and fluency, we saw that fluency measures are relevant indicators of perceived fluency. However, the impact of each individual measure is subject to variation across language pairs, tasks, and methodological choices (Suzuki et al., 2021), and listeners' perception of fluency can be influenced by other factors than temporal variables, such as intonation errors (Trofimovich et al., 2017). Fluency measures most strongly correlated to proficiency levels are composite measures, speed measures, and breakdown measures (Baker-Smemoe et al., 2014; Saito et al., 2018; Tavakoli et al., 2020). Results are more contrasted on repair measures. It also emerges that fluency development in a language pair might differ from another pair, and it is therefore difficult to draw common traits across languages (Baker-Smemoe et al., 2014; Préfontaine et al., 2016 for L2 French).

Lastly, we addressed the relation between L1 and L2 macro-level and fluency patterns. Through the studies reviewed and the presentation of a study we conducted ourselves on L1 and L2 French and English that included a within-speaker and between-group analysis, we highlighted that these aspects are strongly influenced by L1 patterns and the target language specificities. It appeared that overall, breakdown measures are strongly correlated to L1, speed measures moderately so, and repair measures weakly.

We also highlighted the fact that the variability between two test times in L1 is never looked at in relation to the same variability in L2. Consequently, in addition to the role L1 plays in L2 patterns, when employing a multiple test-time design, we could obtain even more accurate measures if they were also shaved off of the L1 between-time variability. This will be discussed in the analysis of the results of the study presented in Chapter V & VI.

In the next and third chapter, we turn to the perception of L2 speech by native speakers, and the listening abilities of L2 speakers towards the target language.

CHAPTER III - L2 SPEECH PERCEPTION

INTRODUCTION

The literature reviews presented so far have been focused on studies using acoustic measures taken from L2 speech production. In the field of language teaching and testing however, research employs - instead, or in combination with acoustic measures - listeners' judgements in the form of ratings. These types of measures have the undeniable advantage of putting perception in the foreground.

Indeed, taken by themselves, there is no way to know if what is observed through objective measures is relevant to the human ear and speech perception processing system. Section 2.3.1. (p. 84) already highlighted the relationship between objective measures and their impact on listeners' perception of fluency (utterance fluency & perceived fluency).

In this chapter, we present three other constructs listeners are frequently asked to evaluate in L2 speech acquisition studies: accentedness, comprehensibility, and intelligibility. After defining them, we will see what empirical evidence tell us about what they are related to in L2 speech. Through a literature review, we will present how these constructs have been used and operationalised in L2 speech acquisition studies.

In the second half of this chapter, we turn to what L2 speaker face when they hear the target language spoken by natives, i.e., L2 speech segmentation. We will first discuss the challenges L2 learners face in that regards from a theoretical point of view, and highlight the role of prosody. Then, a review of experimental studies on the matter will provide more concrete information on the cues L2 learners rely on, and on the origin of segmentation errors.

1. DEFINITIONS AND CONTEXT

1.1. FOREIGN ACCENT

A foreign accent is a very obvious and commonly experienced consequence of speaking an L2. However, giving a precise definition of the phenomenon is not as straight forward. Els & de Bot (1987) list a few definitions of foreign accent that diverge in what they highlight:

> "The label 'foreign accent' is applied to a speech-pattern by the listener on the basis of the way he hears the sounds of the talker in terms of his own background" (Chriest, 1969, p. xvii)
>
> "Phonological cues either segmental or suprasegmental, which identify the speaker as a non-native user of the language." (Scovel, 1969, p. 38)
>
> "[...] complex of interlingual or idiosyncratic phonological, prosodic, and paralinguistic systems which characterizes a speaker of a foreign language as non-native." (Jenner, 1976, p. 167)
>
> "[...] the perceived effect of many discrete and general differences in pronunciation between native and non-native speakers" (Flege, 1987, p. 162)

Chriest and Flege's definitions clearly mention the *perceived* nature of the phenomenon. This points to the role of the listener in defining their own norm (L1), and the degree of deviation they perceive from that norm. Conversely, definitions by Scovel and Jenner are more focused on the nature of foreign accented speech cues. Both refer to the segmental and suprasegmental levels.
More recently, Rasier & Hiligsmann (2007) give a definition that encapsulate both aspects:

> "[...] the perception of general and discrete deviations from the generally accepted norm of pronunciation of a language that are reminiscent of another language" (p. 43)

Even though this definition is more general regarding the nature of the phonetic cues to non-native speech, the term "pronunciation" - also used in Flege's definition - includes both segmental and suprasegmental levels. This definition also adds the notion of L1 transfer which is absent in the definitions above, apart from Chriest who refers to it indirectly referring to the listener's "own background".

Ultimately, we find Rasier & Hiligsmann's definition to be the most comprehensive as it covers the perceived nature of the phenomenon, the fact that it relates to pronunciation, and it acknowledges the influence of the speaker's L1.

In fact, Munro (2008) considers foreign accent as the "clearest evidence" of the influence of the L1 on an L2. Munro reminds us of the negative connotations that used to be associated with foreign accent in early definitions such as that of Greene & Wells (1927; cited in Munro, 2008):

> "Foreign accent, being of the nature of imperfect speech, is the result of incorrect articulation and enunciation and is therefore classified, from our therapeutic viewpoint, as stammering speech." (p. 24)

However, foreign-accented speech is only a very normal and common consequence of L2 acquisition, especially for late-learners - when the L2 is learned after childhood. This claim is supported by many research on the critical period hypothesis (see Azieb, 2021 for a recent review). While native-like pronunciation is not unattainable and can be achieved even when learning an L2 after childhood, it is definitely rare and seems to be conditioned by a combination of factors such as aptitude (Ioup et al., 1994), motivation (Moyer, 2004), and social factors (Hansen Edwards, 2008).

Paradoxically, the native norm generally remains the reference and can still represent the ultimate goal of L2 pronunciation acquisition for some keen learners. Yet, from a pedagogical point of view, the modern take is to focus on the attainment of intelligible speech and effective communication rather than the native form per se.

Clearly, a native-like pronunciation is far from necessary for effective communication. The fact is that millions of L2 speakers get by totally fine in their everyday life using foreign-accented speech (Hansen Edwards, 2008). This does not mean however, that foreign-accented speech does not have any consequences for the native listener.

Studies on the perception of L2 speech by native listeners have revealed a range of potential outcome such as diminished acceptability and intelligibility (Flege, 1988), low credibility (Lev-Ari & Keysar, 2010), and overall negative evaluation.

As pointed out by Flege and Jenner's definitions of foreign accent above, both the segmental and suprasegmental levels are involved. However, historically, early studies on the perception of foreign accent have overwhelmingly been focused on segmental aspects, and these were at the base of the development of speech learning models such as the SLM (Flege, 1995) and PAM (Best, 1995), presented in section 1.2.3., (p. 64).

Inaccuracy at the segmental level translates to omission, insertion, or substitution of phones, and/or divergent sub-phonemic realisation (Zampini, 2008). At the suprasegmental level, the realisation and distribution of intonation contours, stress patterns, timing, speaking rate, and pauses can all differ from that of a native speaker and therefore contribute to the perception of foreign accent (Ulbrich & Mennen, 2016).

Research has been carried out with the aim of disentangling the effects of segmental and suprasegmental inaccuracy on the perception of foreign accent, and their respective weight. Yet, the conclusions are inconsistent. On the one hand, some studies find that suprasegmentals have a stronger impact on the perception of foreign accent than segmentals (amongst others Anderson-Hsieh et al., 1992; Boula de Mareuil & Vieru-Dimulescu, 2006). On the other hand, others point to the opposite (e.g., Sereno, Lammers, & Jongmann, 2014; Ulbrich & Mennen, 2016).

While segmental and suprasegmentals effects on the perception of foreign accent have been attested independently of one another, there is an undeniable interplay between the two. For instance, Ulbrich & Mennen (2016) showed that segmental inaccuracy can blur the perception of fine prosodic deviances.

In a similar way, within suprasegmental features, the perception of differences on one aspect (e.g., speech rate) is influenced by the realisation of other aspects (e.g., intonation contours). Polyanskaya, Ordin, & Busa (2017), using synthetised speech samples where segments and intonation were neutralised, investigated the effect of speech rate and speech rhythm and found that while both had an effect on foreign accent perception, speech rhythm had a larger effect. When adding an imposed intonational contour, fine differences in rhythmic pattern were perceived, while the perception of differences in speech rate diminished.

Today, enough empirical evidence is available to assert that both segmental and suprasegmental deviations impact the perception of foreign accent. However, the respective weight, and the way these two levels interact is not yet fully understood, and most certainly varies according to the language pairs, the speaking contexts, and the listeners sensitivity.

As important as the identification of what contributes to foreign accented speech, is its impact on communication. A large body of research has yielded empirical data regarding this topic through perceptual judgement paradigms. Most notably, Munro & Derwing (1995a, 1995b) and Derwing & Munro (1997) investigated the relationship between the degree of foreign accent (henceforth *accentedness*) and the intelligibility and comprehensibility of L2 English speech samples, as judged and rated by *unsophisticated* native listeners, i.e., untrained, non-specialist listeners. Intelligibility was defined as:

> "The extent to which a speaker's message is actually understood by a listener" (Munro & Derwing, 1995a, p. 76)

Intelligibility was operationalised as the number of words correctly transcribed by the listeners. Comprehensibility however, referred to the estimated difficulty of understanding utterances, and was operationalised as a rating on a 9-point scale, similarly to accentedness. Results showed that the three dimensions, while correlated, also varied independently.

Correlation between comprehensibility and accent ratings varied across listeners suggesting that these two dimensions cannot be taken as strictly co-linear. While comprehensibility ratings were associated with intelligibility scores, accentedness was not. As a matter of fact, in several cases very strong degrees of accentedness were associated with high intelligibility (correct transcription).

Derwing and Munro's work has inspired a myriad of studies on the relationship between accentedness, intelligibility, comprehensibility, and also fluency; along with their correlates in speech. The next sections present a review of such work.

1.2. LISTENERS' JUDGEMENTS OF ACCENTEDNESS AND COMPREHENSIBILITY

Studies using listeners' judgement of foreign accent in the speech of L2 speakers go back at least as far as the mid-20th century, most notably in research concerning the *critical period* for language acquisition, such as Scovel (1969). In these studies, native listeners' assessment of foreign accents is a key methodological component. In the 1990's, it became also a common practice amongst researchers in L2 acquisition of pronunciation - especially in North America - where an important research community has developed around the teaching and learning of English as a second language (ESL).

As mentioned in the previous section, the study conducted by Munro & Derwing (1995a) is considered as pioneering work. These authors introduced their definitions of the three key constructs of accentedness, comprehensibility and intelligibility and their inter-connections, as well as a methodology that has since then been replicated in countless L2 pronunciation studies.

According to Munro & Derwing, **accentedness corresponds to the degree of perceived foreign accent** and is assessed - as in previous studies - through the use of a Lickert scale ranging from 1 (*no foreign accent*) to 9 (*very strong foreign accent*). Intelligibility and comprehensibility are both related to the general understanding of the message, but concern slightly different aspects.

Intelligibility is defined as the degree to which a listener understands the intended message. While some authors have used Lickert scale to measure it (Fayer & Krasinki, 1987; Palmer, 1976; cited in Munro & Derwing, 1995a), Munro & Derwing use a transcription task based on Gass & Varonis (1984; cited in Munro & Derwing, 1995a). Listeners are asked to write out the sentences they hear, and a score is calculated based on the accuracy of the transcription.

Lastly, **comprehensibility corresponds to the effort required for the listener to reach understanding**. It therefore relates to the ease or difficulty experienced by the listener in the decoding process. In this 1995a study, it is measured just like accentedness, with a 9-point Lickert scale from 1 (*extremely easy to understand*) to 9 (*impossible to understand*).

The first question addressed by the authors concerns the relationship between the three constructs. To that effect, they ran correlation tests between all three measures for each individual listener (n=18). They found that for a vast majority of them, the correlation was significant between comprehensibility and accentedness scores, and between comprehensibility and intelligibility. However, the strength of the correlations varied greatly across listeners. Conversely, a correlation between accentedness and intelligibility was found for only five listeners (out of 18).

The authors also coded the speech samples collected for phonemic errors (substitution, insertion or deletion of segments), phonetic errors (mispronounced segments), intonation errors, and grammatical errors. In order to understand which type of error contributed to the listeners judgement of all three dimensions, they ran correlation tests. It was found that phonemic, phonetic, intonation, and grammatical errors all correlated with accentedness for a majority of listeners. However, only a small proportion of listeners showed a correlation between phonetic and phonemic errors and comprehensibility, suggesting that segmental errors impact more the perception of foreign accent than the comprehensibility of the message. This was also reflected by the absence of correlation between the segmental errors and the intelligibility scores.

The main conclusion of the study is that a strong foreign accent, although correlated to comprehensibility, does not necessarily weaken the intelligibility and comprehensibility of the message. This finding had strong implications in the teaching and assessment of ESL where the strength of the accent was traditionally associated to the comprehensibility of the message. However, as the authors conclude:

> "If comprehensibility and intelligibility are accepted as the most important goals of instruction in pronunciation, then the degree to which a particular speaker's speech is accented should be of minor concern, and instruction should not focus on global accent reduction, but only on those aspects of the learner's speech that appear to interfere with listeners' understanding." (p. 93)

This study paved the way for the investigation of these aspects of L2 speech that impact comprehensibility and intelligibility. It also supported the pedagogical shift from a native-like pronunciation goal to L2 speech intelligibility and comprehensibility, which Levis (2005) conceptualised as the *nativeness principle* vs the *intelligibility principle*.

1.3. FOCUS ON COMPREHENSIBILITY

An important takeaway from Munro & Derwing's study is the differentiation between intelligibility and comprehensibility, and their relationship to accentedness. Their results showed that even when the L2 speech samples were perfectly transcribed - therefore perfectly intelligible - by the native listeners, they could be assigned poor comprehensibility scores. It is therefore assumed that accented speech can lead to an increased difficulty in processing, which translates to an increase in processing time, itself associated with comprehensibility.

In order to investigate the relationship between processing time, comprehensibility and accentedness, Munro & Derwing (1995b) asked 20 native English listeners to evaluate as true or false sentences read by non-native and native English speakers. They also had to rate the sentences for comprehensibility and accentedness. Processing time corresponded to the latency between the presentation of the sentence and the response given by the listener (true or false).

It was found that sentences read by non-native speakers (L1 Mandarin) yielded significantly longer processing time than those uttered by native speakers. Results on the correlation between accentedness and comprehensibility replicated those of their previous study, with a strength of 0.6 overall, but still an important variation between listeners was found.

Finally, they found that longer processing time were associated with lower comprehensibility scores and higher accentedness scores. However, only the relationship between processing time and comprehensibility was significant. In conclusion, **this study demonstrated (once again) that accentedness and comprehensibility are partially independent dimensions, and that listeners take processing time into account when judging the comprehensibility of speech samples.**

Fast forward 20 years, in a recent publication, Kennedy & Trofimovich (2019) advocate for the use of comprehensibility measures in L2 pronunciation research, highlighting the value of this construct in their review article.

The first argument in favour of comprehensibility measures is its practicality and ease of operationalisation in an experimental design. In its most common form, comprehensibility is measured as described in Munro & Derwing's works, through the collection of listeners' ratings on a scale. This simple task means that multiple judgements of multiple speech samples can be collected in a limited amount of time, and in a single testing phase. In contrast, intelligibility requires more time for listeners to transcribe the content of the sample heard, and intelligibility scores have been found to lack reliability across tasks (Kang, Thomspon, & Moran, 2018; Kennedy, 2009; cited in Kennedy & Trofimovich, 2019).

Secondly, comprehensibility ratings reflect the effort needed by the listener in the processing of L2 speech to reach understanding. While comprehensibility scores have been shown to be connected to intelligibility scores (amongst others Kennedy & Trofimovich, 2008), they actually give more information on the nature of the listeners' reaction to L2 speech. Studies have shown that the amount of effort required to understand L2 speech has an impact on the listeners' emotional reaction to the L2 speaker, as well as their credibility judgements (Dragojevic & Giles, 2016; cited in Kennedy & Trofimovich, 2019; Lev-Ari & Keysar, 2010).

Lastly, the authors insist on the influence of the listeners' potential social bias towards L2 speech. Studies have shown that raters who report negative attitude towards L2 speakers' proficiency tend to give harsher comprehensibility scores than raters who report a positive attitude (Sheppard, Elliott, & Baese-Berk, 2017). The perceived ethnicity of the L2 speaker can also influence the reported understanding of the listeners (Rubin & Smith, 1990; cited in Kennedy & Trofimovich, 2019). Thus, researchers should be aware of these aspects when selecting their listener-raters.

This section presented the constructs of accentedness, intelligibility and comprehensibility through the pioneering work of Munro & Derwing. The concept of comprehensibility was developed further as it is the least straight forward amongst the three, and comprehensibility measures are used in the study presented in Chapter V & VI.

In the following section, we bring the focus onto the elements of L2 speech that have an influence on each dimension. We will see that all dimensions of speech rhythm as we understand it (see Chapter I) are involved, as well as lexicogrammar aspects, and that their influence on each dimension differs, and can also depend on the listeners' profile, target language, and task type.

2. LINGUISTIC CORRELATES OF COMPREHENSIBILITY AND ACCENTEDNESS

In a follow-up study to their 1995 ones, Derwing & Munro investigated further the relationship between accentedness, comprehensibility and intelligibility, and different types of errors in L2 English spoken by L1 Cantonese, Japanese, Polish, and Spanish speakers (of intermediate L2 proficiency level).

Participants were recorded on picture-based narrative task. Native listeners rated the speakers' accentedness and comprehensibility on 9-point scales, and also transcribed the speech material to obtain an intelligibility score. The experimenters extracted the number of grammatical and phonemic errors, and rated the *global goodness of prosody* on a 9-point scale by listening to the speech samples in a low-pass filtered version to remove the segmental information and focus on prosodic aspects only. The speech rate was also calculated as syllables per seconds.

Results showed that, similarly to Munro & Derwing (1995a), **accentedness was rated more harshly than comprehensibility**, which was itself rated more harshly compared to the intelligibility scores obtained, showcasing once again the partial independence of the three dimensions.

As to the impact of each type of errors, it was found that grammatical errors (which were quite numerous overall) correlated with both accentedness and comprehensibility ratings for half of the listeners, prosodic scores and speaking rate for about a third of the listeners, and phonemic errors for only 15% of the listeners. All kinds of errors were correlated to intelligibility scores only for a minority of listeners.

Interestingly, it seems like each type of error had a similar effect on both accentedness and comprehensibility. However, listeners were also asked to judge which factor they found to impact more their accentedness and comprehensibility scores, and differences appeared. Accentedness was mainly associated with segmental errors, then grammatical errors and enunciation (mumbling), but less so with prosodic features and fluency.

Conversely, comprehensibility was not at all associated with segmental errors but mainly enunciation (which corresponded to a negative correlation found between speech rate and comprehensibility) and grammatical errors. Fluency was also more strongly associated with comprehensibility than with accentedness, although in small proportion in both cases.

In a later study involving utterance fluency measures (Derwing et al., 2004), **fluency was found to be more strongly correlated with comprehensibility than accentedness**, supporting the hint that emerged from Derwing & Munro's study.

Unfortunately, the effect of the speakers' L1 on the ratings and error type correlations was not tested in this study (listeners were tested on their ability to recognise the speakers' L1 and on their familiarity with them).

The impact of suprasegmental aspects of L2 speech on comprehensibility and proficiency ratings were the focus of a study conducted by Kang, Rubin, & Pickering (2010). L2 English speech samples from an iBT TOEFL® speaking task belonging to 26 native speakers of Chinese, Spanish, Korean, and Arabic, were analysed. 188 native listeners rated the samples for language proficiency and comprehensibility on 7-point scales. 29 acoustic variables - considered by the authors as suprasegmental measures - were extracted from the recordings. These included common measures of speed and breakdown, stress measures (pace and space), pitch measures (e.g., f0 value on prominent/non-prominent syllables), and paratone measures (e.g., quantity and average height of low termination tones). **Results showed that comprehensibility and proficiency ratings were correlated**, and that a cluster of acoustic variables was specifically associated with both ratings. These variables included all speed measures, pace (number of stressed syllables per run), and an intonation measure (quantity of mid-falling tones). On the basis of these variables clustering together, the authors argue that "fluency is an intonational phenomenon as well as a temporal one" (p. 562). This illustrates the cross-level nature of fluency we have described in Chapter I. Boundary marking features (silent pauses and termination tones) were also found to have a strong impact on both comprehensibility and proficiency ratings. Overall, suprasegmentals accounted for 50% of variance in ratings.

Based on the results of the study, the authors conclude that **the contribution of suprasegmental errors is at least as important, if not more, as segmental errors in foreign accented speech**. **From a pedagogical perspective, this supports the need to insist on prosody in pronunciation instruction, in order to improve comprehensibility and perceived oral proficiency.** The study presented in Chapter V & VI circles back to this.

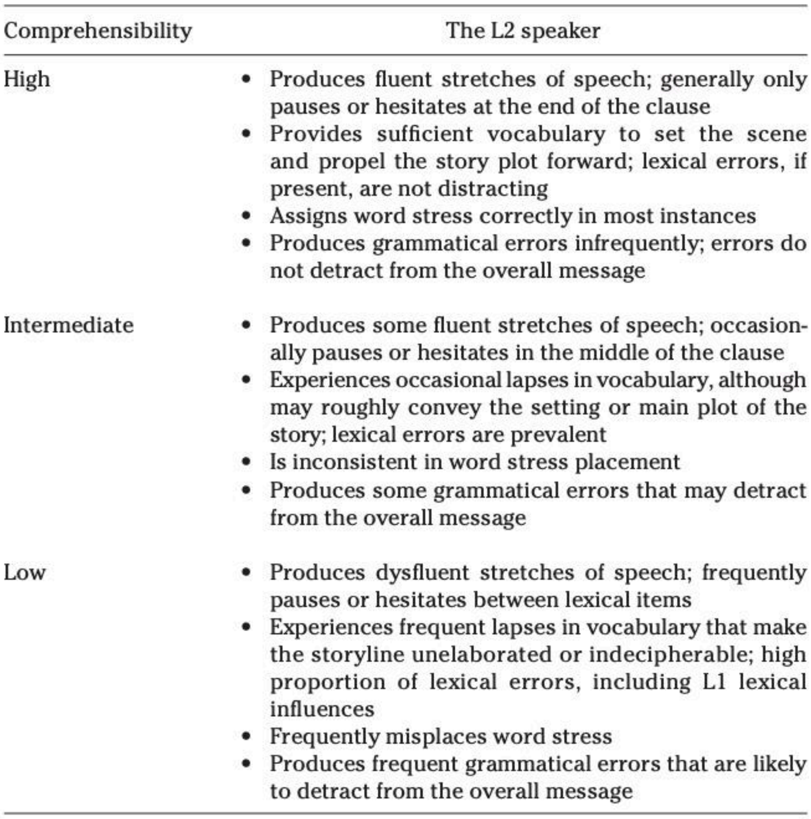
Yet, we have seen that comprehensibility is also influenced by segmental and lexicogrammar factors (Munro & Derwing, 1995a; Saito et al., 2017). The work of Isaacs & Trofimovich (2012) shed some light regarding the impact of different linguistic variables on comprehensibility, and their role in distinguishing speakers at different comprehensibility levels.

The study involved 40 Francophone Canadian speakers at L2 English levels varying between beginner and advanced. They were recorded on a picture-based narrative task, and four categories of measures were extracted: phonological (segmental and suprasegmental), fluency measures (speed, breakdown, repair), linguistic resource measures (grammatical accuracy, vocabulary size and errors), and discourse measures (adverbs used as cohesive devices, proposition number and characteristics). 60 native English speakers and novice raters (students from non-linguistic disciplines) rated the speech samples for comprehensibility on a 9-point scale. Correlation coefficients were computed for each of the 19 speech measures and comprehensibility scores.

The strongest correlations were spread over the four categories: vocabulary richness for the linguistic resource category; stress errors and vowel reduction ratio for the phonological category; mean length of run for fluency, and richness of types of propositions for discourse. Conversely, the weakest correlations were found for syllable structure error ratio, pruned syllable per second, and number of unfilled pauses. No correlation was found for the measure of pitch range.

In order to see which measures were best for distinguishing between comprehensibility levels, three groups were formed based on the collected ratings: low, intermediate, and high. It was found that **word stress distinguished between all three comprehensibility levels**. Vocabulary richness and mean length of run both distinguished between the low level on one side, and the intermediate and high levels on the other. Richness in proposition type and grammar both distinguished between the low and intermediate levels on one side, and the high level on the other.

Based on these findings, the authors propose a summary of the descriptors for each comprehensibility levels as guidelines for L2 comprehensibility development, presented in Table 6 below.



Comprehensibility	The L2 speaker
High	• Produces fluent stretches of speech; generally only pauses or hesitates at the end of the clause • Provides sufficient vocabulary to set the scene and propel the story plot forward; lexical errors, if present, are not distracting • Assigns word stress correctly in most instances • Produces grammatical errors infrequently; errors do not detract from the overall message
Intermediate	• Produces some fluent stretches of speech; occasionally pauses or hesitates in the middle of the clause • Experiences occasional lapses in vocabulary, although may roughly convey the setting or main plot of the story; lexical errors are prevalent • Is inconsistent in word stress placement • Produces some grammatical errors that may detract from the overall message
Low	• Produces dysfluent stretches of speech; frequently pauses or hesitates between lexical items • Experiences frequent lapses in vocabulary that make the storyline unelaborated or indecipherable; high proportion of lexical errors, including L1 lexical influences • Frequently misplaces word stress • Produces frequent grammatical errors that are likely to detract from the overall message

Table 6 - Suggested guidelines for L2 comprehensibility scale development. From Isaacs & Trofimovich, 2012, p. 497.

In conclusion, this study showed that **factors other than pronunciation-related ones have a strong impact on comprehensibility ratings of non-expert listeners, namely, richness of vocabulary and proposition type diversity**. In fact, the authors also collected qualitative data from three expert raters (teachers of ESL) who indicated that grammatical errors weighted a lot on their evaluation. However, notable differences appeared between the three raters, suggesting that the construct of comprehensibility can be interpreted in different ways.

Nevertheless, **certain suprasegmental and fluency measures were still found to correlate strongly with the ratings of non-expert listeners, more so than segmental aspects.** Notably, word stress was found to be the only measure to distinguish between speakers at low, intermediate, and high comprehensibility levels. However, it should be noted that the important impact of this measure might be related to the language pair in question. Indeed, as we have previously outlined, French and English are known to differ significantly in stress pattern, making it a crucial aspect of L2 speech development when learning one of these two languages from the other.

In a follow-up study using the same speech samples (Trofimovich & Isaacs, 2012), the same 19 measures were tested for correlation with accentedness scores. In contrast with the results obtained with comprehensibility**,** segmental accuracy was found to correlate strongly with accentedness ratings, and was also mentioned as determinant in the accent judgement of expert ESL teachers (along with "naturally sounding rhythm", p. 913). Word stress and vowel reduction were also strongly associated with accentedness ratings from the novice listeners.

The difference between comprehensibility and accentedness in terms of linguistic variables involved in both kinds of rating was replicated by Saito, Trofimovich, & Isaacs (2017). Using again the same speech samples from the same 40 French Canadian learners of L2 English, this time the linguistic variables to test for association with comprehensibility and accentedness ratings were operationalised differently. The authors selected 11 linguistic variables also related to all categories of pronunciation, fluency, lexis, grammar, and discourse. However, this time around, these variables were also rated by the listeners rather than extracted from the speech samples. Basically, raters first listened to the speech samples and rated global accentedness and comprehensibility. In a second phase they listened to the samples again and rated five pronunciation-related variables. In the last phase they rated six lexical, grammatical, and discourse level features from the orthographic transcriptions.

It was found that pronunciation variables were associated to both accentedness and comprehensibility. However, the latter was also associated to the lexis, grammar, and discourse variables. **Within the pronunciation variable, all five (segmental accuracy, word stress, intonation, prominence alternation between content and function words, and speech rate) were associated to comprehensibility with comparable strength; while for accentedness, segmental accuracy and word stress had the strongest correlations, and speech rate the weakest**.

As per usual, a lot more data is available on L2 English and the generalisation of the results of such studies onto other L2s remains an open question. Turning now to (some of the rare) recent studies focusing on L2 French, it seems like similar results than for L2 English are found - particularly, the role of pronunciation features in both accentedness and comprehensibility ratings, and the role of lexicogrammar and discourse features exclusively for comprehensibility.

However, some differences emerge in the specificities of the pronunciation features that weight on one or the other construct.

Bergeron & Trofimovich (2017) analysed speech samples from 40 L1 Spanish learners of L2 French at an intermediate proficiency level, using ratings of global accentedness and comprehensibility as well as specific linguistic variables, from 20 native Quebec-French listeners. The study also included a comparison between a narrative task (the one most commonly used in previous research) and an interview (judged as more cognitively demanding). The specific variables concerned phonology and fluency (segmental errors, intonation *naturalness*, speech rate), and lexicogrammar (lexical richness and accuracy, grammatical accuracy and complexity, discourse coherence).

The authors found similar results than in L2 English in their narrative task, i.e., **the distinction between accentedness and comprehensibility based on the lexicogrammar aspects affecting only the latter while all pronunciation variables correlated with scores of both constructs. However, interestingly this distinction disappeared in the interview task** as lexicogrammar and discourse aspects were also correlated with accentedness scores. The authors explain this as a result of the more demanding nature of the interview task which involves the use of more linguistic resources on the speakers' side, and no predictability on listeners' side (whereas in the narration task, listeners already know the story told by the image prompts). This results in a weaker distinction between comprehensibility and accentedness since a larger array of L2 speech features are drawn upon in the perception of each construct.

Another study looked more specifically at pronunciation variables and their link to accentedness and comprehensibility in L2 French. Trofimovich, Kennedy, & Blanchet (2017) analysed read-aloud speech and picture-based narrations before and after a 15-week course focused on segmental and supra segmental aspects. The speakers were from various L1 backgrounds and all were enrolled in an intermediate level course. Native Quebec-French listeners rated the speech samples for accentedness, comprehensibility and fluency, and the ratings were tested for correlations with expert-coded segmental errors, intonation errors, *enchainement* and *liaison,* f0 range, mean length of run, and number of disfluencies (all kinds).

It was found that even though learners significantly improved after the course for all measured pronunciation variables, only a few of them were significantly correlated to ratings of accentedness, comprehensibility, and fluency. Intonation errors were correlated to all three constructs in both tasks. f0 range was also correlated to all three dimensions but only in the narration task (a narrower pitch range was associated to better scores). **Mean length of run was associated to comprehensibility and fluency ratings but only in the reading task. And number of disfluencies were strongly correlated to comprehensibility, and more moderately so to fluency**.

In summary, accentedness was (surprisingly) only associated to intonation features, while improvement in comprehensibility and fluency was associated to those as well as longer speech runs and fewer disfluencies. However, the task type nuanced these correlations.

3. CONCLUDING REMARKS ON COMPREHENSIBILITY & ACCENTEDNESS

The three constructs of intelligibility, accentedness and comprehensibility appear to be at the same time overlapping and partially independent. **Intelligibility corresponds to the understanding of the intended message and is usually measured through transcriptions. It seems to be the least influenced by accentedness and comprehensibility since strongly accented speech and low scores of comprehensibility can be associated to successful transcriptions of the message**. **This means that a strong accent does not de facto prevent understanding, however it can slow down the process**. **Which is why comprehensibility gives a more precise picture of the processing of L2 speech.** Comprehensibility relates to the difficulty or ease encountered by the listener, in the process of understanding the message heard. It has been shown that foreign-accented speech tends to require more processing time than native speech (Munro & Derwing, 1995b).

We have summarised the linguistic correlates to accentedness and comprehensibility reported by the studies mentioned in the previous section in Table 7 below. We have associated different colours to easily distinguish suprasegmental and fluency variables (in orange), segmental variables (in green), and lexicogrammar and discourse variables (in black).

	ACCENTEDNESS	COMPREHENSIBILITY	Languages & proficiency	Reference
Linguistic correlates	- segmental - intonation - grammar	- intonation - grammar	L1 Mandarin L2 English (adv.)	Munro & Derwing, 1995a
	- grammar - segmental - enunciation	- grammar - enunciation - speech rate	L1 Cantonese L1 Japanese L1 Polish L1 Spanish L2 English (int.)	Derwing & Munro, 1997
	NA	- speed measures - pace - intonation - silent pauses (as boundary markers)	L1 Chinese L1 Spanish L1 Korean L1 Arabic L2 English (NS)	Kang, Rubin, Pickering, 2010
	- segmental - rhythm - stress - vowel reduction	- vocabulary - stress - vowel reduction - mean length of run - propositions richness - grammar	L1 Canadian-French L2 English (beg. to adv.)	Isaacs & Trofimovich, 2012 Trofimovich & Isaacs, 2012
	- segmental - stress	- segmental - stress - intonation - speech rate - vocabulary - grammar - discourse	L1 Canadian-French L2 English (beg. to adv.)	Saito, Trofimovich, & Isaacs, 2017
	- segmental - intonation - speech rate - lexicogrammar (interview task only)	- segmental - intonation - speech rate - lexicogrammar	L1 Spanish L2 French (int.)	Bergeron & Trofimovich, 2017
	- intonation	- intonation - mean length of run - number of disfluencies	Various L1 L2 French (int.)	Trofimovich, Kennedy, & Blanchet, 2017

Table 7 - Summary of linguistic correlates associated with accentedness and comprehensibility. For ease of legibility, suprasegmental and fluency variables are coloured in orange, segmental variables are coloured in green, vocabulary, grammar, and discourse related variables are in black.
NA = non applicable; NS = non specified.

As can be seen in Table 7, **segmental aspects tend to be associated with accentedness but less so with comprehensibility. Suprasegmental features such as stress, rhythm and intonation are associated with both constructs. Fluency measures tend to be more often associated with comprehensibility than with accentedness. And finally, pronunciation-unrelated variables are more associated to comprehensibility than accentedness.**

What this table tells is that, if the main goal of an L2 speaker is to improve their comprehensibility level, then prosody and fluency should be their focus in L2 pronunciation. However, if the goal is more geared towards reaching native-like pronunciation, segmentals should also be worked on as they seem to participate as much as suprasegmentals in the perception of foreign accent. In a classroom setting where, nowadays, the intelligibility principle prevails, pronunciation activities should emphasise suprasegmental and fluency aspects, perhaps before segmental aspects. This constitutes the base of the choices we made when designing our study presented in Chapter V & VI.

So far, we have focused on the L2 speaker experience in terms of L2 speech production and how L2 speech is perceived by native listeners. The next and final section of this chapter is dedicated to how L2 speakers perceive the target language when spoken by natives, and mechanisms at play in L2 speech processing from the L2 speaker's perspective.

4. L2 SPEAKERS' PROCESSING OF THE TARGET NATIVE LANGUAGE

Although the main focus of our work concerns speech production, speaking an L2 necessarily involves also listening to it. The "L2 picture" would not be complete without addressing the L2 speaker's perception and processing of the target language as spoken by natives. This is especially relevant to speech rhythm since its fundamental function is to structure the stream of speech, facilitating processing on the listener's side.

In anticipation of the presentation of our experimental study (Chapter V, p. 179) which includes a speech segmentation task, this section briefly presents theories about L2 speech processing, focusing on the role of prosody and mainly rhythm in the segmentation of continuous speech. In a second part, we will review studies testing L2 learners' segmentation abilities, and the learnability of new prosodic segmentation cues.

4.1. L2 SPEECH SEGMENTATION

Speech segmentation relates to the human capacity to identify boundaries in the continuous stream of speech, in order to segment it into words and access meaning. Indeed, speech is continuous and words are not separated by spaces like in written language (although some languages do not use spaces in their written form either, e.g., Japanese).

In L1, this process is seamless, extremely fast and efficient. Yet, it involves several cognitive and perceptual mechanisms such as: phoneme recognition and the use of phonotactic cues, the use of prosodic cues for identification of syllables, words and higher-level units, lexical activation from memory, knowledge of grammatical structure rules, and deduction of semantic and pragmatic aspects from context.

In L2 however, every rung of the ladder leading to successful comprehension constitutes a difficulty. L2 listeners also tend to perceive L2 speech as being spoken at a faster rate than their L1. This *Gabbling Foreigner Illusion* - as Cutler puts it (Cutler, 2012, p. 338) - is directly related to segmentation difficulties L2 listeners experience (Snijders et al., 2007; cited in Cutler, 2012).

As we have seen in section 1.2. of Chapter II (p. 62), phoneme recognition and distinction in L2 comes with challenges pertaining to the L1 phonological filter.

Lexical activation is subject to the vocabulary size which is much smaller in

L2 than in L1, especially at the initial stages of learning. Phonotactic and prosodic cues to segmentation differences lead to the ineffective use of probabilistic strategies that work in the L1 but are inadequate in the L2. And finally differences in syntax structure, semantic and pragmatic encoding, and cultural discourse codes also represent yet another layer of potential confusion and barrier to comprehension (Cutler, 2012). Thus, speech segmentation and comprehension are multi-layered processes that invoke several speech dimensions. For the purpose of this dissertation, we will focus solely on the role of prosody, and more specifically rhythm.

Investigating the role of stress in the segmentation strategies used for English, Cutler & Carter (1987) started by analysing stress position in English words. It emerged that 70% of words in the English dictionary started with a stressed syllable. When analysing a large corpus of spoken English, that proportion went up to 90%, which led the authors to conclude that English listeners could rely on the assumption that a stressed syllable signals the beginning of a word to segment speech, with a very high success rate.

Evidence of preference for segmentation based on initial stress also came from observations of segmentation resetting of misheard sentences: *descriptive prose* heard as *the script of prose* or *she's a must to avoid* heard as *she's a muscular boy* (Cutler, 2012, p. 123). Laboratory experiments confirmed this tendency for English listeners (Cutler & Butterfield, 1992). Listeners also tend to segment words based on the initial stressed syllable rule in Dutch, a language that shares similarities with English in terms of rhythmic structure (Vroomen & De Gelder, 1995; cited in Cutler, 2012). This finding implies that listeners' segmentation strategies are tied up to their language's rhythmic structure.

Yet, not all languages are stressed languages, and French is a common example of a non-stressed language (see section 4 of Chapter I, p. 46). French therefore calls for a different segmentation strategy than stress-based languages. Or does it? The work of Mehler, Dommergues, Frauenfelder, & Segui (1981; cited in Cutler, 2012) brought evidence to the fact that French listeners - giving the syllabic rhythmic structure of the language - rely on the syllable to segment words. This type of strategy for syllabic languages was also found for Spanish and Catalan (Sanchez-Casas, & Garcia-Albea, 1993; Sebastián-Gallés et al., 1992; Bradley; cited in Cutler, 2012). However, the supposed dichotomy between stress-based strategy vs syllable-based strategy is not as clear-cut. French listeners also rely on accentual cues such as the final accent (Banel & Bacri, 1994), and successfully perceive the presence of an initial accent when it occurs, suggesting that a metrical segmentation strategy is also used in French (Astésano & Bertrand, 2016).

The Metrical Segmentation Strategy (MSS) as originally proposed by Cutler & Norris (1988) - although associated to the stress-based segmentation strategy for English - mainly claimed that:

> "[...] whatever type of structure best describes the metrical forms a language prefers to use - for example in poetry - that structure will also be useful for listeners in segmenting continuous speech." (Cutler, 2012, p. 132)

The MSS was actually later renamed the *Rhythmic Segmentation Hypothesis* to avoid it be necessarily associated with English's stress-based strategy, but rather with its more general claim about the importance of languages' rhythmic structure in speech segmentation (Cutler, 2012). In any case, enough empirical evidence supports the fact that in L1, prosody is essential for speech processing especially for the segmentation of the continuous flow of speech.

Meanwhile, within the field of Second Language Acquisition (SLA), segmentation strategies and the role of prosody in speech processing in L2 remain under-researched topics (Calhoun et al., 2023). Yet learners often struggle with limitations of their listening skills, even at high proficiency levels (Charles et al., 2015; Tremblay et al., 2012). The following section gives an account of the current state of L2 listening (to broaden the scope of segmentation) research in SLA.

4.2. LEARNABILITY OF NEW PROSODIC SEGMENTATION CUES IN L2

The previous section highlighted the role of prosody in the segmentation strategies adopted by the listeners. Prosody being language-specific, L2 speech segmentation represents an important challenge to the L2 listener, who most likely will use an ill-adapted strategy from their L1. Therefore, it can be assumed that the degree of similarity or difference between two languages' prosodic systems will affect the listener's strategy's success.

This hypothesis was tested in several studies. For instance, Sanders, Neville, & Woldorff (2002) tested how L1 Japanese and Spanish learners of L2 English would make use of the English stress cue (stressed syllables most likely signal a word initial vs unstressed syllables signal a word medial) to determine the location of a target sound within words in L2 English sentences.

The authors predicted potential outcomes: if both L1 groups are not able to use different rhythmic cues than that of their L1, then neither will use the stress cue in English; if they are able to use the English stress as a cue to segmentation, they might do so differently according to their L1 stress-pattern habits. Contrary to English, in Japanese loudness and duration are not cues to lexical stress. In Spanish they are, but the stress pattern differs. So, for L1 Japanese the acoustic parameters differ (i.e., the systemic dimension in the L2 Intonation Learning Theory (LILt) - see Chapter II, section 1.2.6., p. 68), whereas for L1 Spanish, the pattern does (i.e., the realisational dimension in LILt).

Stimuli sentences were controlled for lexical and syntactic information by using non-words (replacing both content and grammatical words), such that sentences corresponded to three conditions: acoustic information only, acoustic and syntactic, acoustic lexical and syntactic. Results indicated that both the Spanish and Japanese groups were able to use the English stress as a cue to word segmentation, and that they relied on the stress cue more strongly when lexical information was missing. These results suggest that the different L1 stress patterns for each group did not have an effect on their use of the L2 English stress pattern. However, this study fails to mention the role of f0 in stress realisation in English, Japanese and Spanish.

Contrastively, the use of F0 as a segmentation cue in L2 was the focus of a study conducted by Trembley, Broersma, Coughlin, & Choi (2016). These authors investigated the use of an f0 rise as a cue to word-final boundary in L2 French by L1 speakers of Korean and English. The L1 groups were chosen because of their differing degree of similarity with French in terms of the segmentation prosodic cues they rely on.

The authors focused on the f0 rise (H*) occurring on the final (full) syllable of words in final position of an accentual phrase (AP) in French. Similarly, in Korean f0 rise also marks a word-final boundary in the AP domain, however H* is aligned differently as it occurs earlier than in French. In English though, H* usually signals a word-initial boundary.

Based on the predictions described in Best's (1995) PAM-L2 model, and Flege's (1995) SLM, the authors postulated the *Prosodic-Learning Interference Hypothesis* (p. 2), which predicts that learning a new segmentation cue will be more difficult if the L1 and L2 prosodic systems are similar, yet non-identical; than if they are radically different. To put it differently, the assumption is that phonological similarity with phonetic difference (French-Korean) will be more difficult to learn than phonological difference (English-French/Korean). This hypothesis draws upon the assimilation phenomenon occurring when L1 and L2 features are very similar, leading the learner to assimilate the L2 feature to a pre-existing L1 category (see Chapter II section 1.2.3., p. 64).

Following this reasoning, the similarity between French and Korean in the use of an f0 rise to signal a word-final boundary (with a slight difference in alignment) should cause the Korean learners to perform worse than L1 English speakers, for whom the f0 rise signals a different boundary (word-initial), in a word-final boundary location task. The results confirmed the hypothesis. The Korean group indeed performed worse than the English group and the native French group. Koreans seemed to have difficulty using the f0 rise as a cue to word-final boundary and even tended to interpret it as the reverse: a cue to initial-word boundary.

In a follow-up study (Tremblay et al., 2021), the Prosodic-Learning Interference Hypothesis was tested again in a reverse situation, i.e., with L1 English and L1 French learners of Korean L2. A similar experiment was conducted to test the participants' use of AP-initial and AP-final tones as word boundary cues. The L1 French were expected to have more difficulty than the L1 English group once again because of the phonological similarity between French and Korean and according to the Prosodic-Learning Interference Hypothesis.

However, the results contradicted the authors' prediction. L1 French and L1 English speakers showed a similar capacity to use tonal cues to word boundary in L2 Korean. The discrepancies between the 2016 study and those results might be explained by the difference in L2 exposure the learners received. As a matter of fact, in the 2021 study, both groups of L2 learners of Korean had been immersed in the country for a minimum of 1.5 years. Conversely, L2 French learners in the 2016 study lived in their home country and did not have any such L2 immersion experience.

Thus, the authors concluded that difficulties stemming from L1-L2 similarity in prosodic segmentation cues can be alleviated by an intensive exposure to the L2 (such as an immersion). Other studies have supported the benefiting (and unsurprising) effect of exposure for the learning of L2 segmentation cues (Namjoshi et al., 2012; Gilbert et al., 2016; cited in Calhoun et al., 2023).

Another study compared the impact of systemic vs frequency dimension (from LILt) differences between languages on segmentation capacities. Morrill (2016) tested monolingual English speakers' capacity to identify word boundaries in Japanese and in Finnish, two languages they had never been exposed to. English's lexical stress differs from Japanese's lexical pitch accent in the systemic dimension; whereas English and Finnish differ in the frequency dimension since word-initial stress is mandatory in Finnish vs frequent in English.

The results showed that English speakers were better at identifying word boundaries in Finnish than in Japanese. The author concluded that differences in the systemic dimension constitutes a greater difficulty than differences in the frequency dimension. In fact, these results suggest that the similarity between English and Finnish was an advantage, which contradicts Tremblay et al.'s hypothesis.

Overall, studies on L2 segmentation abilities show that **learning to use new prosodic cues is absolutely feasible, that ease or difficulty in doing so depends on the similarity or difference between the L1 and L2, and on the LILt dimension they concern. It seems that the systemic dimension is particularly susceptible to predict learning difficulties.** Exposure has been shown to have a positive effect, although it is stating the obvious. But is increased exposure the only way for learners to improve their L2 segmentation skills? What about teaching L2 learners about prosodic cues to segmentation? Although this area of research is still fairly young, studies investigating the effect of prosody instruction on learners' listening skills have emerged recently, showcasing promising results. This literature is presented in Chapter IV (section 4.5., p. 171) of this dissertation, a chapter specifically focused on pronunciation instruction and its effects on speech production and perception.

CHAPTER SUMMARY

This chapter focused on L2 speech perception, first, from the point of view of native speakers, and then from the perspective of L2 learners.

We started by defining the construct of **foreign accent**, and noted Rasier & Hiligsmann's (2007) definition, which we found the most comprehensive. **Foreign accent is primarily a perceived phenomenon, related to pronunciation, and influenced by the speaker's L1**. Empirical evidence supports the fact that **both segmental and suprasegmental deviations impact the perception of foreign accent** (e.g., Anderson-Hsieh et al., 1992; Sereno, Lammers, & Jongmann, 2014). However, **the respective weight, and the way these two levels interact remains somewhat unclear**, and most certainly varies according to the language pairs, the speaking contexts, and the listeners sensitivity (Ulbrich & Mennen, 2016; Polyanskaya et al., 2017).

Three essential constructs in research involving native listeners' perceptual judgements of L2 speech were defined following the pioneering work of Munro & Derwing (1995a):

Accentedness corresponds to the degree of perceived foreign accent and is assessed through the use of a Lickert scale ranging from 1 to 9.

Comprehensibility corresponds to the effort required from the listener to reach understanding. It relates to the ease or difficulty experienced by the listener in the decoding process. It is commonly measured similarly to accentedness, on a 9-point Lickert scale.

Intelligibility is defined as the degree to which a listener understands the intended message. It is usually assessed through a transcription task.

Studies have explored the relationship between the three constructs and it emerges that they appear to be at the same time overlapping and partially independent. Intelligibility seems to be the least influenced by accentedness and comprehensibility, since strongly accented speech and low scores of comprehensibility can be associated to successful transcriptions of the message. This means that **a strong accent does not de facto prevent understanding, however it can slow down the process**. Which is why **comprehensibility gives a more precise picture of the processing of L2 speech**.

Because we are particularly interested in accentedness and comprehensibility (measures we use in our study), we have compiled data on the linguistic correlates of both constructs from the literature (Table 7). Segmental aspects tend to be associated with accentedness but less so with comprehensibility. Suprasegmental features such as stress, rhythm and intonation are associated with both constructs. Fluency measures tend to be more often associated with comprehensibility than with accentedness. And finally, pronunciation-unrelated variables are more associated to comprehensibility than accentedness.

From a pedagogical perspective, this implies that **prosody and fluency should be in focus in L2 pronunciation classes, order to improve the learners' comprehensibility. Conversely, in the pursuit of native-like pronunciation, work on segmentals should be added as they seem to participate as much as suprasegmentals in the perception of foreign accent**. In a classroom setting where, nowadays, the intelligibility principle prevails, pronunciation activities should emphasise on suprasegmental and fluency aspects, perhaps before segmental aspects. This constitutes the base of the choices we made when designing our study presented in Chapter V & VI.

In the second half of this chapter, we gave an overview of the processes at play in speech segmentation and highlighted the challenges they represent for an L2 speaker listening to the target language spoken by natives. **Phoneme recognition and distinction are hindered by the L1 phonological filter, lexical activation is subject to the vocabulary size which is much smaller in L2 than in L1, and phonotactic and prosodic cues to segmentation differences lead to the ineffective use of probabilistic strategies that work in the L1 but are inadequate in the L2**.

We then presented the ***Rhythmic Segmentation Hypothesis* which insists on the importance of languages' rhythmic structure in speech segmentation** (Cutler, 2012), and a review of studies investigating L2 learners' speech segmentation abilities. Findings suggest that **learning to use new prosodic cues is very much feasible, that ease or difficulty in doing so depends on the similarity or difference between the L1 and L2, and on the LILt dimension they concern**. It seems that the systemic dimension is particularly susceptible to predict learning difficulties.

In the last section, we posed the question of whether instructing L2 learners on prosodic cues could assist them in developing L2 speech segmentation skills.

This interrogation constitutes one of the research questions addressed in our study (Chapter V & VI).

Pronunciation instruction and its effectiveness on L2 speakers' performances in production and perception is the central topic of the next chapter.

CHAPTER IV - TEACHING L2 PROSODY

INTRODUCTION

After having reviewed in the preceding chapters the different challenges associated with L2 learning in terms of phonological aspects (both production and perception), this chapter focuses on how L2 pronunciation instruction can help L2 learners overcome their difficulties.

The chapter begins with an overview of the place given to oral language in classroom contexts. It briefly retraces the major shifts in pedagogical trends over the last century, which affected the place given to speaking skills in L2 instruction.

The second part of the chapter references the main methods and techniques currently in use for the teaching of pronunciation, and their theoretical bases.

Finally, the last part presents a literature review of studies testing the effect(s) of - and comparing different teaching methods or techniques. Based on the available empirical data, we have selected several tools and teaching techniques to include in the L2 French prosody course designed for our experimental study.

1. ORAL IN THE CLASSROOM

In the past forty years, research showing the importance and efficacy of dedicated pronunciation instruction in L2 classes has multiplied. However, many authors - if not all - from all over the world, introduce their book chapter or article by a common statement: pronunciation gets significantly less attention than other language aspects such as lexicon and grammar in L2 classes (Alazard, 2013; Darcy, 2018; Detey & Durand, 2021). Despite the prolific data supporting the necessity to teach pronunciation, the gap between research and practice in the classroom still persists.

Several surveys regarding English as a Second Language (ESL) teaching practices across Canada report similar outcomes with very little evolution over the course of sixteen years. For instance, Breitkreutz et al. (2001) surveyed 67 ESL university programs, and found that less than half of them included standalone pronunciation classes, less than a third used language labs for pronunciation instruction, and only 30% of teachers in these programs had received pedagogical training for teaching pronunciation. In fact, a lot of the respondents commented on the lack of training for pronunciation teaching, insisting on the fact that they wished to include more pronunciation exercises in their classes/courses but did not feel competent and/or legitimate enough to carry it out.

10 years later, in a follow-up study where 160 ESL instructors for adult classes were surveyed on the same topic, Foote et al. (2011) found that the proportion of respondent who indicated that some instructors in their institution had received pronunciation training increased by 20%. However, once again less than half reported the integration of pronunciation exercises in their ESL classes. The authors concluded that even though more training opportunities are available to teachers, a majority of them still feel under-trained and uncomfortable teaching pronunciation to their students.

Similar results were found in a study involving the observation of three teachers of ESL to grade 6 francophone learners in Quebec (Foote et al., 2016). The teachers were observed for 40 hours over the course of a 400-hour program. Pronunciation training represented only 10% of all language related sequences and solely targeted individual sounds. Moreover, authors reported that pronunciation-related content mainly took the form of student-specific corrective feedback rather than a specific sequence embedded in the pedagogical progression.

In Europe, Henderson et al. (2012) conducted a large-scale survey on ESL teaching practices across seven countries (Finland, France, Germany, Macedonia, Poland, Spain, Switzerland). A vast majority of the 459 respondents (mainly teachers in the public school system and some in private language schools), reports near absent training in phonetics and pronunciation. Responses indicated that although teachers give great importance to pronunciation in general, it is often not equal to other language skills. Communication being the main goal for their learners, pronunciation is viewed as a mean of communication and is not a focus in itself, making it a low priority.

As for the teaching of French as a Foreign Language (FFL), survey data on teachers' practices is not available. However, teacher-researchers report similar tendencies in their publications (Abel, 2018; Alazard, 2013, Detey & Durand, 2021). For Billières (2014), phonetics is the black sheep of L2 didactics for the same reasons mentioned above: teachers lack training therefore legitimacy, but they also lack resources.

Aside from the teacher-training issue, language programs' goals regarding speaking skills bare an important responsibility. Oral production was the main point of focus in Audio-Oral (in the 1940's) and Audio-Visual (in the 1960's) methods. These relied heavily on repetition and imitation exercises, individual practice in language labs, and sessions of phonetic correction - the goal being the attainment of native-like pronunciation.

Since the 1970's-1980's, the communicative approach had become the norm and with it, oral activities shifted focus from pronunciation skills to effective communication and interaction skills. Through this approach, the native model is left behind in favour of intelligibility (Levis, 2005). Exercises take the form of role play and specific situation simulation, the goal for the learner being, above all, to communicate effectively in any given situation. In that context, pronunciation instruction is viewed as being too demanding to implement in class (from the teachers' perspective) for limited results on the learners' side (Billières, 2008; MacDonald, 2002).

In line with the pedagogical aims of the communicative approach, nowadays the development of speaking skills is regarded a lot of the time as the observable expression of the learners' capacity to communicate, rather than an object of study itself - a mean rather than an end. In the context of foreign language classes in primary and secondary schools (in Switzerland), Lauzon et al. (2009) talk about the fact that spoken language is a learning goal but the learning process rests entirely on the learners' side. Teachers provide space for practice through activities that require interactions (role play, debate, presentations...) but the technicalities of speech - the underlying grammar of oral language, i.e., the prosodic structure and rules, and the phonemic specificities - are not explicitly worked on.

This connects with the fairly old assumption that spoken language is learned implicitly, through exposure and practice in interactions, following the same processes as L1 acquisition (Darcy, 2018; De Pietro & Wirthner, 1996; Lauzon, 2009). In the end, the conjugation of such belief with the emphasis on interaction in the communicative approach, and the intelligibility principle (Levis, 2005) partly explain the fading of explicit and dedicated pronunciation instruction in language classes. As Breitkreutz et al. (2001) put it:

> "The advent of the communicative approach to language teaching marked the decline of pronunciation instruction." (p.52).

Authors have also mentioned the lack of clear descriptions of the oral language systems, arguing that oral language has not been studied as much as written forms (De Pietro & Wirthner, 1996; Dolz et al., 1998). While that argument could stand in the 90's, research on oral language and speech has blossomed since then, and the issue does not seem to be so much the lack of descriptions anymore, but rather the absence of a transfer between speech-related research findings and practitioners - whether language teachers or speech therapists.

Overall, Burgess & Spencer (2000) sum up the issues and questions teachers and program coordinators face when designing a pronunciation course:

"1. the selection of features of pronunciation;

2. the ordering of the features selected;

3. the type(s) of discourse in which to practice pronunciation;

4. the choice of methods which will provide the most effective results; and

5. the amount of detail to go into at different stages."

In the next section, we present an overview of pronunciation teaching methods available to teachers.

2. METHODS AND TOOLS FOR TEACHING PRONUNCIATION

Walking through the aisles of a university library in the foreign language method section is one way to witness the under-representation of resources focusing on pronunciation as compared to any other linguistic aspects (grammar and vocabulary mainly). But however small, a selection of pronunciation manuals is available. The size of this selection depends in part on the target language. As always, it seems that a greater volume of resources is available for English than for any other language.

When comparing two of the most used pronunciation manuals of similar editions for English as a Second Language (ESL) and French as a Foreign Language (FLE)[10] (at least in France), we notice similarities and differences in the contents. Both manuals include a unit for each sound of the phonemic inventory of each language, and both present vowels before consonants. However, the differences lie in the sections that concern suprasegmentals.

In the ESL manual, there are as many units on prosodic aspects, as there are on segmentals. Prosodic aspects include syllables, word-stress, intonation and rhythm in sentences, links between prosody and syntax, and pragmatic functions of intonation. Conversely, in the FLE manual, there are three units on rhythm and accents, intonation, and between-word linking, totalling to 15 pages.

This observation effectively illustrates the disparity between pronunciation instruction in English, which is supported by extensive research, and in French, which is far less studied.

Another difference is the order in which the segmental vs suprasegmental units are presented. In the ESL manual, suprasegmentals come after segmentals, which might suggest a priority given to individual sounds. In contrast, in the FLE manual the three units on prosody are presented before those on segmentals. This is in line with one of the principles of the Verbo-Tonal Method (Guberina, 1956, 1975) which gives priority to prosody (this method is presented in the next section).

[10] Marks, J. (2007). *English pronunciation in use: Elementary: self-study and classroom use*. Cambridge University Press.
Abry-Deffayet, D., & Chalaron, M.-L. (2010). *Les 500 exercices de phonétique: Niveau A1-A2*. Hachette français langue étrangère.

These two manuals also have in common the type of practice exercises: a mix of perceptive tasks such as sounds discrimination, and production such as repetition tasks. At the beginning of each section introducing a new sound, an illustration explains how to position the lips and tongue to achieve a successful production of the target sound. This shows that these pronunciation manuals mostly rely on probably the most wide-spread pronunciation teaching approach: the articulatory approach.

2.1. EXISTING METHODS

2.1.1. The articulatory approach

This method emerged from the work of phoneticians such as Paul Passy, Henry Sweet, and Wilhelm Viëtor, who created the International Phonetic Alphabet (IPA) at the turn of the 20th century. The foundation of the International Phonetic Association brought together phoneticians who greatly influenced the *Reform Movement* in foreign language teaching by putting pronunciation in the foreground, and providing the IPA as a tool for teaching the sounds of languages (Celce-Murcia et al., 2010).

The articulatory method brakes away from Intuitive-Imitative approaches where the learning process is implicit and conditioned only by the learners' exposure to the language sounds, and their ability to replicate them. On the contrary, the articulatory approach advocates for an explicit instruction of sounds formation. Manuals use illustrations and charts (Figure 12 below) to explain how to position the vocal apparatus in order to produce the target sound, and teachers are encouraged to demonstrate.

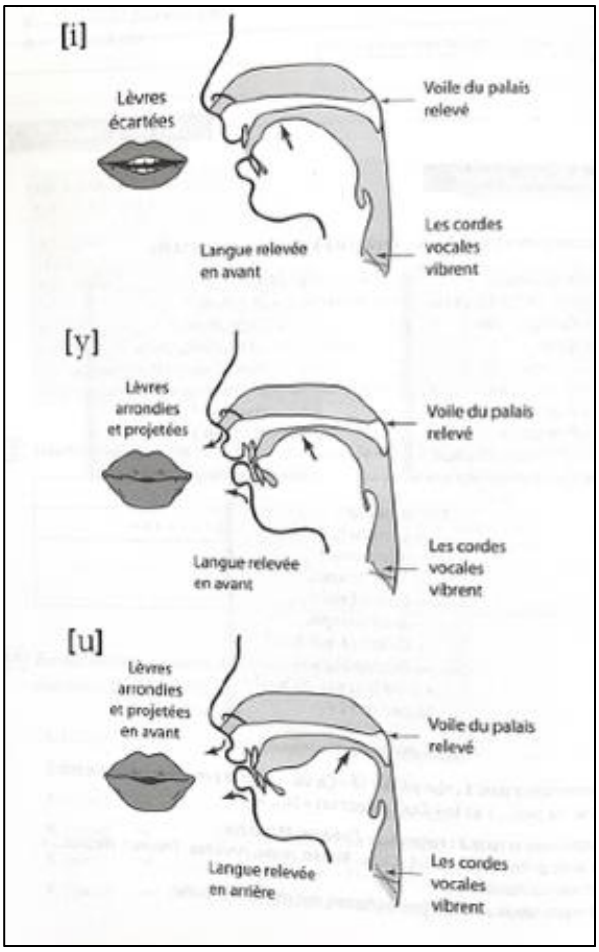

[i]
Lèvres écartées
Voile du palais relevé
Les cordes vocales vibrent
Langue relevée en avant
[y]
Lèvres arrondies et projetées
Voile du palais relevé
Les cordes vocales vibrent
Langue relevée en avant
[u]
Lèvres arrondies et projetées en avant
Voile du palais relevé
Les cordes vocales vibrent
Langue relevée en arrière

Figure 12 - Articulatory shape of specific sounds in French. (From Abry-Deffayet & Chalaron, 2010, p. 36).

Once the learners have become familiar with the articulatory movement associated with the target sound, they are encouraged to practice. First the sound is produced in isolation, then inside words, and finally sentences. Exercises always involve discrimination tasks between the target sound and another resembling phoneme. For example, if the target sound is the French /y/, it will be contrasted with the neighbouring phoneme /u/ in the minimal pair "sûr - sourd" [*certain - deaf*]. Then the two sounds will be presented in a sentence such as "il est sûr d'être sourd." [*He is certain to be deaf.*]. This type of minimal-pair drill has been used since the 1950's, and is still a very common exercise found in a majority of pronunciation manuals.

In essence, this method presents the advantage of being easily implementable in the classroom, and it also does not require the teacher to have an extensive training in phonetics. While using the IPA might facilitate the process, it is by no means indispensable. However, this method solely focuses on individual sounds and their articulatory shape. By working on the phonemes only, it does not take into account the reality of the syllable as the minimal unit of production (Massaro, 1972, 1974; Mehler et al., 1981; cited in Alazard, 2013), and therefore is not concerned with the phenomena of co-articulation, linking and such.

Moreover, this method does not involve suprasegmental elements and does not address the relationship between the segmental and suprasegmental level. In sum, if the goal is to improve the learners' pronunciation as a whole (segmentals and suprasegmentals) while taking into account realistic speech production constraints, the articulatory approach does not suffice and should be completed or extended to suprasegmental aspects and work on the syllable unit.

2.1.2. The Verbo-Tonal Method (Guberina, 1956, 1975)

The Verbo-Tonal Method (MVT[11]) contrasts with the articulatory one on many grounds. It was developed in the 1960's by Guberina and Rivenc, in the framework of the Structuro-Global Audio Visual (SGAV) methodology for learning languages. The guiding principles of this methodology recommend to give priority to oral communication and its vocal, verbal, and gestural aspects; and to delay the presentation of the language's written form in order to create an initial phase of oral language only, using audio recording and picture sequences (Renard, 1979; Cuq, 2003; cited in Alazard, 2013).

The MVT was initially developed as a form of speech therapy for ear-deficient people. But a parallel with language learners' experience with L2 phonetics soon emerges. Both populations process the sounds through their own system (c.f. the phonological filter from Troubetzkoy, 1939). Therefore, the main principle of the MVT is to re-shape the learners' perception, a necessary pre-requisite to attain a correct production of the target sound.

The learners' pronunciation errors are interpreted as deviations on either the tension continuum (neuromuscular activity necessary to produce sounds) and/or the light/dark continuum. The latter concerns the harmonic composition of the sound, which corresponds to the distribution of the energy over the

[11] From its French version: Méthode Verbo-Tonale

frequencies. A light timber is when more energy is directed at higher frequencies, a dark timber is the opposite.

Learners' errors can be diagnosed as too/not enough tense, and too light/dark. The correction that will be proposed is based on this diagnosis. As such, the MVT offers a paradigm where the individuality of the learner is acknowledged, and the correction adapted.

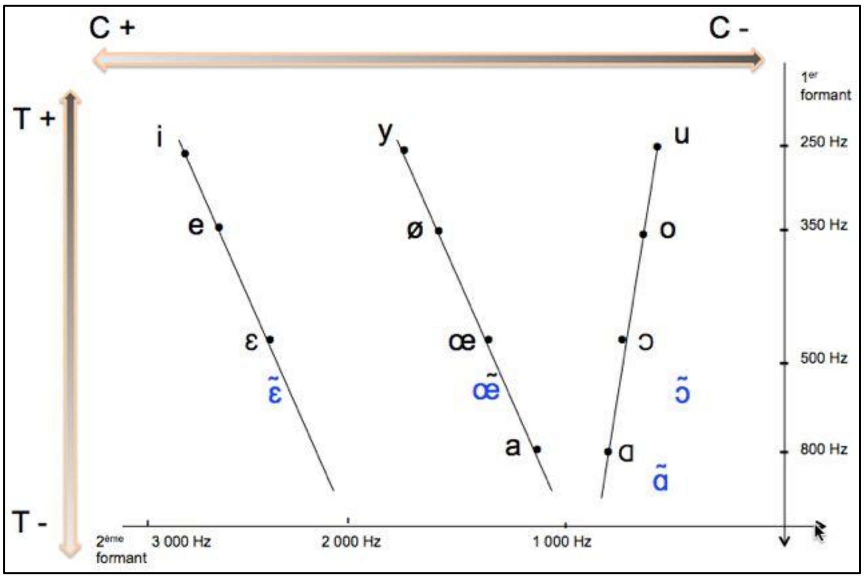


Figure 13 - Distribution of French vowels on the tense/lax (T+/T-) and light/dark (C+/C-) axes. From Billières, M. (n.d.). ***Méthode Verbo-Tonale: diagnostic des erreurs sur l'axe clair-sombre.*** **https://www.verbotonale-phonetique.com/methode-verbo-tonale-diagnostic-erreurs-axe-clair-sombre/**

A major difference with the articulatory approach is that the MVT considers prosody as a priority, since accentuation, rhythm and intonation influence how phonemes are realised. For instance, a light timber will be better perceived when placed in a stressed syllable or associated with a rising intonation contour. The MVT includes an introduction to the L2 prosodic system through exercises on rhythm and intonation, with the use of logatomes (replacing words by a single repeated syllable such as "dadada"), and visual representations of intonation contours drawn with the hand. The use of gestures to materialise suprasegmental elements is thought to facilitate their perception and production.

Lastly, the MVT relies on a non-explicit form of training where sounds and their articulatory shape are not intellectualised prior to produce them. Non-explicit learning would favour proceduralisation and automaticity (Paradis, 2004; Germain

& Netten, 2005; cited in Alazard, 2013). The multimodality this method uses reflects that of speech, and therefore constitute a rather holistic approach.

However, because the learner's error in its individuality is at the base of the process, this method seems to be only suited for working one-on-one or with small groups of students. It would not be adapted for a classroom filled with learners from a diversity of backgrounds. This method also requires teachers to be properly trained, and could not be implemented like the articulatory method, from just a manual. In addition, even though prosody is taken into account, it is used as a mean to facilitate the accurate production of phonemes, rather than an end in itself.

2.1.3. The Silent Way (Gattegno, 1972, 1976, 2010)

The Silent Way approach was developed by Gattegno in the 1970's for teaching ESL. This method made pronunciation accuracy a priority as it introduced sound instruction - both segmentals and suprasegmentals - right from the initial stages of L2 learning. One of the core principles is that the teacher's role is to unable and help learners develop their inner resources, rather than positioning themselves as the source of information and therefore placing the learners in a passive state (Stevick et al., 1998). Hence the name of the method, which refers to the attitude of the teacher: they should speak as little as possible. Instead, the teacher relies heavily on gestures, and visual materials to which they point.

Gattegno designed a set of visual charts and materials to be used in class: a *Sound Color Chart* (see Figure 14 below), *Fidel Charts* (colour/sound/spelling correspondence charts), *Word Charts* (selection of words written with the corresponding colours of each sound), *Cuisenaire Rods* (small rectangular blocks of different colour and length used to represent syllables, words etc.; see Figure 15), and a metal pointer to point at the different elements of the charts (Celce-Murcia et al., 2010).

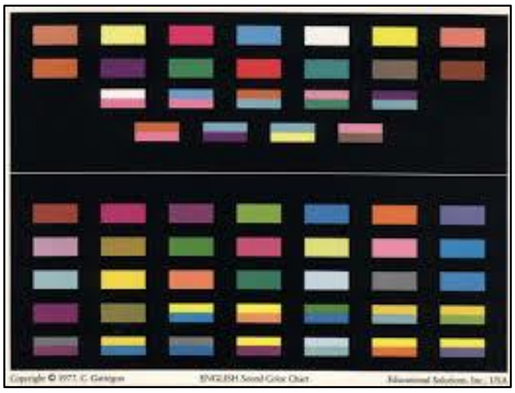

Copyright © 1977, C. Gattegno
ENGLISH Sound Color Chart
Educational Solutions, Inc. USA

Figure 14 - Sound Colour Chart used in the Silent Way method. Each block of colour is associated with a phoneme, two-colour blocks are diphthongs.

Figure 15 - Cuisenaire Rods used in the Silent Way method to represent syllables, words, or chunks.

In the sound colour chart in Figure 14, each block of colour is associated to a phoneme. The top part contains vowels, the bottom part consonants, and two-colour blocks represent diphthongs. The teacher would for instance ask the students to produce a sound, and would point to the corresponding block so students can gradually learn the colour-sound associations. Then by pointing at the blocks, learners would produce the corresponding sounds, which, in combination would start to form words. The rods can be used to teach suprasegmentals, a sentence can be materialised with one block for each word and the teacher can ask the students to place the stressed words higher than the other blocks for example.

The Silent Way method presents multiple advantages: it can be used for any language, by using the sound colour chart at initial stages of L2 learning it bypasses the potential influence of spelling, the attitude of the teacher promotes the learners' autonomy and involvement in the learning process. However, it requires for the teacher to be properly trained, and to acquire the necessary set of materials. It would be interesting to gain some insight into students' reaction and feedback on this method, as well as supporting empirical evidence.

2.2. TOOLS AND TECHNIQUES

2.2.1. Gestures and embodiment

Co-speech gestures

In Chapter I (p. 27), we have mentioned the link between speech rhythm and motor rhythm, and introduced some of the fundamental theories describing the close relationship between speech and body-movements. Considering gestures and facial mimics alongside speech as essential in the language production system leads to a conception of speech as being multimodal. One of the first advocate of this conception, Kendon (1972), considered that both language production and processing included a co-verbal dimension.

For some authors, co-verbal elements are mainly useful to the speaker as they help the encoding process. In reference to language production models (such as those presented in Chapter II, section 1.1, p. 58), movement activation occurs late in the process, at least after the conceptualisation phase (Krauss et al., 1995, 2000; Hadar & Butterworth, 1997; cited in Coletta, 2007).

Conversely, others consider that co-verbal elements also bear a communicative function, in addition to their function in the encoding process. According to this view, motor activation is programmed early in the language production process, possibly in the conceptualisation phase (McNeill, 1992, 2000; Kita, 2000; cited in Colletta, 2007). Several empirical studies support this latter conception, showing that auditors in an interaction integrate both visual (gestures and mimics) and auditory (speech) cues (McNeill, 1992; Thompson & Massaro, 1994; Beattiee & Shovelton, 1999; cited in Colletta 2007; Kendon, 2004; Zwaan, 2004).

McNeill's (1992) hand-gesture typology has become a reference in research concerned with speech multimodality. It allows to distinguish between different types of gesture and their associated functions. *Iconic* gestures serve as a direct representation of concrete objects or ideas, *metaphoric* gestures are associated with abstract ideas, *deictic* gestures depict spatial relations (such as pointing to a real/abstract object), and finally *beat* gestures are associated with speech rhythm and prosodic prominence (see Table 8 below). In contrast with all other categories, beat gestures are not associated with the semantic aspects of the message, rather:

> "The semiotic value of a beat lies in the fact that it indexes the word or phrase it accompanies as being significant, not for its own semantic content, but for its discourse-pragmatic content." (McNeill, 1992, p. 15)

Gesture dimension	Definition	Example
Iconic	Represent what is talked about	Flapping arms like wings when talking about a bird
Metaphoric	Represent an image of an abstract concept	Making a movement with straight parallel hands when describing someone as strict or square
Deictic	Pointing at abstract or concrete entities	Pointing while giving driving directions
Beats	Rhythmical movements, typically biphasic, with a highlighting function	Flick of the hand/arm up and down or back and forth

Table 8 - Summary of McNeill's gesture typology.
Compiled from McNeill (1992) and Berry (2009)

Gestures for teaching

It is likely that language teachers use gestures when teaching pronunciation more or less intuitively (Baills, Rohrer, et al., 2022). Drawing a contour with the hand matching the intonation of speech (metaphoric gestures) for instance, or tapping in synchronisation with accented syllables (beat gestures) to enhance the perception and production of prominences can come to mind fairly spontaneously to some.

Gilbert (1978)'s publication on tools and gadgets for teaching pronunciation was quite influential. The author advocated for a multimodal approach to pronunciation teaching, encouraging the use of rubber bands to materialise syllable or vowel length, and kazoos to illustrate pitch variations (Brinton, 2017).

For teaching the pronunciation of French as a Foreign Language (FFL) , Briet et al. (2014) published a manual compiling classroom activities - some of which seem to be inspired by Gilbert's - that heavily rely on iconic and beat gestures, as well as body movements. It also includes activities that use songs, and the book covers both segmentals and suprasegmentals. This manual is one of the resources used to build the prosodic training for our study, therefore examples of such activities are presented in section 2.1.2. of Chapter V (p.190), and in Appendix 1 (p. 303).

Gestures used by the teacher have been described and categorised by Tellier (2006, 2008) - amongst others - who refers to them as *pedagogical gestures*. These include head, hands and arms movements, and facial mimics, and assume three main functions: *manage* the class, *assess* the learners, and *inform* on linguistic elements. Information gestures include *phonological and phonetic gestures* which are used specifically for teaching pronunciation.

Audiovisual corpus analyses have shown that teachers use horizontal hand gestures with upward/downward/flat trajectories to depict melodic contours (Tellier, 2008), lateral hand or body movements can be used to materialise vowel length (Hudson, 2011; cited in Baills et al., 2022), and beat gestures such as clapping or tapping are used to mark rhythm, syllables, and prominences (Chan, 2018; Baker, 2014; Hudson, 2011; cited in Baills, 2022).

Body-movements

Hand gestures are something that everybody uses on a daily basis, which makes their implementation in the classroom fairly seamless as teachers are only re-purposing and bringing awareness to something that is already familiar to the learners. However, movements involving the whole body can also be used as a learning tool. Instead of tapping (on the knees, or just with a hand on the desk) the syllables or prominences, we can engage the body into walking (on the spot or moving if space allows it) on the syllables, or walking/stopping according to different level of prosodic edges for instance. Such exercises seem to be more rarely implemented than hand/arm gestures.

The latter can easily be used in a sitting-behind-a-desk position, the usual classroom setting; whereas body movements require an adapted space. Some examples of exercises engaging the whole body can be found (but not exclusively) in the MVT framework (see https://www.fonetix.fr/pas-a-pas/ for a video example), in a method developed for English pronunciation: the Essential Haptic-Integrated English Pronunciation framework (EHIEP; presented below), and in the work of individual teacher-researchers such as Chan (2018) for English, and Llorca (2001) for French. It is still unclear, however, if engaging the whole body vs using only hand gesture makes a difference in terms of successful learning. Nevertheless, studies have shown the benefits of embodiment for learning

Embodied learning

The positive influence of action and body engagement for cognitive growth and learning has long been recognised in the field of developmental psychology (Held & Hein, 1963; Piaget, 1952; cited in Kontra et al., 2012). Indeed, the development of young children is deeply connected to their sensorimotor experience of the world. Nonetheless, the benefits of involving gestures and movement in learning continues throughout the lifespan (Kontra et al., 2012). Studies have shown that encouraging motor involvement of students enhance learning outcomes (Bahnmueller et al., 2014; Smith et al., 2014; cited in Baills, 2022).

Essentially, the use of gesture and more generally engaging the body can benefit the learning process in two ways: when the teacher uses gestures to materialise prosody, adding the visual modality helps the learners' awareness and perception of the target features.

When the learners reproduce the gestures along with the vocal material, they are themselves creating a pathway between their gestural materialisation and the speech features, gaining control and enhancing intelligibility and memorisation - as posited by the embodied cognition and learning theory (Kushch, 2018; McCafferty, 2006; Smotrova, 2017).

The benefits of using gestures to teach L2 prosody and pronunciation are supported by recent work, which display promising pedagogical implications (Baills, 2022; Cavicchio & Busa, 2023; Kushch, 2018; Maastricht et al., 2019; McCafferty, 2006; Smotrova, 2017), these are presented in section 3.3 of this Chapter, p. 158.

The Essential Haptic-Integrated English Pronunciation framework (Acton et al., 2012)

Acton et al. (2012) developed a method that feeds from the embodied cognition and learning principles, as it puts body movement and touch in the foreground. The Essential Haptic-Integrated English Pronunciation framework

(EHIEP) calls for the use of *pedagogical movement pattern* (PMP; similar to Tellier's pedagogical gestures) to present, correct, and practice new sound features, and to enhance their recall and integration into spontaneous speech.

> "'Haptic' in this context refers to systematic hand movement across the visual field accompanying speech that typically terminates in a touch of some kind, like one hand touching the other. That touch occurs simultaneously with the articulation of a stressed syllable of a word, focal stress of a phrase or a prominent word in discourse." (Acton et al., 2012, p. 234)

The method focuses on a selection of "essential features of basic English pronunciation" (p.237), however the authors suggest that haptic-based exercises can be extended to any phonological structure, included that of languages other than English. The EHIEP covers vowels and consonants, stress, intonation, rhythm grouping, conversational rhythm, and discourse features; much like most pronunciation instruction syllabus (such as Celce-Murcia et al., 2010).

Each phonological target is associated to a specific PMP which the main author has recorded in video[12]. Interestingly, movements are not limited to the hands and arms, some exercises include stepping, dancing, or shaking an object in time with pre-defined linguistic targets, therefore engaging the whole body.

However promising, it does not seem like this method has been officially formalised and published, and to the best of our knowledge, it has not been tested experimentally.

2.2.2. Musical activities

The use of songs and music is not uncommon in L2 classrooms, and we have already mentioned some authors who include rhythmical exercises and songs into their practice, usually coupled with some form of embodiment (Acton et al., 2012; Briet et al., 2014; see previous section).

A quite well-known ESL method based on musical activities, still used today, is the one developed by Graham (1978): *Jazz Chants.* It includes practice of

[12] It does not seem that the video bank has been made accessible, the only resource we could find online was the main author's blog https://hipoeces.blogspot.com/2013/03/?m=0 on which demonstration videos are no longer available.

little dialogues in the form of poem and songs, with a special emphasis on rhythmic and intonation patterns.

Indeed, music and language have a lot in common: their processing at the neural level share common resources (Koelsch et al., 2002; cited in Jekiel, 2022), they both rely on melody and rhythm and share the same acoustic parameters of duration, frequency, intensity, and timber (Fadiga et al., 2009; cited in Jekiel, 2022; Chobert & Besson, 2013), and they are both composed of sound units organised hierarchically and allow an infinity of structures (Fenk-Oczlon, 2009; cited in Jekiel, 2022). These similarities led to the development of research around the relation between musical aptitude and language learning.

In their review, Chobert & Besson (2013) present several studies supporting the positive effect of a musical training on L2 sounds processing and production. It appears that musicians outperform non-musicians in the perception of subtle pitch and duration variations, the production of unfamiliar phonological contrasts, and the segmentation of an unknown (artificial) language (Marques et al., 2007; Besson et al., 2007; Sadaka & Sekiayama, 2011; Slevc & Miyake, 2006; Milanov et al., 2010; François & Schön, 2011; cited in Chobert & Besson, 2013).

The OPERA hypothesis proposed by Patel (2011, 2012, 2014) explains the relationship between musical abilities and speech processing as the result of the shared nature of processes involved in both domains. Furthermore, Chobert & Besson point out:

> "It may be that musical expertise refines the auditory perceptive system (bottom-up facilitation), but it may also be that years of intensive musical practice exert top-down facilitatory influences on auditory processing [...]." (p. 928)

Thus, it is one thing that musically trained people - that is, people who already have a certain level of music expertise prior to learning an L2 - have an advantage over non-trained ones regarding L2 sounds acquisition, it is another to show that a musical training parallel to or within the language class could help the L2 acquisition process. To our knowledge, very few studies have addressed this question, but those that have generally indicate that a musical training benefits linguistic skills.

François et al. (2013) conducted a 2-year longitudinal study on 8-year old children and showed that children following a musical training improved their speech segmentation skills whereas children following a painting training did not. Similarly, Linnavalli et al. (2018) found that 5-year old children who attended a once-a-week musical session for two years improved their phoneme processing skills as compared to children who did not attend the sessions. As of today, studies looking at the effect of a musical training on the production of L2 sounds in older population seem scarce.

Addressing this gap, Zhang et al. (2022, forthcoming) recently conducted a study testing the effect of an embodied musical training as opposed to a musical training without embodiment, on L2 phonological productive skills of Chinese adolescents learning L2 English. The results showed that the embodied-musical training group performed significantly better than the non-embodied-musical training group on a foreign language imitation task and an L2 reading task (this study is described in more details in section 3.4. of this Chapter, p. 167).

The Jaques-Dalcroze method

The embodied-musical activities used in this study were inspired by a method developed by composer, musician and music pedagogue Émile Jaques-Dalcroze at the beginning of the 20th century. At the time, music theory was taught primarily in a scholar fashion, through music reading and the teaching of musical concepts in abstract ways, without involving much of the students' senses.

Jaques-Dalcroze felt that these techniques did not allow students to develop their ear and inner feel for melody and rhythm (Juntunen, 2016). He was convinced of the necessity to involve the body in order to build a sensorial appreciation of music, especially rhythm. He developed exercises that integrated mind and body, for a multisensorial experience of the music, a decidedly embodied approach to learning music and musicality: the *Méthode Jaques-Dalcroze* (1906).

Movements in a Dalcroze sequence would include functional movements (using the hand to represent pitch variations), and rhythmic, creative, dramatic and dance moves. Movements are in connection with listening, and analysing different aspects of music (tempo, meter, phrase, harmony etc.).

Exercises sometimes use props such as balls, sticks or scarves, and some include vocalisation and singing (Juntunen, 2016; Y. Zhang et al., forthcoming). Students can be asked to walk/clap/jump to a beat, mark some aspects of the melody with a move (e.g., turn when you hear a C-note), improvise a sequence of moves on a specific phrase (Mead, 1994; cited in Juntunen, 2016).

Studies have brought supporting evidence that Dalcroze-like activities enhance the development of musical skills (Leman, 2007; Leman et al. 2018; cited in Zhang, forthcoming). Considering the relation between musical abilities and L2 phonological skills, Zhang et al. (2022, forthcoming) went one step further and demonstrated how Dalcroze-like musical exercises can boost L2 pronunciation. Based on these findings, we decided to incorporate a sequence of Dalcroze-like musical activities in the prosody course we designed for our study (see next Chapter, p. 179).

2.2.3. Explicit vs implicit instruction

The type of instruction can also be characterised in terms of explicit vs implicit teaching/learning. In an influential meta-analysis on the effectiveness of L2 instruction methods, Norris & Ortega (2000) define explicit and implicit instruction based on DeKeyser (1995): **explicit instruction** involves the explanation of rules (deduction) and/or for the teacher to directly ask the students to pay attention to specific forms and find a generalisation rule themselves (induction); **implicit instruction** conversely, corresponds to the absence of rule presentation and explanation, as well as the absence of any kind of direction of the students to particular forms.

While L1 acquisition is predominantly implicit (at least until children start school), L2 acquisition can be either or, depending on the context. Implicit acquisition of an L2 could occur if the L2 is learned solely from exposure, in a full immersion experience for instance. More often, L2 acquisition is mostly explicit, in a classroom setting and/or with the use of resources such as text books.

Another, and closely related way to define the type of instruction is the terminology introduced by Long (1991; cited in Norris & Ortega, 2000), later adapted by Doughty & Williams (1998)[13] and relayed in recent publications on the matter (e.g., Saito et al., 2017; Saito & Plonsky, 2019), which distinguishes three types of instruction: *focus-on-form, focus-on-formS,* and *focus-on-meaning.*

Focus-on-form (also referred to as *form-focused* or abbreviated into *FonF*) refers to a type of instruction where the teacher draws the learners' attention to specific forms in controlled contexts and also integrates those forms into communicative activities (e.g., working on the intonation of yes/no questions in isolation, with logatoms for instance, and then practice it in a dialogue).

Focus-on-formS (abbreviated into *FonFS*) corresponds to working on specific forms but in isolation, without integrating them into meaningful communication contexts. (e.g., working on a phoneme with the articulatory approach only).

Lastly, **focus-on-meaning** instruction corresponds to listening and speaking exercises that do not draw the attention of the learners on any specific forms (e.g., listening comprehension activity, role pay...).

If we think of all the methods and approaches we have described in the previous sections, it appears that the explicit/implicit opposition is not a strict one, but rather should be understood as a continuum:

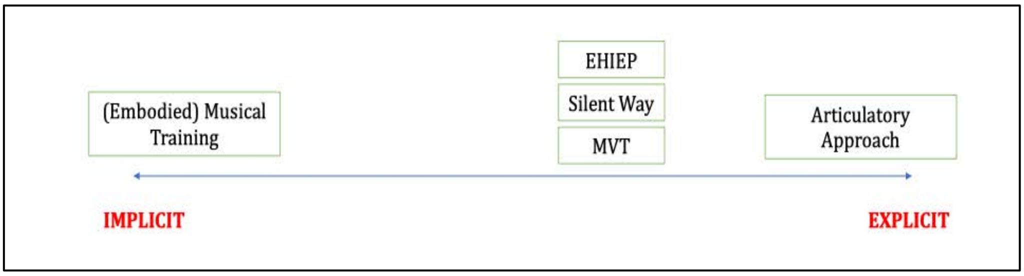


Figure 16 - Implicit/explicit continuum of pronunciation instruction methods. (According to DeKeiser's definition of the concepts)

The embodied musical training used in Zhang et al. (forthcoming) is fundamentally implicit since it does not even focus on language itself, let alone phonological rules and forms. The MVT, Silent Way, and EHIEP methods however, could be categorised as form-focused, as they involve drawing the learners'

[13] Other authors such as Spada (1997) and Ellis (2006) have used the same terminology with slightly different definitions, especially for the focus-on-form instruction type, as to what kind of practice it includes or not.

attention to forms without necessarily including the explanation of the underlying rules. The articulatory approach combines both attention-drawing and rule explanation, however in its purest form it does not integrate the forms into contexts, it therefore corresponds to Long's focus-on-formS instruction type.

A fair number of studies have compared more or less explicit L2 pronunciation teaching methods (e.g., Schwab & Dellwo, 2022; R. Zhang & Yuan, 2020), and recent meta-analyses of pronunciation training studies have concluded that **explicit instruction that include form-focused exercises integrated into communicative activities yield the best results** (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019). More detail on such studies is given in the following section of this Chapter (p. 158).

2.3. CONCLUDING REMARKS

The above review of methods, techniques, and tools that are currently available and used for teaching L2 pronunciation demonstrates the variety of approaches to choose from and/or combine, and the dynamism of L2 teacher-researchers around this topic.

Because our focus is on teacher-students classroom instruction, our presentation left aside student-autonomous learning techniques such as computer assisted methods (e.g., High Variability Phonetic Training; Thomson, 2011, 2012, 2018) , or mirroring/shadowing (repeating a passage from a native speaker while imitating them as faithfully as possible; Brinton, 2017).

Using gestures and encouraging learners to engage their body through specific movements or object manipulation is something that several approaches have in common. Indeed, the **benefits of a kinaesthetic approach for learning are supported by the embodied cognition and embodied learning principles**.

Furthermore, L2 sounds and especially suprasegmentals can be very abstract for the L2 learner initially, **materialising prosody through gestures, mimics, or working with music gives learners other means - maybe more tangible - to grasp and integrate these aspects of speech**. Besides, **movements and music bring energy to the class and boosts learners' attention,**

motivation, and involvement (Lo & Li, 1998; Fonseca-Mora & Gant, 2016; Kao & Oxford, 2014; cited in Zhang et al., forthcoming).

The goal of this exploration was for us to get inspired for **building an L2 French prosody course** for our experimental study. The resulted **pedagogic progression combines several of the above-mentioned techniques into a multimodal course**, which is described in the next Chapter (p. 190).

The following section focuses on experimental studies comparing different types of L2 pronunciation instruction/training, in order to present the reported effects of such instruction/training on L2 learners phonological productive and perceptive skills.

3. EFFECTS OF L2 PRONUNCIATION INSTRUCTION

3.1. META-ANALYSES: FOCUS AND TYPE OF INSTRUCTION, ELICITED SPEECH TASKS

In the early 1980's, the goal of L2 instruction research shifted from answering the question of whether L2 instruction made a difference in comparison to learning from naturalistic exposure, to the question of which type of instruction would be most effective for L2 learning (Doughty, 1991; Long, 1991a; cited in Norris & Ortega 2000).

Narrative and meta-analytic reviews on general L2 instruction (i.e., without distinguishing the different areas of competence of grammar, vocabulary, speaking skills etc.) revealed that explicit types of instruction are more effective than implicit types, and contextualised form-focused instruction is more effective than decontextualised (focus-on-formS) methods (Ellis, 2002; Norris & Ortega, 2000; Spada & Tomita, 2010).

Saito (2012) was the first to compile pronunciation instruction studies' results to give a synthesis on L2 pronunciation instruction effectiveness. 15 studies published between 1993 and 2012 were selected, all following a pretest-instruction/training-posttest design. 12 studies were conducted in pre-existing university classes, the remaining three specifically recruited participants and assigned them to either a control or experimental group. Nine studies concerned the acquisition of L2 English, four studies L2 Spanish, and one study L2 French. Saito looked at the effectiveness of L2 pronunciation instruction as a function of the focus of instruction (segmentals vs suprasegmentals), instruction type (focus-on-form vs focus-on-formS), and elicited speech style for measurement (controlled vs spontaneous speech).

Results showed that all studies reported improvement of learners' performance at posttest (T2) both on specific segmental and suprasegmental aspects, and comprehensibility ratings, except from two studies: one with a very short training phase (15 to 30 minutes), the other with learners achieving very high scores at pretest (T1), not leaving much room for improvement. Control groups who received meaning-focused instruction did not improve at T2. Furthermore, whereas focus-on-formS instruction led to improvement mostly on controlled elicitation tasks, it appears that focus-on-form instruction led to improvement also on spontaneous speech. Of all studies focusing on segmental aspects (n=5), only one showed improvement on spontaneous speech (Saito & Lyster, 2012; cited in Saito, 2012), all others only on controlled task. Suprasegmental-focused studies all showed improvement on controlled task too, and the only one that included a spontaneous speech task demonstrated improvement as well (Derwing et al., 1998).

Similar findings emerged from the meta-analysis conducted by Lee et al. (2015) which included 86 studies published between 1982 and 2013. In addition to supporting Saito's conclusions, it was found that learners' proficiency levels did not have an effect on the effectiveness of pronunciation instruction, laboratory-based trainings yielded stronger effects than in-class instruction studies, longer training phases led to larger effects, and instruction type that included feedback were more effective.

However, the authors raised the issue of the lack of details regarding the type of activities and material included in the trainings, which made it impossible to precisely analyse the effect of type of instruction. Finally, it was found that instruction blending both segmental and suprasegmental aspects led to stronger effects on learners' overall pronunciation than when only one of the two was targeted. As in Saito (2012), studies using controlled task (usually read speech) produce larger effects than those using spontaneous speech. The authors encouraged the use of more authentic speech samples in research in order to gain in ecological validity.

The most recent meta-analysis on L2 pronunciation instruction was conducted by Saito & Plonsky (2019). In contrast with the two previous reviews, the authors focused on the type of measurements employed instead of the instruction type. They argue that all studies use explicit instruction, whether focus-on-form or focus-on-formS, and that most studies do not provide enough detail regarding the teaching methods employed in order to form categories. Therefore, this review focuses on L2 pronunciation instruction studies' methodology rather than pedagogical approaches.

The authors hypothesised that the reported effectiveness of an L2 pronunciation training can vary depending on the type of outcome measures used. They looked in details at studies' results as a function of the measurement focus, i.e, global ratings (accentedness, comprehensibility, perceived fluency and intelligibility) vs specific measurements (such as acoustic measures or expert ratings of accuracy, number of errors etc.), the scoring method: subjective (human ratings) vs objective (acoustic measures), and the task type (controlled vs spontaneous speech).

77 primary studies published between 1982 and 2017 were included in the analysis. It was found that very few studies combine global and specific measures in order to gage the impact of pronunciation instruction. Similarly to Saito and Lee et al.'s conclusions, benefits from pronunciation instruction were especially robust when measured on controlled tasks and specific measures. This makes sense since controlled tasks such as sentence reading or repetition are usually what is used in the classroom to practice, and such tasks allow participants to focus on their performance. However, such results fail to inform on the impact of pronunciation training on participants' "real life" performance. Weaker effects were reported when the impact was measured through global ratings and spontaneous speech samples.

All three review confirm that **L2 pronunciation instruction does indeed benefit L2 learners, both on specific segmental and suprasegmental aspects, and global ratings**.

From a pedagogical perspective, it appears that **L2 pronunciation instruction seems to be most effective in the case of an explicit style of teaching (focus-on-form) that includes contextualisation of the targeted features and corrective feedback**, as opposed to meaning-oriented only activities (control groups). Moreover, **pronunciation proficiency as measured by global ratings seems to be most positively impacted when both segmentals and suprasegmentals are targeted**. Lastly, it seems that **instruction focused on suprasegmentals leads to improvement that carries over to spontaneous speech**, more so than segmental-focused instruction. However, these reviews also report an **under-representation of suprasegmental studies**, and encourage research to address suprasegmental aspects more extensively (see also Thomson & Derwing, 2015).

From a methodological standpoint, all three reviews conclude that **controlled tasks and specific measures lead to more robust results**. **However, all authors draw attention to the weak ecological validity of such results, and encourage future research to include spontaneous speech tasks, and combine global ratings with specific measures**. Lastly, Saito & Plonsky point out the lack of precision in the description of teaching methods used, which prevents the comparison of different types of instruction's effectiveness.

3.2. SEGMENTAL VS SUPRASEGMENTAL INSTRUCTION ON GLOBAL PRONUNCIATION

Since intelligibility and comprehensibility became the main goal of L2 pronunciation instruction (see Levis, 2005), one of the questions addressed in pronunciation instruction studies has been the respective weight of segmental vs suprasegmental errors on global evaluation of speakers' pronunciation (intelligibility, comprehensibility, accentedness, fluency).

Derwing et al., (1998) tested the effects of three types of instruction on 48 ESL learners: segmental-focused instruction, suprasegmental-focused instruction, no particular pronunciation instruction (i.e., meaning-focused). All participants followed a 20-hour per week ESL course during 11 weeks, which included a 20-minute per day special session for the segmental and suprasegmental groups. The segmental group worked on individual sounds using language-lab material and teacher-led exercises such as discrimination, identification, and repetition tasks. The suprasegmental group worked on intonation, rhythm, stress, fluency, using the *Jazz Chants* method (Graham, 1978), as well as exercises with logatomes and embodiment. Participants were recorded on a sentence-reading task and a picture-based narrative task at pretest and posttest. Native English speakers rated the recordings for comprehensibility, fluency and accentedness.

The results showed different outcomes depending on the task and instruction type. On the sentence-reading task, both segmental and suprasegmental groups improved in comprehensibility, whereas the meaning-focused group did not. All groups improved their accentedness, however the segmental group did so significantly more than the other two groups. Nevertheless, this tendency was reversed in the narrative task: only the suprasegmental group significantly improved in comprehensibility and fluency, while none of the groups improved in accentedness.

This seems to indicate that post-training segmental gains are difficult to transfer to spontaneous speech because the task requires attention in a variety of linguistic aspects, and not only specific forms. Therefore, learners cannot allocate enough resources to segmental realisation. In contrast, it does not seem to be the case for prosodic aspects, which apparently are successfully transferred over spontaneous speech. However, in addition to the focus of instruction, the instruction style differed too as the suprasegmental group benefited from musical and embodied activities, when the segmental group did not.

Gordon & Darcy (2016, 2022) showed the superior benefits of explicit L2 pronunciation instruction focused on suprasegmentals on the comprehensibility of L2 learners. In their 2016 study, they compared three groups of learners who received pronunciation instruction for three weeks, with 25-minute sessions three days a week. One group received explicit instruction on segmentals (four specific vowels), one group received explicit instruction on suprasegmentals (rhythm, stress, linking, intonation), the last group received non-explicit instruction on both segmentals and suprasegmentals (such as listen-repeat exercises without any explanations or guided practice). Participants took a delayed sentence-repetition task before and after training, and their recordings were rated for comprehensibility by native judges. Results showed that only the suprasegmental group significantly improved their comprehensibility after training.

In the 2022 study, the authors expanded the length of the training and included a wider range of pronunciation features (both segmentals and suprasegmentals). This time ESL students received a 30-minute training session per week during 10 weeks. Three groups were formed with one focused on segmentals, another on suprasegmentals, and the third one a mix of both. In contrast with their previous study, the three groups were matched in terms of instruction style. All received explicit instruction and practiced the forms studied in communicative activities (focus-on-form instruction). Therefore this time, the comparison concerned solely the focus of instruction.

Participants were recorded on a video retelling task before and after the training, and native listeners rated them for comprehensibility, accentedness and fluency. It was found that only the suprasegmental group significantly improved in comprehensibility and fluency. None of the groups improved in accentedness.

Similarly, Zhang & Yuan (2020) examined the outcome of pronunciation instruction focused on segmentals, suprasegmentals, and meaning-focused-only activities. 90 L1 Chinese students followed ESL classes twice a week for a total of 3 hours and 20 minutes per week. Embedded in this course, the segmental and suprasegmental groups received 35 minutes of pronunciation training twice a week during the 18 weeks. The instruction style was the same in the two groups who received pronunciation instruction, which consisted in focus-on-form activities and practice in communicative activities.

Students' speech samples were collected before the training, at the end of the course, and also 20 days after the end of the course. Students were recorded on a sentence-reading task, and a picture-based narrative task as in Derwing et al. (1998). Expert native English speakers (ESL teachers) rated the recordings for comprehensibility. Results showed that both the segmental and suprasegmental groups differed significantly form the meaning-focused group in the sentence reading task. In that condition, they both improved from pretest to posttest, pretest to delayed-posttest, and posttest to delayed posttest. However, in the narrative task, only the suprasegmental group significantly improved, both from pretest to posttest and from pretest to delayed-posttest (but not from posttest to delayed-posttest).

Taken together, these studies' results are in accordance with the conclusions from the previous section (p. 158). While both segmental and suprasegmental trainings lead to pronunciation improvement, it seems like suprasegmental gains are more transferrable to spontaneous speech than segmental ones. However, authors of these studies are not trying to argue that pronunciation instruction should focus only on suprasegmentals, rather, **these findings constitute an argument in favour of a more systematic and early integration of prosodic aspects, jointly with segmentals, in pronunciation instruction**, as highlighted in Wang's (2022) recent review.

3.3. EFFECTS OF PRONUNCIATION INSTRUCTION USING GESTURES

Research around the effect of using gestures to teach L2 pronunciation - especially suprasegmentals - has blossomed in the past 10 years, yielding mixed results. Studies focused on the use of metaphoric pitch gestures to teach L2 intonation and tones have either shown positive results (Baills et al., 2019; Hannah et al., 2016; Kelly et al., 2017; Yuan et al., 2019; Zhen et al., 2019), or limited effects (Morett & Chang, 2015; Zheng et al. 2018).

Since our work is focused on speech rhythm rather than intonation, this section focuses on studies comparing gesture-based instruction with non-gesture-based instruction on the perception and production of L2 rhythmic patterns.

Maastricht et al. (2019) investigated the effect of using metaphoric gestures vs beat gestures on the acquisition of lexical stress. 62 adult L1 Dutch speakers participated in one training session on L2 Spanish lexical stress. The participants were all complete beginners in L2 Spanish and were separated into three groups. One group received audio-visual only training, one group received an audio-visual training with metaphoric gestures, and the last received an audio-visual training with beat gestures. All were informed of the rules governing lexical stress realisation in Spanish and all training modalities were conducted through video-watching.

Participants were asked to read the same 28 short sentences in Spanish at pretest and posttest. Two expert raters coded the read sentences for correct of incorrect stress realisation and position. The results showed that the two groups trained with gestures performed better than the group who received audio-visual only training. Furthermore, the metaphoric gesture group performed better than the beat gesture one. However, between-group differences were found statistically non-significant. The authors concluded that the limited length of the training and the scoring method might explain the weakness of the results.

In contrast, Zhang, Baills and Prieto (2020) found a significant effect of a training involving hand-clapping to the rhythm structure of words using an acoustic measure (syllable duration). 50 L1 Chinese adolescents were tested on a word-repeating task in L2 French after following a 10-minute video training including 20 French words, either with or without hand-clapping matching the rhythmic pattern. Participants in the hand-clapping condition were asked to repeat the words and hand-clap, whereas participants in the other condition just repeated the words.

For the pretest and posttest, participants were asked to listen and repeat (without gesture) 14 French words, 10 of which were included in the training. Recordings were then submitted to two expert raters who assigned accentedness scores to each item, and a measure of syllable duration was extracted. Results showed that the hand-clapping group performed better than the non-clapping group. Specifically, the hand-clapping group obtained better accentedness scores although the between-group difference was only near-significant. However, the difference in syllable duration for the target final stressed syllables was significant between-group in favour of the hand-clapping group. This study shows that even a short training involving stress-synchronised gestures has a positive effect on stress realisation, and that effects are better reflected on acoustic measures than global scores of accentedness.

Similarly, Gluhareva & Prieto (2017) found that a short video training with beat gestures helped participants reach better (lower) accentedness scores on difficult items (longer, more complex sentences). 20 Catalan undergraduate students at an upper-intermediate level of L2 English participated in the study. The pre and posttest consisted of recording 12 sentences elicited from image prompts. In the training, participants watched videos of a US-English native speaker uttering the 12 sentences corresponding to the image prompts, half of the items with beat gestures falling on nuclear accents and intermediate stressed syllables, the other half without any gesture.

Five naive US-English native speakers, unfamiliar with Spanish and Catalan rated the recordings for accentedness. Results of the study showed that post training scores were significantly less accented than pre training scores, meaning that the training did lead to improvement no matter the condition. Additionally, beat gestures were shown to have a positive effect as they improved the participants performance, however only on prompts that were categorised as difficult (less usual situation depicted in the prompts, longer and more complex associated sentences). These findings suggest that seeing beat gestures may help L2 learners' perception and in turn production of prosodic prominence.

In a follow-up study using a similar design, Kushch (2018) investigated the effect of a training that involved observing beat gestures (as in Gluhareva & Prieto 2017), in comparison to a training where participant were asked to imitate the beat gesture. 18 native Catalan students, at an upper-intermediate L2 English level were separated in two groups. One received a video training with beat gesture (beat observation group), the other group saw the same training videos but were asked to imitate the beat gestures (beat production group). The same kind of prompt material as Gluhareva & Prieto was used. 10 prompts were used for the pretest, and the same 10 prompts with an additional 10 new items were used for the posttest.

Participants' recordings were rated for accentedness by five US-English native speakers.

It was found that participants' accentedness scores were lower at posttest in both conditions, and even more so in the beat production group. However, scores on the new items were similar to those at pretest which shows that benefits of the training did not transfer over to new items. These results suggest that producing beat gestures is beneficial for pronunciation learning, and has a stronger impact than only observing beat gestures.

The benefits of using (producing) gestures on L2 pronunciation were also supported by the results from Llanes-Coromina et al. (2018), again on L1 Catalan speakers learners of L2 English, after a short training.

All of the above-mentioned authors acknowledged that the training used in their design was very short, carried in laboratory settings, and the posttest directly followed the end of the training session (with a 10-minute break), which prevented from gaging any lasting effects of the training.

In order to test the impact of gesture-based pronunciation instruction in a more ecological setting, Baills et al. (2022) conducted a study where the training was embedded into a university intermediate-level French course. The study also intended to test the effect of using logatomes, a technique commonly used in the MVT framework (see section 3.1.2 of this Chapter, p. 143).

The intervention took place over five weeks with a total of three training sessions. 75 bilingual Catalan-Spanish students were separated into three groups. One group received pronunciation instruction by mean of speech only, one group with logatomes, and one group with both logatomes and gestures (observation only). The pretest and posttest consisted of a dialogue reading task, the four same dialogues were used. Participants recordings were rated by three expert native French speakers for accentedness, comprehensibility, fluency, segmental accuracy, and suprasegmental accuracy.

The results showed that all groups significantly improved at posttest in all five measures. However, higher effect sizes were found for the logatome-gesture group, and the same group improved significantly more than the speech group in regards to accentedness and suprasegmental accuracy, while the logatome group did not.

The effectiveness of the MVT - which includes segmental and suprasegmental training with gestures and logatomes - was in fact previously tested by the second author of the above-mentioned publication: Alazard (2010) found that an 8-week training with the MVT helped L1 English participants improve their reading fluency in L2 French, more so than participants who received an equivalent training with oral comprehension and production activities.

Later, the same author (Alazard, 2013) compared a 8-week MVT training with an articulatory training. Once again, the MVT group improved their reading fluency more than the articulatory one, especially so in the case of beginner learners, and before the introduction of written language in the course (i.e., after 4 weeks). As a matter of fact, both groups' performance declined after the introduction of written language in the course, indicating that introducing written forms alter learners' pronunciation.

Most research around **the impact of using gesture in L2 pronunciation teaching - either as a visual tool only, or as an embodiment technique for the learners - report positive outcome**. It seems like adding **visual and kinaesthetic elements helps learners improve their overall pronunciation** (accentedness, fluency) **and more specific suprasegmental features** (stress realisation, accuracy).

More empirical evidence is needed in order to determine what type of gesture might be most effective, in association with what speech feature, and how to implement them in various teaching contexts. The development of this field could eventually lead to the creation of pedagogical resources available to teachers around the globe.

3.4. EFFECTS OF PRONUNCIATION INSTRUCTION USING MUSICAL ACTIVITIES

Despite the close relation between musical abilities and L2 phonological skills (see section 3.2.2 of this Chapter, p. 151), to our knowledge, only a few studies have compared the effect of L2 pronunciation instruction based on musical exercises with non-musical instruction.

Fischler (2009) built a 4-week intensive pronunciation course designed to help L2 English intermediate-level adolescents improve their pronunciation, especially word and sentence stress. The course included explicit instruction of stress rules, and used rap songs with rhythmic and singing exercises to practice different stress patterns. Embodied rhythmic exercises were also included such as stretching rubber bands on vowel duration, sit/stand on stressed syllable, and beat drums.

The six participants (from various L1) were recorded on a reading task and a picture-based narrative task before and after the training. The ESL teacher coded the recordings for number of correct stress and overall intelligibility score, and three native English speakers also rated the recordings for overall intelligibility. The latter scores were better at posttest for all except one student in the reading task, and participants also increased their number of correct stress assignments. However, on the narrative task only two students improved their intelligibility scores. The author argues that this might be explained by the limited length of the training, which would not be enough to carry improvement over to spontaneous speech.

Most importantly, this study does not include a control group and uses a blend of musical and embodied activities. These results therefore do not permit to attribute the students' improvement solely to the use of music.

Good et al. (2015) compared the effect of learning a text in L2 English with practicing it as a poem vs as a song, in elementary school L1 Spanish children. Children participated in three practice sessions and were recorded afterwards on the same text. Their performance was coded for correct/incorrect vowels and consonants. The singing group clearly outperformed the poem group in the realisation of vowels, but no between-group difference appeared for the consonants. It should be noted as well that most children of the singing group actually performed the text singing in the posttest, whereas the poem group did not.

Nakata & Shockey (2011) also found that a pronunciation training based on singing was more beneficial than a training with listening activities in the mastery of L2 English syllable structures for L1 Japanese speakers. Trainings consisted of 20-minute sessions once a week for 10 weeks. The group who trained with songs significantly reduced the rate of epenthesis phenomenon (vowel insertion into consonant clusters) on a sentence-reading task.

In contrast, Nemoto et al. (2016) found that singing did not improve L1 Japanese's performance in L2 English. 30 students participated in a 10-minute training session where they practiced a short text (lyrics from a song) either by singing it or simply reading it aloud. Their performances were rated by 108 listeners for accentedness, intonation, and pronunciation. The singing group obtained worse scores than the reading group.

Ludke (2018) compared the impact of beginner L2 French classes supplemented with either songs and singing activities, or visual art and drama, in a school setting with L1 English pupils aged 12-13 years old. The sessions were spread over six weeks for a total of 8.5 hours for each group. Students were tested on a range of linguistic skills before and after training, by means of curriculum-based language tests. While both groups overall improved, more important progress was made in the singing group. Specifically, the largest differences in favour of the singing group were found for grammar, listening comprehension, conversational skills, and intonation. However, the two groups performed similarly on the pronunciation of isolated words.

More recently, Baills et al. (2021) compared the effects of a short video training involving either a song, or rhythmic speech (recited lyrics like a poem) on the learning and pronunciation of new French words. In both conditions the video showed the instructor accompanying words with gestures and facial expressions. 50 L1 Chinese high school students participated in the study and were recorded on a word-imitation task that included 14 words and four short sentences. Five expert native French speakers (teachers and translators) rated the recordings for accentedness. Both groups significantly improved at posttest, but the singing group showed a greater improvement than the rhythmic speech group.

Furthermore, in a second experiment using the same video material for the training (but different participants), the authors tested the effect of simply watching the singing video vs watching the same video and singing along. Accentedness ratings improved for both groups in a similar fashion, showing no differential effect of the treatments.

The results of the first experiment were replicated by Zhang et al. (2023), this time in L2 English. L1 Chinese adolescents participated in three 30-minute training sessions where they learned three English songs, either by listening to and reciting the lyrics without music (speech group), or listening to the lyrics sung and singing along (singing group). The pretest and posttest consisted of a word-reading task which was rated for accentedness, and a sentence-reading task which was rated for accentedness too, as well as comprehensibility, fluency, segmental accuracy, and suprasegmental accuracy. Results showed that the singing group outperformed the speech group in the word-reading task and the sentence-reading task, in all five measures.

In a recent study, Zhang et al. (forthcoming) went one step further and investigated the effect of a music-only training (i.e., instrumental music, without any focus on language) on foreign language pronunciation skills. Considering the relation between musical ability and phonological skills (see section 3.2.2 of this Chapter, p. 151), and the benefits of embodiment for learning, the authors hypothesised that an embodied musical training could have a positive impact on foreign language phonological productive skills through a transfer effect.

48 L1 Chinese adolescents were split into two groups. They all attended three 40-minute training sessions on three consecutive days. One group received a musical training that followed the Chinese music curriculum. The other received an embodied music training based on Dalcroze-inspired exercises (see section 3.2.2 of this Chapter, p. 151). The pretest and posttest tasks were carried out on the first and last day of training. They consisted in a speech imitation task where participants were asked to listen and imitate as faithfully as possible 12 sentences in six unfamiliar foreign languages, and reading aloud three sentences in English. Recordings were rated by three native speakers of each language.

Results showed that the embodied group improved their accentedness score significantly more than the non-embodied group on the foreign language imitation task. Moreover, in the read-aloud task, only the embodied group improved significantly in accentedness, comprehensibility, fluency, segmental and suprasegmental accuracy. These findings reveal the presence of a transfer effect between the gains from an embodied music training, and foreign language pronunciation skills. The encouraging results of this study inspired us to include embodied musical exercises in the prosody course designed for our study (presented in the next Chapter, p. 179).

Overall, **the integration of songs and singing activities seems to boost L2 pronunciation productive skills**, as compared to more "traditional" methods or even other types of artistic activities. This is not really surprising as the inclusion of music in the classroom has been shown to boost students' overall motivation and attention (Duarte Romero et al. 2012; Garcia Marrama, 2014; Wolfe & Noguchi, 2009; cited in Baills, 2021).

Furthermore, the results of Zhang et al.'s most recent study indicate that even a non-linguistic embodied musical training has a positive impact on non-native pronunciation skills. However, in most studies, trainings' outcomes are evaluated on controlled tasks only. This field of research would benefit from additional empirical evidence based on more spontaneous speech tasks.

3.5. EFFECTS OF PRONUNCIATION INSTRUCTION ON L2 LISTENING SKILLS

In Chapter III (p. 103) we presented the theoretical foundations explaining the essential role of prosody for speech segmentation, and the difficulties L2 learners face regarding listening to the target language. Based on those considerations, it is fair to assume that instruction on suprasegmentals might help L2 learners' listening abilities, and more specifically their capacity to identify word boundaries, by refining their bottom-up processing skills. This section presents recent studies addressing this question.

Kissling (2018) tested the impact of segmental vs suprasegmental, and perception-focused vs production-focused pronunciation instruction on L2 Spanish learners' perception of the target language, in terms of intelligibility and comprehensibility. 116 L1 English students, beginners in L2 Spanish were assigned to four groups corresponding to the four pronunciation instruction conditions. They followed four 20-minute sessions embedded in their Spanish course over one semester. Participants were tested before and after the training sessions with a listen-transcribe task. 9 sentences uttered by L1 Spanish speakers were used. The participants were asked to transcribe them (intelligibility) and to rate their comprehensibility.

After training, all groups improved their listening skills as shown by higher intelligibility scores and comprehensibility ratings. However, some differences appeared between groups. Intelligibility was especially reinforced for the group who received suprasegmental and perception-focused instruction. In contrast, comprehensibility ratings were higher in the group who received segmental and production-focused instruction. These findings suggest that, at the initial stages of L2 learning, focusing on suprasegmental features through perception-based practice better improves intelligibility of the target language, even though a segmental and production-based practice leads learners to find the target language more comprehensible.

Yenkimaleki et al. (2023) used a similar design to Kissling. However, they also added a training condition that blended segmental, suprasegmental, perception-based and production-based activities. They conducted the study on 120 L1 Iranian students, at an intermediate level of L2 English. All groups attended 10 sessions of one hour each, spread over five weeks. Six groups were formed. For the control group, the sessions consisted of listening comprehension exercises with no specific pronunciation instruction. Four groups also engaged in similar exercises but for one third of each session (20 minutes), they received specific instruction: either segmental or suprasegmental focused, coupled with either perception or production exercises. The last group received a blend of segmental, suprasegmental, perception-based, and production-based exercises (holistic group) equivalent in volume to the others.

Students were tested prior to the start of the training, immediately after the end of the training, and one month later. Different versions of the TOEFL listening comprehension task were used for all testing sessions. All groups improved at posttest and retained their gain at the delayed posttest. However, the holistic group significantly outperformed all other groups. Interestingly, the groups who worked with perception-based exercises improved more than the ones who worked with production exercises (although the difference was not significant). That variable was found more impactful than the focus of instruction (segmental vs suprasegmental).

Another study also found that instruction on suprasegmental features better improved learners' listening skills than instruction based on general listening exercises only. McAndrews (2023) tested the effect of the two types of instruction on 64 L1 Chinese (and one Japanese) learners of L2 English enrolled in a 16-week ESL intensive course. Students attended two 50-minute sessions within their course, either a training on prosodic paratone and phrasing, or listening-comprehension exercises. Students were tested on a paratone boundary identification task, a phrase boundary identification task, and a general listening proficiency test.

The results showed that the prosody-trained group outperformed the listening-trained group on both prosodic boundary identification task, and on the listening proficiency test, at posttest and four weeks later at a delayed-posttest. However, it should be noted that the type of task used in the prosodic boundary tests were similar to those included in the prosody training. Therefore, the prosody group had a familiarity advantage over the listening group. Nevertheless, the prosody group still outperformed the listening group on the general listening proficiency test, indicating the benefits of suprasegmental instruction for improving general listening skills.

Luu et al. (2021) also tested the impact of prosody-focused training on ESL learners listening skills. The authors designed their prosody course through an online platform which included small dialogues that students had to practice with accompanying gestures, inspired by the MVT method. They also practiced the melody of the sentence by hearing low-filtered pass audio to remove any semantic content and focus solely on the prosody. The course included shadowing exercises where students had to imitate the audio simultaneously.

65 L1 Vietnamese students, beginners in L2 English were split into two groups. One group followed the prosody course, while the other followed a course with general listening comprehension activities. Both courses lasted 10 weeks. Students took a listening comprehension test before and after their course. Both groups improved at posttest but the prosody-trained group significantly outperformed the control group.

Other authors have tested other types of training that did not involve pronunciation instruction, on L2 listening skills.

François et al.'s (2013) study (mentioned in section 3.2.2 of this Chapter, p. 151) is one of the rare ones that has focused on the link between music training and the development of foreign language speech segmentation skills. The study used a longitudinal design where 24 non-musician L1 French children (8-year-olds) followed 45-minute classes of either painting or music for two years, twice a week the first year and once a week the second year. They were tested before the start of the classes (T0), one year into it (T1), and after two years (T2).

The speech segmentation task used for all testing phase consisted in spotting trisyllabic pseudo-words in a stream of non-sense syllables. The target pseudo-words had been previously assigned a specific melodic contour and were presented in a learning phase where they were sung in a continuous way. Children then listened to two spoken words: a target pseudo-word, or another made up word containing the last syllable of one and the first two of another. Children had to choose which one of the two was more familiar.

The music-class group was the only one to improve their performance at the segmentation test, both at T1 and T2. These results corroborate those of Zhang et al. (forthcoming) in showing the transfer between musical and linguistic skills, here on the perception side.

Lastly, Charles et al. (2015) took a different approach. Their study looked at the impact of an intensive exposure through watching TV programs in the target language, either including subtitles (what they called *bi-modal input*) or not, on learners' segmentation skills. Instead of using a listening-comprehension test, they used a shadowing task.

The authors question the test construct validity of listening comprehension tests for measuring listening skills. They argue that because the comprehension questions in the tests are usually written, the test involves the students' reading abilities as much as their listening skills (Rost, 2002; cited in Charles et al., 2015). Furthermore, a listening comprehension task heavily relies on recall. Answers to test questions are not a direct access to what the student heard, but rather what they remember they have heard, or what they are able to reconstruct (Brown, 1990; cited in Charles et al., 2015).

Instead, a shadowing task does not involve any reading nor memorisation (at least only very short term): participants listen to short excerpts and have to immediately repeat whatever words they have just heard. The task is therefore not based on comprehension at all, only the capacity to extract individual words from a continuous speech stream, i.e., the ability to segment speech.

In their study, Charles et al. recruited 12 L1 Chinese ESL students at an intermediate-advanced level of English. They were separated into three groups, and all attended four 30-minute training sessions over four weeks, in which they watched documentaries narrated by a native English speaker. One group watched the documentaries with sound and subtitles, one group with sound only, and the last group with no sound and subtitles only. The latter group was introduced in order to control if a potential advantage in the bi-modal group would come from the subtitles only. Participants took a pretest, immediate posttest, and delayed posttest, consisting in a shadowing task based on excerpts from documentaries either included in the training, or new.

It was found that the bi-modal group outperformed both sound-only and no-sound groups, on known and new items. These findings suggest that exposure to bi-modal input in the form of video-subtitles helps improve learners' segmentation abilities, more than exposure without subtitles.

This literature suggests that pronunciation instruction, and especially **instruction focused on prosody, does help L2 learners improve their listening skills**. Besides suprasegmental instruction, François et al.'s study also demonstrates that **foreign language segmentation skills can be enhanced by a music training**. Lastly, while the benefits of exposure to the target language have been documented (Calhoun et al., 2023), Charles et al.'s study has shown the added value of bi-modal input exposure.

Given the scarcity of studies investigating this question, and even though our main focus was the effect of prosody instruction on speech rhythm in designing our experimental study, we did not want to miss the opportunity to test the effect of the training on our learners' segmentation abilities as well. Thus, **we replicated Charles et al.'s shadowing task using our own material**, and included it in our pre and post training tests (see Chapter V, section 2.3.3, p. 198).

CHAPTER SUMMARY

This Chapter's aim was to present a state of the art of today's practices in the classroom in regards to the development of L2 learners' speaking skills, and especially pronunciation instruction. **Even though there is a general consensus on the necessity to include pronunciation instruction in foreign language courses and classes, in reality it still tends to lag behind other linguistic aspects considered a higher priority** (Alazard, 2013; Billières, 2014; Darcy, 2018; Detey & Durand, 2021).

Furthermore, teacher trainings often fail to include sufficient knowledge and pedagogical techniques specific to phonological aspects and the teaching of pronunciation. **Language teachers have repeatedly reported to feel a lack of competence and therefore confidence to teach phonetics, as well as a lack of available resources on which to draw** (Breitkreutz et al., 2001; Foote et al., 2011, 2016; Henderson et al., 2012).

Notwithstanding, a variety of methods and tools have been developed over the past 50 years. **The articulatory approach** usually constitutes a base in pronunciation manuals, blending perception and production exercises, and explicit explanations. **The Verbo-Tonal Method** (MVT, Guberina, 1956, 1975) in contrast, uses a form-focused and embodied approach centred on the learner's error, with an emphasis on prosodic features of the target language.

As a matter of fact, **quite a few pronunciation teaching methods call for the use of gesture and embodiment** (e.g., the Silent Way, Gattegno, 1972, 1976, 2010; and The Essential Haptic-Integrated English Pronunciation framework, Acton, 2012). **Embodied cognition and learning theories have been at the base of research showing the enhancer effect of embodiment on learning**. Using gesture is particularly appropriate to work on sound patterns that can feel quite abstract for the learners (Kontra et al., 2012; McCafferty, 2006).

Generally speaking, **multimodal teaching of pronunciation makes a lot of sense since it reflects the multimodal nature of speech itself**. In addition, the bridges between music and language and especially **the close relation between musical abilities and L2 phonological skills have motivated the use of music in pronunciation teaching, with positive impacts on L2 phonological**

productive and perceptive skills (Baills et al., 2021; Chobert & Besson, 2013; François et al., 2013).

While empirical research on pronunciation instruction comparing the efficiency of different methods is still young, **we already dispose of solid data showing the benefits of explicit instruction, embodiment, and musical activities**. **We drew on these techniques to build a multimodal L2 French prosody course and tested its effects on speech rhythm, comprehensibility, accentedness, and also L2 segmentation skills of L1 English learners at an A2-B1 level.** The next two Chapters of this dissertation present our study's design and results.

CHAPTER V - STUDY METHODOLOGY

INTRODUCTION

This first section introduces our second study's theoretical background and exposes our research questions and hypotheses. Then, we explain the choices made in terms of methodology, specifically regarding the inclusion of data on L1 in our analysis, ecological aspects of this methodology, and finally the selection of measures used.

1.1. THEORETICAL BACKGROUND, RESEARCH QUESTIONS, HYPOTHESES

The last chapter presented the problematics of pronunciation teaching in language classes in general, and more specifically in French as a foreign language (FLE). What was presented regarding the state of pronunciation in language teaching constitutes the grounds for our second study's *raison d'être*. In essence, this study was motivated by:

- the will to contribute to filling a gap in the literature regarding the acquisition and teaching of L2 French prosody
- the desire to build an L2 French prosody course prototype, which could be a base for future research and pedagogical resources development
- the need to compare the effects of prosody instruction against what is commonly done in FLE classes i.e., communicative activities (oral expression and listening comprehension)
- the need to combine different types of measures to look at the effects of the type of instruction on several aspects of the learners' performance (acoustic measures of speech rhythm, native listeners' judgements, learners' segmentation abilities) which in turn allows to look for potential correlations between these measures and aspects.

The study's design involved the creation of two courses corresponding to the two training modalities: one prosody course vs one oral production/comprehension course. It also involved two distinct data collection phases.

Phase 1 consisted in the pre and posttest sessions (hereafter T1 and T2) that took place the week before and after the training. Participants were asked to record themselves on a reading aloud task in L1 and in L2, and on a free speech task in L1 and L2 as well. These recordings constitute our speech data. They also took a listen-and-repeat task, with the purpose of testing their speech segmentation skills (hereafter referred to as the segmentation task).

Phase 2 was carried out seven months later, and consisted of a comprehensibility and accentedness judgement task by native French listeners.

Several research questions guided this work:

RQ1: what is the effect of a prosodic training on L2 speech rhythm, and how does it compare to that of common oral expression and comprehension training?

RQ2a: what is the effect of a prosodic training on comprehensibility and accentedness, and how does it compare to that of common oral expression and comprehension training?
RQ2b: are comprehensibility and accentedness scores correlated?

RQ3: what is the effect of a prosodic training on segmentation abilities, and how does it compare to that of common oral expression and comprehension training?

The operationalisation of speech rhythm follows our vision of speech rhythm as a multifaceted construct (see Chapter I). Measures include aspects from all four levels of analysis: micro-level, meso-level, macro-level, fluency. They are detailed in section 2.4.1 of this Chapter (p. 202).

Before going into the details of the experiment design, we elaborate on the decision process behind our methodological choices.

1.2. METHODOLOGICAL CHOICES

1.2.1. Inclusion of L1 data

Since our first steps into second language speech research (see Drouillet et al., 2023), we have made a point of including L1 data to L2 analyses. This is especially relevant for acoustic measures which are very sensitive and subject to intra-individual variation. We believe that looking at the evolution of acoustic features in L2 across proficiency levels or time should be paralleled with the same measures in the subjects' L1 for two reasons.

First, when assessing a progression in L2, L1 data gives us a more complete picture as it represents the learners' baseline. As exposed in Chapter II section2.4., (p. 92), the fluency dimension of speech rhythm is not an isolated L2 phenomenon, but instead reflects L1 fluency habits of the speaker.

Furthermore, we have also seen that the macro-level of rhythmic organisation can be language-dependent. Indeed, L1 French and L1 English speakers differ in their macro-rhythmic patterns (Grosjean & Deschamps, 1975; Judkins et al., 2022b), and these L1 observations help us interpret and understand what happens in the speakers' L2 productions.

Second, in the case of repeated measures such as in longitudinal designs, collecting L1 data at the different time-points allows to determine how the within-subject/between-times variation observed in L2 relates to that in L1. To the best of our knowledge, no study on pronunciation instruction has ever addressed this point.

Second language studies concerned with the description of the interlanguage and/or L1-L2 transfer processes rest on a contrastive analysis paradigm. Such experimental protocols have been described by Selinker (1972, 1992) and later taken up again by Rasier & Hiligsmann (2007) in a study focused on prosodic transfer and acquisition. They advocate for an "*Integrative Contrastive Model*" (p.7) which encompasses comparisons between L1-L1, L1-L2, and L2-L2 (with two differing L1 groups learning a common target language). The purpose of comparing the interlanguage of two different L1 groups is for them to determine if the observed features in L2 speech can be accounted for by L1 transfer phenomena vs universal acquisition processes. This last question is beyond the scope of our work for now, but the integration of L1 data in our acoustic analysis of rhythmic correlates in L2 will enable a more comprehensive interpretation of our results (although the L1-L2 comparison is not addressed as a direct research question).

In addition, a majority of L2 speech studies using acoustic measures are based on an L2/native-target speech analysis and only a small portion include an L1-L2 comparison (but see Chapter II, section 2.4., p. 92). Consequently, it is difficult to attribute a value to the intra-subject variations observed in L2 (in the case of repeated measures) without information on the same subjects' intra-individual variation in L1. Yet, the relation between L1 variation across time and that of L2 is not the main focus of this work. As such, it will be addressed as a potential limitation to the interpretation of our results on acoustic measures, in the discussion section (Chapter VI, section 1.3., p. 248).

1.2.2. Ecological validity

In the process of designing this study, several choices were made in favour of the ecological validity of the data collected. The lack of ecological validity is a common limitation in L2 studies, and it lies mainly in the choice of the language elicitation task(s), and the timing and conditions of the testing phase(s) (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019).

As exposed in Chapter IV (section 3., p. 158), a vast majority of L2 studies use controlled tasks such as reading aloud or a picture narrative task (Derwing & Munro, 2005) to elicit language. Although these types of tasks present the advantages of ease to administrate and being comparable between studies, they fail to reflect the real situation of communication learners are confronted to in their daily lives. Much less data is available on free speech, whether monologic or interactive (although Crowther, 2020). This has been pointed out by several meta-analysis articles reviewing pronunciation instruction studies (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019).

Even though our experiment includes a reading aloud task, our analyses focus on the free speech samples only. The decision to include a reading task stems from the future project of running a comparison between the results obtained in both tasks, in order to contribute data on this methodological issue. However, the core of this study is concerned with spontaneous speech. We are aware that this choice potentially diminishes our chances to observe a significant impact of the training provided. Indeed, as Saito & Plonsky meta-analysis (2019) concluded:

> "Their [the learners] interlanguage L2 pronunciation performance tends to show a great deal of variation as a function of speech styles or different task types. Namely, L2 learners' pronunciation forms tend to be more target-like when their performance is elicited from formal controlled tasks (e.g., word reading) than from free speech tasks." (p.665)

Even so, our choice is to favour a close-to-reality speech style.

Further, to reinforce the naturalness of the speech samples collected, participants recorded themselves at home. The conditions in which speech samples are collected necessarily impacts the degree of naturalness. Recording one's voice is already an unusual task that can impact the level of confidence and anxiety of the participant. Setting the participant in a sound-proof room, with specific recording equipment such as professional microphones or headphones makes the environment even more atypical and more prone to create discomfort.

Nowadays, most adults own a computer or at least a phone and therefore have access to a recording device. We believe that having the participants record themselves at home, in a familiar space, and without anyone listening, creates a safe space in which they can express themselves comfortably, and therefore renders a more faithful picture of their speaking skills. In addition, administrating the tests at home facilitate the process of delaying the posttest as the participants do not have to come back to the research location to complete it.

Such out-of-hands testing conditions represent yet another risk and have been criticised in the past. Derwing & Munro (2005) for instance, advise to control the test phases conditions to ensure that participants follow the instructions given. The issue lies in finding the right balance between providing comfortable settings for the participants to increase speech naturalness and avoid participant loss due to an over-demanding design, vs controlling testing conditions to ensure the adequation of the data collected.

Since our protocol already required participants to come to the research facility eight times to attend the training sessions, we thought better to avoid adding another two for the testing phases. To make sure participants would follow the instructions during the test phases, we insisted that instructions needed to be carefully followed, we provided them with the experimenter's phone number in case they were confused or had questions about the instructions at the time of the test, and the instructions they received were written with the maximum precision.

The timing of the post-test was another issue to be settled. As Nagle (2022) points out in his brief review of pronunciation training studies on the subject, posttests are in vast majority taken immediately after the end of the training. Yet immediate post-tests involve that participants' performances are necessarily biased by recency effects. Consequently, it does not allow to make predictions on the training effects' durability.

The lack of a delayed post-test is a limit commonly stated in pre-post-test studies that several meta-analyses have pointed out (Lee et al., 2015; McAndrews, 2019; Saito & Plonsky, 2019; Sakai & Moorman, 2018). Nonetheless, scheduling a delayed post-test represents a non-negligeable risk at least at three levels: it increases the chance to lose participants (they might not come back for it), it diminishes the chances of obtaining striking results as the effects of the training might fade away, and lastly the more time passes, the more exposure participants get, which lessens the possibility to attribute the delayed post-test performances to the original training.

However, isn't improving learners' performance beyond the setting of the classroom the goal of all type of instruction? Here again, to avoid over-loading the experience for participants, we decided on one posttest only, which we asked participants to take during the week following the last training session. They all sent in their recordings between the fifth and seventh day.

1.2.3. Choices of measures

A majority of L2 pronunciation instruction studies assess learners' progress using human ratings of comprehensibility, intelligibility, accentedness, and perceived fluency, rather than acoustic measures (Saito, 2012; Saito & Plonsky; 2019). However, using only human global ratings involves bias in that raters rely on more than just phonetics, but also the use of syntax, lexicon etc. and across studies, the effectiveness of instruction can be slightly more varied when measured through expert ratings than acoustics.

Meanwhile, taking the opposite route and using only acoustic measures prevents from knowing if the changes observed from T1 to T2 have any relevance in terms of human perception. Therefore, as Saito (2012) argues, the combination of both acoustic measures and perceptual judgements constitutes a more robust paradigm.

Consequently, in the study presented here, we use a combination of acoustic measures of speech rhythm representative of aspects from the micro, meso, and macro level of analysis of rhythm, as well as fluency measures.

As for global ratings, we use judgements of accentedness and comprehensibility from native listeners. As exposed in Chapter III (p. 112 & 114), comprehensibility relates not only to comprehension of the message (intelligibility) but to the effort needed in the process of reaching understanding. Thus, comprehensibility ratings give a more precise picture of the listener's perception and processing speed of L2 speech. However, we have seen that this measure is correlated to linguistic aspects beyond pronunciation, such as vocabulary and syntax richness and accuracy. Accentedness however, relates to pronunciation aspects only. Therefore, coupling a measure of accentedness with a measure of comprehensibility enables to see the relationship between the two, and thus to assess focus on the effect of pronunciation aspects on both measures.

Following these methodological remarks, the next section provides a detailed description of the experimental protocol of Phase 1 of this study.

2. PHASE 1 - PRETEST/TRAINING/POSTTEST

2.1. COURSES CONTENT

Testing the effect of a type of instruction ideally implies a control group or another type of instruction to compare it against. Given the crucial role input and exposure bear in the development of L2 pronunciation (Saito & Hanzawa, 2018), it is important to ensure equivalency in that regard in all conditions. In the case of a comparison made between a group that follows a training and a group that does not, the argument could be made that the difference in outcome can be attributed to the increased exposure the training induced to the group who followed it, rather than to the actual content of the course.

In our study, the goal was to compare a novel kind of training focused on prosody solely, with the most common oral activities used in FLE classes. As such, we divided our participants into two groups: the Prosody group who followed a prosodic training (new method to be tested), and the Oral group who followed an oral expression and comprehension course (common method, acting as a control group). To control for input equivalency, the same teacher (the experimenter) taught both groups and spoke exclusively French in class.

The volume, length, and pace of the course was determined according to feasibility constraints. Sufficient class time was needed to cover the material, and we also wanted to provide enough hours of training to ensure an effect. Previous studies have shown that short trainings (under 3 hours) tend to yield limited impacts on the participants' performance (MacDonald et al., 1994; Baills et al., 2022). However, the length and volume of the trainings could not be too demanding a commitment for the participants to ensure their presence for the whole course. We did our best to strike a balance and landed on a 12-hour training, divided into eight 1.5-hour sessions, spread over four consecutive weeks, with two sessions a week. Figure 17 below summarises the experimental design for Phase 1 of the study.

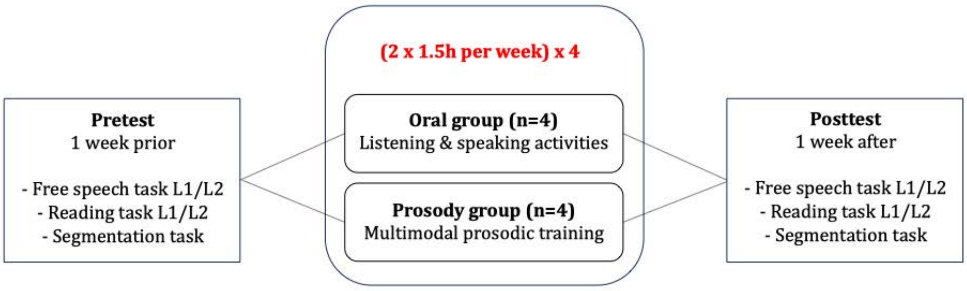


Figure 17 - Experimental design of Phase 1

Both the Oral and Prosody group attended their training sessions between March 21 and April 12, 2023. The trainings' contents are detailed in the next section. All course and experimental material are available on OSF.

2.1.1. Oral course

The oral course was built with the aim of recreating oral activities commonly used in FLE classes. As such, it represents a certain norm when it comes to working on oral skills (production & perception) in FLE classes. Following the terminology used by Long (1991), Doughty & Williams (1998), and many authors in Second Language Acquisition (SLA) since then, the activities of this course are meaning-focused. That is to say they are oriented towards a communicative goal through the use of specific lexicon determined by the topic of the lesson, and speaking practice through discussions or presentations.

The very first session was dedicated to an ice-breaker activity where everyone got to introduce themselves. It consisted of a guessing game where participants had to write on the board two numbers, two places, two verbs, and two objects that meant something to them. Then everyone else had to ask questions to guess the meaning of these things to the person interrogated. The purpose of this playful activity was also to install a comfortable and safe atmosphere for the participants to feel at ease when speaking French in the class.

After that, each session followed more or less the same format: the teacher introduced a topic to the participants (such as cinema, gastronomy...) and encouraged an informal conversation to get started. This was followed by a listening comprehension exercise (either audio or video) which provided an introduction to relevant vocabulary. The teacher asked guiding questions (orally only) to help comprehension. Answers to those questions were discussed in group.

Finally, participants were asked to prepare an expression exercise, either by themselves or in pairs, and present it to the class. The session generally ended with a broader discussion on the subject. Participants were free to take notes during classes but were never asked to produce anything in writing. A new topic was introduced and developed in each session. During class, the teacher gave minimal pronunciation-related corrective feedback, only when necessary, and did not provide any explicit rule or exercise related to segmentals or suprasegmentals.

The course was solely focused on oral practice and listening comprehension. Authentic material was used such as extract from radio shows and video clips from French TV channels, as well as activities from TV5 Monde (https://enseigner.tv5monde.com/) and French as Foreign Language textbook L'Atelier (L'atelier B1, Ed. Didier 2020). The detail of each session and the material used is provided in Appendix 1 (p. 303).

2.1.2. Prosody course

The course built for this study was a first attempt at designing a course covering mainly rhythmical aspects of French prosody. Since the goal of this study was not to test different method to teach prosody but to compare a training focused on prosody vs a training focused on oral expression in general, the choice was made to use a combination of teaching techniques rather than a unique one.

As presented in the previous chapter, using embodiment and gestures as well as musical activities to teach pronunciation yields encouraging results. Both these means were used in the course.

The first session started with an introduction to speech prosody with a perception exercise (discrimination of two audio samples based on prosody only), and a discussion around the sound characteristic of a French accent in English. The teacher then presented some key concepts and basic theory on prosody, prosody in L2, and the link between music and prosody as a transition into the first musical exercises.

Since for most people rhythm is linked to music (rather than language), we started the course with musical activities. The goal was to get our participants to pay attention and connect to rhythm in music, and to synchronise body movement to it. We used exercises inspired by the Dalcroze method presented in the previous chapter (section 2.2.2, p. 151) that were kindly provided by Florence Baills from the University of Lleida[14] (see Zhang, Baills & Prieto, forthcoming & 2022).

Exercises ranged from synchronising your walk to the beat of a tune, to more complex tasks such as anticipating an accent (strong beat) in the melody and mark the accent with landing a jump on it. The most intricate exercise combined a choregraphed sequence of movement that had to be executed onto a specific tune where participants also had to coordinate their movements collectively as one of the movements involved passing a small bag of sand to the next person in the circle.

After these introductory musical activities (sessions 1 & 2), we moved on to exercises designed to raise awareness on the accent's final position in French, and the syllabic rhythm (sessions 3 & 4). Perception exercises included locating the accented syllable in a given rhythmic pattern, and matching a logatom rhythmic pattern with a sentence heard[15]. Production exercises included repetition of rhythmic patterns in logatoms tapping on their knees at the same time, and tapping along a dialogue first uttered in logatoms then in words. Participants also worked in pairs to practice the dialogue. Apart from tapping or walking to mark the syllabic rhythm, participants were also asked to embody final accents by stretching a rubber band between their hands while saying words or practicing a dialogue.

Sessions 5 & 6 were dedicated to French syllable structure, linking, and cohesion within the rhythmic group. Perception and production exercises were used, and gesture was added to emphasise continuation between linked syllables (pointed index doing a U-shape movement).

Finally, the last two sessions (7 & 8) were focused on intonation and pause placement and realisation. There also, a combination of perception and production exercises was used, as well as gestures to accompany melodic contours. An emphasis was made on intonation's functions of demarcation and sentence mode (declarative, interrogative, exclamative etc.).

In each session, exercises progressed from controlled conditions (perception or repetition exercise of short utterances, executed one participant

[14] F. Baills has been using a combination of Dalcroze exercises to test the impact of a musical training on second language imitation abilities (publication forthcoming). She kindly offered to share the exercises she had used for her experiment, we used a selection of them.

[15] Most exercises used in sessions 3 to 8 were taken from either La Prononciation du Français en Classe, G. Briet, V. Collige & E. Rassart, 2014, Ed. PUG; *Les 500 exercices de Phonétique*, Hachette Langue Etrangère; and the online platform Fonetix.org (Berdoulat et al., 2018) to which Laure Fesquet & Sebastien Palusci (founders) gave us access. See Appendix 1 (p. 303) for details on exercises and sources.

after the other with personal feedback from the teacher) to more organic speaking exercises where participants working in pairs, writing their own practice dialogue. The idea behind being to carry the original exercise to a more natural speech practice. Detailed course and material are provided in Appendix 1 (p. 303)

2.2. PARTICIPANTS

To recruit participants, the study was advertised in international student networks within Toulouse universities, English speaking venues in the city, and on social media. The inclusion criteria were the following:

- be at least 18 years of age
- be a native speaker of English
- have a level in French between A2 and C2 of the CEFR
- be able to commit to 1.5hour classes twice a week for four weeks

We were hoping for 20 participants total to form two 10-participant groups. Because the courses were designed to favour speaking practice and individual feedback in the prosody course, the number of participants per group had to be limited.

Initially, we opened the selection to L2 French levels ranging from A2 to B2. Upon receiving responses, it became clear that we had to narrow down the selection to more homogenous levels to ensure better cohesion in the classes. The decision was made to favour elementary levels (A2-B1) as instruction would be most beneficial in the beginning stages of language acquisition.

The question of the variety of L1 English spoken by the participants was also at stake. While many varieties of English share similar rhythmical properties, others - because of cross-linguistic influences of other languages spoken locally - differ significantly (e.g., Singapore English, Indian English; Fuchs, 2023). Consequently, in order to ensure the comparability of participants, we excluded such varieties.

In the end, given our inclusion criteria and the participants' availability constraints, 11 participants constituted our sample. However, three of them had to be excluded later on. One revealed that English was in fact not their first language, and two dropped out of the course. Consequently, we were able to collect data on 8 participants. On the first session of the training, they were given a background questionnaire to fill out. We collected information on demographics, linguistic profile, experience and exposure to French, and an L2 French self-evaluation. This information is presented in Table 9, Table 10, and Table 11 below and commented upon. Participants identified with the letter "B" in their anonymous ID number were assigned to the Prosody group, those with the letter "A" the Oral group. The assignment to either group was purely random as it rested upon participants' schedule openings to a morning class (Oral group) vs. an evening class (Prosody group).

Table 9 below summarises the participants' demographic characteristics. Their age ranged from 21 to 66 with an average of 37, and they came from England, Scotland, the USA, and South Africa.

Participant ID.	Gender	Age	Birth Country	Education	English Variety	Other Language	Music practice
B37	M	66	England	High School	South-Western English	-	flute
B33	F	63	England	High School	South-Western English	-	-
B31	F	21	Scotland	College	Scottish	-	-
B35	F	21	Scotland	College	Scottish	-	drums
A20	F	36	USA	College	American	-	-
A21	M	24	South Africa	Masters	South African	Afrikans (int.)	-
A22	F	21	Scotland	College	Scottish	Spanish (beg.)	flute
A24	F	45	South Africa	College	South African	-	-

Table 9 - Participants' demographic characteristics. A22 is greyed because her free speech samples did not follow our instructions and she was excluded from the analyses (see section 2.3.2 of this Chapter, p. 197)

Table 10 below presents the information collected on their experience with L2 French. The age at which they started learning French (age of onset) is quite variable, and they all - except from A21 - had taken French classes before, mostly during their middle or high school years. For B37 and B33, the years of French classes do not match with their indicated age of onset and age at the time of the experiment. We suspect that they indicated the age of onset as their most recent experience in French class but that in fact the years in French class include classes they had in primary or secondary school.

The total time spent in France is also very variable. Some participants have been in the country for 3 or 5 years, when others have only been there for 7 months (these are Erasmus students). However, their exposure to French and use of the language is in majority fairly limited.

All of the participants reported using French less than 30% of the time in their social circle (friends and family) - apart from A20 whose partner is French. However, three of them reported an 80% to 100% usage of French in their work/school environment (here school refers to university). As mentioned before, the distribution of the participants into each of the two groups depended only on their schedule, it was therefore impossible for us to control for equivalence of exposure when forming the two groups. These inter-group differences are taken into account in the interpretation of the results.

Participant ID.	Age of Onset	Years in French class	Months in France	Months working in French	Weekly use w/ friends	Weekly use w/ family	Weekly use work/school
B37	58	10	36	6	20%	0%	0%
B33	57	10	36	0	20%	0%	0%
B31	13	8	7	5	0%	0%	100%
B35	13	8	7	6	0%	0%	10%
A20	28	3	64	4	30%	50%	80%
A21	22	0	18	0	10%	0%	10%
A22	15	5	6	4	30%	0%	90%
A24	43	4	36	3	30%	0%	30%

Table 10 - Experience with and exposure to L2 French

To determine the L2 French level of the participants, they were asked to rate themselves in the questionnaire. Because the proficiency is not an independent variable in our study but rather, we wanted to constitute homogenous groups in terms of level and ensure that instruction would benefit everyone, we decided it was not necessary to include an objective proficiency test in our design.

Table 11 below summarises the ratings on 1 to 6 point scales each participant attributed themselves for the four main language competences. They were also asked to report which level of the CEFR they thought they were at. A21 and A24 both placed themselves at the A1 level, however it was clear from interacting with them during the training that they were at an A2 level[16]. All participants reported lower scores for the speaking competence (1.8 over 6 on average) and mainly judges themselves to be better at listening (3.25 over 6 on average).

Participant ID.	Speaking	Listening	Reading	Writing	CEFR
B37	2	3	3	2	A2
B33	2	2	2	2	A2
B31	1	5	4	3	B1
B35	2	3	4	2	B1
A20	3	3	4	4	B1
A21	2	3	3	2	A1
A22	2	4	4	2	B1
A24	1	3	2	1	A1

Table 11 - Participants' self-assessed level in L2 French

In order to gage the predominant characteristics of each group and see how they compare, we have summarised the three tables above into Table 12 below. The between-group differences are commented in the discussion of our results (Chapter VI, section 1.2.4., p. 245).

[16] This judgement is based on 7+ years of experience as a trained FLE teacher.

	PROSODY GROUP	ORAL GROUP
Number of participants included in the analysis	4	3
English variety	South Western English (2) Scottish (2)	American (1) South African (2)
Age	66, 63, 21, 21	36, 24, 45
Musicians	2 (B37, B35)	0
Years in French class	Mean = 9	Mean = 2.3
Months in France	Mean = 21.5	Mean = 39.3
Mean weekly use of French (as a % of time)	37.5%	40%
Mean self-assessed speaking proficiency	1.75/6	2/6
Mean self-assessed listening proficiency	3.25/6	3/6
Self-reported CEFR levels	A2, A2, B1, B1	B1, A1, A1

Table 12 - Summary of predominant characteristics of each group.

2.3. TASKS[17]

Prior to the beginning of the tests and training sessions, the consent form was sent to the selected participants and they were invited to an online meeting with the experimenter to receive information about the study, clarify what was expected of them, and answer their questions. They were told the study was about the acquisition of speaking skills but were unaware of the two different teaching methods put to the test.

Participants completed the pre and posttest in full autonomy, in the comfort of their home. They were sent detailed instructions on the tasks to complete and were provided with an online voice recording app[18] in order to record themselves and save the resulting sound files. They were then asked to upload the recorded speech samples onto a secured online platform provided by the university. As explained in section 1.2.2. (p. 184), two reasons motivated the decision to administer the tests remotely: spare the participants from having to come to the research location twice on top of the eight times required by the training schedule;

[17] All tasks' material are available on OSF

[18] https://online-voice-recorder.com/

and increase their comfort during the tasks so as to collect speech samples as natural as possible. The pre and posttests consisted of five tasks each: reading aloud and free speech in L1 and L2, and a segmentation task. The order of the tasks was imposed and similar for all participants at T1 and T2. They started with the first two tasks in L1 so that they could get acquainted with the format before moving on to the tasks in L2. The segmentation task came last.

2.3.1. The reading task

Participants were asked to read aloud small texts in English and French. In total, four texts were used. For the pretest, the English text was 262 word long on the topic of "Christmas in New Zealand". The text in French was 258 word long on the topic of the typical French breakfast: "Le petit déjeuner des français". For the posttest, the English text was 234 word long on the topic of tourism, and the French text was 261 word long on French holidays.

All texts were chosen and adapted to ensure an equivalent level of complexity. They are all A2-level descriptive and informative short articles that use frequent vocabulary and simple sentence structure. Although read speech is not analysed in the study presented in this dissertation, we thought including it would not overload the test phases, and would allow us to test the effect of speech style as a factor in future analyses.

2.3.2. The free speech task

To elicit natural speech, we provided participants with different topics to discuss. For the pre and posttest tasks in L1 English, participants were asked to talk about the last book they had read, or movie/series they had seen. For the L2 French tasks, they were asked to talk about what they think about life in France in the pre-test, and about a memorable vacation in the posttest. They were instructed to speak for a minimum of two minutes and a maximum of five. In L2: 225s on average; in L1: 185s.

Figure 18 below summarises the entirety of the speech data collected. The free speech samples are highlighted since the analysis presented here concerns those only.

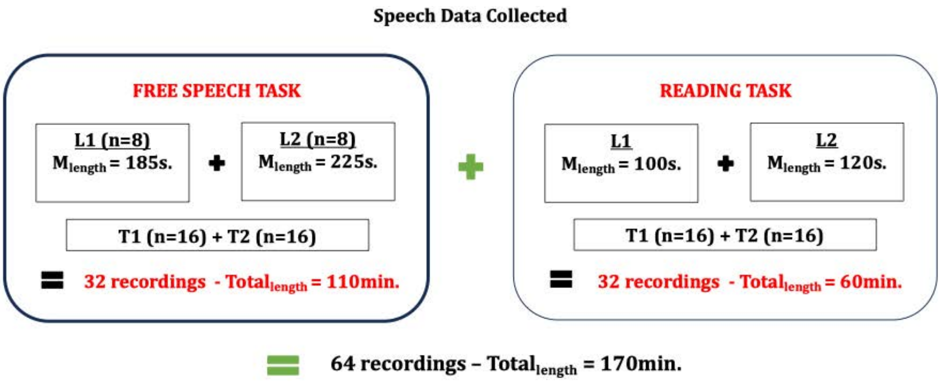


Figure 18 - Summary of the speech data collected.
s. = seconds, min. = minutes

2.3.3. The segmentation task

This last task was designed following the work of Charles et al. (2015). The purpose of this task is to test the participants' performance in the first stages of speech perception processing: speech segmentation. As exposed in Chapter III, section 4.1. (p. 123), prosody bears a fundamental role in speech segmentation, word retrieval and access to meaning. As such, it is tempting to imagine that teaching learners about the prosodic system of the target language could help them make better use of prosodic cues to segment speech when exposed to native speakers of the target language.

If it is the case, teachers could provide their students with a solution that goes beyond the usual advice that only exposure can help develop their listening skills. Not to say it is untrue, but when facing discouraged students who ask how they could improve their listening skills, it can feel disconcerting to have nothing else to offer than just "you just need more exposure". Despite this tight relation between prosody and segmentation skills, to this day very few studies have looked at the effect of a prosodic training on segmentation abilities (see Chapter IV, section 3.5, p. 171). Even though the development of listening skill is not our main focus, we did not want to miss the opportunity to add a listening task to our protocol.

Thus, the segmentation task was designed to test if the prosodic training would have a positive impact on participants' speech segmentation abilities. It consisted in listening to short excerpts of speech, and repeating immediately after as many words heard as possible. The repetition served only as a window into what the participant was able to retrieve. It is different than an imitation task since the production of the participant itself is not being analysed beyond the number of correctly repeated words.

When designing the task material, the goal was once again to get as close as possible to a real-life situation where learners are confronted to French native speakers talking in their usual way (that is, not slowed down or adapted to a non-native speaker). We selected excerpts from eight French native speakers in the B-FREN3 corpus (Drouillet et al., 2023; Judkins et al. , 2022a & 2022b). For each speaker, we selected four excerpts from the video retelling task (VR) and four from the conversation task (C). We made sure to select short enough excerpts to avoid short-term memory overload. Excerpts selected corresponded to IPU less than two seconds long, ranging from four to ten words, 5.6 words on average, as shown in the examples below:

"Ça fait des grosses semaines" (*Resulting in busy weeks*)

"Elles ne peuvent plus sortir de la maison" (*They cannot escape the house anymore*)

We were careful to avoid repetitions of words as much as possible from one excerpt to another. In total, 64 excerpts were manually extracted from the B-FREN3 corpus using Praat. They were divided in two in order to present different stimuli for the pretest and the posttest. With each 32-item sets, we created one set comprising seven excerpts for the training phase, and one set comprising 25 excerpts for the test phase. The order of the excerpts was randomised and three sets with different randomised orders were created for T1 and T2 respectively. Figure 19 below sums up the distribution of the stimuli and the creation of the different sets for the task.

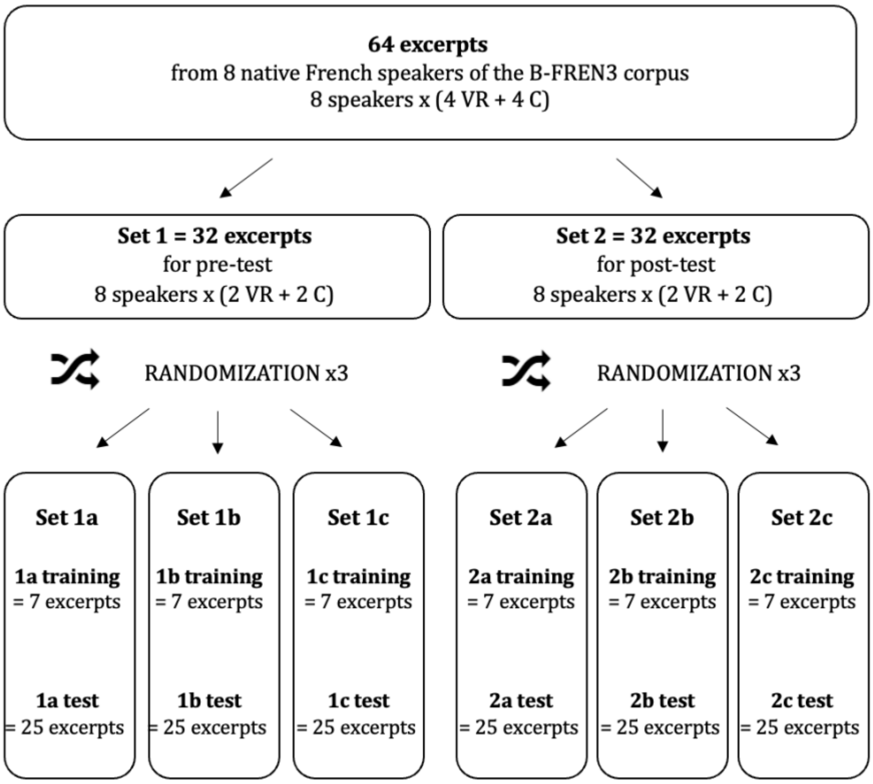


Figure 19 - Distribution of stimuli in the different sets created for the segmentation task. VR=video retelling elicitation task, C=conversation elicitation task.

Each set was arranged into a single sound file using a Praat script. In the resulting audio files, each excerpt was preceded by a warning tone and followed by seven seconds of silence during which participants repeated the stimulus heard.

As both the T1 set and the T2 set were created from the same corpus with the same speakers on the same tasks, and the selection of excerpts followed the same criteria, we considered them to be equivalent in terms of complexity.

To ensure equivalency, as well as to constitute a control group, we submitted the two sets back-to-back to four native French speakers (two males, two females). Two of them heard version *a*, *b* or *c* of set 1 then set 2, and the other two set 2 then set 1. They were given the same instructions as the participants i.e., to repeat as many words of what they had just heard as possible. The resulting scores showed no difference in performance between the two sets, confirming their equivalency. However, for five excerpts, one - and in one case two - native participants omitted one word in their responses. We interpreted these omissions as a sign of difficulty associated with the repeated excerpt so they were excluded from the final analysis.

Figure 20 below summarises the data collected for the segmentation task (after exclusion of the faulty excerpts).

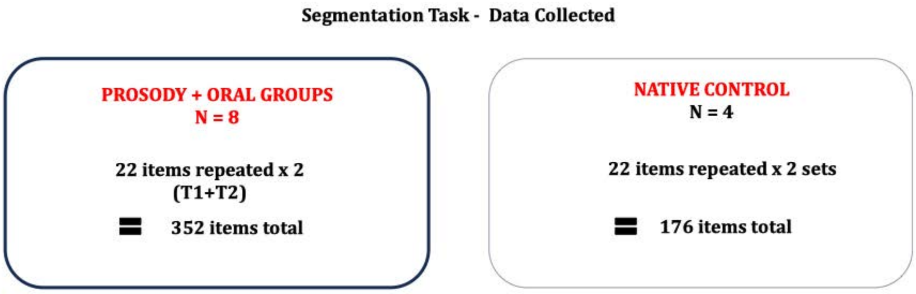


Figure 20 - Summary of data collected for the segmentation task.

2.4. DATA EXTRACTION

At the end of Phase 1 of the experiment, we collected a total of 80 sound files: 10 sound files per participants, five at T1 and five at T2, corresponding to the tasks described above. The read speech samples were left aside for future analyses.

When listening to the free speech samples, it appeared that one participant (A22) did not fully respect the instructions of the task. Most likely with the intention to do well, they clearly prepared their speech, and it was obvious to the three annotators (the main author and her two supervisors) that they were reading instead of speaking spontaneously. Consequently, it was determined that A22's free speech samples were not comparable to the others, and had to be excluded from the analysis.

The free speech samples (n=32) were all semi-automatically segmented into syllables using the Praat Plugin EasyAlign (Goldman, 2011) for the L2-French speech files, and WebMaus[19] (Kisler, Reichel, Schiel, 2017; Riechel, 2012) for the L1-English speech files. The alignment of syllable boundaries was manually checked and corrected, and audio files were manually annotated in Praat (Boersma & Weenink, 1992-2023). The syllabification for both the L2-French and L1-English speech samples followed the principles explained by Delattre (1940). Mainly:

[19] https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface

- A single intervocalic consonant was considered as the onset of the second syllable.
- For consonant clusters, preference was given for splitting the cluster (except if positioned in coda followed by a pause, if the first consonant has a lower sonority than the following one, and if the first consonant is articulated in the front of the mouth and the following one further back in addition to the previous rule).
- In the case of *liaison* (e.g., *nous sommes allés* = nu/som/za/lE) and *enchainement* (e.g., a book I read = @/bU/kaI/rEd) the same rules were applied.

From the EasyAlign and WebMaus automatic segmentation we kept the syllable tier and the orthographic transcription tier. Three tiers were added to annotate accents, disfluencies, and Inter Pausal Units (IPU) respectively, as shown in Figure 21 below. More detail on the annotation process for each tier is given in the following sections.

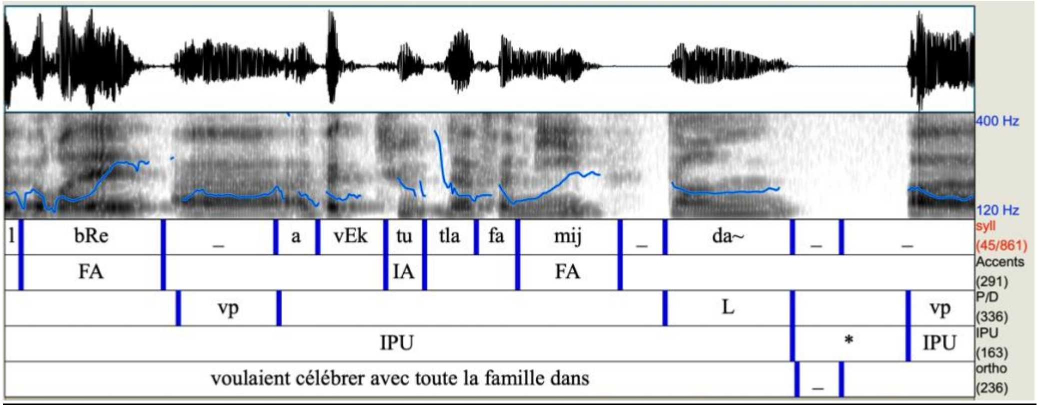


Figure 21 - Illustration of the annotation tiers on an L2-French textgrid in Praat. (IA=initial accent; FA=final accent; L=lengthening; vp=voiced pause)

2.4.1. Speech rhythm measures

a. Micro-level

The micro-level of speech rhythm analysis as defined in Chapter I (section 3.3.1., p. 30) corresponds to the phonemic or segmental level. So called rhythm metrics such as %V (proportion of vowels), delta measures (standard deviation of vocalic or consonantal intervals); and measures of durational variability (PVI) belong to this level. The latter presents an interest in studies involving French and English as it captures the variability in duration of the chosen unit (for instance but not limited to: vocalic interval, consonantal interval, syllable), and is therefore an indicator of the difference produced by the high syllable complexity and vowel reduction in the case of English, vs syllables of a lesser complexity in structure, and no reduction phenomenon in French. In addition, we saw in Chapter II (section 2.1., p. 72) that measures of nPVI reflect differences between L2 proficiency levels, especially so when calculated on syllables (Ordin et al., 2011).

The PVI was originally developed by Low & Grabe (1995) to account for rhythmical differences between Singapore English (assumed to be syllable-timed) and British English (assumed to be stress-timed). The idea was to measure the durational variability of pairs of successive consonantal and vocalic intervals, and to obtain a global mean durational variability. Therefore, the PVI, calculates the difference in duration between each pair of successive intervals in a string of speech and compile them all into a mean following this formula:

$$rPVI = 100 \times \sum_{k=1}^{m-1} |d_k - d_{k+1}|$$

Where *m* is the number of intervals and d_k the duration of the $_k$th interval. This measure is commonly called raw PVI (rPVI) as opposed to its normalised version, the nPVI detailed below.

It has been shown that vocalic interval duration is highly influenced by the speech rate (Grabe & Low, 2002; Ramus, 2002). Consequently, it is preferable to use the raw PVI on consonantal intervals or if used on vocalic intervals, it should be on a controlled task that yields a fairly stable and similar across-participant speech rate.

To overcome this issue though, Low (1998) later proposed a normalised version of the PVI: the nPVI. This version adds the division of the durational difference between two intervals by the mean duration of the pair. Consequently, the nPVI takes into account the speech rate influence at the local level instead of the global level. This is of particular interest when analysing spontaneous speech where speech rate evolves through the utterances. This normalisation procedure also allows for inter-speaker comparability. Dellwo (2010) demonstrated the robustness of the nPVI on speech samples with changing speech rate.

The Normalized Pairwise Variability Index (nPVI) was calculated on our data using Grabe & Low (2002) equation:

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

The resulted scores usually range between 0 (perfect regularity, absence of durational variation) and 100, although the upper limit is entirely dependent on the amount of variation within the dataset.

Previous studies have shown that the nPVI of vocalic intervals and syllables is consistently higher in English than in French. Table 13 below reports a selection of nPVI values as points of reference.

	L1-French		L1-English	
nPVI-V (vocalic intervals)	50	(White & Mattys, 2007 - read speech)	73	(White & Mattys, 2007 - read speech)
	43.5	(Grabe & Low, 2002 - read speech)	57.2	(Grabe & Low, 2002 - read speech)
	42.5	(Obin et al., 2012 - read speech)	66	(Arvaniti, 2012 - free speech)
	46	(Vieru et al., 2011 - read speech)	75	(Ordin & Polyanskaya, 2015 - directed speech)
	35.9	(Patel et al., 2006 - read speech)	55	(Patel et al., 2006 - read speech)
nPVI-S (syllables)	28	(Guilbaud, 2002 - free speech)	50	(Guilbaud, 2002 - directed speech)
	49.4	(Mok & Dellwo, 2008 - read speech)	69.6	(Mok & Dellwo, 2008 - read speech)
			69	(Ordin & Polyanskaya, 2015 - directed speech)

Table 13 - Selection of nPVI values for L1 French and English found in the literature.

It is clear from this table that values can vary quite a lot within the same language. This variation can be explained by the difference of language elicitation method, the speech material, the number of speakers involved, the volume of the speech sample studied, and the definition of the intervals measured. As shown in this table, a majority of studies use vocalic intervals on read speech. The use of the PVI on spontaneous speech is much rarer.

Another aspect that may distinguish between studies is the inclusion or not of the phrase final interval which is usually lengthened (at least in English and in French). Although a majority of studies such as those presented in Table 13 do include final lengthening, some decide to exclude final lengthening in order to obtain an indication of the durational variability of phrase-internal intervals only (Bertinetto & Bertini, 2008; Grabe et al., 1999; Guilbaud, 2002). In our case, given the fact that phrase final lengthening is present in both French and English, there is no reason to exclude it. This also allows for a comparison with previous studies such as the ones presented in Table 13 that all included final lengthening in their analyses (except from Guilbaud, 2002).

It is worth noting that the study from Vieru et al. (2011) also included a measure of the nPVI of L2-French spoken by L1-English speakers (the only one to our knowledge), the resulted score was 62.

In a majority of studies, PVI measures are carried out on vocalic and/or inter-vocalic intervals. However, it can also be used on syllables or other types of intervals depending on the research question. Several studies on the acquisition of speech rhythm in L2 have used the syllable as their base unit to calculate the nPVI (Guilbaud, 2002; Ordin et al., 2011; Ordin & Polyanskaya, 2015; Mok & Dellwo, 2008). Ordin et al. (2011) actually conclude that metrics calculated on syllable duration are better suited than those calculated on vocalic or consonantal intervals to discriminate between proficiency levels.

Since we are not interested in classifying languages here, but rather measure the progress of our participants in L2; and since our work on French rhythm in the Prosody course heavily relied on the syllable unit, the nPVI was calculated on syllables.

After careful segmentation, the nPVI was calculated on the syllables of the L2-French and L1-English free speech samples collected using a Python script[20]. Phrase-final lengthenings were included, all disfluencies (detailed in section2.4.1.d below) were excluded, and syllables belonging to the end of an IPU and the beginning of the next one (separated by a silence of at least 250ms) were not paired.

As mentioned at the beginning of this section, **in previous studies the nPVI is consistently higher in English in comparison to French. Since the prosody group received a specific training on the syllabic structure and rhythm of French, we would expect the nPVI scores in L2-French to be lower than in L1-English, and more so at T2 after training.**

b. Meso-level

[20] Available on OSF

For measures related to the meso-level (accentuation), the author and her two supervisors individually annotated the perceived prominences on the 16 recordings in L2-French and the 16 recordings in L1-English. The main author then centralised the three annotations and identified points of divergence. The three authors met, listened to the recordings again, and discussed to find a consensus. Perceived accents were then annotated onto the corresponding textgrids. In French, two types of accents were identified: initial accent (IA) and final accent (FA). In English, all perceived accents were marked "A". The distinction between primary accents and secondary accents was not made.

A note on the perception of prominences

It appeared that the perception of accents in the L2-French samples was not homogenous amongst the three annotators. Even though the three authors are experienced prosody experts, the individual annotation of prominence revealed different sensitivities to different cues, as well as the interference of knowledge on the accentual system of French (sometimes we really want to hear a prominence somewhere because it would make sense in theory!). However, listening again together and being able to discuss the nature of what was heard allowed us to come to a consensus, and an effort was made to be as faithful as possible to acoustic cues only.

In the L1-English samples, accented syllables were more obvious and their perception more consistent across the three annotators. However, the three annotators were surprised to notice a high number of occurrences of consecutive accented syllables. It was not rare to find three and sometimes four consecutive stressed syllables (more often on monosyllabic words). This observation questions the stress clash rule (Liberman & Prince, 1977).

The purpose of this annotation was to allow the extraction of a measure of accented and non-accented syllables' durations. This measure is complementary to the nPVI described in the section above. But whereas the nPVI indicates the variability in duration between successive syllables, this measure allows to grasp in more detail the accented syllable to non-accented syllable ratio - that is, by how much the accented syllables are lengthened compared to unaccented syllables.

In L1 French, lengthening is the most prevalent acoustic cue to accented syllables in final position. Studies have shown that on average, accented syllables are close to twice as long as unaccented syllables (Delattre, 1966), although that relation is subject to around 20% variation (positive or negative) depending on the speech rate (Pasdeloup, 2004).

Conversely, in L1 English, accented syllables are realised with a combination of increased intensity, f0, and potentially duration (Wenk, 1985) but in smaller proportion than in French. In Delattre's (1966) reference study, a syllable duration ratio of 1.78 was found in French and 1.6 in English on the speech of radio interviews, considered as spontaneous speech. Astésano (2001) also found a 1.7 ratio in L1 French in read speech, radio interviews, and radio news.

In the present study, syllables' duration was extracted through a Praat script[21]. **We expected the ratio between accented syllable and unaccented syllable to increase at T2, as a sign of a progression from the English rhythmic pattern to the French one.**

c. Macro-level

The macro-level refers to the organisation of speech into chunks of uninterrupted speech (Inter Pausal Units, IPU), delimited by silent pauses of at least 250ms (Bosker et al. 2013; Kahng, 2018; Kormos & Denès, 2004). A specific tier in the textgrids was dedicated to the annotation of IPU and silent pauses. This layer of annotation was carried out manually, through listening to the audio file, along with the visual inspection of the spectrogram in Praat, similarly to the method employed in our previous studies, Judkins et al. (2022a, 2022b).

As explained in Chapter I (section3.3.4., p. 39), the 250ms threshold is not an absolute but has been found to be the most adapted to the study of L2-speech (De Jong & Bosker, 2013; Kahng, 2018). Silent pauses in between IPU were tagged with a star ("*"). We also found pauses between IPU that also included a voiced pause and/or a false start. For example, there could be a silence >250ms after the end of an IPU, followed by a voiced pause, itself followed by another silence >250ms. In these cases, the interval between the two IPU was tagged with two stars ("**") in order to distinguish empty IPU-external pauses from filled ones. For that reason, we prefer the term "external pauses" to "silent pauses" to refer to pauses between IPU.

From this tier we extracted the quantity and duration of IPU, the duration of IPU-external pauses, and the proportion of filled external pauses over the total of external pauses.

[21] Available on OSF

Since we used the same criteria on these measures than those used in Judkins et al. (2022b), we expect to see similar results in the within-subject analysis. **From T1 to T2, we expected less and longer IPU and filled external pauses in less proportion as a sign of improved fluency.**

d. Fluency

The last annotation tier included all kinds of disfluencies, namely: voiced pauses, lenthenings, false starts, and code-switching.

Voiced pauses were identified through the careful listening of the audio files and inspection of the spectrogram. They are usually realised with a [œ:] sound in French, and [m:] in English (Laver, 1994). Lengthenings are a similar phenomenon but are realised on the last syllable of a word or on a monosyllabic word such as "en" or "et" in French, "the" or "and" in English for instance.

The perception threshold of a lengthened syllable fluctuates according to the context, i.e., the duration of the preceding syllables. Because syllable lengthening is tied up to the speech rate of the speaker, using an arbitrary threshold is not adapted, especially in spontaneous speech where the speech rate can vary from an utterance to the other. Also, not all lengthenings are hesitation markers, they can bear semantic and/or pragmatic functions (Di Cristo, 2016; Johnsen & Avanzi, 2020). For these reasons, lengthenings were tagged as such according to the perception of the main author only, as to avoid non-relevant instances.

False starts are characterised by an interruption and/or a repetition of a syllable, word or group of words. Each repetition was tagged as an individual false start. If a group of word was interrupted or repeated, the entire group was tagged as one false start (rather than multiple false starts corresponding to each word).

Some participants used a few words in English in their L2-French free speech (rarely) which were tagged as "code switching". We did not run any specific analysis on these but they were considered as disfluencies and were excluded from the analyses on syllables and nPVI.

From the annotations on the disfluency tier, we extracted measures of quantity of disfluencies, and duration of voiced pauses located within IPU. The articulation rate was calculated as the count of syllables per seconds of phonation, excluding voiced pauses but including all other type of disfluencies (false starts, lenthenings, code switching). Table 14 below presents a summary of all acoustic measures extracted from the annotated textgrids.

ACOUSTIC MEASURES EXTRACTED			
Micro-level	**Meso-level**	**Macro-level**	**Fluency**
nPVI	Syllable duration Durational ratio of accented to unaccented syllables	Quantity of IPU IPU duration External pauses duration	Quantity of disfluencies Voiced pauses duration External voiced pauses over external pauses proportion Articulation rate

Table 14 - Summary of acoustic measures extracted from the annotated textgrids of the free speech task in L1-English and L2-French.

2.4.2. Segmentation scores calculation

To calculate the segmentation scores, the participants' audio files for this task were compared to the original items. The main author listened to the recordings with high quality headphones and reported for each item the number of correctly repeated words.

The final score corresponds to the proportion of the total correct words repeated by the participant, over the total number of words of all items together. Table 15 below gives an example of the scoring process.

	Item	Word count	Repetition	Correct words repeated
Example 1	"quand il essaie de couper"	5	"continuer de se couper"	2/5
Example 2	"on a eu un peu de beau temps"	8	"un peu beau temps"	4/8
Example 3	"j'enseignais à Toulouse"	4	"les gens signaient à Toulouse"	2/4
		[...]		
Total item words		5+8+4+[...] = **129**	**Total correctly repeated**	2+4+2+[...] = **76**

FINAL SCORE: 76/129*100 = 58.9%

Table 15 - Segmentation task scoring process, example extracted from participant B31 at T1.

In addition to the five items removed because of the incomplete responses they yielded from our native control participants, two more items had to be removed. These items were in last position of a set (respectively), and several participants skipped them. In total, seven items were excluded from the analysis: three belonging to the sets used at T1, and four belonging to the sets used at T2. The final scores for each participant were calculated from 22 items (130 words) at T1 and 21 items (125 words) at T2.

<u>*A note on segmentation error types*</u>

The examples displayed in Table 15 *show different types of error in the repetitions. Example 1 shows a situation where the participant completely replaced the original words with others, except for the last one "couper". We can imagine that "quand il essaie" was re-worded as "continuer" because of the liaison and enchainement phenomena between "quand", "il" and "essaie" which results in "kan.ti.lé.ssé". Apart from the last syllable "ssé" which disappear, "kan.ti.lé" and "continuer" are in fact phonetically close. This shows a segmentation error that seems to be imputable to a lack of knowledge of the liaison and enchainement rules in French.*

Example 2 shows an omission of several words but the ones resituated are correct and here too, the end words are preserved. We noticed that in the case of word omission, grammatical words are often the ones that disappear.

Example 3 is another instance of re-wording which shows a segmentation error since "j'enseignais" becomes "gens signaient" but this time, it is not a case of enchainement or liaison, rather, the initial accent on "j'en-" is interpreted as a final one. A detailed analysis of the type of error in the participants' repetitions could tell us more about how they process the input but it is beyond the scope of this work.

However, we note that overall, the last word or couple of words are usually preserved. This could be due to a recency effect and/or a prominence effect. Since the last word of a rhythmic group bears stress in French (and potentially a nuclear accent), it makes sense that it is uttered more clearly and stands out from the previous string of speech.

Before presenting the results of the above-described Phase 1 of our study, the following section presents Phase 2, which relates to the comprehensibility and accentedness judgements collected on our participants' free speech recordings.

3. PHASE 2 - COMPREHENSIBILITY AND ACCENTEDNESS JUDGEMENTS

3.1. FIRST ATTEMPT

Before describing the design used to collect the comprehensibility and accentedness scores presented in the next result section, we would like to acknowledge the fact that there was a first attempt that was considered too flawed to be reliable.

Originally, following the ecologic principle we have tried to put forward in this whole experiment design, participants recruited for this judgement task were non-experts (specifically non-linguists nor language teachers), native speakers of French, ranging in age from 18 to 70.

Two online surveys were put up to collect judgements on the free speech recordings collected in Phase 1. From each of the 16 recordings (8 at T1, 8 at T2) we extracted a one-minute sample from the middle of the recording. The decision was made to split the items to be judged into two surveys, so as to keep the length of the task decent for our naive judges. This decision implied that half the items were judged by a group of people and the other half were judged by different people. In total, we collected the responses of 35 judges for one half of the recordings, and 28 for the second half. When we looked at the results, scores were very inconsistent.

As commonly done in studies using global ratings, we used Cronbach's Alpha to determine the inter-rater consistency of ratings (e.g., Munro & Derwing, 1995a; Trofimovich & Isaacs, 2012). The alpha coefficient ranges from 0 to 1. A higher value (closer to 1) suggests greater reliability or agreement among the raters. Typically, an alpha between 0.7 and 0.8 is considered acceptable, 0.8-0.9 is considered good, and above 0.9 excellent for inter-rater reliability (Tavakol & Dennick, 2011).

Cronbach's Alpha on our data for the comprehensibility ratings came out at $\alpha = .72$ which is barely acceptable, and at $\alpha = .82$ for the accentedness ratings, which is ok but not great. We also collected feedback from the survey participants who, for the most part, indicated that they found the task difficult and they felt very unsure of their responses because they had no point of reference.

Looking back at how we designed this first attempt, we realised that dividing the items into two surveys was a mistake as it introduced an extra layer of variability amongst judges. In addition, native-speaker items were not included in that version, and because it was administered entirely online the concepts of comprehensibility and accentedness could not be properly explained to the participants. All in all, it became clear that this design was flawed and that the results were not reliable. It was decided to re-design the experiment in the hope that correcting our mistakes would lead us to collect more reliable results. The experiment described below is therefore the edited second version.

3.2. JUDGES

Ten students from the French as a Foreign Language program (3rd and 4th year) at the University of Toulouse II Jean Jaurès were recruited to judge the Phase 1 recordings in terms of comprehensibility and accentedness. As these students were training to become French teachers, we consider them as semi-experts regarding the task of judging non-native speech. They are certainly not trained linguists or phoneticians but their exposure to non-native speakers makes them familiar with the diversity and plurality associated to non-native speech, as well as with the exercise of evaluating learners' speaking skills. Because our first attempt with completely naive listeners was unsuccessful, we made the choice of going a little further up in the expertise scale without going all the way.

Judges were all females, aged from 19 to 56 (M_{age} = 34.1 years). Two of them were non-native but highly proficient speakers of French. We decided to include them since it has been shown that there is no difference in ratings between native speakers of the target language and highly proficient non-native speakers (Derwing & Munro, 2013). The other eight listeners were all native speakers of European French. Basic information on their profile was collected and is presented in Table 16 below.

Judge ID	Age	Native French	Languages spoken	Familiarity with non-native speech	Familiarity with English	Familiarity with English-accented French
1	39	yes	Spanish-adv. English-int. Italian-deb.	5	4	5
2	41	no	Portuguese-L1 English-adv. Spanish-beg.	5	5	3
3	23	yes	Spanish-adv. English-int. Catalan-int. Corean-beg.	3	3	3
4	19	yes	English-adv. Spanish-int. Russian-beg.	2	4	3
5	19	yes	English-int. Spanish-int.	3	2	3
6	56	no	Arabic-L1 English-NS Italian-NS	4	3	3
7	34	yes	English-adv. Italian-adv. Spanish-int.	3	4	3
8	22	yes	English-adv. Spanish-beg. Corean-beg.	2	4	2
9	36	yes	English-beg. Spanish-beg. Japanese-beg.	3	5	2
10	52	yes	English-beg. Italian-beg.	2	2	1

Table 16 - Information collected on the judges' profile. In the "languages spoken" column, adv. = advanced level, int. = intermediate level, beg. = beginner level, L1 = native language, NS = level not specified. Familiarity is coded from 1 (very unfamiliar) to 5 (very familiar). Judge 6 is greyed because they were later excluded from the results' analysis (see section 3.5., p. 219).

To our surprise, a majority of judges indicated a low to average degree of familiarity with non-native speech. However, they all spoke English, for the most part at an intermediate level or above, and they indicated to be fairly familiar with English-accented French. From this information, we conclude that in the end, our judges' profile is closer to naive listeners than to experts.

3.3. MATERIAL

The selection of the audio material to include in the judgement task raised a lot of questions. Namely: the number of samples per participant and condition, the length of the samples, and the rules guiding the selection of the sample within the whole recording (at the beginning, the middle, the end?). The underlying debate relates to the degree to which the selected samples will accurately reflect the speakers' overall L2 speech performance on the one hand; and on the other, the ideal sample length for the listener to be able to form an impression and decide on a score for that sample.

There is no clear consensus in the literature, and authors rarely explain or justify their decision. Sample length vary from short utterances (10 words on average) to 150 to 290 seconds in Nagle et al. (2019). The latter study actually questioned the underlying mechanisms at play in listeners' judgements and considered comprehensibility as a dynamic construct. They used longer samples for the purpose of testing a dynamic measure of comprehensibility. Instead of a fixed scale which is commonly used in comprehensibility studies, listeners could move a curser indicating the degree of comprehensibility as they were hearing the samples. The study demonstrated that comprehensibility judgement could indeed evolve over the course of a speech sample and that listeners were able to position the curser after 30 seconds of speech.

As a matter of fact, a vast majority of studies use 20 to 30 seconds samples, usually from the beginning of the recording, and excluding any initial disfluencies (Crowther et al., 2015; French et al., 2020; Isaacs & Thompson, 2022). These, and studies using longer samples (45 seconds in Zielinski & Pryor, 2022 for instance) systematically use a single sample per participant and condition, whereas studies using shorter samples (utterance or phrase-sized) use two or three samples for each participant and condition (Derwing & Munro, 1997; Munro & Derwing, 1995a; Nagle & Huensch, 2022).

The recordings to be judged in our study are free speech, and as such the performance of the speakers varies a lot across the whole span of the recording in terms of content and fluency. In order to render this variability, we decided to use multiple samples per recordings. This meant that whatever size we would land on, it couldn't be much longer than 30 seconds otherwise the total audio material would not be sufficient to multiply the excerpts (the free speech task yielded two-to-five-minute recordings). To decide on the sample length, we ran a small pilot study where the two author's supervisors performed the comprehensibility and accentedness rating task with samples of either IPU size (two to five seconds long, five words on average), utterance size (10 seconds long, 15 words on average), or 20 to 30 seconds excerpts. They both found the 20-30 second samples to be the most comfortable format for the rating task, in accordance with Nagle et al. (2019). Therefore, three 20 to 30 second samples were selected within each recording - one in the first third, one in the middle part, one in the final third. We made sure samples started and ended at IPU boundaries but we did not exclude initial or final disfluencies.

To give a baseline to the judges and to make sure ratings were not randomly assigned, we included in the final set 12 excerpts from native French speakers (6 from a female, 6 from a male) belonging to the B-FREN3 corpus (Drouillet et al., 2023). The inclusion of native speech samples is a common practice in studies with similar design for the same reasons stated above (Gordon & Darcy, 2022; Munro & Derwing, 1994 & 1995a).

In total, the audio set for the task included 48 samples from the Phase 1 free speech recordings (3 samples X 8 participants X 2 times) and 12 native French samples for a total of 60 samples to be rated. We also added a training phase with 4 samples belonging to 2 native French and 2 L2-French speakers. Those samples belonged to speakers that were not featured in the test audio set (2 were from two participants who completed the pretest but dropped out of the study, the other 2 were from two other native speakers from the B-FREN3 corpus).

To optimise samples equivalence in loudness and quality, all excerpts were converted to 16000Hz sampling frequency sounds and normalised for peak amplitude using the scale peak function in Praat set at 99dB. To control for the potential order of presentation effect, the 60 excerpts were then arranged into three different randomised sets, creating three single sound files containing the 60 excerpts in different order. In a revision of their 1995a study, Munro & Derwing (2022) commented on the fact that they did not find differences in ratings between using different randomisation for each listener vs grouping listeners and assigning a randomisation to each group. In the three sets created, each excerpt was preceded by a warning tone and followed by 15 seconds of silence.

A questionnaire was created using the online tool Lime Survey to collect the judges' ratings. For each excerpt, judges were provided with two 9-point Lickert scales, one for comprehensibility, one for accentedness.

The type and length of scales used for this kind of judgement task has been debated. A majority of studies have used fixed point Lickert-type scales with descriptors at the scale endpoints (such as "very easy/difficult to understand" for comprehensibility and "very/not at all accented" for accentedness). The scale size has varied from 5 points (Isaacs & Thomson, 2013; Zielinski & Pryor, 2022), to 7 points (Derwing et al., 2008; Southwood & Flege, 1999), and for a vast majority of studies to 9 points (Derwing & Munro, 2013; French et al., 2022; Munro & Derwing, 1995a). The number of scale points should reflect the different levels a rater is able to distinguish. A too large span might provoke precision loss if the raters are not able to clearly differentiate between levels of the scale. A too narrow scale might induce a ceiling-effect (Isaacs & Thomson, 2013).

To answer this issue Southwood and Flege (1999) compared accentedness ratings made on a 7-point scale to direct magnitude estimation (DME) ratings. The latter type of rating does not involve any endpoints, but rather, raters indicated the difference between a reference speech sample and the speech samples to be judged as a ratio. If the reference sample was 1, a speaker with an accent twice as strong would be rated 2, and a speaker with an accent half as strong would be rated 0.5. By analysing the dispersion of the DME scores obtained, they concluded that the 7-point scale was too narrow to capture all the distinctions the raters percieved, and that 9 or 11-point scales would be better adapted.

Following this work, Munro (2018) replicated the study on comprehensibility ratings. He compared DME ratings with 9-point scale ratings and found that the latter was valid and reliable for measuring L2 speech comprehensibility. In the last 10 years, authors have also tested larger continuous scales (1,000-point sliding scale in Crowther et al., 2015 and Saito & al., 2017), and dynamic rating on a computer interface which allows the rater to click a "decrease comprehensibility" or "increase comprehensibility" button in real time (Nagle et al., 2019). Still, so far the use of a 9-point scale has been validated by several studies, and it is widely used in the field of L2 pronunciation research making results comparable. It also presents the advantage of being adapted to measures of both accentedness and comprehensibility, and has been shown to be easy to use for raters.

In order to test the experiment and make sure it was adapted to our participants, two native French speaker and fellow researchers in the lab tested the experiment and validated the material.

3.4. PROCEDURE

Seven of the 10 judges recruited undertook the experiment at the same time in a computer equipped classroom on the university premises. High quality headphones were provided. Because of schedule constraints, the 3 remaining judges did the experiment individually at different times, with similar equipment, in a quiet room at the research lab. Judges were told they had to evaluate L2-French speech samples. They were provided with one sound file containing the training samples, and a second sound file containing the 60 excerpts to rate. They were also given a link to the online questionnaire. The experimenter (the author) told them to carefully read and follow the instructions provided in the questionnaire, carry out the training phase, and then ask questions if anything had to be clarified before starting on the testing phase.

The first page of the questionnaire included a definition of the comprehensibility and accentedness concepts. Comprehensibility and accentedness were defined as follow (translated from French by the author):

Definitions of comprehensibility and accentedness

"Comprehensibility relates to the degree of ease or difficulty with which you understand what the speaker says. You might understand everything the speaker says but understanding might require more attention and effort (think about these moments when you squint your eyes and frown). What we are interested in is specifically this effort, this little time span between hearing the words, and understanding them."

"Accentedness refers to a pronunciation that differs from that of a native speaker. You have to evaluate the strength of the accent you perceive, meaning the gap between the speaker's pronunciation and that of a native speaker of French. It is different from comprehensibility: it is possible that you understand a speaker effortlessly despite a very strong accent."

Since the whole experiment took approximately an hour, judges were instructed to take a short break after the first 30 excerpts as to limit fatigue. Apart from this pause, judges were asked to let the sound file play and not listen to the excerpts more than once.

After rating all 60 excerpts, the questionnaire's final section included a few questions related to the judges' profile (see Table 16, section 3.2., p. 213). Judges were compensated for their time with two tickets to a local cinema each.

3.5. DATA EXTRACTION

The ratings collected from the Lime Survey questionnaire were exported, and were first examined to verify if the ratings on native French samples were accurate. One judge (judge 6 in Table 16) presented very inconsistent and inaccurate ratings, suggesting they did not understand how to use the scales. This judge was therefore excluded from further analysis. Following this, ratings of the native French samples were also excluded from the dataset so as to avoid skewing the inter-rater agreement score. The final dataset was constituted of 432 items (9 judges X 3 excerpts X 2 times X 8 speakers).

As in many judgement-task based studies, Cronbach's Alpha was used to assess the reliability of the ratings collected as a function of the inter-rater variability. A α = .98 score was obtained for the comprehensibility ratings and α = .95 for the accentedness. These results confirmed that the changes applied to the design of the experiment compared to our first attempt reinforced its validity as we had hoped.

The ratings were then pooled by participant and time, and averaged in order to get a unique score for each.

The following and last Chapter of this document presents our results and the discussions on their interpretation.

CHAPTER VI - RESULTS & DISCUSSIONS

INTRODUCTION

The following section presents the results of the study, and is organised by type of measure. First, we present the acoustic measures collected on the L2-French samples in order to address our main research question:

RQ1: what is the effect of a prosodic training on L2 speech rhythm, and how does it compare to that of common oral expression and comprehension training?

After the discussion of the results on L2 acoustic measures, we present a selection of L1 data and address the issue of the T1-T2 variability in L1 in relation to that in L2.

The results on the comprehensibility and accentedness judgement task will then be exposed to address the following questions:

RQ2a: what is the effect of a prosodic training on comprehensibility and accentedness, and how does it compare to that of common oral expression and comprehension training?
RQ2b: are comprehensibility and accentedness scores correlated?

Finally, we will present the results of the segmentation task to address our last research question:
RQ3: what is the effect of a prosodic training on segmentation abilities, and how does it compare to that of common oral expression and comprehension training?

1. SPEECH RHYTHM MEASURES

1.1. RESULTS ON SPEECH RHYTHM L2

Data visualisation and statistics

The small number of participants in this study (n=3 for the Oral Group, n=4 for the Prosody Group) limits the possibility for relevant and appropriate statistic testing, especially for what concerns the between-group analysis. Consequently, most of our analyses rely on the description of our data in raw forms, through plot visualisation. Tables with data summaries are provided in the Appendix. All plots in the following sections present the data collected for each participant, at pretest (T1) and posttest (T2). Bar plots were chosen to present measures of quantity, with an indication of the exact value in red at the top of each bar.

Measures of duration are presented in box plots. The box represents the interquartile range (IQR), i.e., the 50% of data points situated above the first quartile (Q1, the 25th percentile) and below the third quartile (Q3, the 75th percentile). The horizontal line inside the box represents the median (Q2, 50th percentile). Lines that extend at either side of the box show the minimum and maximum values of the dataset, excluding outliers. The minimum is defined as Q1 - 1.5 * IQR, and the maximum: Q3 + 1.5 * IQR. Data points beyond these limits are considered outliers as they are significantly distant from the typical range of the data.

Despite the small number of participants in our study, the decent length of the speech samples (on average, 225 seconds in L2; 185 seconds in L1) enabled us to obtain substantial datasets for measures of intervals duration, suitable for the use of mixed-effects models.

Mixed models allow to test the predictive power of independent variables (fixed effects) on a dependent variable, while taking into account the effects of variables that are not measured, such as the inter-individual variation, or the effects of items presented in a task when it is the case (random effects).

Therefore, for each measure of duration (IPU, external pauses, voiced pauses, and syllables) mixed models were run with Time (T1 and T2), Group (Prosody and Oral), and Time x Group as factors. Given that measures of duration such as ours are susceptible to co-variate with the speed of speech delivery, Articulation Rate (AR) was entered as random intercept. Because the Participant variable and the Articulation Rate were redundant, we only entered Articulation Rate as random effect. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

For the IPU, external pauses, and voiced pauses duration (each are separate dependent variables), the following formula was used:

Dependent Variable ~ Time + Group + Time : Group + (1 | AR)

For the duration of syllables, an additional factor was added: Accent (non-accented and final accent). The following formula was used:

Syllable Duration ~ Time + Group + Accent + Time : Group + Time : Accent + Group : Accent + Time : Group : Accent + (1 | AR)

Models were computed in R Studio (2023) using the *lmerTest* package (Kuznetsova et al., 2017). Where fixed effects or interactions were found significant, post-hoc analyses were conducted using the *emmeans* package (Lenth, 2024), and p-values were adjusted using the Tukey method. Models' outputs are provided in the Appendix. As commonly recommended to improve the reliability of the model, our datasets excluded outliers as defined by the interquartile method explained above (Field et al., 2012). For consistency purposes, the box plots presented below also do not display outliers.

The reader is reminded once again that the data collected and commented below is by no means inferential-statistic friendly. The results from the models are merely indicative of the strength of trends observed in the raw data. In any case, generalisation is precluded.

Results are presented in four sub-sections corresponding to the four level of analysis of speech rhythm described in Chapter I & II (p. 44; p. 71). Sections below follow a progression from the micro-level, gradually moving towards the broader dimension of fluency.

1.1.1. Micro-level

Figure 22 below presents the nPVI scores in L2 French for each participant at T1 and T2. As explained in section CHAPTER V - 2.4.1 of the preceding Chapter (p. 202), the normalised Pairwise Variability Index (nPVI) is a measure of the durational variability between pairs of successive intervals, here syllables. In terms of rhythm type, the higher the score, the more stress-timed the rhythm, the lower, the more syllable-timed (see Chapter I, p. 30).

Since our speakers are native English speakers, we expected their scores to be in between around 60 and 40 (reported nPVI for native English and French, see Table 13, p. 202). Therefore, from L1-English to L2-French, and as the speakers develop proficiency in L2, scores should follow a decreasing tendency. Further, since the Prosody group received a specific training on the syllabic rhythm of French, we would expect the nPVI scores in this group to decrease more than in the Oral group.

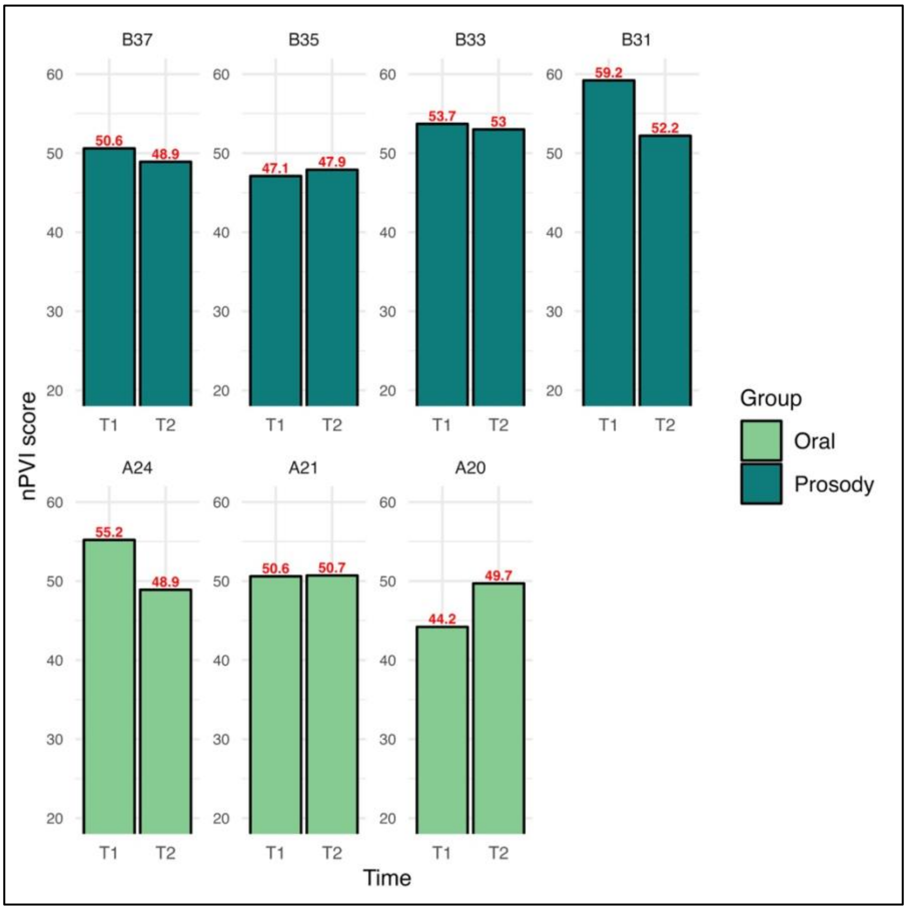


Figure 22 - nPVI scores in L2 French per Participant (participants' ID are shown at the top of each bar graph) and Time. Numbers in red indicate the exact nPVI score.

First of all, all scores are, as expected, between 40 and 60 which confirms that our results are relevant at least in regards to the literature. Scores range from 44.2 (A20) to 59.2 (B31) at T1, and 47.9 (B35) and 53 (B33) at T2. We observe that the score range at T2 is much more compressed than at T1. Interestingly, all scores at T2 seem to centre around 50 (+/-3). Participants who were the furthest from this value at T1, whether in positive or negative, all converged towards it at T2, while participants who were already close to 50 at T1 show minimal change at T2. This suggests that both trainings had a convergence effect on speakers' nPVI at T2, regardless of their starting point.

Indeed, the differences between T1 and T2 within each participant are quite variable: 0.1 to 2.3 for the four participants showing slim differences (A21, B37, B35, B33), and from 5.5 to 7 for the 3 participants showing the biggest gaps (A24, A20, B31).

When comparing the two groups, the main difference lies in the direction of that gap. In the Oral group, participants all go different ways and only one decreases at T2. The Prosody group shows more consistency: all but one speaker decreases their nPVI at T2. This seems to indicate that the Prosody training might have indeed been more helpful for participants to acquire a more syllabic rhythm, closer to that of native French speakers, as we would expect.

However, because the nPVI - as we computed it - is a unique value for each participant and condition, we could not test the significance of the differences observed within-speaker nor between-speaker. The nPVI results on the L1-English samples presented in section 1.3 (p. 248) will at least enable us to compare the intra-individual variation of nPVI scores between T1 and T2 in L2 and in L1. Table 19 in the Discussion section also provides a reference guide to help contextualise our nPVI scores in L1 and L2 amongst results reported in the literature.

We expected nPVI scores to decrease in T2, so as to show a transition from L1 English stress-time rhythm to L2 French syllable-time rhythm. It is the case for most participants in the Prosody group (three out of four), however to variable degrees. Contrastively, participants in the Oral group all show a different evolution. However, **the most striking observation is that both groups converged towards the same score at T2 (50, +/-3) regardless of their scores at T1. This suggests that both types of training had a similar outcome on the speakers' nPVI.**

1.1.2. Meso-level

Finer grained measures related to the accentuation level and based on the syllable unit are presented below. Figure 23 and Figure 24 allow us to look more closely at the accented and non-accented syllables' durations. The figures also indicate in red the duration ratio of the accented syllables to the non-accented syllables. That is, for example, for participant B31 at T1, accented syllables are twice as long as non-accented syllables.

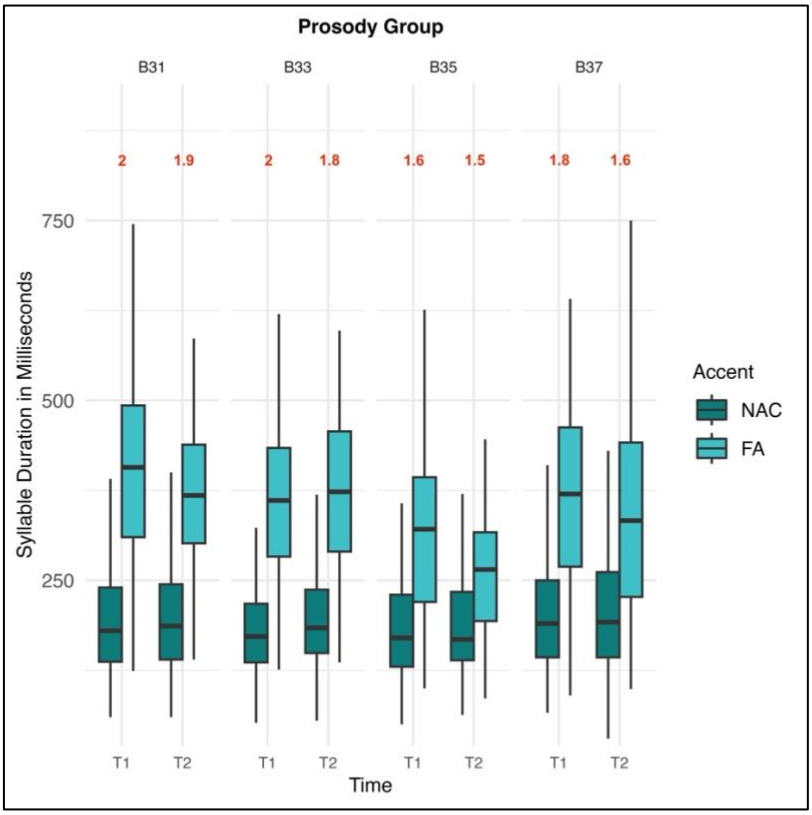


Figure 23- Duration of non-accented (NAC) and accented (FA) syllables, and duration ratio of the FA to the NAC syllables (indicated in red at the top) per Participant and Time for the Prosody group.

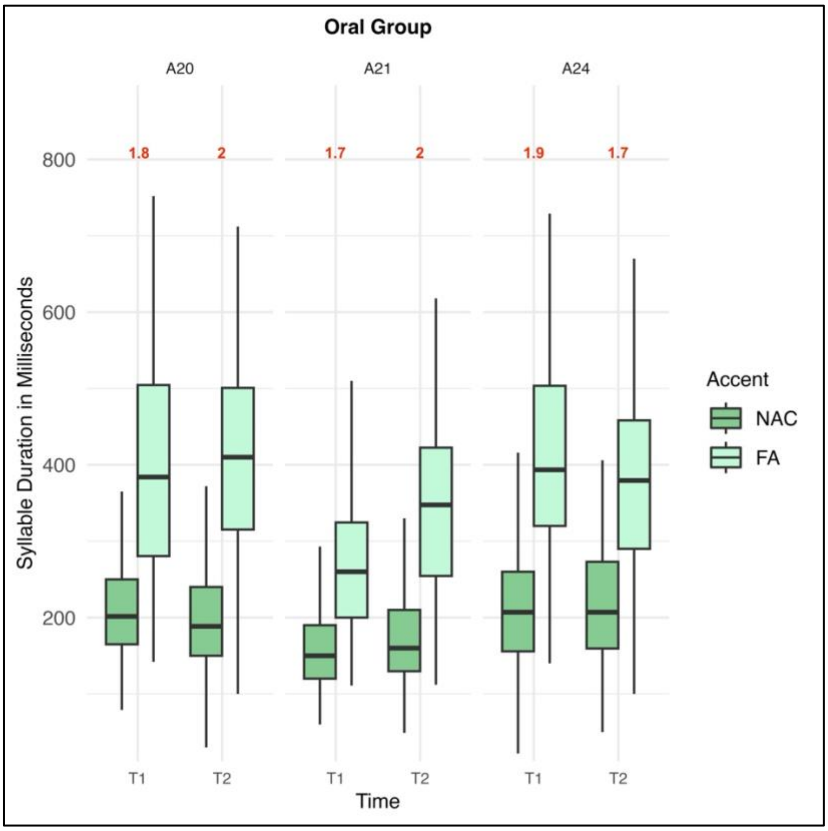


Figure 24 - Duration of non-accented (NAC) and accented (FA) syllables, and duration ratio of the FA to the NAC syllables (indicated in red at the top) per Participant and Time for the Oral group.

In both groups, the durations of the non-accented syllables between T1 and T2 are quite stable, and their dispersion limited. Conversely, the accented syllables are subject to greater durational variation and wider distribution. This is in line with previous work on French syllable duration showing that accented syllables present a greater elasticity than non-accented syllables (Astésano, 2001; Pasdeloup, 2004).

In the Oral group, accented syllables tend to lengthen at T2 (except for A24), whereas we see the opposite tendency in the Prosody group (except for B33). Focusing now on the duration ratio and taking as a reference point Delattre (1966)'s 1.78 for L1-French free speech, we note that two out of three participants in the Oral group are above this ratio at T2 as the ratio increased from T1 to T2.

Conversely, in the Prosody group, this ratio is consistently decreasing at T2, and for participants B35 and B37, it actually falls to 1.5 and 1.6, the latter being the reported ratio for L1-English in Delattre (1966).

The mixed model (see Appendix 2, p. 303) fitted to the data showed that the durational difference between non-accented syllables (NAC) and final-accented syllables (FA) is significant (β=-132.64, SE=6.25, p<0.001, 95%CI [-144.89 - -120.38]). In addition, there is a significant interaction between Time and Accent (β=-25.76, SE=8.64, p<0.003, 95%CI [-42.70 - -8.82]), reflecting the fact that non-accented syllables stay very stable across Time while accented syllables differ, in line with Pasdeloup (2004).

Most importantly, the interaction between Time, Group, and Accent is significant (β=42.54, SE=11.69, p<0.001, 95%CI [19.61 - 65.46]), indicating that the difference between the Prosody and Oral groups regarding the duration of accented syllables at T2 is significant. Since the Articulation Rate (AR) was entered as random effect in the model, these fixed effect and interactions are significant beyond the variance explained by the AR.

Still, we propose to take a closer look at the relation between syllable duration ratio and articulation rate in the search of elements of explanation for the decreasing ratio observed in the Prosody group. Following Pasdeloup's (2004) description of the correlation between AR and syllable duration ratio, where AR increases, ratio should decrease and vice versa. However, what we observe here is not as clear-cut. For ease of reading, we have summarised the AR and syllable ratio tendencies in the following Table 17:

	Participant	**AR**	**NAC to FA ratio**
Prosody Group	**B31**	↗ 0.27	↘ 0.1
	B33	↘ 0.05	↘ 0.2
	B35	↗ 0.03	↘ 0.1
	B37	↗ 0.2	↘ 0.2
Oral Group	**A20**	↗ 0.09	↗ 0.2
	A21	↗ 0.06	↗ 0.3
	A24	↗ 0.11	↘ 0.2

Table 17 - Articulation rate and syllable duration ratio (NAC to FA) tendencies from T1 to T2 in L2 French per participant.

In the Prosody group, three out of four participants (all but B33) show a tendency consistent with that described by Pasdeloup (2004) where a decrease in syllable ratio is consistent with an increase of AR. However, in the Oral group, A20 and A21 increase their AR while also increasing their syllable duration ratio; and only A24 follows the expected collinearity tendency with having an increased articulation rate coupled with a decreasing ratio.

Overall, the relationship between articulation rate and syllable duration ratio is quite inconsistent. Therefore, **these observations along with the output of the mixed model indicate that syllable duration ratio variations are partially independent from AR variations. Thus, it is possible that the decreasing tendency on syllable ratio observed in the Prosody group might be imputable to the training.** The work on the syllabic rhythm of French in the Prosody training might have led to the slight reduction of the accented syllable duration and to a more compressed duration ratio between non-accented and accented syllables. This is also reflected by the nPVI scores which also decrease (except for one participant) in the Prosody group, showing a reduction of the overall variability of syllable length.

1.1.3. Macro-level

This section presents the results on IPU quantity and duration, and on the duration of pauses in between IPU (external pauses hereafter).

Since participants produced recordings of variable lengths, the raw number of IPU is not comparable across participants. The IPU quantity is therefore expressed relatively to the total IPU time (sum of all IPU durations) following the formula: IPU total count*100/IPU total duration, similarly to Derwing et al. (2009), and Bosker et al. (2013). Figure 25 below presents the results on IPU quantity for each Participant and Time.

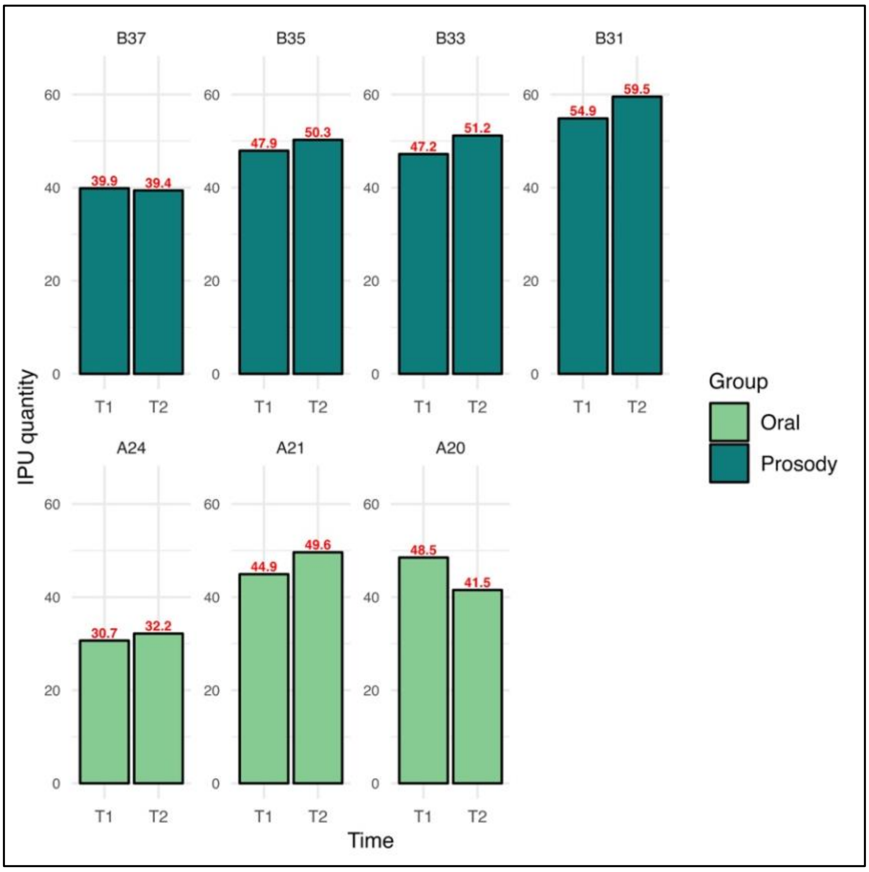


Figure 25 - IPU quantity per Participant and Time for the Prosody group and the Oral Group. The exact value is indicated in red, participants' ID are indicated above each barplot.

For a majority of participants, the quantity of IPU is greater at T2. However, A20 in the Oral group goes the opposite route, as well as B37 in the Prosody group for whom the difference is quasi-null. Our expectation was to see the number of IPU decrease at T2 as it is usually associated with learners' improvement, especially going from L1-English to L2-French as we demonstrated in Judkins et al. (2022, see Chapter II, p. 92). Another tendency that emerged in Judkins et al.'s study is the relationship between quantity and duration, in the sense that more IPU is associated with shorter IPU, and less with longer IPU.

We can therefore assume that since most participants produce more IPU at T2, they should be shorter too. Figure 26 below presents the data on IPU durations.

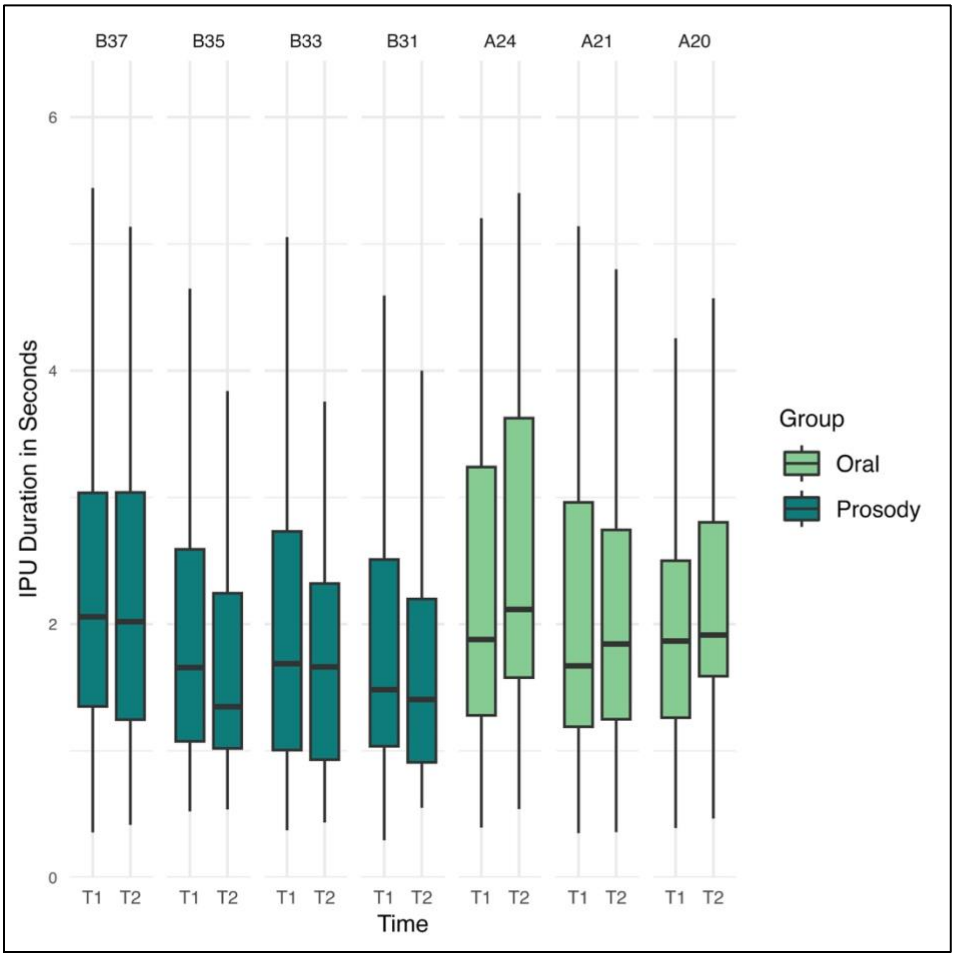


Figure 26 - IPU duration per Participant and Time for both groups.

The quantity-to-length relation stands for most participants in the Prosody group. B37 is the only one showing almost no difference in both quantity and duration across Time. In the Oral group, A24 and A21 contradict the quantity-to-length relation as they both make longer and more IPU at T2. A24 also stands out with a notably more extended boxplot than the others, showing a higher degree of variability and a wider range of IPU length in their speech. A20 however makes slightly longer and less IPU at T2, following the expected tendency usually associated with improvement.

Overall, the two groups are showing opposite tendencies with a decrease in IPU duration for all participants in the Prosody group, and an increase in the Oral group. The model fitted on the IPU duration data showed that none of the fixed effects (Time, Group, Time:Group) were statistically significant (see Appendix 2, p. 303).

Taking together IPU quantity and duration (Figure 25 and Figure 26), the tendencies observed are contrary to what was expected. Indeed, studies have shown that improvement in learners' speech usually translates to less and longer IPU, however it is not what we observe here, especially in the Prosody group. Fluency measures presented in the next section will help explain these observations.

We now turn to the duration of external pauses, i.e., pauses of at least 250ms of silence between IPU (Bosker et al., 2013). External pauses length can vary a lot - as shown by the size of some boxplots in Figure 27 below - because some of these between-IPU pauses can also contain voiced pauses or false starts (see Figure 31 for the proportion of filled external pauses). In that case, the voiced pause or false start is preceded and followed by a silence of at least 250ms. There are also occurrences of multiple voiced pauses and/or false starts within the same between-IPU stretch, which significantly lengthens the total duration of the external pause.

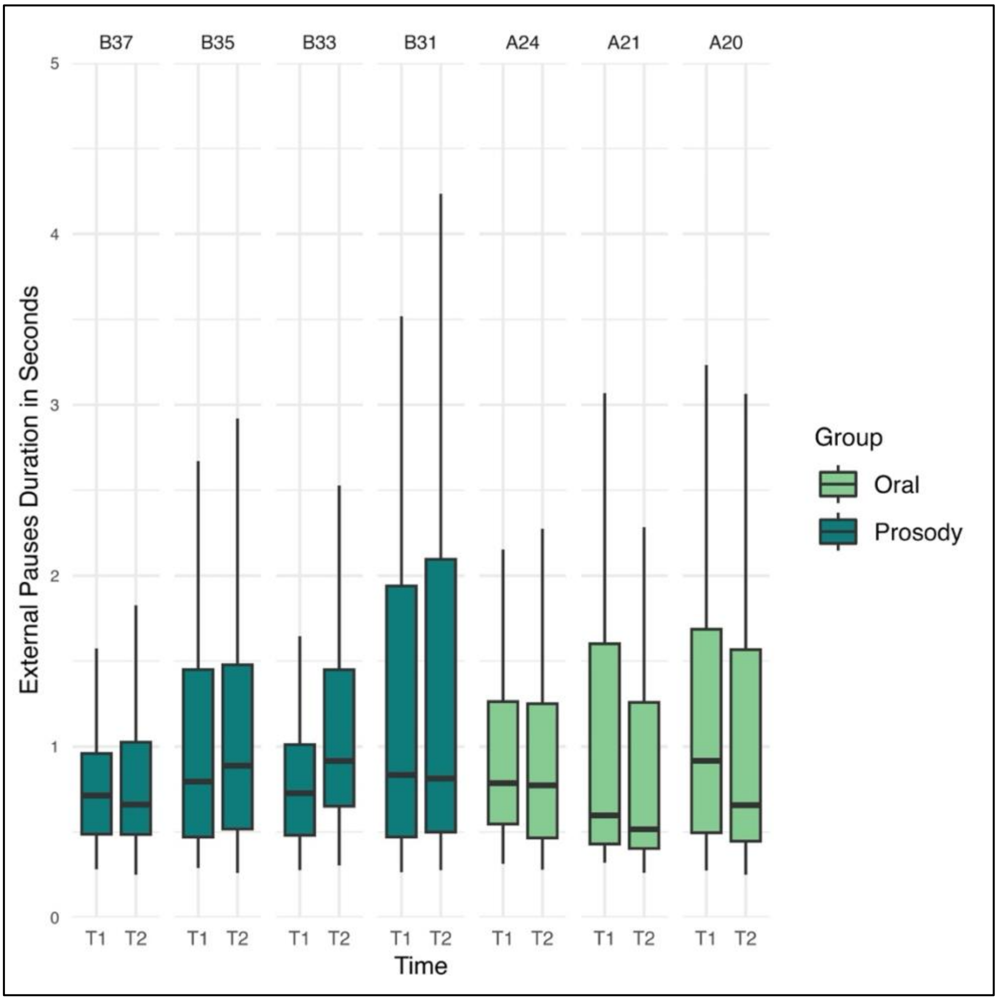


Figure 27 - External pauses duration per Participant and Time for both groups.

The results are quite inconsistent across participants. In the Prosody group, B35 and B33 show an increase at T2, B37 a decrease, and B31 nearly no change. In the Oral group A21 and A20 show a decrease while A24 stays near equal. The model fitted on the duration of external pauses shows that there is no significant direct effect nor interaction of Time or Group.

Overall, the two groups show opposite tendencies, the Prosody group tends to increase the duration of external pauses, the opposite for the Oral group. These trends are more pronounced in the mean values presented in Appendix 2 (p. 310) than in the medians shown on this box plot.

The evolution of the macro-level of speech rhythm from T1 to T2 in the two groups is rather puzzling. **Our results do not follow the expected improvement generally reflected by less and longer IPU associated with shorter external pauses. Instead, the Prosody group produces more and shorter IPU, and longer external pauses, while the Oral group mainly produces more and longer IPU, with shorter external pauses**. The next section on fluency measures helps understand what is going on.

1.1.4. Fluency measures

Figure 28 below presents the articulation rate for each participant in each condition. The articulation rate corresponds to the total number of syllables (excluding voiced pauses but including false start and lengthenings and code-switching) divided by the total speech time excluding IPU-external pauses. It is expressed in number of syllables per second. We would expect the articulation rate to increase from T1 to T2 as a sign of improvement in speech delivery and automaticity.

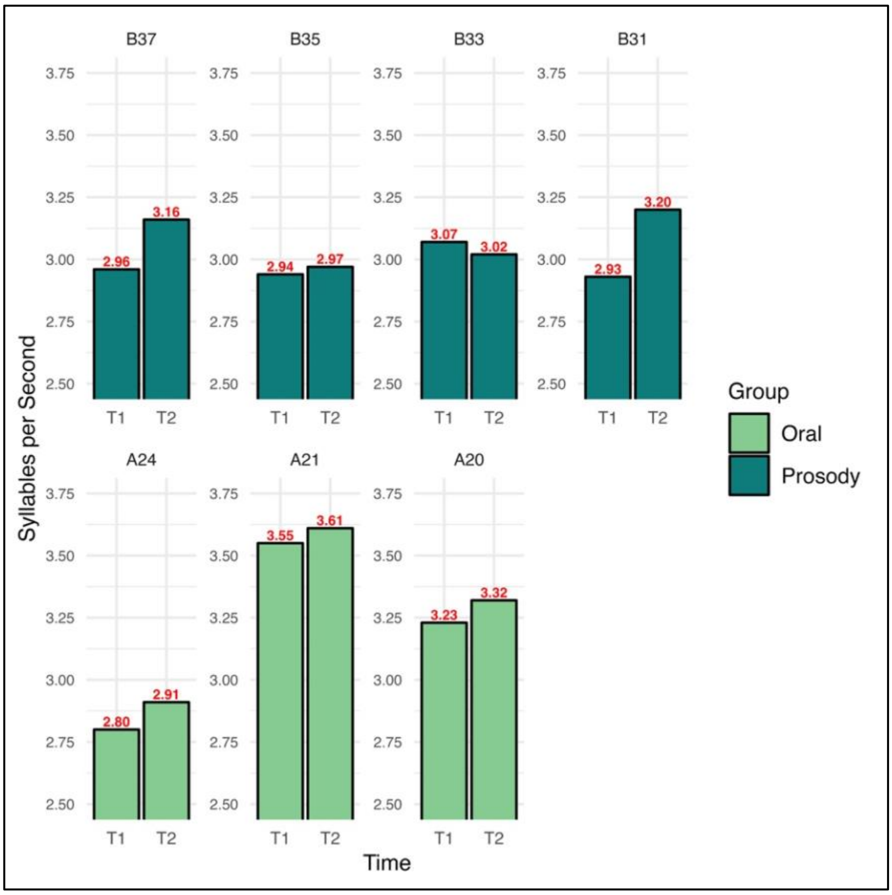


Figure 28 - Articulation rate (syllables/second) per Participant and Time. Numbers in red indicate the exact value.

All participants but one (B33) increase their speech rate at T2. The greater improvements belong to participants B37 (+0.2) and B31 (+0.27) from the Prosody group, whereas in the Oral group, the average increase is of 0.08. While the degree to which the articulation rate increases remain quite small, both groups still improved.

Figure 29 below presents the results on the total quantity of disfluencies within IPU. These include all types of disfluencies: voiced pauses, false starts, lengthenings. The total count of disfluencies for each participant in each condition is multiplied by 100 and divided by the total time of the corresponding speech sample (sum of IPU and external pauses durations) similarly to Bosker et al. (2013).

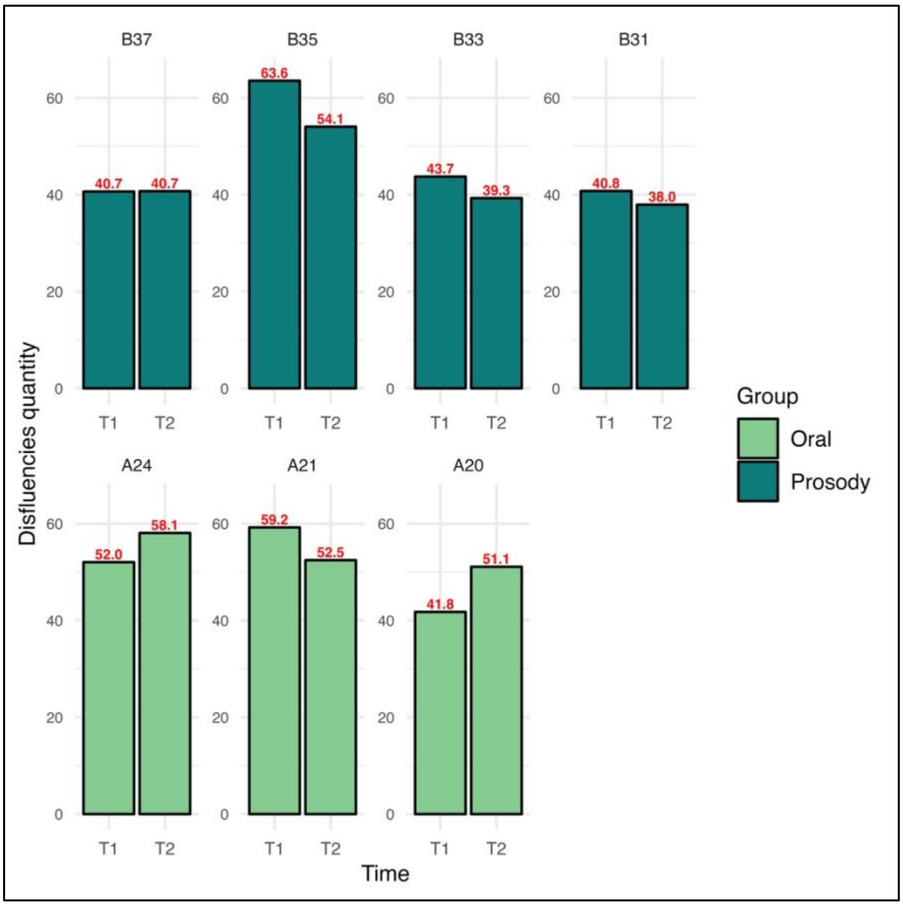


Figure 29 - Quantity of disfluencies within IPU (total disfluencies / total time *100) per Participant and Time. Exact values are indicated in red.

In the Prosody group, participants tend to produce less disfluencies at T2. B37 stays perfectly constant across times. In the Oral group, all participants but one (A21) produce more disfluencies at T2. This indicates an improvement for most participants in the Prosody group, and along with the increase in articulation rate, it helps explain the decrease in IPU duration observed above. Surprisingly, two participants in the Oral group actually worsen after training, despite having increased their articulation rate.

Figure 30 below shows the duration of voiced pauses located within IPU.

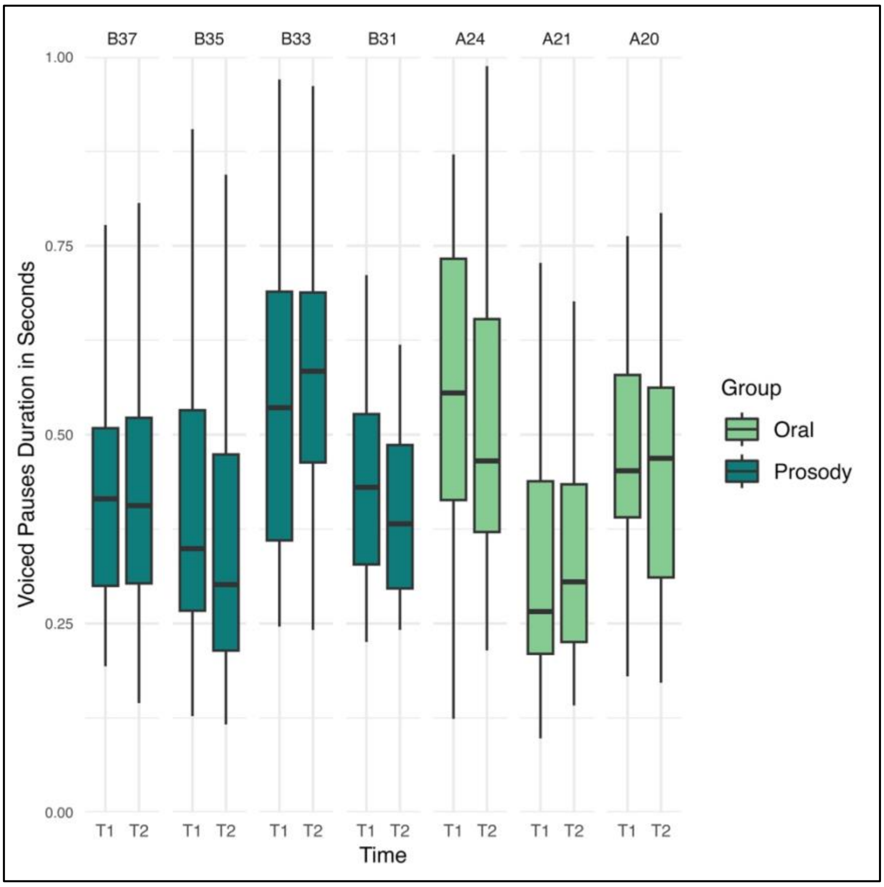


Figure 30 - Within-IPU voiced pauses duration per Participant and Time.

In the Prosody group, all participants but one shorten their pauses at T2. The participant (B33) who lengthened their voiced pauses at T2 is also the one that decreases slightly in articulation rate, which suggest a link between the two phenomena. As a matter of fact, the model fitted on the duration of voiced pauses shows a significant intercept (β=0.46, SE=0.04, p<0.001, 95%CI [0.37 - 0.55]) suggesting that the articulation rate impacts the voiced pauses' duration. However, none of the fixed effects were found significant (see Appendix 2, p. 310). Yet, the shortening of voiced pauses for most of the Prosody group indicates an improvement.

In the Oral group, only A24 shows a decrease, while A21 and A20 tend to increase their voiced pauses' duration at T2, which is contrary to the expected improvement.

After looking at disfluencies located within IPU, the following Figure 31 presents the proportion of external pauses that contain disfluencies (usually voiced pauses). This proportion is expressed as a percentage of filled external pauses over the total of external pauses. That is, we look at the extent to which pauses that actualise IPU boundaries are also used as a location for hesitations, in relation to programming the next chunk.

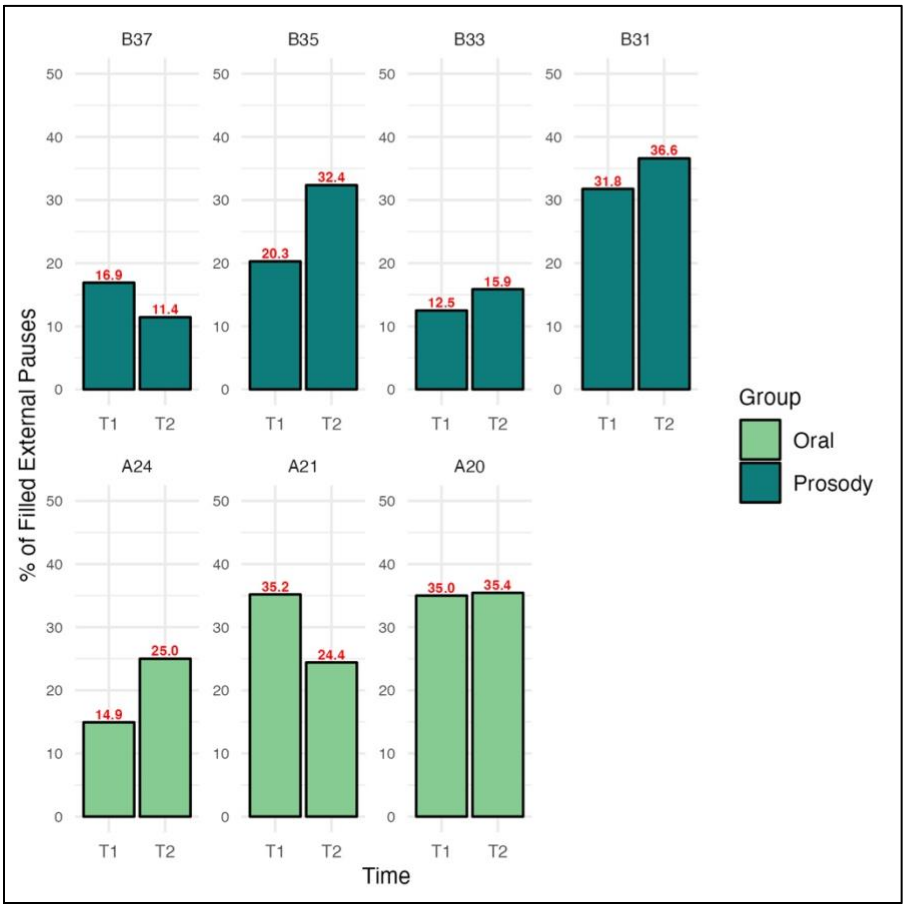


Figure 31 - Proportion of filled external pauses over the total of external pauses per Participant and Time. Exact valued are indicated in red.

In the Prosody group, all participants but one increase the proportion of filled external pauses by 6.76% on average. So, whereas participants in this group tend to produce less disfluencies within IPU in T2, they tend produce more in between IPU.

In the Oral group, there is no convergence between participants: A20 stays almost constant across Time, A21 reduces the number of filled external pauses at T2, and A24 increases it. This suggests that inter-individual variability is most impactful in the Oral group.

Overall, most participants in the Prosody group (three out of four) follow the expected tendencies associated with improved fluency for measures taken with IPU: faster articulation rate, less disfluencies, shorter voiced pauses. However, this group tend to increase the quantity of hesitations placed outside IPU.

In contrast, the Oral group only improves in articulation rate. Two out of three participants produce more disfluencies and longer voiced pauses at T2. In regards to disfluencies located outside IPU, participants all go in different directions.

This suggests that the Prosody group improved their fluency at T2, and especially the within-IPU cohesion, whereas the Oral group did not.

1.2. SUMMARY & DISCUSSION

In order to grasp the overall tendencies within each group, we have summarised the results presented in the sections above in Table 18 below. For each measure, we report the tendency shown by the majority of participants within each group, based on the mean values and expressed the original unit of each measure. In brackets, we indicate the number of participants following the main trend over the total of participants in that group. As an example, in the Prosody group, the main trend for articulation rate is an increase by 0.16 syllable per second on average, for three participants out of four. When no trend can be identified because each participant follows a different tendency, we use the term *scattered*.

MEASURE	ORAL GROUP	PROSODY GROUP
nPVI	Scattered	Decrease by 3.1 (3/4)
Syllable duration ratio NAC/FA	Increase by 0.25 (2/3)	Decrease by 0.15 (4/4)
IPU quantity	Increase by 3.1 (2/3)	Increase by 3.6 (3/4)
IPU duration	Increase by 0.262 (2/3)	Decrease by 0.189 (4/4)
External pauses duration	Decrease by 0.105 (3/3)	Increase by 0.093 (3/4)
Articulation rate	Increase by 0.08 (3/3)	Increase by 0.16 (3/4)
Disfluencies quantity	Increase by 8.6 (2/3)	Decrease by 6.9 (3/4)
Voiced Pauses duration	Increase by 0.035 (2/3)	Decrease by 0.03 (3/4)
Filled external pauses proportion	Scattered	Increase by 6.76% (3/4)

Table 18 - Summary of acoustic measures tendencies in L2 French by Group (brackets indicate the number of participants following the main tendency over the total)

We will now discuss the results presented in relation to our first research question:

RQ1: What is the effect of a prosodic training on L2 speech rhythm, and how does it compare to that of common oral expression and comprehension training?

We will comment and compare the results of both groups, following the same progression from the micro level to fluency measures.

1.2.1. T1-T2 evolution of the micro-level of rhythm

We have seen in Chapter II (p. 72) that L2 nPVI scores usually fall between the L1 and the target native language values (Carter, 2005; Ordin et al., 2011; Ordin & Polyanskaya, 2015). Since L1 English nPVI scores are consistently higher than L1 French ones (see Table 19 below for a summary of nPVI scores reported in the literature and in our study), we expected participants to gradually decrease their nPVI scores as they get closer to the syllable-time rhythm of French.

nPVI Reference Guide	
L1-English average from previous studies (Table 13)	63
L1-English average from both Oral & Prosody groups	49
L1-French average from previous studies (Table 13)	40
L1-French average from B-FREN3[22]	43
L2-French average (L1-English speakers) from Vieru et al. (2011)	62
L2-French average from Oral & Prosody groups	51

Table 19 - Summary of nPVI scores for L1 English, L1 French, and L2 French from previous studies, and from our data (in bold).

Our results show that they indeed follow the expected tendency of reducing their nPVI scores - that is the overall durational variability of syllables. However, the difference between T1 and T2 is very slim and might not be significant. Conversely, the Oral group's scores are very inconsistent in the direction of the evolution between T1 and T2. We could stop there and conclude that the prosodic training led participants in this group to follow the expected tendency associated with improvement, whereas the Oral group is too inconsistent to conclude anything.

But looking at both groups' scores at T2, a convergence appears. All participants at T2 obtained scores centred around 50 (+/-3) regardless of their scores at T1, and regardless of the distance separating their T1 score to 50. This suggests that both types of trainings had the effect of aligning participants on a common interlanguage score, which falls halfway between L1 English (around 60) and L1 French (around 40). Since the type of training does not seem to make a difference, it might be that the exposure and practice both groups benefited from during the training - albeit in different formats - could explain this. It is also possible that the familiarity with the posttest task at T2 had a homogenising effect, however this effect would be limited to the nPVI since it is not found on other measures.

Nevertheless, we must interpret the results on this metric with caution since the nPVI has not been used often in similar experimental design, and most of all, for speakers going from a stress-time language to a syllable time language. Moreover, a number of studies have shown the sensitivity of this measure to factors

[22] In order to have an L1 French nPVI reference score that matches our methodology of extraction and speech task to contextualise our results, we have computed nPVI scores on 1-minute samples of the conversation task from 2 speakers of the B-FREN3 corpus (see Judkins et al., 2023 & 2022 for details on B-FREN3). This is the mean of the 2 speakers' scores.

such as inter-individual variation, speech material, and word frequency (Arvaniti, 2012; Harris & Gries, 2011; Wiget et al., 2010).

We will come back to the interpretation of nPVI scores in a further section (p. 250), in light of the data collected in our participants' L1.

1.2.2. T1-T2 evolution of the meso-level of rhythm

Regarding the duration of non-accented and accented syllables, our results show that while the non-accented syllables remain stable across Time, the duration of accented syllables decreases in the Prosody group while they tend to increase in the Oral group. This is in line with Pasdeloup (2004) who also found that accented syllables are more susceptible to variate while non-accented syllables remain stable.

Consequently, the durational ratio of accented to non-accented syllable decreases in the Prosody group and increases in the Oral group. This contradicts the results from Alazard (2013) who found that after an MVT training, the ratio of beginner L1 English-L2 French learners increased. Yet, her measures were taken on read speech at an immediate posttest, making our results are not directly comparable.

There are very few available data in the literature regarding this ratio in L1 English and French. Delattre (1966) found a 1.6 ratio for L1 English and a 1.78 ratio for L1 French, Astésano (2001) also found a ratio of 1.7 in French L1. From these values, we would expect more of an increasing tendency from T1 to T2 as a sign of improvement. However, we found that the ratio at T1 in both groups was already of 1.8 on average which is slightly above the L1 French ratio from Astésano and Delattre.

This could be due to the tendency to over-articulate reported in the L2 literature (Barry, 2007; Gut, 2003), however this literature concerns the acquisition of a stress-time language, and to our knowledge, we do not dispose of empirical data on learners of a syllable-timed language in that regard. Nonetheless, if over-articulation might explain the high ratio at T1, the decrease of the ratio at T2 in the Prosody group can be seen as an improvement, as all participants reach an L1-French like ratio of 1.7 on average after training. Contrastively, the Oral group increase their ratio even more at T2. This suggests that the type of training might have a different impact on this ratio, and that only the Prosodic training has positive outcomes.

Furthermore, we saw that the evolution of this ratio across Time is partially independent from the effect of articulation rate (see Pasdeloup 2004). Lastly, the results of the mixed model indicates that the difference between the Prosody and Oral groups regarding the duration of accented syllables at T2 is significant (see Appendix 2, p. 310).

The results on the nPVI and syllable ratio in the Prosody group are for the most part consistent, since both indicate a reduction of syllabic durational variability. This tendency is in line with an interlanguage variety evolving towards L1 French rhythmic patterns, which the Oral group does not follow.

1.2.3. T1-T2 evolution of the macro-level of rhythm & fluency

Since the results on the macro-level of rhythm and fluency measures are co-dependent, we shall discuss them together. In regards to these two aspects of speech, the Prosody group and the Oral show divergent trends. In the Prosody group, most participants produce more and shorter IPU, longer external pauses more often filled with hesitations, an increase in articulation rate, less disfluencies, and shorter filled pauses after training. Previous studies have reported that improved fluency/proficiency tend to be correlated with less and longer IPU, and faster articulation rate (Judkins et al., 2022; Kormos & Dénes, 2004; Lennon, 1990; Préfontaine et al., 2016; Saito et al., 2018; Tavakoli et al., 2020).

Therefore, we might interpret the increase in IPU quantity and shorter IPU duration as contrary to improvement. However, the increased articulation rate along with a decrease of disfluencies and shorter filled pauses within IPU explain the decrease in IPU duration in a positive way. As a matter of fact, a decrease in number of disfluencies has been associated with improvement in L2 French (Trofimovich et al., 2017). This contrasts with findings from studies on other target languages, which reported mixed results on the relationship between fluency/proficiency improvement and number of disfluencies (Kormos & Dénes, 2004; Saito et al., 2018). However, Baker-Smemoe et al. (2014) have shown that findings regarding fluency development in a pair of language cannot be generalised to other language pairs.

Interestingly, external pauses duration increased, in relation to the increased proportion of external pauses filled with hesitations. While several studies associate shorter external pauses with improved fluency/proficiency

(Bosker et al., 2013; Cucchiarini et al., 2002; Lennon, 1990; Suzuki & Kormos, 2020; Tavakoli et al., 2020), in L2 French Préfontaine et al. (2016) found that the longer the pauses, the higher the fluency score. This makes sense in light of findings regarding pause pattern in L1 French (Grosjean & Deschamps, 1975; Judkins et al., 2022), which showed that French native speakers make less but longer external pauses than L1 English speakers.

Taking into account the specificities of the L1 French pattern regarding the macro-level of rhythm and perceived fluency, the overall results of the Prosody group support an improvement as patterns get closer to that of L1 French. Moreover, the improvement of the Prosody group seem to indicate that their training helped enhanced their speech programming processes, as suggested by theories on the link between utterance fluency and cognitive fluency (Goldman-Eisler, 1968; Segalowitz, 2010).

Conversely, the Oral group shows overall opposite trends, except for the articulation rate which also increases at T2, although in smaller proportion than the Prosody group. The Oral group tends to produce more and longer IPU, however they also increase the number of disfluencies within it, and the length of filled pauses, most likely accounting for the increased duration of IPU, rather than constructing more elaborated chunks. In addition, the duration of external pauses decreases, but the proportion of external pauses filled with hesitations differs across participants.

Therefore, considering L1 French patterns as exposed above, the Oral group does not seem to improve overall after training in regards to macro-level rhythm, and fluency.

1.2.4. Limitations

The results discussed above point to the superior benefits of the Prosodic training over the Oral training on participants' speech rhythm, as we had hypothesised. However, we must acknowledge some limitations to this tempting conclusion.

First of all, the small number of participants in each group prevented robust statistical tests of significance of Time and Group differences. The mixed models we have fitted on durational data showed that none of our fixed effects reached significance, except from the model on syllable duration, however the strongest effect comes from the difference between accented and non-accented syllable, which is hardly surprising. All of which is to say that all trends observed on the acoustic measures should be interpreted with caution and cannot be generalised.

Secondly, because of the schedule constraint on the assignment of participants into the two groups, we were unable to control for equivalence. We present again Table 12 below to discuss how between-group differences in terms of individual characteristics might have impacted our results.

	PROSODY GROUP	**ORAL GROUP**
Number of participants included in the analysis	4	3
English variety	South Western English (2) Scottish (2)	American (1) South African (2)
Age	66, 63, 21, 21	36, 24, 45
Musicians	2 (B37, B35)	0
Years in French class	Mean = 9	Mean = 2.3
Months in France	Mean = 21.5	Mean = 39.3
Mean weekly use of French (as a % of time)	37.5%	40%
Mean self-assessed speaking proficiency	1.75/6	2/6
Mean self-assessed listening proficiency	3.25/6	3/6
Self-reported CEFR levels	A2, A2, B1, B1	B1, A1, A1

Table 12 - Summary of predominant characteristics of each group.

The different number of participants in each group is definitely an issue, but beyond that, the two groups also differ in their L1 English variety. We have seen in Chapter II (section 3.4, p. 92) that L1 fluency patterns have an influence on L2 fluency patterns, and that such patterns might differ across L1 languages (Derwing et al., 2009; Huensch & Tracy-Ventura, 2017). It is therefore possible that the difference in English variety between English and Scottish in the Prosody group, and American and South African in the Oral group has an effect on L2 patterns. However, to our knowledge, the effect of English variety on L2 speech rhythm pattern has never been tested, and we cannot know if it would be greater than inter-individual differences.

Another potential influential difference between the two groups is the presence of two musicians in the Prosody group vs none in the Oral group. Considering the relation between musical aptitude and L2 phonological skills (see Chapter IV, section 2.2.2, p. 151), it is possible that B37 and B35 had an advantage and positively influence the group's overall performance. However, when looking specifically at these participants' performances, we notice that B37 is actually the one that often shows opposite tendencies as compared to the rest of the group, furthering them from L1 French patterns (less IPU of the same duration, shorter external pauses, no improvement on disfluency quantity). As for B35, they are the only one increasing their nPVI score at T2, and also has the most minimal increase in articulation rate (0.03 syl./s.). Therefore, the potential advantage from their musical background is far from being straight forward.

Regarding the participants starting level and experience with French, there is a considerable difference between the two groups as regards to the total of years during which participants were enrolled in a French course (which include primary & secondary school, university, and classes in language schools). The Prosody group reports an average of 9 years while the Oral group 2.3 years. However, as mentioned in Chapter V (section 2.2, p. 192), B37 and B33 from the Prosody group indicated years of French classes that do not match with their indicated age of onset and age at the time of the experiment (see Table 10). We suspect that they indicated the age of onset as their most recent experience in French class, but that the years in French class they reported include classes they had in primary or secondary school. Given their age (66 and 63), years of French in primary or secondary school might not be relevant.

In any case, there is also an unbalance, this time in favour of the Oral group who on average has spent more time in France than participants in the Prosody group. However, we do not dispose of detailed information on the actual exposure and practice they had in their time in France. Nevertheless, we do know that at the time of the experiment, both groups had a comparable weekly use of French (about 40% of the time).

As for their self-reported level in French L2, both groups report similar levels of speaking and listening skills, even though two participants in the Oral group reported lower CEFR levels than the others. As mentioned in Chapter V (section 2.2, p. 192), from the point of view of the experimenter/teacher, these two participants actually demonstrated A2 level abilities. Moreover, the experimenter/teacher noticed that overall, the Prosody group were more at a B1 level, whereas the Oral group was an A2 group.

But, even considering these differences between groups mostly indicating a higher proficiency level of the Prosody group, it does not explain their better performance at T2. In fact, several studies have shown that progress in fluency and suprasegmentals after instruction are most striking at beginner levels (e.g., Alazard, 2013). Therefore, following this reasoning, the Oral group should show more important progress than the Prosody group, and it is not the case.

The one argument that stands to explain the between-group difference in the sense of a worse performance and less consistency in the Oral group, is that two participants in that group (A21 and A24) missed two out of eight training sessions. However, they end up performing better on some measures than A20 who attended all sessions.

Overall, the absence of robust statistics, and between-group differences prevent us from asserting any firm conclusion on our results. However, after closely looking into the between-group difference, nothing striking emerges that could completely rule out the effect of the training as being responsible for the better performance of the Prosody group.

1.3. T1-T2 VARIABILITY IN L1 VS L2

Longitudinal designs in Second Language Acquisition (SLA) are common practice and rely entirely on the comparison of collected measures at different times. However, beyond the effect of a treatment, time itself is a factor of intra-subject variability, even in L1. L1 intra-subject variability of speech rhythm or prosodic aspects has been documented mostly across speech style (e.g., Astésano, 2001; Simon et al., 2010). A branch of research has also brought empirical evidence of the effect of emotional state on speech register variability within-subject (Révis, 2013; Scherer & Oshinsky, 1977). However, to our knowledge, there is little information on within-subject speech rhythm (including fluency) variability across time in L1.

Most SLA studies that include L1 data compare performances of participants at different times in L2, with a unique performance in L1. In Chapter II (section 2.4., p. 92) we saw that in L1-L2 fluency studies, L1 fluency is systematically taken as a set of measures at one given time, however just like any other aspect of speech, it is highly likely to be subject to a certain amount of intra-subject variability from one time to another.

In this section, we propose to look at the intra-subject variability across time in L1, and compare it to the intra-subject variability observed in L2. Since some acoustic measures in L2 yield very thin variability between T1 and T2, we can ask ourselves if the difference observed is relevant and accounted for by the training, or if it is within the margins of a variability occurring in L1 repeated measures as well. We also take this chance to look at our L1 data and make some comments on the comparison between L1 and L2 measures.

For each measure, we have compiled tables that show the variability between T1 and T2 both in L1 and in L2, for each individual, and also per group. The "Var" columns present the variability across Time as the subtraction of the T1 value from the T2 value. The "Total" rows correspond to the mean. Note that Var means are calculated on absolute values.

Tables for all measures are available in Appendix 3 (p. 315), but we only present and comment on a selection here. First, we look at the nPVI and syllable duration ratio, secondly the articulation rate and the IPU duration, lastly the number of disfluencies.

As a general assumption, if we consider the inherent intra-subject variability of speech rhythm aspects from one point in time to another in L1, then the variability between T1 and T2 in L2 should also include a part of inherent variability, disconnected from the effects of training. If the training has indeed an effect, then the overall T1-T2 variability in L2 should be greater than that in L1 because the effect of training would be superimposed on the variability of time itself. Furthermore, if the Prosody training had a stronger impact than the Oral training, then the difference between the variability in L1 and in L2 should be enhanced in the Prosody group. However, that only stands under the assumption that the time-only variability is comparable in L1 and L2, which is far from being an established fact.

1.3.1. nPVI & Syllable Duration Ratio

Table 20 below presents our data on nPVI. The last line of the table shows that the overall variability in L2 is greater than that in L1. However, this is only caused by the Oral group who indeed shows greater variability in L2 than L1. However, in the Prosody group, it is the opposite. Since the variability in L1 and in L2 are very comparable, and without enough data to run significance tests, it is impossible at this stage to know if the changes in L2 are caused by the training.

As discussed in section 1.2.1 of this Chapter (p. 241), we noticed that in L2_T2, nPVI scores converge towards 50 regardless of the T1 value, and in both groups. This suggests a similar effect of both trainings, and implies that the Prosody training did not enhance the progression of participants in that regard.

nPVI							
Participant	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	48.11	48.07	-0.04	50.61	48.88	-1.73
B35	Prosody	42.83	50.41	7.58	47.09	47.9	0,81
B33	Prosody	54.79	54.57	-0.22	53.69	52.97	-0.72
B31	Prosody	48.61	52.58	3.97	59.23	52.2	-7.03
Total Prosody		48.59	51.41	**2.95**	52.66	50.49	**2.57**
A24	Oral	46.42	49.7	3.28	55.18	48.87	-6.31
A21	Oral	47.76	44.83	-2.93	50.56	50.7	0.14
A20	Oral	49.72	47.47	-2.25	44.21	49.75	5.54
Total Oral		47.97	47.33	**2.82**	49.98	49.77	**4**
TOTAL ALL		48.35	49.88	**2.90**	51.65	50.22	**3.11**

Table 20 - nPVI and variability across Time, Language, Participant, and Group.

Aside from the variability comparison between L1 and L2, what is most surprising in this table is the nPVI values obtained in L1. We have seen from previous studies that L1 English nPVI usually centres around 60, whereas L1 French around 40 (see Table 13 and Table 19). The values obtained in L2 French are consistent with an interlanguage variety sitting in between the L1 English and French values from the literature.

However, the L1 English nPVI values of our participants are centred around 49, that is, they are pretty much equivalent to the L2 French scores. One of the differences between our scores and those from the literature could pertain to the speech style. Previous studies have mostly extracted nPVI of intervocalic intervals, on read speech or directed speech and on very small samples (a few sentences), whereas our study is based on syllables, free speech and samples lasting 185 seconds on average. Only Guilbaud (2002) found an nPVI calculated on syllables of 50 in L1 English free speech. The only other nPVI on syllable we found in the literature is Ordin & Polyanskaya (2015) who found a score of 69, however on directed speech, not free speech. More empirical data are needed to establish a reliable nPVI baseline on free speech and calculated on syllables.

Crucially, the L1-L2 comparison in our study completely contradicts previous findings in regards to the acquisition of an L2 rhythm belonging to a different type (i.e., stress vs syllable). Most studies on the acquisition of a stress language report L2 scores that sit between the two L1 in question, and nPVI scores increase with L2 proficiency as they get closer to the target. In one of the rare studies involving the acquisition of a syllable-time language from a stress-time one (L1 Russian L2 Latvian; Stockmal et al., 2005), it was found that nPVI scores in L2, instead of decreasing towards the target, exceeded the L1 Russian scores. That is, learners showed an increase in nPVI beyond that of their L1 (Stockmal et al., 2005). However, authors explain this tendency in relation to the specificity of the Latvian quantity system, which is clearly not applicable in our case.

Together, our results and those from Stockmal et al. disprove Ordin & Polyanskaya (2015)'s theory of speech rhythm acquisition. They propose that L2 rhythm acquisition should follow a similar pattern as L1 rhythm acquisition: an evolution from syllable-time rhythm to stress-time rhythm (in the case of a stress-time target language). Therefore, when the L2 is syllable-timed their prediction is to see a drop in nPVI from L1 to L2 and very little variation as proficiency increases since the universal tendency to lend on a syllable-time like pattern at the beginning stages of L2 learning already brings the learners close to the target (see Chapter II, section 3.1., p. 72). However, this theory rests on the assumption that durational variability is mainly caused by the presence or absence of vowel reduction phenomenon, but seem to leave aside the effect of accentual and phrasal lengthening.

Table 21 below shows the evolution of the final-accented (FA) to non-accented (NAC) syllable durational ratio across Language and Time. Before focusing on the Time variability, let us look at the ratios in L1 vs L2 in relation to the observations just made on the nPVI.

Syllable Duration Ratio FA to NAC							
Participant	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	1.3	1.4	0.1	1.8	1.6	-0.2
B35	Prosody	1.3	1.4	0.1	1.6	1.5	-0.1
B33	Prosody	1.4	1.5	0.1	2	1.8	-0.2
B31	Prosody	1.4	1.4	0	2	1.9	-0.1
Total Prosody		1.35	1.43	**0.08**	1.85	1.7	**0.15**
A24	Oral	1.3	1.1	-0.2	1.9	1.7	-0.2
A21	Oral	1.4	1.4	0	1.7	2	0.3
A20	Oral	1.2	1.3	0.1	1.8	2	0.2
Total Oral		1.3	1.27	**0.1**	1.8	1.9	**0.23**
TOTAL ALL		1.33	1.36	**0.09**	1.83	1.79	**0.19**

Table 21 - Syllable Duration Ratio FA to NAC and variability across Time, Language, Participant, and Group.

Ratios in L1 are overall centred around 1.35 whereas they are quite a bit higher in L2, around 1.8. Since the nPVI captures both the variability caused by vowel reduction phenomena and accentual and phrasal lengthening, and considering the ratio difference just pointed out, **it is possible that the nPVI in L1 and L2 ends up being similar because the absence of variability explained by vowel reduction phenomena in L2 French is compensated by the increase of the accented to non-accented syllable ratio.**

From a methodological point of view, this illustrates **the necessity to associate measures of accentual/phrasal lengthening - such as this ratio - to nPVI scores, in order to understand the origin of the variability in syllable duration observed**.

Turning to the ratio in L1 English, we notice that our values (1.35 on average) are much lower than the 1.6 ratio reported by Delattre (1966). This could be due to differing methodologies in the extraction of the measure and the speakers involved. We are not aware of other and more recent research reporting on that measure in the languages of interest here.

Shifting focus to the T1-T2 variability in L1 vs L2, Table 21 shows that the variability in L2 is greater than that in L1. This goes towards supporting an effect of the training. However, the L2 variability is higher in the Oral group and in the opposite direction as in the Prosody group: two participants in the Oral group increase their ratio, whereas all participants in the Prosody group decreases. As discussed in section 2.2.2 of this Chapter (p. 243), because participants in the

Prosody group get closer to the L1 French target, we might consider that the prosody trainings led to better outcomes than the Oral training in regards to this measure.

While the nPVI variability across time is comparable in L1 and L2, it appears that the syllable duration ratio is more stable in L1. We now turn to the articulation rate and disfluency quantity.

1.3.2. Articulation rate

Table 22 below shows the articulation rate in L1 and L2, and the T1-T2 variability within each language. The articulation rate has been found to correlate strongly with perceived fluency and to be a reliable cue to L2 speech proficiency. However, the table shows that the variability in L1 is greater than that in L2, which indicate the sensitivity of this measure to intra-subject factors other than the language and level of proficiency in it.

Articulation Rate							
Participant	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	3.63	3.55	-0.08	2.96	3.16	0.2
B35	Prosody	4.37	4.23	-0.15	2.94	2.97	0.03
B33	Prosody	3.95	3.88	-0.07	3.07	3.02	-0.05
B31	Prosody	4.61	4.55	-0.06	2.93	3.2	0.27
Total Prosody		4.14	4.05	**0.09**	2.98	3.09	**0.14**
A24	Oral	4.1	3.52	-0.59	2.8	2.91	0.11
A21	Oral	5.44	4.85	-0.59	3.55	3.61	0.06
A20	Oral	3.56	3.96	0.4	3.23	3.32	0.09
Total Oral		4.37	4.11	**0.53**	3.19	3.28	**0.09**
TOTAL ALL		4.23	4.07	**0.25**	3.06	3.16	**0.12**

Table 22 - Articulation Rate and variability across Time, Language, Participant, and Group.

Since the participants carried out the L1 and L2 free speech tasks back-to-back at each Time, we can assume that they recorded both L1 and L2 free speech samples in the same environment, mood, physiological and emotional state. Interestingly, all participants except one have a lower articulation rate at T2 in L1. Maybe because they felt more comfortable with the task that second time around, maybe they felt that less was at stake at the end of the experiment and after having spent four weeks in class with the experimenter. In any case, if the lower articulation in L1 was a consequence of their general state or attitude towards the task, it most likely also affected their L2 production. Yet, all except for one increase their speech rate at T2 in L2, which supports the conclusion that the training did have an effect that overpowered time-induced intra-subject variability.

It is also interesting to see how elastic the T1-T2 variability is in L1. While it remains tiny in the Prosody group, it is much greater in the Oral group. This might be related to the English variety difference between the two groups - American & South-African in the Oral group vs English and Scottish in the Prosody group (see Jacewicz et al., 2010; Robb et al., 2004).

Lastly, we must acknowledge that variability in speed delivery in L1 and L2 may not be caused by the same factors. While the speakers' general physiological and emotional state, along with their attitude and motivation towards the task might have an impact in both L1 and L2, variability in L1 can also be consciously induced, for pragmatic or stylistic purposes. In L2 however, especially at a pre-intermediate level such as our participants, speed of delivery is mostly bound by speech programming and process constraints (Segalowitz, 2010).

This also goes for the production of disfluencies, IPU, and IPU duration. As can be seen in Appendix 3 (p. 315), the variability in L1 exceeds that in L2, and once again these measures are most likely influenced by similar and different factors in each language.

1.3.3. Conclusion

The purpose of this section was to complete the picture of our L2 results with looking at L1 data, and to address a methodological issue regarding the time-induced variability intrinsic to longitudinal studies. The simple exploration of our data clearly highlights the non-negligeable variation between times in L1. This raises the question of how reliably can a T1-T2 change be attributed to a treatment, given the necessary variability induced by such experimental design.

Similarly to De Jong et al. (2015) who proposed two different ways of studying L2 fluency while excluding the influence of L1 fluency patterns on the data (see Chapter II section CHAPTER II - 2.4., p. 92), it would be interesting to find a way to strip off T1-T2 variability from training-independent factors, in order to isolate the effect of the training solely. The task is of complex nature since, as discussed above, between-time variability in L1 and L2 might share some factors, but also different ones specific to each language.

To our knowledge, this issue has not been investigated and research in this area would be welcome to gain knowledge on time-related variability within-subject, and to refine longitudinal approach methodology and findings.

One way to approach this in the future would be to take the statistic-testing route, however this is not necessarily compatible with ecology-forward experimental designs such as ours. Still, another way to find out if the T1-T2 changes in L2 have any relevance, and can be tied to the type of training, is to use perceptual measures such as human ratings of comprehensibility and accentedness (see Chapter III, section 1., p. 106). The next section presents our results on perceptual measures, and discuss the links with our findings on acoustic measures.

2. COMPREHENSIBILITY & ACCENTEDNESS

We now turn to the results of Phase 2 of our study which consisted in the collection of perceptual judgements from native French raters on the free speech samples of our participants.

These results will allow to address the following research questions:

RQ2a: what is the effect of a prosodic training on comprehensibility and accentedness, and how does it compare to that of common oral expression and comprehension training?

RQ2b: are comprehensibility and accentedness scores correlated?

After exclusion of the data belonging to participant A22 (see Chapter V, section CHAPTER V - 2.4, p. 201) a total of 378 scores of comprehensibility and accentedness respectively were collected (6 samples x 7 participants x 9 judges).

We used mixed effects models (R Studio, 2023; *lmerTest* package, Kuznetsova et al., 2017) on each of the dependent variables with Time and Group as factors, and Participant and Judge were entered as random effects. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The following formula was used:

Dependent Variable ~ Time + Group + Time : Group + (1 | Participant) + (1 | Judge)

Models' outputs are presented in Appendix (p. 310, p. 318, p. 320).

2.1. RESULTS

Table 23 below presents the mean comprehensibility and accentedness scores given by the nine judges for each group.

	COMPREHENSIBILITY		**ACCENTEDNESS**	
Group	T1	T2	T1	T2
Prosody	5.65 *(2.2)*	5.76 *(2.12)*	4.71 *(1.57)*	4.85 *(1.67)*
Oral	6.28 *(1.51)*	5.81 *(1.62)*	4.2 *(1.42)*	4 *(1.5)*
Native	8.7 *(0.77)*		8.87 *(0.4)*	

Table 23 - Mean scores (and standard deviations) of comprehensibility and accentedness for each group.

As expected, the Native samples were rated with very high scores and low variability which confirms that judges understood the task. The Prosody group improve both their comprehensibility and accentedness scores in T2, although in small proportion. The improvement is a little more important in accentedness (+0.14) than in comprehensibility (+0.11).

Conversely, the Oral group worsen their scores at T2, both in comprehensibility and accentedness. The drop is especially remarkable in comprehensibility (-0.47), while it is smaller in accentedness (-0.2).

Interestingly, the Oral group obtains overall better comprehensibility scores than the Prosody group both at T1 and T2. However, the decline at T2 indicates that the training did not help them at all in that regards. Regarding accentedness however, we see the opposite: the Prosody group gets better scores at T1 and T2, than the Oral group. In addition, in both groups, accentedness is rated more harshly than comprehensibility. This suggests a partial independence of the two constructs in the sense that low accentedness scores do not necessarily involve low comprehensibility scores, which is consistent with the literature (Munro & Derwing, 1995a, 1995b; see Chapter III, section 2., p. 106).

The model fitted on comprehensibility scores reveals a significant effect of Time (β=-0.47, SE=0.23, p=0.043, 95%CI [-0.92 - -0.01]), and a marginal significance of the interaction between Time and Group (β=0.58, SE=0.31, p=0.059, 95%CI [-0.02 - 1.18]). The model on accentedness shows that none of the fixed effects reached significance (see Appendix 4, p. 318).

The following figures allow us to look into individual trajectories in order to explore the groups' consistency.

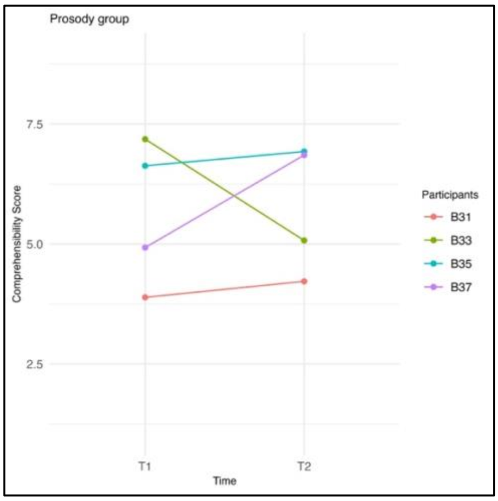


Figure 32 - Mean comprehensibility score for each participant of the Prosody group at T1 and T2.

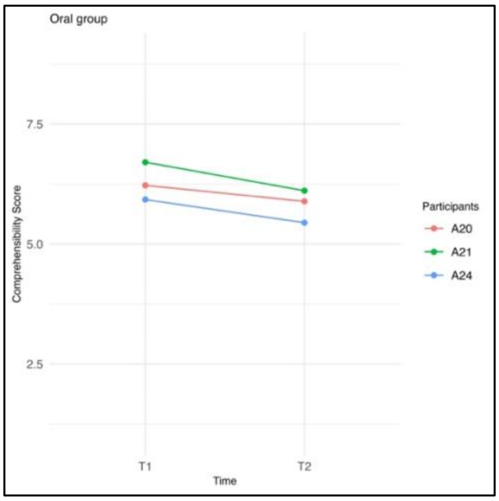


Figure 33 - Mean comprehensibility score for each participant of the Oral group at T1 and T2.

While all three participants of the Oral group follow the same tendency, the Prosody group is a lot more dispersed. Crucially, B33 is the only one to obtain worse scores at T2, and quite dramatically so.

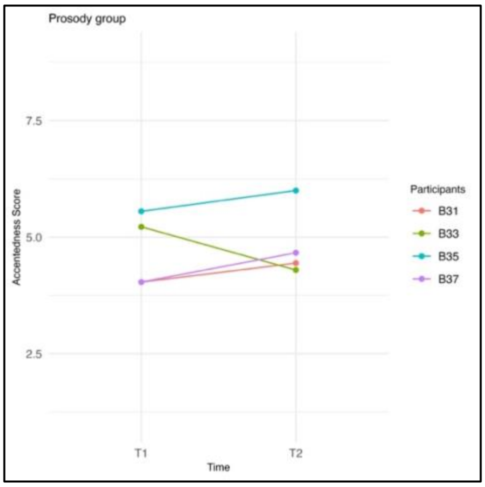


Figure 34 - Mean accentedness score per Participant and Time in the Prosody group.

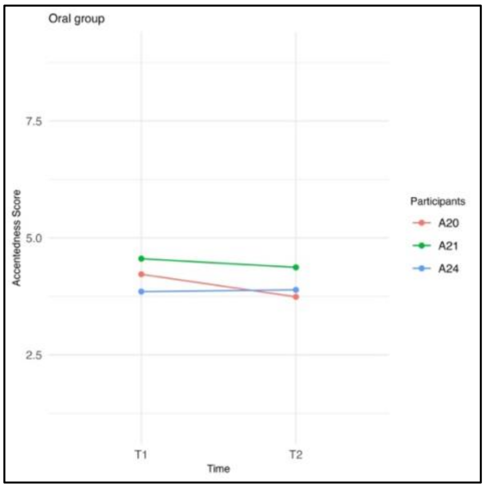


Figure 35 - Mean accentedness score per Participant and Time in the Oral group.

The same tendency appears for the accentedness scores in the Prosody group: B33 is the only one worsening at T2. In the Oral group, A24 is the only participant to improve at T2, however the difference is extremely tiny.

There is an explanation for B33's behaviour. This participant's speech sample at T2 was of far lesser sound quality than everyone else's. The participant seemed to be a little far from the microphone, and while they were still perfectly audible, the sound volume and overall quality did not match the other recordings.

Despite all of our efforts to compensate this default using tools in Praat, we could not match the quality of B33 T2 speech sample to the others. Therefore, it is fair to assume that B33's decline in both comprehensibility and accentedness is at least partially due to this problem.

Taking this technical issue into account, we ran the mixed models again, this time excluding B33. The model on comprehensibility showed once again a significant effect of Time (β=-0.47, SE=0.21, p=0.029, 95%CI [-0.89 - -0.05]) and this time a significant interaction of Time and Group (β=1.32, SE=0.30, p<0.001, 95%CI [0.73 - 1.92]). The model on accentedness also showed a significant interaction between Time and Group (β=0.70, SE=0.29, p=0.016, 95%CI [0.13 - 1.28]). This indicates that the Prosody group significantly improved in both comprehensibility and accentedness at T2 while the Oral group did not.

We shall now address the second part of our research questions:

RQ2b: are comprehensibility and accentedness scores correlated?

The between-group comparison illustrated the partial independence of the two constructs since the higher comprehensibility scores of the Oral group were associated with worse accentedness scores than the Prosody group. This means that a strong foreign accent does not necessarily induce lower comprehensibility, in line with previous findings (e.g., Munro & Derwing, 1995a, 1995b) .

However, when looking at individual trajectories, the progression between T1 and T2 in both comprehensibility and accentedness is consistent for all participants, i.e, if they improve in comprehensibility, they also improve in accentedness and vice versa. This suggests a collinearity of the two measures. We ran a Spearman's correlation test which confirmed a moderate positive correlation between comprehensibility and accentedness ratings (*r*= .414, *p* <.0005).

In conclusion, our results show that **only the Prosody group significantly improved in comprehensibility and accentedness, while the Oral group did not.** This indicates that **the prosodic training had a positive effect on the overall performance of our participants, as reflected by L2 French native judges' ratings of global measures. On the other hand, it seems that speaking and listening activities did not help participants improve.**

Moreover, **the two constructs appear to be colinear at the individual level, but partially independent at the group level**, as illustrated by a moderate positive correlation. The next section discusses these findings further.

2.2. DISCUSSION

2.2.1. Effect of the training modality

Our findings are in line with previous studies that have shown the importance of suprasegmental aspects on both constructs of comprehensibility and accentedness. Several meta-analyses support the positive effect of pronunciation instruction on global ratings of comprehensibility and accentedness (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019; see Chapter IV, section CHAPTER IV - 3.1, p. 158).

In addition, previous studies have also highlighted that suprasegmental instruction helped learners improve their accentedness and comprehensibility more so than no specific pronunciation instruction - i.e., speaking and listening exercises such as the training followed by our Oral group - especially on spontaneous speech (Derwing et al., 1998; Gordon & Darcy, 2016; R. Zhang & Yuan, 2020).

It is surprising however, that the Oral group did not improve at all, and even worsened after training, regarding both constructs. Yet, similar outcomes have been found in the past (Derwing et al., 1998; Gordon & Darcy, 2016).

Still, we must acknowledge that the two types of trainings did not only differ in the focus of instruction, but also in the tools and techniques employed in the classroom. We have seen in Chapter IV (section 3., p. 158) that using gestures and musical exercises boosts L2 learners phonological skills (Baills, Santiago, et al., 2022; Gluhareva & Prieto, 2017; Good et al., 2015; Y. Zhang et al., 2020, 2023, forthcoming). In our study, only the prosodic training included the use of gesture and musical activities, therefore, we are not able to determine the differential effects of the focus of instruction and that of the multimodal nature of the training. This will be addressed in the perspectives.

2.2.2. Linguistic correlates of comprehensibility and accentedness

We have seen in Chapter III (section CHAPTER III - 2., p. 114) that comprehensibility and accentedness share some linguistic correlates but also differ in some aspects (see Table 7, p. 121). While accentedness is exclusively related to pronunciation aspects - both segmental and suprasegmental, comprehensibility is also influenced by lexical, syntactic, and discursive aspects (Isaacs & Trofimovich, 2012; Saito et al., 2017).

In our study, only the Oral group's training was based on meaning-oriented activities, and included a lot of free-speaking exercises such as oral presentations and debates, and new vocabulary (see Appendix 1, p. 303 for a detailed description of the course). Based on this, we could assume that this group should have improved in those aspects, which should be reflected in their comprehensibility scores. Since it is not the case, we have to consider that: either the group did not improve at the lexical, syntactic and discourse levels, or if they did, not enough to compensate the absence of improvement regarding speech rhythm aspects. One last potential explanation would be that the listener-raters did not rely on linguistic aspects other than pronunciation-related ones.

In fact, the Oral group overall produced more disfluencies at T2 (see Table 12, p. 196). Fluency has been shown to correlate with comprehensibility, and less so with accentedness (Isaacs & Trofimovich, 2012; Saito et al., 2017; Trofimovich et al., 2017; Trofimovich & Isaacs, 2012). Consequently, the disfluency increase at T2 might play an important role in explaining the decline of comprehensibility scores in that group. Moreover, the fact that the decline of accentedness scores is a lot smaller than that of comprehensibility scores suggests that the number of disfluency is more strongly associated with comprehensibility than accentedness in L2 French, as in Trofimovich et al. (2017).

Turning to the Prosody group, acoustic measures of speech rhythm showed that participants in this group adopted patterns at T2 that are closer to native French ones, and that they reduced the number of disfluencies. The fact that only this group significantly improved in comprehensibility and accentedness demonstrate the direct link between speech rhythm and global ratings, in accordance with the literature.

Furthermore, our results on global ratings suggests that the T1-T2 evolution of the Prosody group goes beyond the time-only variability exposed in section 1.3 of this Chapter (p. 248), and supports that the training is responsible for the positive outcomes.

In the next section, we turn to the results concerning the effect of the trainings on participants' speech segmentation abilities.

3. SEGMENTATION SCORES

Finally, we present the results of the segmentation task in order to address our third research question:

RQ3: what is the effect of a prosodic training on segmentation abilities, and how does it compare to that of common oral expression and comprehension training?

As a reminder, participants listened to a series of short excerpts uttered by native French speakers, and were asked to repeat as many words heard as possible. Scores correspond to the proportion of correctly repeated words over the total of item presented.

Similarly to the procedure applied to acoustic measures and global rating scores, a mixed model was fitted on the segmentation scores with the Time and Group variables as factors, and Participant and Item as random effect (R Studio, 2023; *lmerTest* package, Kuznetsova et al., 2017). The following formula was used:

Segmentation scores ~ Time + Group + Time : Group + (1 | Participant) + (1 | Item)

In the analyses of acoustic measures and global ratings which were carried out on free speech samples, we had to exclude A22 because their samples were not comparable to the others. However, in the segmentation task there was no such issue. Therefore, the results presented below include A22 in the Oral group, which makes the two groups equivalent in terms of number of participants.

3.1. RESULTS

Figure 36 below shows the distributions and medians of participants' scores expressed in percentages. In total, each participant listened to and repeated 22 items. Consequently, box plots shown in Figure 36 each represent 22 data points.

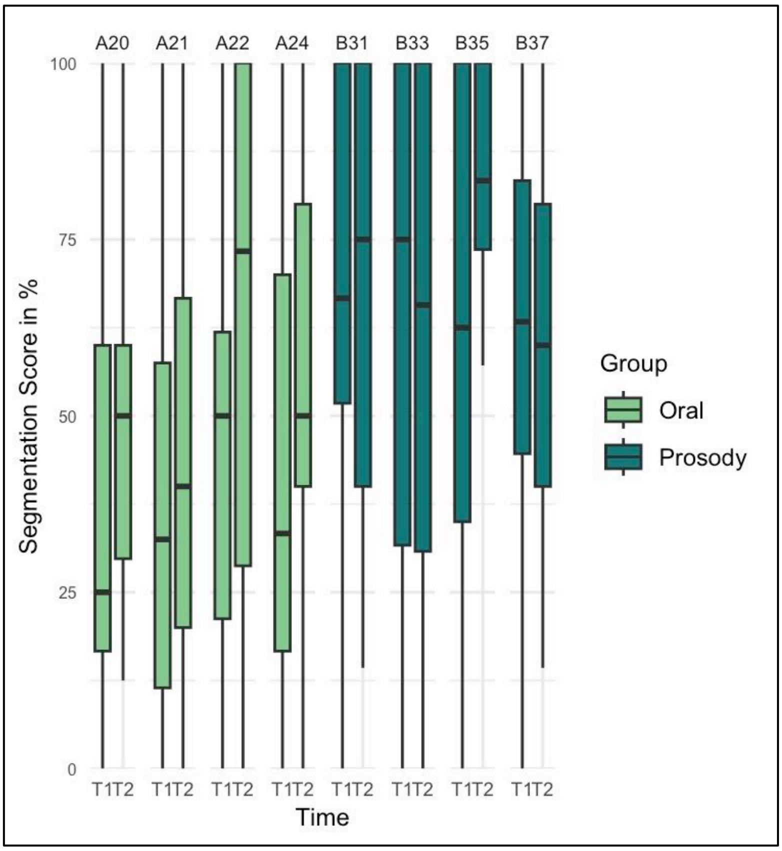


Figure 36 - Box plots of segmentation scores per Participant and Time.

The scores' distributions are quite large, which shows that performance across items is uneven. This makes sense given the various length and complexity of each item[23].

While all participants in the Oral group improve at T2, in the Prosody group B33 and B37 worsen. The model showed that the effect of Time was not significant (see Appendix 5, p. 320).

However, crucially, there is a significant difference between the groups' performance regardless of Time (β=19.08, SE=4.31, p<0.005, 95%CI [10.61 - 27.55].

For added precision, Table 24 gives the means and standard deviations for each participant at each time. While Figure 36 above showed lower medians for B33 and B37, mean scores show that B31 and B33 decline, while B37 actually improves. The T1-T2 difference for B31, B33, and B37 being very thin, they might not be significant in which case we could consider that their performance has not changed at all across times.

[23] Available on OSF

Oral group	T1	T2	Prosody group	T1	T2
A20	36.1 *(29.2)*	50.1 *(26.5)*	**B31**	66.7 (32.2)	65.9 (30)
A21	39.4 *(36.3)*	42 *(31)*	**B33**	63 (35.9)	61.9 (35.3)
A22	50.2 *(30.8)*	63.3 *(37.3)*	**B35**	61.5 (37.1)	74.6 (34.5)
A24	42.3 *(34.7)*	55.4 *(26.4)*	**B37**	60.7 (33.3)	63 (25)

Table 24 - Mean segmentation scores and standard deviations per Participant and Time.

As shown by the means, the Prosody group performed quite a lot better than the Oral group in T1 and T2 with all their scores above 60%. The Oral group remains below 50% at T1 and obtain overall lower scores than the Prosody group at T2.

However, the improvement between T1 and T2 is greater in the Oral group (by 10.7 points on average), and all participants improve at T2. Conversely, in the Prosody group B31 and B33 do not improve, and B35 is the only one showing an important score increase at T2.

Overall, the Oral group improves more than the Prosody group, although not significantly. The Prosody group does not seem to evolve at all from T1 to T2 except from one participant (B35). However, most notably, **the groups significantly differ right from the get go. Therefore, the overall absence of improvement in the Prosody group might be related to a ceiling effect.** This is discussed further in the following section.

3.2. DISCUSSION

Our results indicate that the prosodic training did not help participants improve in the segmentation task. This goes against our prediction, and previous studies which supported the positive effect of a training on suprasegmentals on listening skills (Kissling, 2018; Luu et al., 2021; McAndrews, 2023; Yenkimaleki et al., 2023). Most of these studies included a controlled group who received listening comprehension exercises, and these groups also improved after training, although in smaller proportion than groups trained on suprasegmentals. Therefore, the improvement - albeit not significant - shown by the Oral group is consistent with that at least.

In addition, the studies cited above consistently report the superior effect of instruction based on perception exercises over production ones. Even though the prosodic training included both types, the Oral group was more often exposed to authentic audio documents. Therefore, this difference might play a role in the better performance of the Oral group.

Yet, the main issue in the comparison between the Oral and Prosody group pertains to their significant difference of performance level at T1. Because the Prosody group already reached scores around 60%, their progression margin was reduced in comparison to the Oral group who started with scores below 50%.

Looking into the participants' profile (Table 12) in each group - which we have already discussed in section 1.2.4. of this Chapter (p. 245), we reached the same conclusion which is that groups are fairly equivalent in terms of length of residence, and weekly use of French. The only outlier is B31 who reports using French 100% of the time in their school/work environment, and rated themselves the highest for listening proficiency (see Table 10, p. 194). This is consistent with their performance at T1 which the strongest of all participants.

In order to prevent such unbalance between groups, it would have been ideal to assign participants to groups based on their T1 performances. However, as explained in Chapter V (section 2.2., p. 179) we did not have this luxury.

In any case, it is possible that the absence of evolution in the Prosody group is due to a ceiling effect. Yet, results from Charles et al. (2015) - who inspired the design of our segmentation task - show that their participants (L1 Chinese - L2 English intermediate-advanced) score at 66% at T1, and still reach 80% at a posttest.

Contrastively, our participants being at a pre-intermediate level, it is possible that the progression of segmentation skills varies according to the proficiency level. Unfortunately, to our knowledge this has not yet been investigated. To this day, speech segmentation in L2 remains an under-research area, and more data is necessary to better understand how instruction could help learners improve in that regard.

In the following section, a general discussion is proposed around the entirety of the results found, as well as methodological remarks in relation to the experimental design adopted here.

4. GENERAL DISCUSSION & CONCLUSION

4.1. SUMMARY OF FINDINGS

The goal of our research was to assess the effect of a prosodic training in comparison to a listening and speaking training, on the speech rhythm, comprehensibility and accentedness ratings, and segmentation abilities of L1-English learners of L2 French at a pre-intermediate level of proficiency.

The positive impact of instruction focused on suprasegmental aspects has previously been demonstrated on global ratings of comprehensibility and accentedness (e.g., Derwing et al., 1998; Gordon & Darcy, 2016), less so on acoustic measures of speech rhythm (e.g., Trofimovich et al., 2017), and even more rarely on segmentation abilities (but François et al., 2013; Luu et al., 2021 for instance).

In addition, meta-analyses of pronunciation instruction studies have concluded that designs should more often include other languages than L2 English, spontaneous speech sample, delayed posttest, and the association of both global ratings and acoustic measures (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019).

Furthermore, the use of embodied and musical activities have been shown to enhance the effectiveness of a training focusing in prosodic aspects (Baills, Santiago, et al., 2022; Gluhareva & Prieto, 2017; Good et al., 2015; Y. Zhang et al., 2020, 2023, forthcoming).

Considering all of the above, we designed an experimental protocol, driven by the will to highlight the ecological validity of the study. As such, the conditions for both the testing phases and trainings were created to be as close to natural as possible. Participants performed the testing phases in the comfort of their home, and the training sessions were held on the university premises and resembled regular L2 French classes - albeit with a limited number of students. In addition, the speech samples analysed were spontaneous in nature, and the posttest was delayed (a week after the end of the training).

Globally, our results show that the prosodic training (Prosody group) led to an improvement of the learners' speech rhythm and comprehensibility and accentedness scores, whereas the listening and speaking training (Oral group) did not. However, in regards to segmentation abilities, only the Oral group improved after training. Yet, the between group difference at pretest makes the interpretation of these results difficult.

Aside from the main research questions, we also showed the importance of considering the target language's rhythmic patterns in the interpretation of what constitutes an improvement. Because a vast majority of studies on L2 rhythm and fluency are focused on the acquisition of L2 English, what is commonly considered as an improvement is in fact mostly relevant to English as a target language. In the case of L2 French, some features that have been associated to worse global ratings in most fluency studies (e.g., longer external pauses), are actually positively rated by L1 French judges (Préfontaine et al., 2016).

Furthermore, we also confronted our L2 data to the participants' L1 data. In doing so, we questioned the influence of the between-time variability intrinsic to a longitudinal design on the interpretation of the T1-T2 evolution. We found that the between-time variability is often greater in L1 than in L2, especially for macro-level and fluency measures, highlighting the influence of intra-subject factors other than the language and level of proficiency. This supports the necessity to associate perceptual measures to acoustic ones in order to assess the reliability of the attribution of T1-T2 changes to the training.

The following section discusses the methodological choices made in building the study's experimental design, and the associated limitations.

4.2. ECOLOGICAL CHOICES TRADE-OFF

When building the experimental design for this study, we made the choice to compromise as little as possible on the ecological validity of the outcomes. Indeed, the general aim of this research was to see if classroom-style instruction on speech rhythm aspects would help learners improve their overall pronunciation in real-life settings.

To that effect, not only did we collect free speech samples, but we also did our best to provide testing conditions that would denature the participants' usual speaking habits as little as possible. As such, prior to the beginning of the experiment, participants were invited to a Zoom meeting where they could ask questions about the test phases. The written instructions of the whole procedure were sent to them, and they performed all the tasks at home, without any supervision. This choice was also driven by the idea of sparing participants' time by avoiding to ask them to come to the university premises for the testing phases on top of the training one.

Unfortunately, this led to the exclusion of some speech samples from the analysis. One participant (A22) did not fully respect the instructions for the free speech task. Most likely with the intention to do well, they clearly prepared their speech, and it was obvious to the three annotators (the main author and her two supervisors) that they were reading instead of speaking spontaneously. Another participant (B33) was sitting too far from the microphone for her T2 recording, and we were unable to recover a sound quality comparable to the other samples. This had consequences on the perceptual judgement task as the comprehensibility of this participant was rated much lower at T2. The lack of uniformity and control over the testing environment might also be a source of increased T1-T2 and inter-individual variability.

In addition, as commented in section 1.2.4. of this Chapter (p. 245), the participants' schedule constraints to attend the training sessions prevented us from controlling the between-group balance in terms of relevant language-related characteristics of each individual. This had an important impact on the results of the segmentation task where the two groups greatly differed right from T1, which made the interpretation of the T2 results very fragile.

There is certainly an appropriate middle-ground to be found there. Losing speech samples is not as much of a problem in a large group of participants, it is critical however in our case. One way to maintain the same level of naturalness of the samples collected could be to add more precision to the instructions, and more participants in order to anticipate the loss of some. However, this falls back onto the recruitment issue. In the case of a small sample size, it might be wiser to increase control over the testing conditions as to ensure that all collected data is exploitable.

Nevertheless, while the lack of ecological validity has been pointed out by recent meta-analyses of pronunciation studies (Lee et al., 2015; Saito & Plonsky, 2019), our results - as humble as they are given the small number of participants and limitations of the study - support the benefits of prosodic instruction on learners' speech rhythm and comprehensibility and accentedness. Most importantly, the progress made appear to be robust as they transfer to spontaneous speech, and are still reflected one week after the end of the training.

4.3. PARTICIPANTS FEEDBACK & PEDAGOGICAL IMPLICATIONS

One of the aims of this study was to build a classroom course on French prosody that could be used in an actual language teaching institution. Therefore, it was important to us to collect participants' feedback on their experience. Participants in the Prosody group overall reported that they enjoyed the course. They highlighted the fact that they had never been taught on the aspects covered in the course before, and they felt it was very valuable and helped them understand a lot better how spoken French "works".

However, they also had some criticisms. They all reported enjoying the second half of the training much more than the first. Some participants reported that they felt uncomfortable during the ice-breaker activity of the first session. This exercise involved imitating the French accent in English to illustrate the prosodic differences. For some, it was too exposing and should have been proposed after trust within the group and with the teacher had been established.

The Dalcroze exercises were found troubling for some participants. Even though the teacher explained (briefly) why these exercises were part of the course, participants expected a "more regular" pronunciation class and did not understand the relevance of these activities. They suggested to either spread them over the sessions in alternation with speaking exercises, or to include them in the middle of the course rather than at the beginning.

One of the participants (B33) had a strong negative reaction to the Dalcroze exercises. After the first two sessions, they expressed that rhythmical exercises made them feel very insecure as they felt they did not perform very well (e.g., they had trouble synchronising themselves to the beat). Fortunately, the rest of the course compensated this first negative experience and they were overall happy with the course at the end of it. However, they were the most reserved in the group, and we felt that this first negative impression might have led to an overall feeling of disappointment, which in turn might have lowered the motivation of this participant. This could explain the lack of attention to the quality of the recording at T2, as well as the lower performance at the segmentation task for this participant.

On a more positive note, participants expressed great enthusiasm towards the exercise using the rubber band between the hands to embody word final-lengthening, and they also all really enjoyed the sessions about *liaison* and *enchainement*. Overall, they were all very grateful and felt they had learnt a lot from the course.

The idea behind the design of the prosody course was to make a first attempt at creating a pedagogic progression, covering key aspects of French rhythmic and more broadly prosodic features, using tools and techniques that have been proven to enhance learning outcomes. The feedback from our participants encourages us to pursue this route and re-think the organisation of the different activities and foci.

While the benefits of multimodality for pronunciation teaching is supported by empirical evidence (e.g., Alazard, 2013; Baills et al., 2021; Y. Zhang et al., 2020), implementing activities that require body movement with adult students in a classroom might make some of them uncomfortable. Teachers should be mindful of the possible discomfort students might feel towards engaging their body in front of others. As our participants pointed out, establishing trust beforehand is preferable. Consequently, it is important to carefully design the first class with that objective in mind.

Nevertheless, our study's results go towards supporting the superior benefits of an explicit and multimodal prosodic training, in comparison to what is usually done in L2 French classes, i.e., listening and speaking activities without much specific instruction on suprasegmentals.

4.4. FUTURE WORK

The prospects that emerge from this work are as exciting as they are numerous. On one hand, the speech data gathered for this study, along with the B-FREN3 corpus from prior research have yet to be thoroughly explored. On the other hand, follow-up studies to the one presented here would allow to confront our findings, and/or to test other training and testing conditions.

Exploitation of the collected data

First of all, the speech samples collected for the present study include a reading task. It has been shown that in similar pronunciation instruction studies, controlled task tend to yield more robust results than free speech ones (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019). Running a similar analysis combining acoustic measures and perceptual judgements on the read-speech samples would allow to: a) verify if an effect of the training is found on that speech style - which would reinforce the results found on spontaneous speech, and b) bring additional knowledge on the difference between read and free speech for the analysis of the impact of a pronunciation training. For instance, Kennedy et al. (2017) found that measures of accentedness and comprehensibility were not associated to the same linguistic correlates on their reading task vs narrative task. Conducting a similar comparison between task would bring additional knowledge on this topic in L2 French.

In addition, in the study presented here, we limited our analysis of the meso-level of rhythm to the duration and relation between accented and non-accented syllables, considering only French's primary final accent. The analysis could be completed with data concerning initial accents, phrasal nuclear accents, and the use of f0 in prominence marking, all partaking in the hierarchical instantiation of linguistic rhythm.

Aside from looking at the effect of the trainings on these prosodic aspects, we could for instance investigate the frequency of initial accents and their acoustic realisation in L2 French, in relation to our results on final accents. By doing so we could test the Prosodic-Learning Interference Hypothesis (PLIH) proposed by Trembley et al. (2016) and inspired by Flege's (1995) SLM and Best's (1995) PAM-L2. The hypothesis postulates that a prosodic feature will be more difficult to acquire when it is nearly identical yet different from a an L1 feature, than if it is completely different. The French initial accent and the English word stress share common characteristics: they are both realised on an initial syllable[24] and with an f0 rise. However, the magnitude of the f0 rise tend to be moderate in French whereas it is more pronounced in English (Astésano, 2001; Jun & Fougeron, 2000; Lieberman, 1960). In addition, while English word-stress is lexically determined, French initial accent is rhythmically determined, and marks the left boundary at the level of the prosodic word (Astésano, 2017; and see Chapter I, section 3.3.2., p. 34).

[24] In French, the f0 peak can actually be realised on a medial syllable in the case of a long content word, even though it is consistently percieved on the initial syllable (Astésano & Bertrand, 2016)

Therefore, according to the PLIH the acquisition of the French initial accent should be more difficult than that of the final accent which differs greatly from L1 English patterns (at levels below the intonational phrase). This would contribute to better understand L2 rhythm acquisition in several ways: it would test the relevance of the PLIH on speech production instead of perception; and it would constitute one of the rare contributions about the initial accent in L2 French.

Turning to the data collected with the segmentation task, we gave a few examples in Chapter V (section 2.4.2., p. 209) of the segmentation errors participants produced in the repetition of the items presented in the segmentation task. We saw that errors can be related to enchainement and liaison phenomenon, misinterpretation of an accentual cue, omission of unaccented function words. In order to gain some insight at the development of learners' segmentation skills, we are interested in running a thorough analysis of the type of errors produced at T1 vs at T2. This might reveal hidden between-group differences, and tell us more about the participants' performances.

Lastly, the B-FREN3 corpus (Drouillet et al., 2023) is also a precious source of data since it comprises bi-directional French, English, L1, and L2 speech samples - it also includes free speech and read speech. The extraction of additional speech rhythm measures on this corpus would enable an insightful comparison of L1-L2 production from and to both languages, which in turn would make possible the observation of L1 transfer vs universal L2 speech rhythm patterns.

<u>Follow-up studies</u>

In order to reinforce the reliability of our results, and since the experimental material is ready for use, it would be valuable to replicate the study and expand the sample size. This would enable the use of statistic tests on the acoustic measures, and overall increase the robustness of the conclusions. Additional data would also enable to test correlations between measures from the different speech rhythm levels, and thus to understand better the relationship between them. Our results suggest for instance a relationship between the nPVI and the accented to non-accented syllable ratio. This is not surprising given that the primary accent in French is of durational nature. Yet, this observation illustrates a tight link between the two measures to be considered in French specifically. Disfluencies and articulation rate also have an impact on IPU and external pauses. A more thorough exploration of the connections between each level would allow for a more precise mapping of such inter-dependencies.

An additional training modality - prosodic training without gesture/music - could be added in order to isolate the effect of the multimodal nature of the prosodic training. Prior studies have supported the superior benefits of multimodal instruction (e.g., Alazard, 2010; Baills et al., 2022), but also of explicit instruction on suprasegmentals (e.g., Derwing et al., 2018; Gordon & Darcy, 2016). Yet, comparing the three modalities and on a spontaneous speech task seem rarer. Considering past research, we would expect that both prosodic trainings lead to more improvement than the control group, and for the multimodal training to yield better outcomes than the non-multimodal-prosodic training.

In addition, because our conclusions on the effect of the trainings on segmentation abilities are limited by the between-group difference at T1, it is our wish to properly re-design an experiment with this aim. L2 speech segmentation and how to help learners in that regard is still a very much under-researched topic (Charles et al., 2015), all the more so in French L2.

The observations made on the between-time variability in L1 vs L2 really piqued our curiosity. Since it seems that information regarding this issue is lacking, this constitutes a topic we would like to investigate further, ideally with a large enough sample size to enable the use of statistics, similarly to what De Jong et al. (2015) proposed. In their study on the relationship between L1 and L2 fluency patterns, the authors partialed out the variance explained by the L1 patterns from the L2 measures. In a similar way, we could imagine partialing out the time variability as well as the L1 influence to get "clean" L2 measures. However, as discussed in section 1.3. (p. 248), the question of the origin of the intra-individual time variability in L1 vs L2 must be explored further beforehand.

Another idea that came to mind during the writing of this dissertation concerns the relation between speech rhythm and brain rhythm. As exposed in Chapter I (section 3.2.2., p. 28), an important neural mechanism is at play in speech processing and comprehension (Peelle & Davis 2012). Neural entrainment to the rhythm of the speech heard allows for fast and efficient processing, thus access to comprehension. How fascinating would it be to test if a prosodic training as an effect on the capacity of the L2 learner to synchronise to the rhythm of the interlocutor?

Overall, studies concerned with the acquisition of a language other than English, and more specifically, of so-called syllable-timed languages are still far less numerous than L2 English studies. In order to test current L2 prosody acquisition theories such as that of Ordin & Polyanskaya (2015), or to propose new ones, more data is needed on different language pairs, mixing stress-timed and syllable-timed languages in all possible ways.

Lastly, as mentioned in the preceding section, one of the aims of this study was to design a classroom L2 French prosody course, with the idea of developing pedagogical resources and tools for teachers. We hope that our continuous efforts can eventually lead to the publication of an L2 French prosody manual, in the spirit of bridging research and classroom practices.

BIBLIOGRAPHY

Abel, C. (2018). L’enseignement et l’évaluation de la prononciation en classe de FLE et l’approche par compétences / l’approche actionnelle—Opposition ou synergie ? *Revue TDFLE*, *72*(72). https://doi.org/10.34745/numerev_1279

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press. https://doi.org/10.1515/9781474463775

Abry-Deffayet, D., & Chalaron, M.-L. (2010). *Les 500 exercices de phonétique: Niveau A1-A2*. Hachette français langue étrangère.

Acton, W., Baker, A., Burri, M., & Teaman, B. (2012). Preliminaries to haptic-integrated pronunciation instruction. *Pronunciation in Second Language Learning and Teaching Proceedings*, *4*(1). https://www.iastatedigitalpress.com/psllt/article/id/15218/

Alazard, C. (2013). *Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Etrangère*. [Thèse de Doctorat]. Université Toulouse II Jean Jaurès.

Alazard, C., Astésano, C., & Billières, M. (2010). The implicit prosody hypothesis applied to foreign language learning: From oral abilities to reading skills. *Speech Prosody 2010*, paper 648-0. https://doi.org/10.21437/SpeechProsody.2010-9

Allen, G. D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*(2), 75–86.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native Speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Anderson, J. R. (2014). *Rules of the mind*. Psychology Press.

Archibald, J. (1993). Metrical phonology and the acquisition of L2 stress. *Confluence: Linguistics, L2 Acquisition and Speech Pathology*, 37–48.

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, *10*(3), 215–240. https://doi.org/10.1177/026765839401000303

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Ling*, *35*(1), 167–182. https://doi.org/10.1515/ling.1997.35.1.167

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373.

Astésano, C. (2001). *Rythme et accentuation en français: Invariance et variabilité stylistique*. Editions L’Harmattan.

Astésano, C. (2022). De la supramodalité du rythme: Implications pour la description prosodique, la remédiation linguistique et l’apprentissage des langues. *XXXIVe Journées d’Études sur la Parole -- JEP 2022*, 1–14. https://doi.org/10.21437/JEP.2022-1

Astésano, C. (2017). Le statut de l'Accent Initial dans la phonologie prosodique du français: enjeux descriptifs et psycholinguistiques. Mémoire d'Habilitation à Diriger des Recherches, UT2J. https://www.researchgate.net/profile/Corine-Astesano/publications

Astésano, C., & Bertrand, R. (2016). Accentuation et niveaux de constituance en français: Enjeux phonologiques et psycholinguistiques. *Langue Française*, *191*, 11–30.

Azieb, S. (2021). The critical period hypothesis in second language acquisition: A review of the literature. *International Journal of Research in Humanities and Social Studies*, *8*(4), 20–26.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, 239–265.

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Variables Affecting L2 Gains During Study Abroad. *Foreign Language Annals*, *47*(3), 464–486. https://doi.org/10.1111/flan.12093

Baills, F. (2022). *Embodied prosodic training boosts phonological learning in a foreign language*. [Doctoral Dissertation]. Universitat Pompeu Fabra, Barcelona.

Baills, F., Rohrer, P. L., & Prieto, P. (2022). Le geste et la voix pour enseigner la prononciation en langue étrangère. *Mélanges CRAPEL*, *43*, 1.

Baills, F., Santiago, F., Mairano, P., & Prieto, P. (2022). The effects of prosodic training with logatomes and prosodic gestures on L2 spontaneous speech. In *Proceedings of Speech Prosody 2022* (pp. 802-806). https://doi.org/10.21437/SpeechProsody.2022-163

Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to Songs and Singing Benefitted Initial Stages of Second Language Pronunciation but Not Recall of Word Meaning. *Language Learning*, *71*(2), 369–413. https://doi.org/10.1111/lang.12442

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Variables Affecting L2 Gains During Study Abroad. *Foreign Language Annals*, *47*(3), 464–486. https://doi.org/10.1111/flan.12093

Banel, M.-H., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication*, *15*(1–2), 115–126.

Barlow, J. S. (1998). *Intonation and Second Language Acquisition: A study of the acquisition of English intonation by speakers of other languages* [PhD Thesis]. University of Hull.

Barry, W. (2007). Rhythm as an L2 problem: How prosodic is it? In J. Trouvain & U. Gut (Eds.), *Non-Native Prosody* (pp. 97–120). Mouton de Gruyter. https://doi.org/10.1515/9783110198751.1.97

Bergeron, A., & Trofimovich, P. (2017). Linguistic Dimensions of Accentedness and Comprehensibility: Exploring Task and Listener Effects in Second Language French. *Foreign Language Annals*, *50*(3), 547–566. https://doi.org/10.1111/flan.12285

Berry, M. (2009). The Importance of Bodily Gesture in Sofia Gubaidulina's Music for Low Strings. *Music Theory Online*, *15*(5). https://mtosmt.org/issues/mto.09.15.5/mto.09.15.5.berry.html

Bertinetto, P. M. (1977). Syllabic Blood'ovvero l'italiano come lingua ad isocronismo sillabico. *Studi Di Grammatica Italiana*, *6*, 69–96.

Bertinetto, P. M., & Bertini, C. (2008). On modeling the rhythm of natural languages. In *Proceedings of speech prosody 2008* (pp. 427–430). Universidade Estadual de Campinas. https://ricerca.sns.it/handle/11384/902

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience*, *171*. https://cir.nii.ac.jp/crid/1572261549573590656

Best, C. T. (2014). *Cross-language speech perception: Late versus early second-language bilinguals.* [[PowerPoint slides]]. http://lpp.in2p3.fr/presentations/Conferences_Labex/2014/XLang-Bilingual_LABEX_2014.pptx.pdf

Best, C. T., Tyler, M., Bohn, O., & Munro, M. (2007). Nonnative and second-language speech perception. *Language Experience in Second Language Speech Learning*, 13–34.

Billières, M. (2008). Le statut de I'intonation dans I'évolution de l'enseignement/apprentissage de I'oral en FLE. *Français Dans Le Monde. Recherches et Applications*, *43*, 27–37.

Billières, M. (2014). La phonétique, vilain petit canard de la didactique. *Au Son Du FLE*. https://www.verbotonale-phonetique.com/phonetique-didactique/

Bissonnette, S. (2003). Le registre du lecteur de bulletins de nouvelles québécois et français: Un reflet de l'idéal vocal des communautés linguistiques. In M. Demers (Éd.), *Registre et Voix Sociale*, 17–34.

Blanche-Benveniste, C. (1997). La notion de variation syntaxique dans la langue parlée. *Langue Française*, 19–29.

Blanche-Benveniste, C., & Bilger, M. (1999). Français parlé-oral spontané; quelques réflexions*. Revue française de linguistique appliquée, 4(2), 21-30.*

Bloch, B. (1950). Studies in colloquial Japanese IV phonemics. *Language*, *26*(1), 86–125.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In *Handbook of Psycholinguistics.* (M.A. Gernsbacher, Vol. 5, pp. 945–984). Routledge New York.

Bolinger, D. L. (1965). *Forms of English: Accent, Morpheme, Order.* Harvard University Press.

Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology*, *6*(2), 145–238. https://doi.org/10.2307/1410948

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175.

Bot, K. D. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, *13*(1), 1–24. https://doi.org/10.1093/applin/13.1.1

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation Teaching Practices in Canada. *TESL Canada Journal*, 51–61. https://doi.org/10.18806/tesl.v19i1.919

Briet, G., Collige, V., & Rassart, E. (2014). *La prononciation en classe*. https://dial.uclouvain.be/pr/boreal/object/boreal:183629

Brinton, D. M. (2017). Innovations in pronunciation teaching. In *The Routledge handbook of contemporary English pronunciation* (pp. 448–461). Routledge.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge University Press.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, *28*(2), 191–215. https://doi.org/10.1016/S0346-251X(00)00007-5

Calhoun, S., Warren, P., & Yan, M. (2023). Cross-language influences in the processing of L2 prosody. In I. Elgort, A. Siyanova-Chanturia, & M. Brysbaert (Eds.), *Cross-language Influences in Bilingual Processing and Second Language Acquisition* (pp. 47–73). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.16.03cal

Carter, P. M. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, *272*, 63.

Cavicchio, F., & Busa, M. G. (2023). Lending a hand to speech: Gestures help fluency and increase pitch in second language speakers. *Language, Interaction and Acquisition*, *14*, 218–246. https://doi.org/10.1075/lia.22023.cav

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A course book and reference guide.* Cambridge University Press.

Chafe, W., & Tannen, D. (1987). The Relation between Written and Spoken Language. *Annual Review of Anthropology*, *16*, 383–407.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535–544.

Chan, M. J. (2018). Embodied Pronunciation Learning: Research and Practice. *Catesol Journal*, *30*(1), 47–68.

Charles, T., Trenkic, D., Gambier, Y., Caimi, A., & Mariotti, C. (2015). Speech segmentation in a second language: The role of bimodal input. *Subtitles and Language Learning: Principles, Strategies and Practical Experiences*, 173–198.

Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences*, *3*(2), 923–940.

Colletta, J.-M. (2007). Signaux corporels et acquisition du langage: Des relations constantes et étroites. *Langage et Pratiques*, *39*, 20–33.

Cooper, G. W., Cooper, G., & Meyer, L. B. (1963). *The rhythmic structure of music*. University of Chicago press.

Cosnier, J., & Vaysse, J. (1997). Sémiotique des gestes communicatifs. *Nouveaux Actes Sémiotiques (Limoges)*, *52–54*, 7–28.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment—Companion Volume.* Council of Europe Publishing. https://www.coe.int/en/web/common-european-framework-reference-languages

Crowther, D. (2020). Rating L2 speaker comprehensibility on monologic vs. interactive tasks: What is the effect of speaking task type?. *Journal of Second Language Pronunciation*, *6*(1), 96–121. https://doi.org/10.1075/jslp.19019.cro

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a Speaking Task Affect Second Language Comprehensibility? *The Modern Language Journal*, *99*(1), 80–95. https://doi.org/10.1111/modl.12185

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, *111*(6), 2862–2873.

Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. In *Proceedings of Speech Prosody 2002* (pp. 121-126).

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Mit Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3–4), 133–142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113.

Daigmorte, C., Tallet, J., & Astésano, C. (2022). On the foundations of rhythm-based methods in Speech Therapy In *Proceedings of Speech Prosody 2022* (pp. 47-51). https://doi.org/10.21437/SpeechProsody.2022-10

Darcy, I. (2018). Powerful and Effective Pronunciation Instruction: How Can We Achieve It?. *Catesol Journal*, *30*(1), 13–45.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. *The 6th Workshop on Disfluency in Spontaneous Speech (Diss)*, 17–20.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243. https://doi.org/10.1017/S0142716413000210

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism*, *16*(1), 101–116. https://doi.org/10.1177/1367006911405576

DeKeyser, R. M. (1995). Learning Second Language Grammar Rules: An Experiment With a Miniature Linguistic System. *Studies in Second Language Acquisition*, *17*(3), 379–410. https://doi.org/10.1017/S027226310001425X

De Mareüil, P. B., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, *63*(4), 247–267.

De Pietro, J.-F., & Wirthner, M. (1996). Oral et écrit dans les représentations des enseignants et dans les pratiques quotidiennes de la classe de français. *Travaux Neuchâtelois de Linguistique*, *25*, 29–49.

Delais-Roussarie, E., & Di Cristo, A. (2021). L'accentuation du français*,* in *La Grande grammaire du français*. Éditions Actes Sud. (p. 2126). https://shs.hal.science/halshs-00748395

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *IRAL - International Review of Applied Linguistics in Language Teaching*, *4*(1–4). https://doi.org/10.1515/iral.1966.4.1-4.183

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283.

Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence*. [PhD Dissertation]. Universität Bonn https://www.academia.edu/download/8089196/2003.pdf

Dellwo, V., Karnowski, P., & Szigeti, I. (2006). *Rhythm and speech rate: A variation coefficient for deltaC*. https://www.zora.uzh.ch/id/eprint/111789/

Dellwo, V., Wagner, P., Solé, M. J., Recasens, D., & Romero, J. (2003). *Relations between language rhythm and speech rate*. https://www.zora.uzh.ch/id/eprint/111779/

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1–16.

Derwing, T. M., & Munro, M. J. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly*, *39*(3), 379–397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2013). The Development of L2 Oral Language Skills in Two L1 Groups: A 7-Year Study. *Language Learning*, *63*(2), 163–185. https://doi.org/10.1111/lang.12000

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between l1 fluency and l2 fluency development. *Studies in Second Language Acquisition*, *31*(04), 533. https://doi.org/10.1017/S0272263109990015

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning*, *48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second Language Fluency: Judgments on Different Tasks: Language Learning. *Language Learning*, *54*(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Detey, S., & Durand, J. (2021). Chapitre 6. Acquisition de la prononciation en langue étrangère. In *Introduction à l'acquisition des langues étrangères* (pp. 111–126). De Boeck Supérieur.

Dewey, J., & Simon, H. F. (1989). *1934: Art as Experience*. Southern Illinois University Press.

Di Cristo, A. (2013). *La prosodie de la parole*. De Boeck Superieur.

Di Cristo, A. (2016). *Les musiques du français parlé: Essais sur l'accentuation, la métrique, le rythme, le phrasé prosodique et l'intonation du français contemporain*. De Gruyter.

Dolz-Mestre, J., & Schneuwly, B. (1998). *Pour un enseignement de l'oral: Initiation aux genres formels à l'école*. Citeseer.

Doughty, C., & Williams, J. (1998). *Focus on Form in Classroom Second Language Acquisition. The Cambridge Applied Linguistics Series.* ERIC. https://eric.ed.gov/?id=ED419393

Drouillet, L., Alazard, C., & Astésano, C. (2024). Impact of Prosodic Training on Speech Rhythm in L2 French. In *Proceedings of Pronunciation in Second Language Learning and Teaching Proceedings*, *14*(1). https://www.iastatedigitalpress.com/psllt/article/id/17100/

Drouillet, L., Astésano, C., & Alazard-Guiu, C. (2023). Another voice for another language? The impact of language on vocal register. *Language, Interaction and Acquisition*, *14*(2), 247–279. https://doi.org/10.1075/lia.00018.dro

Duez, D. (1976). Etude du débit et des pauses d'un discours politique. *Bulletin de l'Institut de Phonétique de Grenoble*, *5*, 39–53.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315–330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Edwards, I. G. H. (2008). Social factors and variation in production in L2 phonology. *Phonology and Second Language Acquisition*, *36*, 251.

Efron, D. (1941). *Gesture and environment.* King'S Crown Press. https://psycnet.apa.org/record/1942-00254-000

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In *Perspectives on Fluency* (H. Riggenbach, pp. 287–313). University of Michigan Press.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition*, *24*(2), 223–236.

Ellis, R. (2006). Researching the Effects of Form-Focussed Instruction on L2 Acquisition. *AILA Review*, *19*, 18–41. https://doi.org/10.1075/aila.19.04ell

Fauth, C., & Trouvain, J. (2018). Détails phonétiques dans la réalisation des pauses en Français: Étude de parole lue en langue maternelle vs en langue étrangère. *Langages*, *211*, 81–95.

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R/Andy Field, Jeremy Miles, Zoë Field.* London; Thousand Oaks, Calif.: Sage.

Fillmore, L. W. (1979). Individual differences in second language acquisition. In *Individual differences in language ability and language behavior* (pp. 203–228). Elsevier.

Fischler, J. (2009). *The rap on stress: Teaching stress patterns to English language learners through rap music*. https://conservancy.umn.edu/handle/11299/109937

Fitzpatrick, M. (2002). *Theories of Child Language Acquisition, Child Language Acquisition*. Online],[2004, February 4].

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research.* York Press Inc.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70–79.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the Teaching of Pronunciation in Adult ESL Programs in Canada, 2010. *TESL Canada Journal*, 1–22. https://doi.org/10.18806/tesl.v29i1.1086

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, *44*(2), 181–196. https://doi.org/10.1080/09571736.2013.784345

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Foulkes, P. (2020). Phonological Variation: A Global Perspective. In B. Aarts, A. McMahon, & L. Hinrichs (Eds.), *The Handbook of English Linguistics* (1st ed., pp. 407–440). Wiley. https://doi.org/10.1002/9781119540618.ch21

Fraisse, P. (1974). *Psychologie du rythme*. FeniXX.

Fraisse, P. (1982). Rhythm and tempo. *The Psychology of Music*, *1*, 149–180.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music Training for the Development of Speech Segmentation. *Cerebral Cortex*, *23*(9), 2038–2043. https://doi.org/10.1093/cercor/bhs180

Freed, B. (1995). Language learning and study abroad. In *Second Language Acquisition in a Study Abroad Context*, *3*, 34. John Benjamins B.V.

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In *Perspectives on Fluency* (H. Riggenbach, pp. 243–265). University of Michigan Press.

French, L. M., Gagné, N., & Collins, L. (2020). Long-term effects of intensive instruction on fluency, comprehensibility and accentedness. *Journal of Second Language Pronunciation*, *6*(3), 380–401. https://doi.org/10.1075/jslp.20026.fre

Fries, C. C. (1945). *Teaching and learning English as a foreign language.* University of Michigan Press.

Frost, D., & O'Donnell, J. (2018). Evaluating the essentials: the place of prosody in oral production. *The Pronunciation of English by Speakers of Other Languages*, (chapter 12), 228-259.

Fuchs, R. (2014). Towards a perceptual model of speech rhythm: Integrating the influence of f0 on perceived duration. *Interspeech*, 1949–1953. https://www.isca-archive.org/interspeech_2014/fuchs14_interspeech.pdf

Fuchs, R. (2023). Rhythm Metrics and the Perception of Rhythmicity in Varieties of English as a Second Language. In R. Fuchs (Ed.), *Speech Rhythm in Learner and Second Language Varieties of English* (pp. 187–210). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8940-7_8

Gadet, F. (1989). *Le français ordinaire Text*. P.: Armand Colin.

Gadet, F. (1996). Gadet, F. (1996). Une distinction bien fragile: Oral/écrit. *Tranel*, 25, 13-27.

Gao, J., & Sun, P. P. (2024). How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics*, *34*(1), 276–291. https://doi.org/10.1111/ijal.12493

Gattegno, C. (2010). *Teaching foreign languages in schools: The silent way*. Educational Solutions World.

Gilbert, J. (1978). Gadgets: Some non-verbal tools for teaching pronunciation. *TESL Reporter*, *11*, 3–3.

Gindre, A.-F. (2024). *Du rythme à la parole: Effet d'amorces rythmiques langagières, non langagières et musicales modulées par l'engagement moteur sur le temps d'initiation de la parole.* [Thèse de doctorat]. Université de Toulouse II - Jean Jaurès.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609–631. https://doi.org/10.1177/1362168816651463

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music*, *43*(5), 627–640. https://doi.org/10.1177/0305735614528833

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, *2*(1), 56–92. https://doi.org/10.1075/jslp.2.1.03gor

Gordon, J., & Darcy, I. (2022). Teaching segmentals and suprasegmentals: Effects of explicit pronunciation instruction on comprehensibility, fluency, and accentedness. *Journal of Second Language Pronunciation*, *8*(2), 168–195. https://doi.org/10.1075/jslp.21042.gor

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7,* 515–546.

Grabe, E., Post, B., & Watson, I. (1999, August). The acquisition of rhythmic patterns in English and French. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1201-1204). Berkeley, CA: University of California.

Graham, C. (1978). *Jazz chants: Rhythms of American English for students of English as a second language*. Oxford University Press.

Grosjean, F. (1972). *Le rôle joué par trois variables temporelles dans la compréhension orale de l'anglais étudié comme seconde langue et perception de la vitesse de lecture par des lecteurs et des auditeurs* [PhD Thesis]. Université Paris VII.

Grosjean, F. (1980). Temporal variables within and between languages. *Towards a Cross-Linguistic Assessment of Speech Production*, 39–53.

Grosjean, F., & Deschamps, A. (1972). Analyse des variables temporelles du français spontané. *Phonetica*, *26*(3), 129–156. https://doi.org/10.1159/000259407

Grosjean, F., & Deschamps, A. (1973). Analyse des variables temporelles du français spontané. *Phonetica*, *28*(3–4), 191–226. https://doi.org/10.1159/000259456

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, *31*(3–4), 144–184.

Grosser, W. (1993). Aspects of intonational L2 acquisition. *Current Issues in European Second Language Acquisition Research*, 81–94.

Guberina, P. (1956). *L'audiometrie verbo-tonale et son application*. Impr. R. Gauthier.

Guberina, P. (1975). *La méthode audio-visuelle structuro-globale*. Université de l'Etat.

Guilbault, C. P. (2002). *The acquisition of French rhythm by English second language learners*. [Ph.D. thesis]. University of Alberta.

Gut, U. (2003). Non-native speech rhythm in German. *Proceedings of the ICPhS Conference*, 2437–2440.

Gut, U. (2012). Rhythm in L2 speech. *Speech and Language Technology*, *14*(15), 83–94.

Halle, M., & Vergnaud, J.-R. (1987). An Essay on Stress, Cambridge, MA, MITPress. *HalleAn Essay on Stress1987*.

Han, M. S. (1962). The feature of duration in Japanese. 音声の研究 (日本音声学会誌), *10*, 65–80.

Harris, M. J., & Gries, S. T. (2011). Measures of speech rhythm and the role of corpus-based word frequency: A multifactorial comparison of Spanish (-English) speakers. *International Journal of English Studies*, *11*(2), 1–22.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press.

Henderson, A., Frost, D., Tergujeff, E., Kautszch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). *English Pronunciation Teaching in Europe Survey: Selected results*. . *Research in Language*, *10*(1), 5–27. https://doi.org/10.2478/v10015-011-0047-4

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential Temporal Patterns in Spontaneous Speech. *Language and Speech*, *9*(4), 207–216. https://doi.org/10.1177/002383096600900402

Hiligsmann, P., & Rasier, L. (2002). De zinsaccentuering in de tussentaal van Franstalige leerders van het Nederlands. *N/f*, *1*. https://orbi.uliege.be/handle/2268/171002

Huensch, A., & Tracy-Ventura, N. (2017). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, *38*(4), 755–785. https://doi.org/10.1017/S0142716416000424

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*(1), 73–98.

Isaacs, T., & Thomson, R. I. (2020). Reactions to second language speech: Influences of discrete speech characteristics, rater experience, and speaker first language background. *Journal of Second Language Pronunciation*, *6*(3), 402–429. https://doi.org/10.1075/jslp.20018.isa

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, *97*(1), 553–562.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, *128*(2), 839–850. https://doi.org/10.1121/1.3459842

James, A. R. (1988). *The acquisition of a second language phonology.* Narr.

Jannedy, S., & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure: ISIS; Working Papers of the SFB 632*, *3*, 199–244.

Jaques-Dalcroze, É. (1906). *Methode Jaques-Dalcroze: Erster Theil, erster Band. Rhythmische Gymnastik* (Vol. 1). Sandoz, Jobin & Cie.

Jekiel, M. (2022). L2 rhythm production and musical rhythm perception in advanced learners of English. *Poznan Studies in Contemporary Linguistics*, *58*(2), 315–340. https://doi.org/10.1515/psicl-2022-0016

Johnsen, L. A., & Avanzi, M. (2020). Étude des contours prosodiques des listes ouvertes dans le corpus 'OFROM'. *Studia Linguistica Romanica*, *4*, 110–128.

Judkins, L., Alazard, C., & Astésano, C. (2022). Phénomènes de groupement et pause en parole native vs non-native. *XXXIVe Journées d'Études Sur La Parole–JEP 2022*, 910–919. https://hal.science/hal-03980725/

Judkins, L., Alazard-Guiu, C., & Astésano, C. (2022). How do we chunk and pause in non-native vs native speech? *Speech Prosody 2022*, 792–796. https://hal.science/hal-03980742/

Jun, S.-A., & Oh, M. (2000). Acquisition of second language intonation. *Proceedings of International Conference on Spoken Language Processing*, *4*, 76–79.

Juntunen, M.-L. (2016). The Dalcroze Approach. *Teaching General Music: Approaches, Issues, and Viewpoints*, 141–168.

Kaglik, A., & Boula de Mareüil, P. (2009). Perception d'un accent étranger et part de la prosodie selon l'âge de première exposition à la L2: Transfert ou phénomène universel en acquisition. *6es Journées d'Études Linguistiques*, 7–13.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kendon, A. (1972). Kendon, A. (1972). Some relationships between body motion and speech. In A.W. Siegman & B. Pope (Eds.), *Studies in dyadic communication*. Elmsford, NY, Pergamon Press, 177-210.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, Comprehensibility, and Accentedness of L2 Speech: The Role of Listener Experience and Semantic Context. *The Canadian Modern Language Review*, *64*(3), 459–489. https://doi.org/10.3138/cmlr.64.3.459

Kennedy, S., & Trofimovich, P. (2019). Comprehensibility: A Useful Tool to Explore Listener Understanding. *The Canadian Modern Language Review*, *75*(4), 275–284. https://doi.org/10.3138/cmlr.2019-0280

Kissling, E. M. (2018). Pronunciation Instruction Can Improve L2 Learners' Bottom-Up Processing for Listening. *The Modern Language Journal*, *102*(4), 653–675. https://doi.org/10.1111/modl.12512

Konopczynski, G. (1990). Le langage émergent: Caractéristiques rythmiques. *Beiträge Zur Phonetik Und Linguistik*, *60*. https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=6143393

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied Learning Across the Life Span. *Topics in Cognitive Science*, *4*(4), 731–739. https://doi.org/10.1111/j.1756-8765.2012.01221.x

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. *Perspectives on Fluency*, 5–24. https://dialnet.unirioja.es/servlet/articulo?codigo=5732187

Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164.

Kushch, O. (2018). Beat gestures and prosodic prominence: Impact on learning [Ph.D. Thesis]. Universitat Pompeu Fabra, Barcelona. https://www.tdx.cat/handle/10803/463004

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Labov, W. (1981). *Field methods of the project on linguistic change and variation*. Southwest Educational Development Laboratory. http://smjegupr.net/wp-content/uploads/2013/01/Labov1984.pdf

Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures.* University of Michigan Press.

Langer, S. K. (1953). Feeling and form. *Charles Scribner's Sons*. https://www.literatureandemotion.com/s/Langer_FeelingForm.pdf

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119.

Lauzon, V. F., Doehler, S. P., Pochon-Berger, E., & Kohler, F. S. (2009). L'oral? L'oral! Mais comment? *Die Lehrerschaft Bleibt Skeptisch*, *2*, 41–50.

Laver, J. D. (1980). Monitoring systems in the neurolinguistic control of speech production. *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, 287–305.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Lee, J., Jang, J., & Plonsky, L. (2015). The Effectiveness of Second Language Pronunciation Instruction: A Meta-Analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*(3), 253–263. https://doi.org/10.1016/S0095-4470(19)31139-8

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach*. *Language Learning*, *40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Lenth, R. (2024). *_emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.10.0.* (Version 1.10.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377.

Lewandowska-Tomaszczyk, B. (1996). Cross-linguistic and language-specific aspects of semantic prosody. *Language Sciences*, *18*(1–2), 153–178.

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, *36*(2), 223–255.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454.

Liberman, M., & Prince, A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry*, *8*(2), 249–336.

Lin, H., & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *J. Chin. Lang. Comput.*, *17*(3), 127–140.

Linnavalli, T., Putkinen, V., Lipsanen, J., Huotilainen, M., & Tervaniemi, M. (2018). Music playschool enhances children's linguistic skills. *Scientific Reports*, *8*(1), 8767. https://doi.org/10.1038/s41598-018-27126-5

Llanes-Coromina, J., Prieto, P., & Rohrer, P. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. *Speech Prosody 2018*, 498–502. https://doi.org/10.21437/SpeechProsody.2018-101

Llorca, R. (2001). Jeux de groupe avec la voix et le geste sur les rythmes du français parlé. *L'enseignement Des Langues Aux Adultes, Aujourd'hui*, 141–150.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, *7*(2), 179–214. https://doi.org/10.1075/gest.7.2.04loe

Low, E. L. (1998). *Prosodic prominence in Singapore English.* [PhD Thesis]. University of Cambridge. https://www.repository.cam.ac.uk/items/01deeb76-4b52-47a3-9c59-efc2dfcf8f91

Low, E. L., & Grabe, E. (1995). Prosodic patterns in singapore english. *Proceedings of the International Congress of Phonetic Sciences, Stockholm*, *3*, 636–639.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in singapore english. *Language and Speech*, *43*(4), 377–401.

Ludke, K. M. (2018). Singing and arts activities in support of foreign language learning: An exploratory study. *Innovation in Language Learning and Teaching*, *12*(4), 371–386. https://doi.org/10.1080/17501229.2016.1253700

Luu, V. T., Lian, A. P., & Siriyothin, P. (2021). Developing EFL learners' listening comprehension through a computer-assisted self-regulated prosody-based listening platform. *Computer-Assisted Language Learning Electronic Journal*, *22*(1), 254–271.

Maastricht, L. V., Hoetjes, M., & Drie, E. V. (2019). Do gestures during training facilitate L2 lexical stress acquisition by Dutch learners of Spanish? *The 15th International Conference on Auditory-Visual Speech Processing*, 6–10. https://doi.org/10.21437/AVSP.2019-2

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to Improve English L2 Pronunciation: The Variable Effects of Different Types of Instruction. *Language Learning*, *44*(1), 75–100. https://doi.org/10.1111/j.1467-1770.1994.tb01449.x

Macdonald, S. (2002). Pronunciation: Views and practices of reluctant teachers. *Prospect*, *17*(3), 1.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Lawrence Erlbaum Associates.

Major, R. C. (2008). Transfer in second language phonology. *Phonology and Second Language Acquisition*, *36*, 63–94.

Major, R. C., & Faudree, M. C. (1996). Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. *Studies in Second Language Acquisition*, *18*(1), 69–90.

Major, R. C., & Kim, E. (1999). The Similarity Differential Rate Hypothesis. *Language Learning*, *49*(s1), 151–183. https://doi.org/10.1111/0023-8333.49.s1.5

Marks, J. (2007). *English pronunciation in use: Elementary: self-study and classroom use*. Cambridge University Press.

McAndrews, M. (2019). Short periods of instruction improve learners' phonological categories for L2 suprasegmental features. *System*, *82*, 151–160.

McAndrews, M. (2023). The effects of prosody instruction on listening comprehension in an EAP classroom context. *Language Teaching Research*, *27*(6), 1480–1503. https://doi.org/10.1177/1362168821990346

McCafferty, S. G. (2006). Gesture and the materialization of second language prosody. *International Review of Applied Linguistics in Language Teaching*, *44*(2), 197–209. https://doi.org/10.1515/IRAL.2006.008

McNeill, D. (1992). *Hand and mind. What gestures reveal about thought.* Chicago, University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought Chicago: Univ*. Chicago Press.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and Language in Contact* (pp. 171–188). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45168-7_9

Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, *36*(2), 303–329.

Missire, R. (2007). Rythmes sémantiques et temporalité des parcours interprétatifs. *Rythme, Sens et Textualités. Linguistique, Sémiotique Du Discours, Sémantique Des Textes, Rhétorique, Stylistique, Poétique, Toulouse: Editions Universitaires Du Sud*. http://www.revue-texto.net/1996-2007/Inedits/Missire/Missire_Rythmes.pdf

Mok, P., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *ISCA*.

Moyer, A. (2004). Accounting for Context and Experience in German (L2) Language Acquisition: A Critical Review of the Research. *Journal of Multilingual and Multicultural Development*, *25*(1), 41–61. https://doi.org/10.1080/01434630408666519

Morrill, T. H. (2016). A facilitating effect of prosodic transfer on non-native fluent speech listening. *Language, Cognition and Neuroscience*, *31*(6), 801–816. https://doi.org/10.1080/23273798.2016.1167226

Munro, M. J. (2008). Foreign accent and speech intelligibility. *Phonology and Second Language Acquisition,* John Benjamins Publishing Company.

Munro, M. J. (2018). Dimensions of pronunciation. In *The Routledge Handbook of Contemporary English Pronunciation* (O. Kang, R. Thomson, J. Murphy, pp. 413–431). Routledge.

Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, *11*(3), 253–266. https://doi.org/10.1177/026553229401100302

Munro, M. J., & Derwing, T. M. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Munro, M. J., & Derwing, T. M. (2022). Foreign accent, comprehensibility and intelligibility, redux. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *Benjamins Current Topics* (Vol. 121, pp. 7–32). John Benjamins Publishing Company. https://doi.org/10.1075/bct.121.02mun

Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in l2 speech perception-production research: revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, *44*(2), 580–605. https://doi.org/10.1017/S0272263121000371

Nagle, C. L., & Huensch, A. (2022). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *Benjamins Current Topics* (Vol. 121, pp. 51–73). John Benjamins Publishing Company. https://doi.org/10.1075/bct.121.04nag

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*(04), 647–672. https://doi.org/10.1017/S0272263119000044

Nakata, H., & Shockey, L. (2011). The Effect of Singing on Improving Syllabic Pronunciation-Vowel Epenthesis in Japanese. *ICPhS*, 1442–1445.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*(1), 1–19.

Nemoto, S., Wilson, I., & Perkins, J. (2016). Analysis of the effects on pronunciation of training by using song or native speech. *The Journal of the Acoustical Society of America*, *140*(4), 3343.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In *Errors in linguistic performance: Slips of the tongue, ear, pen and hand/ed. By Victoria A. Fromkin* (pp. 87–95). Academic Press Inc.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning*, *50*(3), 417–528. https://doi.org/10.1111/0023-8333.00136

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*, 244–257.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533–544.

Ordin, M., Polyanskaya, L., & Ulbrich, C. (2011). Acquisition of timing patterns in second language. *Twelfth Annual Conference of the International Speech Communication Association*, 1129-1132.

Pasdeloup, V. (2004). Le rythme n'est pas élastique: Étude préliminaire de l'influence du débit de parole sur la structuration temporelle. *Actes Des Journées d'Etudes Sur La Parole, Fés*. http://www.afcp-parole.org/doc/Archives_JEP/2004_XXVe_JEP_Fes/actes/jep2004/Pasdeloup.pdf

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142.

Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. *Annals of the New York Academy of Sciences*, *1252*(1), 124–128. https://doi.org/10.1111/j.1749-6632.2011.06426.x

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108.

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. D. M. (2012). Measuring Child Rhythm. *Language and Speech*, *55*(2), 203–229. https://doi.org/10.1177/0023830911417687

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320.

Peltonen, P. (2018). Exploring Connections Between First and Second Language Fluency: A Mixed Methods Approach. *The Modern Language Journal*, *102*(4), 676–692. https://doi.org/10.1111/modl.12516

Pérez, A., Carreiras, M., Gillon Dowens, M., & Duñabeitia, J. A. (2015). Differential oscillatory encoding of foreign speech. *Brain and Language*, *147*, 51–57. https://doi.org/10.1016/j.bandl.2015.05.008

Pierrehumbert, J. (1980). *The Phonetics and Phonology of English Intonation* [PhD Thesis]. MIT.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure. LI Monograph Series No. 15*. Cambridge, MA: MIT Press.

Pike, K. L. (1945). *The intonation of American English.* University of Michigan Press.

Polyanskaya, L., Ordin, M., & Busa, M. G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Language and Speech*, *60*(3), 333–355.

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia Linguistica*, *54*(2), 136–149. https://doi.org/10.1111/1467-9582.00055

Préfontaine, Y., Kormos, J., & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*, *33*(1), 53–73. https://doi.org/10.1177/0265532215579530

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*(6), 681–702.

Quinting, G. (2019). *Hesitation phenomena in adult aphasic and normal speech* (Vol. 126). Walter de Gruyter GmbH & Co KG.

R Core Team. (2023). *R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. [Computer software]. <https://www.R-project.org/>

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *Proceedings of Speech Prosody 2002*.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.

Rasier, L. (2003). Le système accentuel de l'interlangue d'apprenants francophones du néerlandais. In: A. Mettouchi, G. Ferré (red), *Actes du colloque international 'Interfaces Prosodiques'*, p. 79-84.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Methodological issues and description. *Nouveaux Cahiers de Linguistique Française*, *28*, 41–66.

Révis, J. (2013). *La voix et soi: Ce que notre voix dit de nous*. De Boeck Superieur.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech* (pp. 263–270). De Gruyter Mouton. https://doi.org/10.1515/9783110816570.263

Reynolds, W. T. (1994). *Variation and phonological theory*. [Ph.D. thesis]. University of Pennsylvania.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441. https://doi.org/10.1080/01638539109544795

Robb, M. P., Maclagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics*, *18*(1), 1–15. https://doi.org/10.1080/0269920031000105336

Rohrer, P. L., Prieto, P., & Delais-Roussarie, E. (2019). Beat gestures and prosodic domain marking in French. *19th International Congress of Phonetic Sciences*, 1500–1504. https://hal.science/hal-04141502/

Rosset, T., Ohala, J. J., Boë, L.-J., & Vilain, C.-E. (Eds.). (2010). *Un siècle de phonétique expérimentale, fondation et éléments de développement: Hommage à Théodore Rosset et John Ohala*. ENS éditions.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly*, *46*(4), 842–854.

Saito, K., & Hanzawa, K. (2018). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research*, *22*(4), 398–417. https://doi.org/10.1177/1362168816679030

Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, *39*(3), 593–617. https://doi.org/10.1017/S0142716417000571

Saito, K., & Plonsky, L. (2019). Effects of Second Language Pronunciation Teaching Revisited: A Proposed Measurement Framework and Meta-Analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224.

Sánchez-Alvarado, C. (2022). The acquisition of L2 Spanish intonation: An analysis based on features. *Journal of Second Language Pronunciation*, *8*(1), 40–67. https://doi.org/10.1075/jslp.20041.san

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech Segmentation by Native and Non-Native Speakers: The Use of Lexical, Syntactic, and Stress-Pattern Cues. *Journal of Speech, Language, and Hearing Research*, *45*(3), 519–530. https://doi.org/10.1044/1092-4388(2002/041)

Sauvanet, P. (2000). *Le rythme et la raison: Rythmologiques*. Paris : Kimé.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331–346.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*(4), 357–385.

Schwab, S., & Dellwo, V. (2022). Explicit versus non-explicit prosodic training in the learning of Spanish L2 stress contrasts by French listeners. *Journal of Second Language Studies*, *5*(2), 266–306. https://doi.org/10.1075/jsls.21017.sch

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*(3–4), 245–253. https://doi.org/10.1111/j.1467-1770.1969.tb00466.x

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2). https://doi.org/10.1515/iral-2016-9991

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*(2), 173–199.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, *10*, 209–230.

Selinker, L., & Lakshmanan, U. (1992). Language transfer and fossilization: The multiple effects principle. *Language Transfer in Language Learning*, 197–216.

Selkirk, E. (2014). The prosodic structure of function words. In *Signal to syntax* (pp. 199–226). Psychology Press.

Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics*, *37*(2), 303–322.

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, *9*, 1514.

Sheppard, B. E., Elliott, N. C., & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes*, *26*, 42–51.

Simon, A.-C., Auchlin, A., Avanzi, M., & Goldman, J.-P. (2010). Les phonostyles: Une description prosodique des styles de parole en français. *Les Voix Des Français: En Parlant, En Écrivant, Bern: Lang*, 71–88.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14.

Smotrova, T. (2017). Making Pronunciation Visible: Gesture In Teaching Pronunciation. *TESOL Quarterly*, *51*(1), 59–89. https://doi.org/10.1002/tesq.276

So, C. K. (2010). Categorizing Mandarin tones into Japanese pitch-accent categories: The role of phonetic properties. *Proceedings of Interspeech 2010 Satellite Workshop on Second Language Studies, Tokyo*.

So, C. K. (2012). Cross-language categorization of monosyllabic foreign tones: Effects of phonological and phonetic properties of native language. In T. Stolz, N. Nau, & C. Stroh (Eds.), *Monosyllables* (pp. 55–69). Akademie Verlag. https://doi.org/10.1524/9783050060354.55

So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*, *47*. https://doi.org/10.2478/psicl-2011-0011

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners'assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221.

Sönning, L. (2023). (Re-)viewing the Acquisition of Rhythm in the Light of L2 Phonological Theories. In R. Fuchs (Ed.), *Speech Rhythm in Learner and Second Language Varieties of English* (pp. 123–157). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8940-7_6

Southwood M. H., Flege J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*, *13*(5), 335–349. https://doi.org/10.1080/026992099299013

Spada, N. (1997). Form-Focussed Instruction and Second Language Acquisition: A Review of Classroom and Laboratory Research. *Language Teaching*, *30*(2), 73–87. https://doi.org/10.1017/S0261444800012799

Spada, N., & Tomita, Y. (2010). Interactions Between Type of Instruction and Type of Language Feature: A Meta-Analysis. *Language Learning*, *60*(2), 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Stevick, E. W., Hu, C., & Wang, D. (1998). *Working with teaching methods: What's at stake?* Heinle & Heinle Publishers.

Stockmal, V., Markus, D., & Bond, D. (2005). Measures of Native and Non-Native Rhythm in a Quantity Language. *Language and Speech*, *48*(1), 55–63. https://doi.org/10.1177/00238309050480010301

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143–167.

Suzuki, S., & Kormos, J. (2024). The moderating role of L2 proficiency in the predictive power of L1 fluency on L2 utterance fluency. *Language Testing*, *0*(0). https://doi.org/10.1177/02655322241241851

Suzuki, S., Kormos, J., & Uchihara, T. (2021). The Relationship Between Utterance and Perceived Fluency: A Meta-Analysis of Correlational Studies. *The Modern Language Journal*, *105*(2), 435–463. https://doi.org/10.1111/modl.12706

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tavakoli, P., Nakatsuhara, F., & Hunter, A. (2020). Aspects of Fluency Across Assessed Levels of Speaking Proficiency. *The Modern Language Journal*, *104*(1), 169–191. https://doi.org/10.1111/modl.12620

Tavakoli, P., & Skehan, P. (2005). 9. Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Language Learning & Language Teaching* (Vol. 11, pp. 239–273). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.11.15tav

Tavakoli, P., & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge University Press.

Tellier, M. (2006). *L'impact du geste pédagogique sur l'enseignement/apprentissage des langues étrangères: Etude sur des enfants de 5 ans* [PhD Thesis]. Université Paris-Diderot-Paris VII. https://theses.hal.science/tel-00371041/

Tellier, M. (2008). Dire avec des gestes. *Le Français Dans Le Monde. Recherches et Applications*, *44*, 40–50.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*(3), 744–765.

Thomson, R. I. (2012). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, *62*(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344.

Towell, R., & Dewaele, J.-M. (2005). Chapter 10. The Role of Psycholinguistic Factors in the Development of Fluency Amongst Advanced Learners of French. In J.-M. Dewaele, *Focus on French as a Foreign Language* (pp. 210–239). Multilingual Matters. https://doi.org/10.21832/9781853597688-011

Towell, R., Hawkins, R., & Bazergui, N. (1996). The Development of Fluency in Advanced Learners of French. *Applied Linguistics*, *17*(1), 84–119. https://doi.org/10.1093/applin/17.1.84

Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the Native Language on the Learning of Fundamental Frequency in Second-Language Speech Segmentation. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00985

Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology*, *3*(2), 385-423. https://doi.org/10.1515/lp-2012-0018

Tremblay, A., Kim, S., Shin, S., & Cho, T. (2021). Re-examining the effect of phonological similarity between the native-and second-language intonational systems in second-language speech segmentation. *Bilingualism: Language and Cognition*, *24*(2), 401–413.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Studies in Second Language Acquisition*, *28*(01), 1-30. https://doi.org/10.1017/S0272263106060013

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Trofimovich, P., Kennedy, S., & Blanchet, J. (2017). Development of Second Language French Oral Skills in an Instructed Setting: A Focus on Speech Ratings. *Canadian Journal of Applied Linguistics*, *20*(2), 32–50. https://doi.org/10.7202/1042675ar

Trubetskoy, N. S. (1939). *Grundzüge der phonologie*. Travaux du Cercle Linguistique de Prague. https://pure.mpg.de/rest/items/item_2399346/component/file_2399345/content

Ueyama, M. (2003). Duration and quality in the production of the vowel length contrast in L2 English and L2 Japanese. In *International Congress of Phonetic Sciences (ICPhS), Barcelona*, 1509-1512. https://doi.org/10.13140/2.1.2787.9680

Ueyama, M. (2016). Prosodic transfer in L2 relative prominence distribution: The case study of Japanese pitch accent produced by Italian learners. In Proceedings of Speech Prosody 2016 (pp. 602-605). https://doi.org/10.21437/SpeechProsody.2016-123

Ulbrich, C., & Mennen, I. (2016). When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism*, *20*(5), 522–549. https://doi.org/10.1177/1367006915572383

Vallduví, E. (1990). *The informational component*. [Ph.D. thesis]. University of Pennsylvania.

Van Els, T., & De Bot, K. (1987). The role of intonation in foreign accent. *The Modern Language Journal*, *71*(2), 147–155.

Van Maastricht, L., Krahmer, E., Swerts, M., & Prieto, P. (2019). Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch. *Studies in Second Language Acquisition*, *41*(1), 87–121.

Verdugo, M. D. R. (2003). Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English. *ICAME Journal*, *26*, 115–132.

Vieru, B., Boula de Mareüil, P., & Adda-Decker, M. (2011). Characterisation and identification of non-native French accents. *Speech Communication*, *53*(3), 292–310.

Vogel, I. (2009). Universals of Prosodic Structure. In S. Scalise, E. Magni, & A. Bisetto (Eds.), *Universals of Language Today* (Vol. 76, pp. 59–82). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8825-4_4

Wang, X. (2022). Segmental versus Suprasegmental: Which One is More Important to Teach? *RELC Journal*, *53*(1), 194–202. https://doi.org/10.1177/0033688220925926

Weinreich, U. (1953). *Languages in contact.* Linguistic Circle of New York.

Wenk, B. J. (1985). Speech rhythms in second language acquisition. *Language and Speech*, *28*(2), 157–175.

Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, *10*(2), 193–216.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522.

Wieden, W. (1993). Aspects of acquisitional stages. *Current Issues in European Second Language Acquisition Research. Tuebingen: Gunter Narr Verlag*, 125–135.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, *127*(3), 1559–1569. https://doi.org/10.1121/1.3293004

Woodrow, H. (1909). *A Quantitative Study of Rhythm: The Effect of Variations in Intensity, Rate and Duration.* Science Press.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). Wiley. https://psycnet.apa.org/record/1951-07758-012

Yazawa, K., & Kondo, M. (2022). A comparison of rhythm metrics for L2 speech. *Proceedings of the 11th International Conference on Speech Prosody*, 332–336.

Yenkimaleki, M., Van Heuven, V. J., & Soodmand Afshar, H. (2023). The efficacy of segmental/suprasegmental *vs.* holistic pronunciation instruction on the development of listening comprehension skills by EFL learners. *The Language Learning Journal*, *51*(6), 734–748. https://doi.org/10.1080/09571736.2022.2073382

You, H. (1994). Defining rhythm: Aspects of an anthropology of rhythm. *Culture, Medicine and Psychiatry*, *18*(3), 361–384.

Zampini, M. (2008). L2 speech production research. *Phonology and Second Language Acquisition*, *36*, 219–249.

Zerbian, S. (2015). Markedness Considerations in L2 Prosodic Focus and Givenness Marking. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and Language in Contact* (pp. 7–27). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45168-7_2

Zhang, R., & Yuan, Z. (2020). Examining the effects of explicit pronunciation instruction on the development of l2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905–918. https://doi.org/10.1017/S0272263120000121

Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, *24*(5), 666–689. https://doi.org/10.1177/1362168818806531

Zhang, Y., Baills, F., & Prieto, P. (2022). Training with embodied musical activities has positive effects on unfamiliar language imitation skills. In *Proceedings of Speech Prosody, 2022-May* (pp. 723-727).

Zhang, Y., Baills, F., & Prieto, P. (2023). Singing Songs Facilitates L2 Pronunciation and Vocabulary Learning: A Study with Chinese Adolescent ESL Learners. *Languages*, *8*(3), Article 3. https://doi.org/10.3390/languages8030219

Zhang, Y., Baills, F., & Prieto, P. (forthcoming). Embodied music training may help improve speech imitation and pronunciation skills. *Language Teaching*.

Zielinski, B., & Pryor, E. (2022). Comprehensibility and everyday English use: An exploration of individual trajectories over time. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *Benjamins Current Topics* (Vol. 121, pp. 75–101). John Benjamins Publishing Company. https://doi.org/10.1075/bct.121.05zie

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, *44*, 35–62.

APPENDIX

Appendix 1: Prosody and Oral course detailed

PROSODY TRAINING

Lucie Drouillet

Université Jean Jaurès - Laboratoire de NeuroPsychoLinguistique (LNPL)

Two 1.5 hour classes per week for 4 weeks

All instructions during the classes were given in French.

Class 1 - Introduction & Dalcroze activities

- Introduction with perception exercise: can you distinguish/recognise languages based on prosody only? - Ice-braker exercise: introduce yourself in English but with a French accent (teacher does it first to give example) - Discussion about what we do to have a French accent	**20 min**
- Presentation of some key theory concepts (definition of prosody etc.) in French.	**10 min**
Dalcroze exercises: 1. Walk out the beat - 5 different extracts from tunes are played, students just have to walk on the beat (different each time). Same exercise with 5 new extracts and students have to walk and clap the beat. 2. Students listen to a tune which accelerates, they have to clap the beat and follow the acceleration of the beat. 3. Reproducing rhythm patterns - teacher claps a rhythm pattern, students reproduce it (showing and repeating is done twice on the same pattern), repeat with 6 different pattern total, then every student propose a pattern and everyone else reproduces it (also done twice on the same pattern), then same exercise but doing the rhythm patterns with the feet (on the spot). 4. Move to the accents in the melody: tennis game - students listen to a tune which has clear accents in its melody, they position themselves by pair, facing each other and the do the gesture of throwing a ball with a tennis racket towards their partner on the accents of the melody, alternating. 5. Move to the final accent. Students listen to a tune composed of 4 note phrases with the 4th one always stronger. They then have to jump and land on the 4th note, having to anticipate its timing.	**60 min**

Class 2 - Dalcroze activities

Dalcroze exercises: 1. Warm up with first activity of previous session: walk and clap to the beat of 10 different tunes with different bpm. 2. Move to the change of music scale - first listen to the tune and identify if the music is ascending, descending or being stable. Then	**90 min**

walk and move forward when ascending, backwards when descending, turn on yourself when stable. 3. Body percussion and collaboration on Take 5: students are shown a series of moves corresponding to each count of a 5-count tune (Take 5), they have to reproduce the moves one way, then the other way, then they freestyle but keep one move for one beat. 4. Move to the beat and keep it - they walk out the beat of a tune, then tune stops and they have to clap the beat to maintain it until tune starts again. First done on a piano tune, then on the same tunes used in the warm up.	

Class 3 - syllabic rhythm and final lengthening

- Listen to 4 tunes, in which language are they sung? Identify which rhythm pattern corresponds to which song. Repeat the rhythm pattern (titiTA) of each song, tapping your knees at the same time. Where is the rhythmic accent in each language? - Expression: find 2 cities where these languages are spoken, do they have the same name in English vs French? Placement of the accent on the last syllable in French.	**20 min**
Perception exercise: Fonetix Quizz 1 & 2 de syllabation - How many syllables do you hear? (6 items like "asseyez-vous") - Listen to logatome sentences and match them with an actual sentence (6 items like "dadadada" > "elle est chinoise"	**10 min**
Tapping the syllables on a simple dialogue - First the teacher reads the dialogue in logatome, tapping each syllable. Students repeat + tap one sentence by one sentence, one after the other. Then Teacher shows dialogue with words and tapping, students repeat. - By pair, one student reads one sentence in logatome and the other has to find the sentence that matches and say it with tapping too. - Change pairs, each couple has to write their own little dialogue, and present it to the class with tapping at the same time as saying the words.	**60 min**

Class 4 - final accent

Conversation starter: do you listen to French music? - Listen to Harley Davidson de Brigitte Bardot, underline strong syllables on the lyrics transcription, then clap on the strong syllable, then let's sing it with the music insisting on the strong syllable and tapping it too. Which English words do you hear? What are French words you use in English and English words you hear in French? Difference between saying these words in French and in English.	20 min
English words used in French A list of English words used in French is provided to the students. They are given rubber bands to place around their hands. Teacher shows the example, saying an English word in French, placing the accent on	20 min

the final syllable and at the same time stretching the elastic between their hands. Students repeat. Then do the same but in pair.	
Students have to write a few sentences using words from the list and have to say them in front of the class using the elastic still.	20 min

Class 5 - syllable structure, liaisons, enchainements

Quick theory presentation on syllable structure in French vs English, what is "liaison/enchainement"	**15 min**
500 exo phonétique: p. 98 (perception) work on proverbs, teacher says them, students have to identify where the linking is and which consonant is used, see how the linking consonant gets attached to the beginning of the next syllable, not the end of the previous one, students repeat after teacher	**15 min**
500 exo phonétique exo 2 p. 100: perception - listen to these sentences, identify which consonant is used to do the linking, most common ones are z, t, n	**30 min**
Expression and practice: work on 2 dialogues given by the teacher, identify where the linkings are and mark them, then practice saying the dialogue with linking, switch pair for second dialogue	**30 min**

Class 6 - liaisons, enchainements and rhythmic groups

Fonetix: quizz 1 enchainement consonantique - listen and say if the sentence is pronounced correctly (enchainement or not)	**10 min**
Practice exercise 500 exo phonétique - exo 6 p. 102 - read the sentences with the correct linking	**20 min**
Practice with dialogues - with the same dialogues as in class 5, mark the link and the rhythmic group boundary, present to the class	**30 min**
Listening exercise - radio news broadcast (RFI-journal en français facile), first listen, then with the transcript mark rhythmic groups boundaries, discuss in group	**30 min**

Class 7 - intonation (démarcative + sémantique)

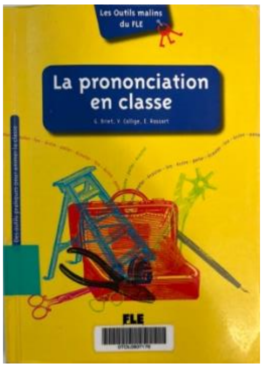
Conversation starter: easter traditions, then what is intonation?	**10 min**
Perception - Fiche 23 la prononciation en classe - listen to these sentences, close your eyes, turn your palms upwards when intonation is rising, downwards when it's falling, first with logatomes, then with real words. Is there a difference? From observation it seemed easier (less hesitation) with logatomes but they reported easier with words.	**10 min**
Listen and repeat: fonetix exo "il chan...ta toulouse, il chantatoulouse" see how intonation rises at the end of the first group but we still have the enchainement.	**10 min**
Boundary function of intonation illustrated with ambiguous sentences: "les gares sont dessinées / les garçons dessinaient" Practice exercise 500 exo phonétique 5 p.16 - read sentences with rising intonation at the end of each group and falling on the last.	**20 min**

Intonation sémantique - Fonétix exo 1 ("ça va? ça va.") listen and repeat. Do we always go up at the end of a question? Listen to exo 17 p.20 500 exo phonétique (quand est-ce qu'il vient... questions qui descendent)	**30 min**
Discrimination exercise: what is the pattern of a question without interrogative word, with interrogative word, assertion, surprise/exclamation - fonetix last exo of intonation sémantique (listen and say which of the 4 it is)	**10 min**

Class 8 - intonation practice, grouping and pauses

Training exercises: repeat like the teacher, then read by yourself 500 exo phonétique - 17 p.20 (listen and repeat), 18 p.20 (read aloud), 6 p.16 (read and reverse the rhythmic groups, see slide)	**20 min**
Expression: by pair, write a dialogue to talk about last weekend, using all forms (questions rising/falling, assertions, exclamations), practice with exagerated intonation, show to the class	**30 min**
Listen to an extract of a French native retelling a film, what do you notice about the pauses? With the transcription, mark each time you hear a pause (voiced or silenced), where are they realised? How?	**20 min**
Go back to the dialogues you've written just before and insert pauses at adequat moments. Present to class.	**20 min**

<u>Books and website used for the course:</u>



Abry-Deffayet, D., & Chalaron, M.L. (2009). *Les 500 exercices de phonétique niveau B1-B2*. Hachette Langue Etrangère.

Briet, G., Collige, V., & Rassart, E. (2014). *La prononciation en classe de langue.* Presses Universitaires de Grenoble.

Berdoulat, H., Fesquet, L., & Palusci, S. (2018). *Fonetix*. [Online platform]. https://fonetix.org/

ORAL EXPRESSION & COMPREHENSION TRAINING

Lucie Drouillet

Université Jean Jaurès - Laboratoire de NeuroPsychoLinguistique (LNPL)

Two 1.5 hour classes per week for 4 weeks

All instructions during the classes were given in French.

Class 1 - Ice breaker

Explanation on how course is going be organised etc.	**10 min**
Ice breaker activity: teacher writes 2 numbers, 2 places, 2 verbs, 2 objects, students have to ask questions to figure out what they mean to the person. Then each student preps their own little list and everyone tries to guess for each one.	**80 min**

Class 2 - Culture, French cinema

Expression: by pair try to find at least 2 things you have in common, then share with class	**15 min**
Conversation starter: do you go to the movies? what genre do you like? do you know any French movie?	**15 min**
Comprehension exercise: listen to 3 presentation of movies (extract from radio show), match each extract with the poster, what genre are they?, can you hear French actors' name you recognize?	**40 min**
Expression: which of these 3 movies would you choose to go see and why?	**20 min**

Class 3 - The news

Conversation starter: how do you keep in touch with the news?	**15 min**
Comprehension activity - L'atelier manual	**20 min**
Comprehension with authentic material	**30 min**
Expression about fake news website le Gorafi, asked to prepare fake title for next class	**25 min**

Class 4 - Gastronomy part 1

Conversation starter: do you have a special diet	**30 min**
Quizz on traditional French dishes + conversation around them	**60 min**
For next class: prepare a little presentation of a restaurant you like in Toulouse	

Class 5 - Gastronomy part 2

Expression: presentation of a restaurant in Toulouse	**20 min**
Comprehension: video about loss of a Michelin star for Michel Sarran + conversation on high gastronomy	**40 min**
Expression: prepare a menu for an imaginary dinner with friends and present to class	**30 min**

Class 6 - Holidays

Conversation starter: what's a good holiday for you, how do you travel etc.	**45 min**
Comprehension: TV5 monde activity	**45 min**

Class 7 - Living spaces

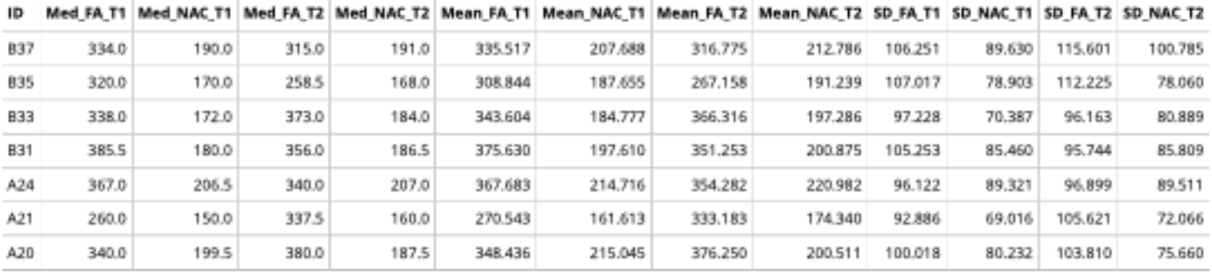
Conversation starter: in what kind of house do you live now? Have you lived in different types of houses?	**30 min**
Comprehension: TV5 monde activity	**30 min**
Expression: what is your dream house like?	**30 min**

Class 8 - Music

Conversation starter: what kind of music do you listen to? Do you go to gigs?	**30 min**
Comprehension: video on fête de la musique	**45 min**
Expression: what will you do for fête de la musique ?	**15 min**

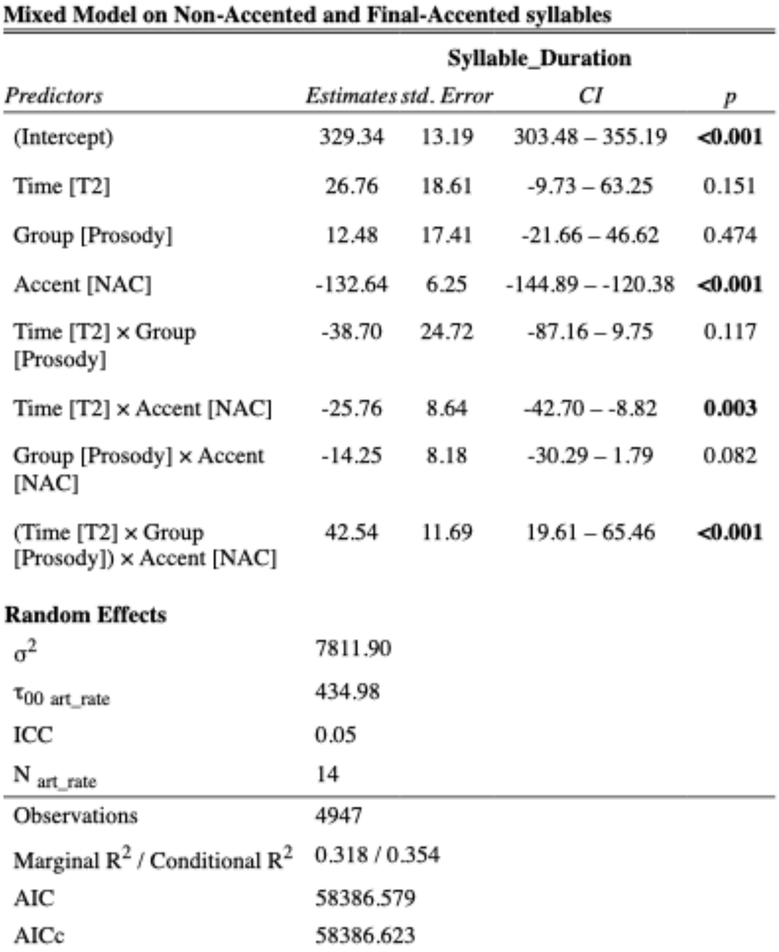
Appendix 2 : Summary tables & mixed models outputs

Appendix 2a - SYLLABLE DURATIONS



ID	Med_FA_T1	Med_NAC_T1	Med_FA_T2	Med_NAC_T2	Mean_FA_T1	Mean_NAC_T1	Mean_FA_T2	Mean_NAC_T2	SD_FA_T1	SD_NAC_T1	SD_FA_T2	SD_NAC_T2
B37	334.0	190.0	315.0	191.0	335.517	207.688	316.775	212.786	106.251	89.630	115.601	100.785
B35	320.0	170.0	258.5	168.0	308.844	187.655	267.158	191.239	107.017	78.903	112.225	78.060
B33	338.0	172.0	373.0	184.0	343.604	184.777	366.316	197.286	97.228	70.387	96.163	80.889
B31	385.5	180.0	356.0	186.5	375.630	197.610	351.253	200.875	105.253	85.460	95.744	85.809
A24	367.0	206.5	340.0	207.0	367.683	214.716	354.282	220.982	96.122	89.321	96.899	89.511
A21	260.0	150.0	337.5	160.0	270.543	161.613	333.183	174.340	92.886	69.016	105.621	72.066
A20	340.0	199.5	380.0	187.5	348.436	215.045	376.250	200.511	100.018	80.232	103.810	75.660

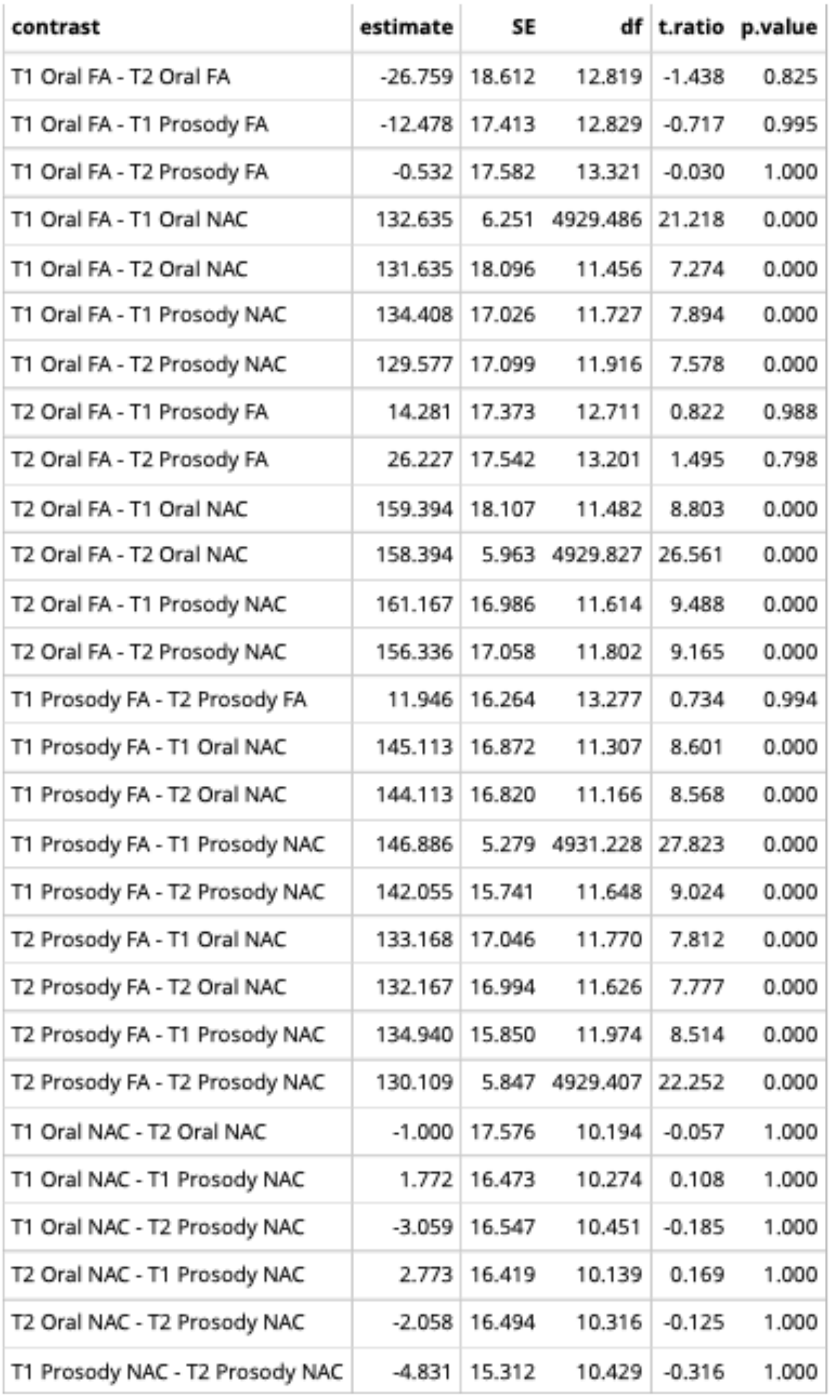
Table A - Summary of syllable duration data.



Mixed Model on Non-Accented and Final-Accented syllables

	Syllable_Duration			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	329.34	13.19	303.48 – 355.19	**<0.001**
Time [T2]	26.76	18.61	-9.73 – 63.25	0.151
Group [Prosody]	12.48	17.41	-21.66 – 46.62	0.474
Accent [NAC]	-132.64	6.25	-144.89 – -120.38	**<0.001**
Time [T2] × Group [Prosody]	-38.70	24.72	-87.16 – 9.75	0.117
Time [T2] × Accent [NAC]	-25.76	8.64	-42.70 – -8.82	**0.003**
Group [Prosody] × Accent [NAC]	-14.25	8.18	-30.29 – 1.79	0.082
(Time [T2] × Group [Prosody]) × Accent [NAC]	42.54	11.69	19.61 – 65.46	**<0.001**
Random Effects				
σ^2	7811.90			
$\tau_{00\ \mathrm{art\_rate}}$	434.98			
ICC	0.05			
$N_{\mathrm{art\_rate}}$	14			
Observations	4947			
Marginal R^2 / Conditional R^2	0.318 / 0.354			
AIC	58386.579			
AICc	58386.623			

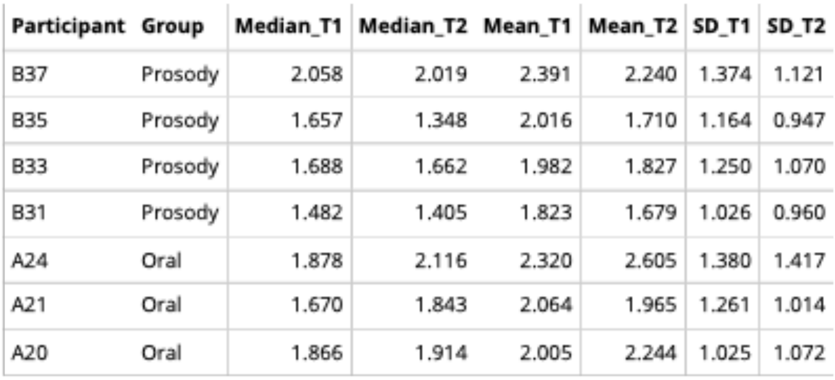
Table B - Mixed model output on syllable durations.



contrast	estimate	SE	df	t.ratio	p.value
T1 Oral FA - T2 Oral FA	-26.759	18.612	12.819	-1.438	0.825
T1 Oral FA - T1 Prosody FA	-12.478	17.413	12.829	-0.717	0.995
T1 Oral FA - T2 Prosody FA	-0.532	17.582	13.321	-0.030	1.000
T1 Oral FA - T1 Oral NAC	132.635	6.251	4929.486	21.218	0.000
T1 Oral FA - T2 Oral NAC	131.635	18.096	11.456	7.274	0.000
T1 Oral FA - T1 Prosody NAC	134.408	17.026	11.727	7.894	0.000
T1 Oral FA - T2 Prosody NAC	129.577	17.099	11.916	7.578	0.000
T2 Oral FA - T1 Prosody FA	14.281	17.373	12.711	0.822	0.988
T2 Oral FA - T2 Prosody FA	26.227	17.542	13.201	1.495	0.798
T2 Oral FA - T1 Oral NAC	159.394	18.107	11.482	8.803	0.000
T2 Oral FA - T2 Oral NAC	158.394	5.963	4929.827	26.561	0.000
T2 Oral FA - T1 Prosody NAC	161.167	16.986	11.614	9.488	0.000
T2 Oral FA - T2 Prosody NAC	156.336	17.058	11.802	9.165	0.000
T1 Prosody FA - T2 Prosody FA	11.946	16.264	13.277	0.734	0.994
T1 Prosody FA - T1 Oral NAC	145.113	16.872	11.307	8.601	0.000
T1 Prosody FA - T2 Oral NAC	144.113	16.820	11.166	8.568	0.000
T1 Prosody FA - T1 Prosody NAC	146.886	5.279	4931.228	27.823	0.000
T1 Prosody FA - T2 Prosody NAC	142.055	15.741	11.648	9.024	0.000
T2 Prosody FA - T1 Oral NAC	133.168	17.046	11.770	7.812	0.000
T2 Prosody FA - T2 Oral NAC	132.167	16.994	11.626	7.777	0.000
T2 Prosody FA - T1 Prosody NAC	134.940	15.850	11.974	8.514	0.000
T2 Prosody FA - T2 Prosody NAC	130.109	5.847	4929.407	22.252	0.000
T1 Oral NAC - T2 Oral NAC	-1.000	17.576	10.194	-0.057	1.000
T1 Oral NAC - T1 Prosody NAC	1.772	16.473	10.274	0.108	1.000
T1 Oral NAC - T2 Prosody NAC	-3.059	16.547	10.451	-0.185	1.000
T2 Oral NAC - T1 Prosody NAC	2.773	16.419	10.139	0.169	1.000
T2 Oral NAC - T2 Prosody NAC	-2.058	16.494	10.316	-0.125	1.000
T1 Prosody NAC - T2 Prosody NAC	-4.831	15.312	10.429	-0.316	1.000

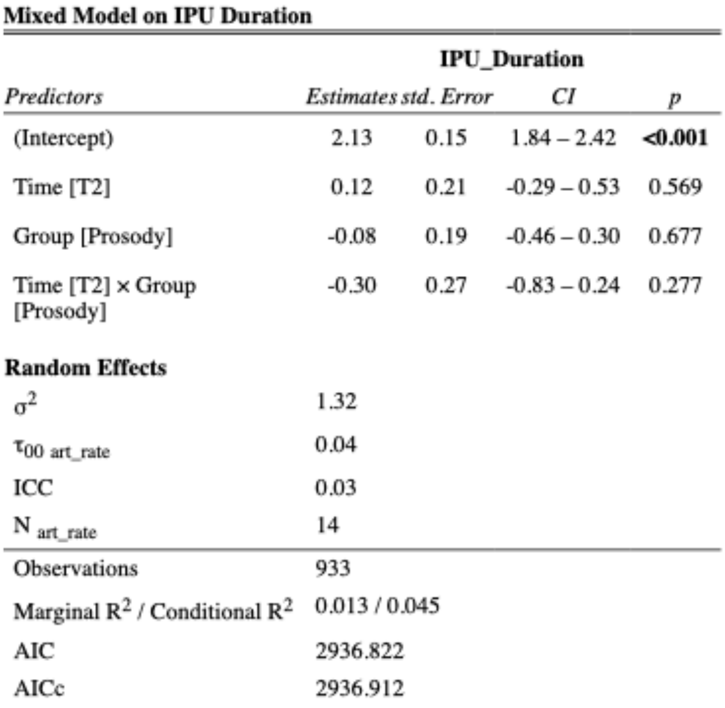
Table C - Post-hoc pairwise comparisons, Tukey method.

Appendix 2b - IPU DURATIONS



Participant	Group	Median_T1	Median_T2	Mean_T1	Mean_T2	SD_T1	SD_T2
B37	Prosody	2.058	2.019	2.391	2.240	1.374	1.121
B35	Prosody	1.657	1.348	2.016	1.710	1.164	0.947
B33	Prosody	1.688	1.662	1.982	1.827	1.250	1.070
B31	Prosody	1.482	1.405	1.823	1.679	1.026	0.960
A24	Oral	1.878	2.116	2.320	2.605	1.380	1.417
A21	Oral	1.670	1.843	2.064	1.965	1.261	1.014
A20	Oral	1.866	1.914	2.005	2.244	1.025	1.072

Table D - Summary of IPU duration data.

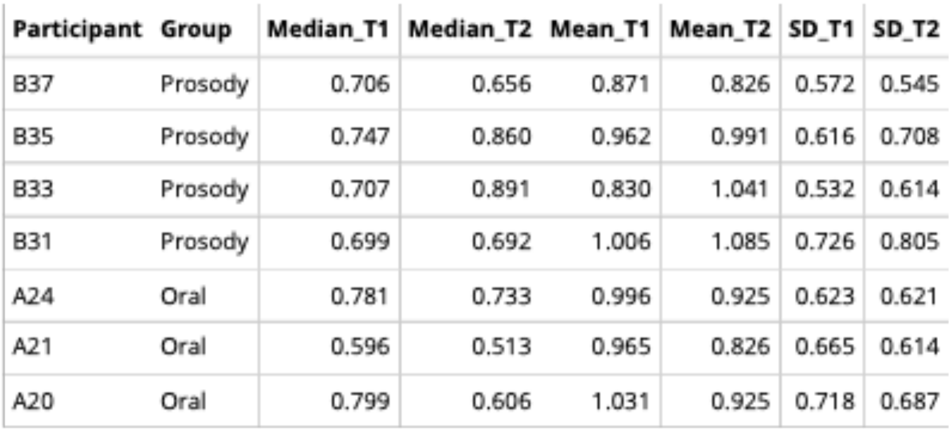


Mixed Model on IPU Duration

	IPU_Duration			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	2.13	0.15	1.84 – 2.42	**<0.001**
Time [T2]	0.12	0.21	-0.29 – 0.53	0.569
Group [Prosody]	-0.08	0.19	-0.46 – 0.30	0.677
Time [T2] × Group [Prosody]	-0.30	0.27	-0.83 – 0.24	0.277
Random Effects				
σ^2	1.32			
$\tau_{00\ art\_rate}$	0.04			
ICC	0.03			
N_{art_rate}	14			
Observations	933			
Marginal R^2 / Conditional R^2	0.013 / 0.045			
AIC	2936.822			
AICc	2936.912			

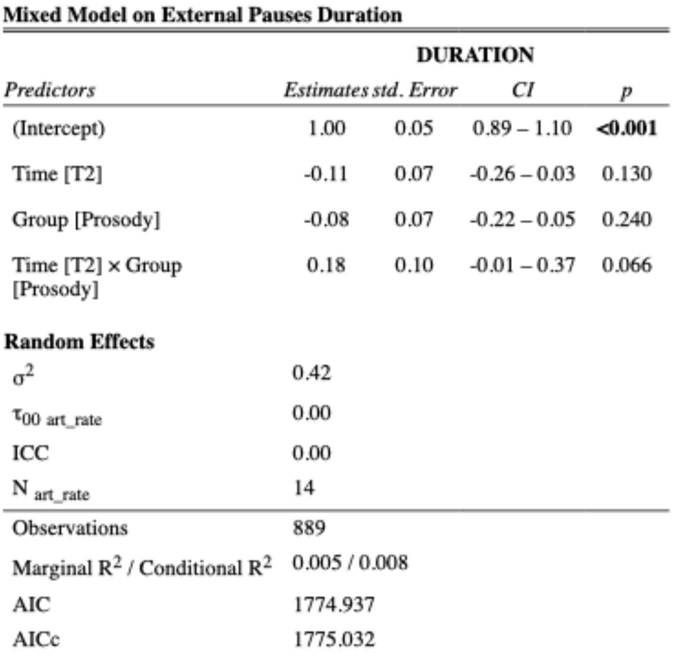
Table E - Mixed model output on IPU durations.

Appendix 2c - EXTERNAL PAUSES DURATIONS



Participant	Group	Median_T1	Median_T2	Mean_T1	Mean_T2	SD_T1	SD_T2
B37	Prosody	0.706	0.656	0.871	0.826	0.572	0.545
B35	Prosody	0.747	0.860	0.962	0.991	0.616	0.708
B33	Prosody	0.707	0.891	0.830	1.041	0.532	0.614
B31	Prosody	0.699	0.692	1.006	1.085	0.726	0.805
A24	Oral	0.781	0.733	0.996	0.925	0.623	0.621
A21	Oral	0.596	0.513	0.965	0.826	0.665	0.614
A20	Oral	0.799	0.606	1.031	0.925	0.718	0.687

Table F - Summary of external pauses duration data.

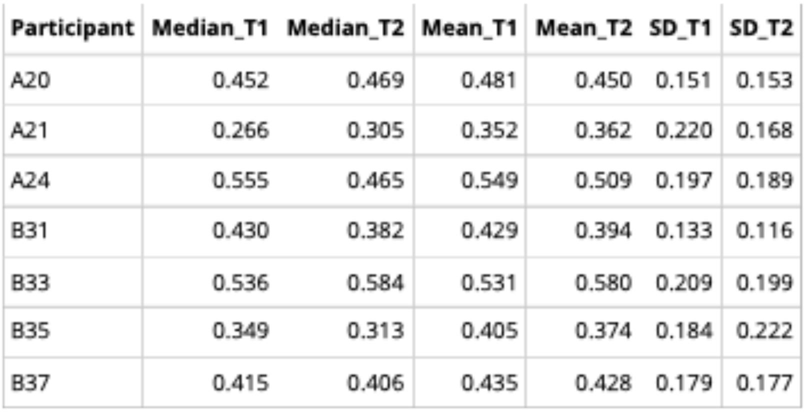


Mixed Model on External Pauses Duration

	DURATION			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	1.00	0.05	0.89 – 1.10	**<0.001**
Time [T2]	-0.11	0.07	-0.26 – 0.03	0.130
Group [Prosody]	-0.08	0.07	-0.22 – 0.05	0.240
Time [T2] × Group [Prosody]	0.18	0.10	-0.01 – 0.37	0.066
Random Effects				
σ^2	0.42			
$\tau_{00\ art\_rate}$	0.00			
ICC	0.00			
N_{art_rate}	14			
Observations	889			
Marginal R^2 / Conditional R^2	0.005 / 0.008			
AIC	1774.937			
AICc	1775.032			

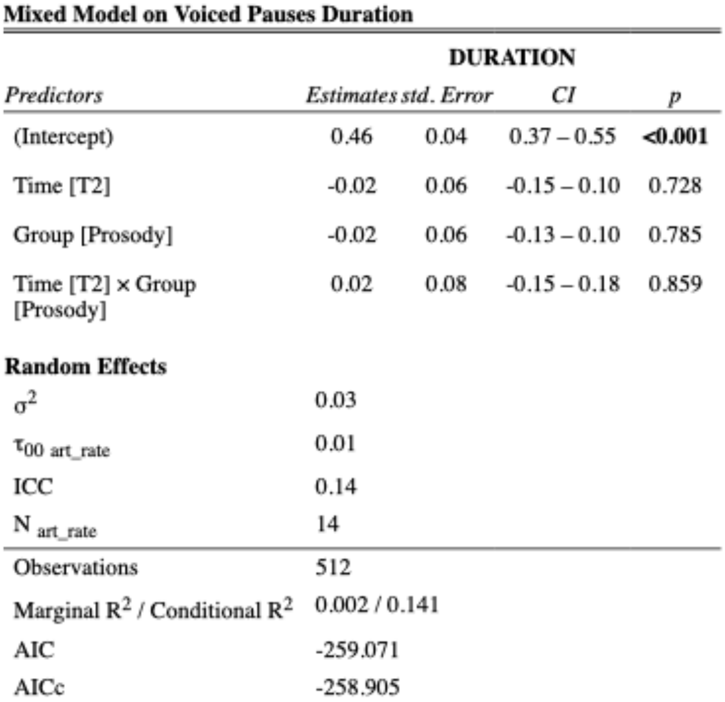
Table G - Mixed model output on external pause durations.

Appendix 2d - VOICED PAUSES DURATIONS



Participant	Median_T1	Median_T2	Mean_T1	Mean_T2	SD_T1	SD_T2
A20	0.452	0.469	0.481	0.450	0.151	0.153
A21	0.266	0.305	0.352	0.362	0.220	0.168
A24	0.555	0.465	0.549	0.509	0.197	0.189
B31	0.430	0.382	0.429	0.394	0.133	0.116
B33	0.536	0.584	0.531	0.580	0.209	0.199
B35	0.349	0.313	0.405	0.374	0.184	0.222
B37	0.415	0.406	0.435	0.428	0.179	0.177

Table H - Summary of voiced pauses duration data.



Mixed Model on Voiced Pauses Duration

	DURATION			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	0.46	0.04	0.37 – 0.55	**<0.001**
Time [T2]	-0.02	0.06	-0.15 – 0.10	0.728
Group [Prosody]	-0.02	0.06	-0.13 – 0.10	0.785
Time [T2] × Group [Prosody]	0.02	0.08	-0.15 – 0.18	0.859
Random Effects				
σ^2	0.03			
$\tau_{00\ \mathrm{art\_rate}}$	0.01			
ICC	0.14			
N $_{\mathrm{art\_rate}}$	14			
Observations	512			
Marginal R^2 / Conditional R^2	0.002 / 0.141			
AIC	-259.071			
AICc	-258.905			

Table I - Mixed model output on voiced pause durations.

Appendix 3: Summary tables of T1-T2 variability in L1 and L2

Disfluency Quantity

Participants	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	24.99	18.1	-6.89	45.41	44.03	-1.38
B35	Prosody	19.15	30.46	11.31	69.87	62.55	-7.32
B33	Prosody	26.73	26.5	-0.23	47.72	44.21	-3.51
B31	Prosody	21.08	32.75	11.67	50.6	49.23	-1.37
Total Prosody		22.99	26.95	**7.53**	53.4	50.01	**3.4**
A24	Oral	27.03	31.22	4.19	55.38	63.95	8.57
A21	Oral	27.75	40.51	12.76	69.65	60.82	-8.83
A20	Oral	31	22.22	-8.78	51.97	60.77	8.8
Total Oral		28.59	31.32	**8.58**	59	61.85	**8.73**
TOTAL ALL		25.09	28.59	**7.92**	55.5	54.45	**5.4**

Table J - Disfluency quantity across Language and Time.

IPU Duration

Participants	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	2.088	2.441	0.353	2.509	2.539	0.03
B35	Prosody	2.031	1.865	-0.166	2.087	-1.989	0.098
B33	Prosody	3.612	2.728	-0.884	2.119	1.954	0.165
B31	Prosody	2.095	2.702	0.607	1.823	-1.679	0.144
Total Prosody		2.457	-2.434	**0.503**	2.135	-2.04	**0.109**
A24	Oral	2.34	-2.227	0.113	3.262	-3.109	0.153
A21	Oral	3.5	3.538	0.038	2.226	-2.015	0.211
A20	Oral	2.306	2.523	0.217	2.062	2.408	0.346
Total Oral		2.715	2.763	**0.123**	2.517	-2.511	**0.237**
TOTAL ALL		2.567	2.575	**0.340**	2.298	-2.242	**0.164**

Table K - IPU duration across Language and Time.

IPU Quantity							
Participants	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	47.895	-40.975	6.920	39.850	-39.385	0.465
B35	Prosody	49.247	53.624	4.377	47.924	50.269	2.345
B33	Prosody	27.688	36.662	8.974	47.200	51.175	3.975
B31	Prosody	47.735	-37.006	10.730	54.865	59.542	4.677
Total Prosody		43.141	-42.066	**7.750**	47.460	50.093	**2.866**
A24	Oral	42.741	44.898	2.157	30.653	32.166	1.513
A21	Oral	28.569	-28.264	0.305	44.923	49.623	4.700
A20	Oral	43.363	-39.642	3.721	48.505	-41.525	6.980
Total Oral		38.224	-37.601	**2.061**	41.361	-41.105	**4.397**
TOTAL ALL		41.034	-40.153	**5.312**	44.846	46.241	**3.522**

Table L - IPU quantity across Language and Time.

External Pause Duration							
Participants	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	0.819	-0.789	0.03	0.676	0.692	0.015
B35	Prosody	0.746	0.844	0.098	0.764	0.75	0.013
B33	Prosody	0.964	-0.845	0.119	0.7	0.935	0.234
B31	Prosody	0.841	0.884	0.043	0.75	-0.649	0.1
Total Prosody		0.842	-0.841	**0.073**	0.723	0.757	**0.091**
A24	Oral	0.807	0.84	0.033	0.836	-0.664	0.172
A21	Oral	0.667	-0.662	0.005	0.581	-0.541	0.039
A20	Oral	0.933	0.965	0.032	0.672	-0.538	0.134
Total Oral		0.802	0.822	**0.023**	0.696	-0.581	**0.115**
TOTAL ALL		0.827	0.834	**0.054**	0.713	-0.691	**0.1**

Table M - External pause duration across Language and Time.

Proportion of filled external pauses over external pause total (in %)

Participants	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	6.06	11.11	5.05	4.97	-4.2	0.77
B35	Prosody	11.94	18.36	6.42	11.84	-4.2	7.64
B33	Prosody	11.11	-9.8	1.3	6.4	-4.48	1.92
B31	Prosody	13.88	20	6.11	24.65	24.65	0
Total Prosody		13.88	14.82	**4.72**	11.96	-9.38	**2.58**
A24	Oral	12.9	17.5	4.59	8.28	-3.92	4.36
A21	Oral	13.33	-8.88	4.44	7.95	14.62	6.67
A20	Oral	17.39	19.23	1.83	13.8	21.6	7.8
Total Oral		14.54	15.20	**3.62**	10.01	13.38	**6.27**
TOTAL ALL		14.28	14.96	**4.31**	11.23	-10.88	**3.96**

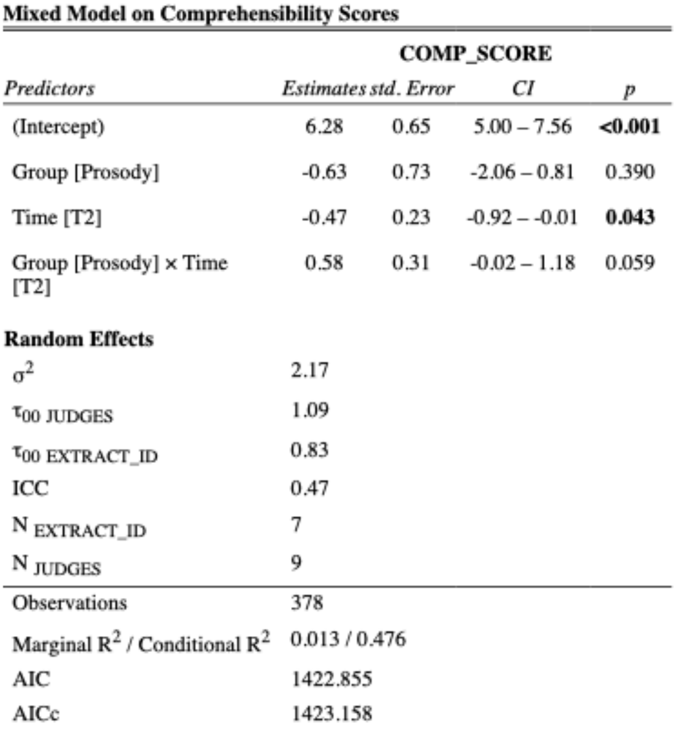
Table N - Proportion of filled external pauses across Language and Time.

Voiced pauses duration

Participant	Group	L1_T1	L1_T2	Var_L1	L2_T1	L2_T2	Var_L2
B37	Prosody	0.408	0.432	0.024	0.459	-0.428	0.031
B35	Prosody	0.401	0.43	0.029	0.432	-0.402	0.03
B33	Prosody	0.412	0.505	0.093	0.531	0.58	0.049
B31	Prosody	0.525	-0.457	0.068	0.429	-0.394	0.035
Total Prosody		0.437	0.456	**0.054**	0.463	-0.451	**0.036**
A24	Oral	0.506	-0.502	0.004	0.583	-0.534	0.049
A21	Oral	0.486	-0.466	0.02	0.397	-0.362	0.035
A20	Oral	0.444	0.537	0.093	0.481	-0.45	0.031
Total Oral		0.479	0.502	**0.039**	0.487	-0.449	**0.038**
TOTAL ALL		0.452	0.473	**0.048**	0.472	-0.45	**0.037**

Table O - Voiced pauses duration across Language and Time.

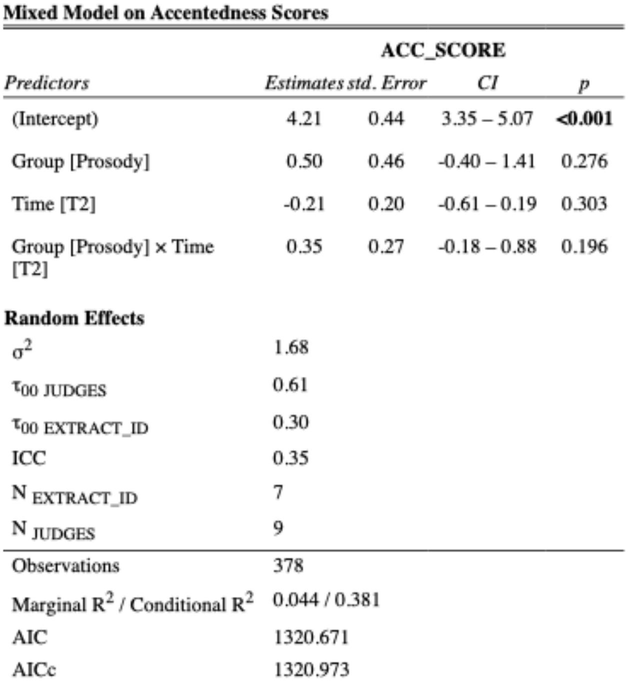
Appendix 4: Comprehensibility and Accentedness mixed-model outputs



Mixed Model on Comprehensibility Scores

	COMP_SCORE			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	6.28	0.65	5.00 – 7.56	**<0.001**
Group [Prosody]	-0.63	0.73	-2.06 – 0.81	0.390
Time [T2]	-0.47	0.23	-0.92 – -0.01	**0.043**
Group [Prosody] × Time [T2]	0.58	0.31	-0.02 – 1.18	0.059
Random Effects				
σ^2	2.17			
τ_{00} JUDGES	1.09			
τ_{00} EXTRACT_ID	0.83			
ICC	0.47			
N EXTRACT_ID	7			
N JUDGES	9			
Observations	378			
Marginal R^2 / Conditional R^2	0.013 / 0.476			
AIC	1422.855			
AICc	1423.158			

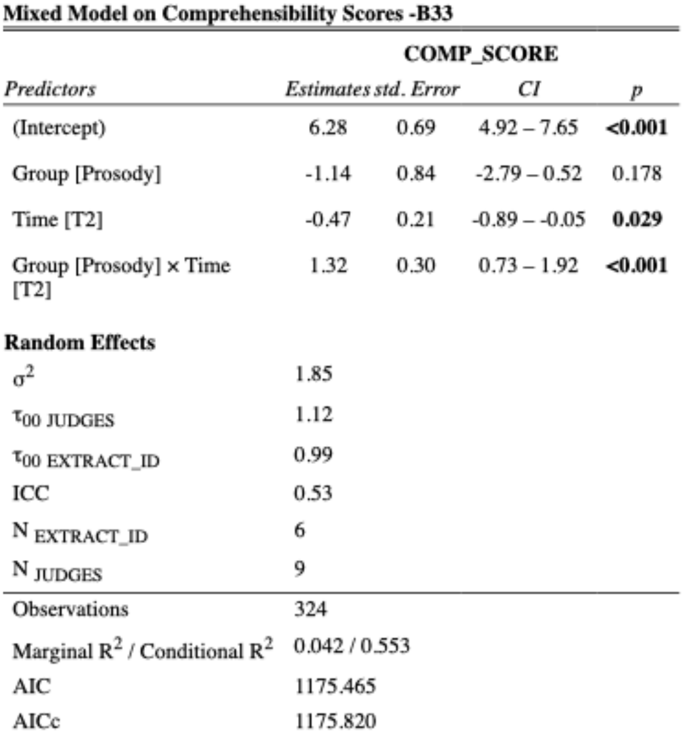
Table P - Mixed model output on comprehensibility.



Mixed Model on Accentedness Scores

	ACC_SCORE			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	4.21	0.44	3.35 – 5.07	**<0.001**
Group [Prosody]	0.50	0.46	-0.40 – 1.41	0.276
Time [T2]	-0.21	0.20	-0.61 – 0.19	0.303
Group [Prosody] × Time [T2]	0.35	0.27	-0.18 – 0.88	0.196
Random Effects				
σ^2	1.68			
τ_{00} JUDGES	0.61			
τ_{00} EXTRACT_ID	0.30			
ICC	0.35			
N EXTRACT_ID	7			
N JUDGES	9			
Observations	378			
Marginal R^2 / Conditional R^2	0.044 / 0.381			
AIC	1320.671			
AICc	1320.973			

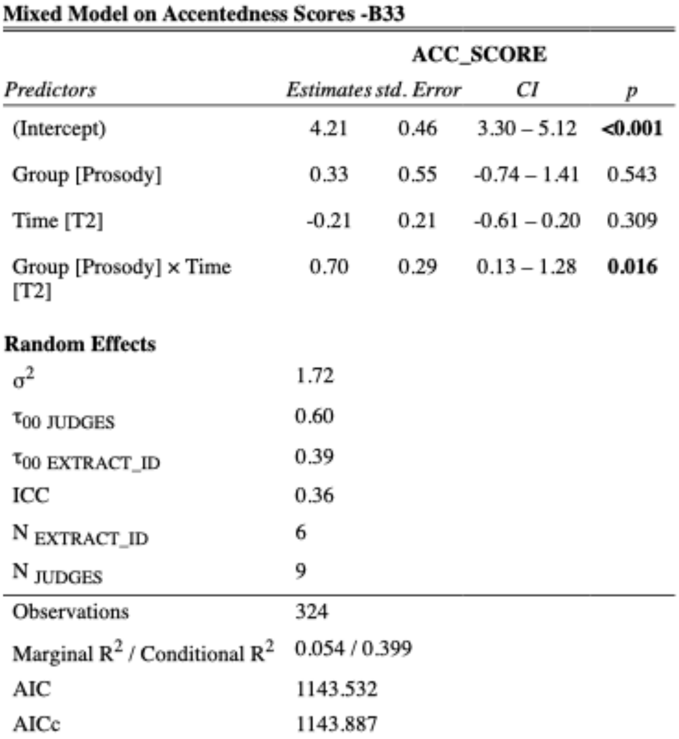
Table Q - Mixed model output on accendtedness.



Mixed Model on Comprehensibility Scores -B33

	COMP_SCORE			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	6.28	0.69	4.92 – 7.65	**<0.001**
Group [Prosody]	-1.14	0.84	-2.79 – 0.52	0.178
Time [T2]	-0.47	0.21	-0.89 – -0.05	**0.029**
Group [Prosody] × Time [T2]	1.32	0.30	0.73 – 1.92	**<0.001**
Random Effects				
σ^2	1.85			
$\tau_{00\ \mathrm{JUDGES}}$	1.12			
$\tau_{00\ \mathrm{EXTRACT\_ID}}$	0.99			
ICC	0.53			
$N_{\mathrm{EXTRACT\_ID}}$	6			
N_{JUDGES}	9			
Observations	324			
Marginal R^2 / Conditional R^2	0.042 / 0.553			
AIC	1175.465			
AICc	1175.820			

Table R - Mixed model output on comprehensibility excluding B33.

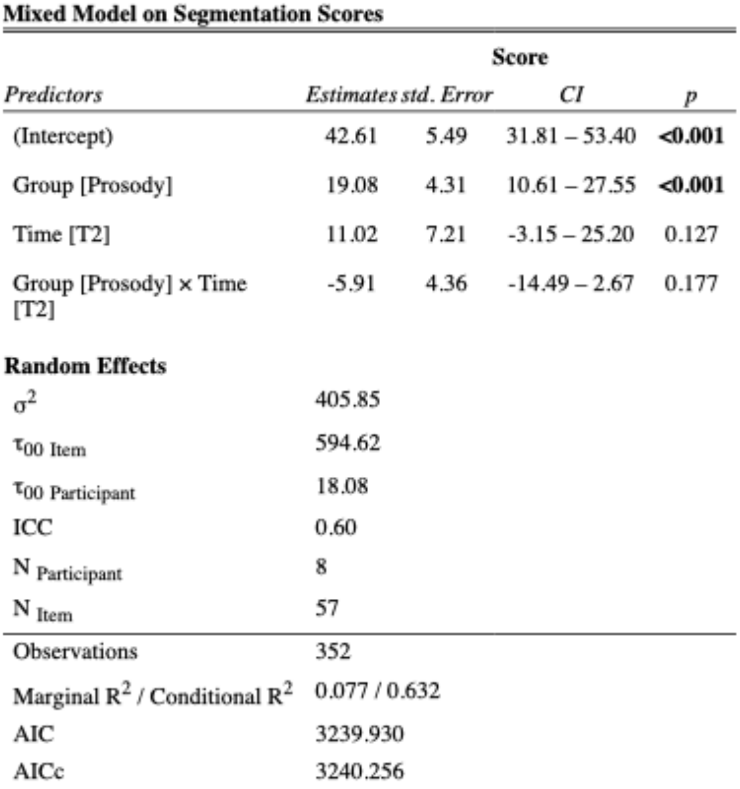


Mixed Model on Accentedness Scores -B33

	ACC_SCORE			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	4.21	0.46	3.30 – 5.12	**<0.001**
Group [Prosody]	0.33	0.55	-0.74 – 1.41	0.543
Time [T2]	-0.21	0.21	-0.61 – 0.20	0.309
Group [Prosody] × Time [T2]	0.70	0.29	0.13 – 1.28	**0.016**
Random Effects				
σ^2	1.72			
$\tau_{00\ \mathrm{JUDGES}}$	0.60			
$\tau_{00\ \mathrm{EXTRACT\_ID}}$	0.39			
ICC	0.36			
$N_{\mathrm{EXTRACT\_ID}}$	6			
N_{JUDGES}	9			
Observations	324			
Marginal R^2 / Conditional R^2	0.054 / 0.399			
AIC	1143.532			
AICc	1143.887			

Table S - Mixed model output on accentedness excluding B33.

Appendix 5: Segmentation scores mixed-model outputs



Mixed Model on Segmentation Scores

	Score			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	42.61	5.49	31.81 – 53.40	**<0.001**
Group [Prosody]	19.08	4.31	10.61 – 27.55	**<0.001**
Time [T2]	11.02	7.21	-3.15 – 25.20	0.127
Group [Prosody] × Time [T2]	-5.91	4.36	-14.49 – 2.67	0.177
Random Effects				
σ^2	405.85			
$\tau_{00\ \text{Item}}$	594.62			
$\tau_{00\ \text{Participant}}$	18.08			
ICC	0.60			
$N_{\text{Participant}}$	8			
N_{Item}	57			
Observations	352			
Marginal R^2 / Conditional R^2	0.077 / 0.632			
AIC	3239.930			
AICc	3240.256			

Table T - Mixed model output on segmentation scores.

Appendix 6: Summary in French - Résumé en français

Résumé en français

Introduction

Il est évident que le monde dans lequel nous vivons aujourd'hui est multilingue. Avec le développement d'internet, des médias sociaux et l'accès à des contenus culturels du monde entier, les langues nous entourent. Mon premier coup de cœur a été pour l'anglais, puis pour le japonais et l'italien. Mais mes sentiments à l'égard d'une langue sont avant tout liés à sa sonorité. Je tombe amoureuse de leur mélodie, de leur rythme, de leur musique. Apprendre une langue, pour moi, c'est être capable de changer la musique de ma parole.

Sans surprise, cet intérêt intuitif pour les langues m'a amenée à devenir enseignante de français langue étrangère (FLE). Cette expérience m'a permis de réaliser que je n'étais pas du tout équipée pour enseigner les aspects de la langue qui me font le plus vibrer. C'est alors qu'a commencé mon voyage dans le monde de la linguistique et de la recherche. J'ai découvert que la mélodie et le rythme que j'aime tant s'appellent la prosodie, et qu'il existe beaucoup plus d'informations, de descriptions, de théories et de modèles à ce sujet que je n'aurais pu l'imaginer. J'ai plongé la tête la première, et le reste appartient à l'histoire.

Le travail présenté dans cette thèse reflète mon intérêt pour l'étude et la mesure des aspects prosodiques, la comparaison inter-langues et le transfert des résultats de la recherche aux méthodes d'enseignement.

Le premier chapitre présente notre objet d'étude : le rythme de la parole. Dans le contexte de la musique, il existe une idée commune de ce qu'est le rythme. Nous avons peut-être du mal à l'expliquer, car c'est quelque chose que nous ressentons plus que nous l'intellectualisons, mais nous sommes tous d'accord pour dire qu'il s'agit des battements sous-jacents, des motifs qu'ils forment, et peut-être aussi du tempo. Mais qu'est-ce que le rythme dans la parole ? Comment la parole est-elle rythmée ? Où et comment se situent les battements ?

Paradoxalement, le rythme de la parole est à la fois l'élément le plus fondamental qui structure le langage parlé, mais aussi le plus difficile à appréhender. Plusieurs approches seront présentées, auxquelles correspondent des corrélats acoustiques distincts du rythme de la parole. Parce que ces points de vue contrastés sont en fait complémentaires, nous proposerons une approche intégrative qui les considère tous comme des niveaux d'analyse différents qui s'entrelacent et créent, au total, le rythme de la parole.

Le chapitre II se concentre sur les spécificités de la parole non native (L2 ci-après). Parler en L2 engage des processus cognitifs et moteurs qui ne sont pas aussi automatiques que parler dans notre première langue (L1), ce qui entraîne des difficultés. En outre, le transfert du système linguistique de la L1 et les processus d'acquisition universels s'entrecroisent et ont un impact sur la production de la parole en L2. Des modèles d'acquisition de la L2 seront présentés ainsi que des études empiriques qui soutiennent leurs hypothèses théoriques. L'analyse de la littérature comprendra des études sur le rythme de la parole en L2 à partir des différentes approches présentées au chapitre I, et abordera également la relation entre les habitudes de parole en L1 et en L2 au sein d'un même sujet.

Dans le chapitre III, nous passons de la production de la parole en L2 à la perception de celle-ci. La définition de ce qui constitue un accent étranger sera interrogée, et son impact sur la perception du discours en L2 par les auditeurs natifs sera exploré. En effet, les auditeurs natifs sont souvent sollicités dans les études sur la prononciation en L2 pour faire le lien entre la performance du locuteur en L2 et la façon dont elle est perçue en termes d'intelligibilité.

Nous nous pencherons ensuite sur la perception et les capacités d'écoute du locuteur L2 à l'égard de la langue cible telle qu'elle est parlée par les natifs. En effet, l'expérience de la L2 implique d'écouter et de comprendre, aussi bien que de parler. La première étape de l'accès au sens est la capacité à segmenter le flux continu de la parole en mots individuels. Ce mécanisme parfaitement fluide en L1 devient un défi de taille en L2.

Le quatrième chapitre passe en revue les méthodes et techniques utilisées pour enseigner la prononciation en L2. Au cours du siècle dernier, les approches pédagogiques sont passées de l'enseignement explicite des sons de la L2 et exercices d'imitation à des objectifs de communication plus holistiques. Certaines méthodes incluent l'utilisation de gestes, d'accessoires et même de musique afin d'aider les apprenants à percevoir et à acquérir les sons et la prosodie de la L2. Une revue de la littérature des études testant et comparant l'efficacité des différentes méthodes d'enseignement de la prononciation sera présentée.

Enfin, les deux derniers chapitres de cette thèse présenteront le dispositif expérimental et les résultats de l'étude que nous avons menée. En français L2, l'enseignement de la prosodie n'est pas une pratique courante. Cependant, les chapitres précédents ont mis en évidence le rôle crucial de la prosodie dans la prononciation et les capacités d'écoute en L2. Par conséquent, l'étude interroge l'impact d'un enseignement spécifique de la prosodie sur les performances des d'apprenants anglais-L1 en français L2. En comparant cet enseignement à des activités générales d'expression et compréhension orale couramment utilisées dans les cours de français L2, nous cherchons à voir si une approche plus directe, explicite et multimodale de la prosodie en français L2 est plus efficace que ce qui est couramment fait dans les cours de français L2. L'effet des deux enseignements sera évalué sur des mesures acoustiques du rythme de la parole, en suivant notre approche intégrative.

De plus, nous évaluerons les progrès des apprenants après la formation grâce à des mesures perceptives de la compréhensibilité (facilité de compréhension) et de l'accentuation (degré d'accent étranger) attribuées par des auditeurs de langue maternelle française. Cela permettra d'évaluer si les changements mesurés dans le rythme de la parole des participants sont pertinents en termes de perception par les auditeurs natifs.

Enfin, les capacités perceptives des participants seront évaluées pour voir si un entraînement à la prosodie peut également aider les apprenants à segmenter le discours des locuteurs natifs français.

Les choix méthodologiques effectués pour construire le dispositif expérimental mettent l'accent sur la validité écologique des résultats. Grâce à un post-test différé et à l'analyse d'échantillons de parole spontanée, nous nous assurons que les changements observés avant et après la formation ne sont pas limités à la salle de classe, mais qu'ils sont transférés dans la parole naturelle, et qu'ils sont toujours visibles une semaine après la fin de la formation. En outre, des échantillons de discours en L1 seront également analysés. Les différences entre la production en L1 et en L2 aideront à interpréter les résultats en L2.

Les implications des résultats en termes de pratique pédagogique, de décisions méthodologiques et de niveau d'analyse du rythme de la parole seront discutées.

Chapitre I

Dans ce premier chapitre, nous avons contextualisé et défini notre principal objet d'étude : le rythme de la parole. Nous avons vu que le rythme est au cœur de toutes les activités et de tous les comportements humains, et parmi toutes les définitions du rythme présentées, nous avons retenu celle de Sauvannet (2000) qui propose que les critères essentiels d'un phénomène rythmique soient la structure, la périodicité et le mouvement.

Dans la parole, la structure émerge du regroupement et de l'organisation hiérarchique des syllabes en constituants plus larges. Le mouvement est créé par l'alternance de syllabes fortes et faibles (accentuées ou non), de sons continus et de pauses, et de contours mélodiques contrastés. Enfin, la récurrence des proéminences, à tous les niveaux de la hiérarchie, crée des motifs et une périodicité.

Nous avons discuté de la relation entre le rythme de la parole et le rythme moteur, et souligné le fait que l'interaction entre les deux est exploitée depuis longtemps dans le contexte de la rééducation de la parole et de l'enseignement de la L2 - bien que de manière essentiellement intuitive. Nous avons également brièvement évoqué le rôle de l'entraînement neuronal au rythme de la parole dans la perception et le traitement du langage. L'étude de ce phénomène dans le développement de la L2 pourrait nous aider à mieux comprendre le rôle du rythme de la parole dans le traitement de la parole en L2.

Après cette introduction générale, nous avons présenté les corrélats acoustiques du rythme de la parole selon quatre points de vue théoriques différents. Du point de vue phonologique, les mesures dites de rythme se concentrent sur des mesures quantitatives de la proportion de voyelles et de consonnes qui reflètent la complexité de la structure syllabique d'une langue, et sur des mesures de la variabilité temporelle des intervalles (vocaliques, consonantiques, syllabiques) qui reflètent le degré de réduction des voyelles. Comme ces mesures ne tiennent pas compte des aspects prosodiques de plus haut niveau, elles sont confinées à un micro-niveau d'analyse du rythme de la parole.

A l'inverse, l'approche prosodique basée sur la théorie métrique considère que le rythme de la parole émerge des règles métriques de la langue, à travers la combinaison et la subordination des constituants, instanciées par la proéminence relative de leurs têtes. Les corrélats acoustiques sont donc ces proéminences, qui se manifestent physiquement par une augmentation de la hauteur, de l'intensité et/ou de la durée. Cette vision du rythme de la parole basée sur le système d'accentuation constitue un niveau intermédiaire d'analyse : le niveau méso.

Un troisième niveau d'analyse concerne les variables temporelles qui se rapportent aux schémas de groupement et de pause. Les durées des séquences de parole et des pauses silencieuses entre les séquences renseignent sur le macro-niveau du rythme de la parole.

Enfin, les mesures de fluidité, telles que la distribution et le nombre de disfluences, et les mesures de la vitesse d'élocution font également partie du rythme de la parole. Cependant, ces mesures ne peuvent être rattachées à un niveau de structuration car elles ne sont pas de nature structurelle, mais plutôt des conséquences de la production vocale. Elles sont donc transversales, puisqu'elles interviennent à tous les niveaux de structuration et de circonscription.

Nous pensons que toutes ces différentes fenêtres sur le rythme de la parole ne s'excluent pas mutuellement, mais qu'elles exercent plutôt une influence les unes sur les autres, et qu'elles devraient toutes être prises en compte pour une approche intégrative du rythme de la parole. C'est ce point de vue que nous adoptons dans cette thèse et, à ce titre, le terme rythme de la parole fera dorénavant référence à cette conception, c'est-à-dire à la combinaison des niveaux micro, méso et macro, ainsi qu'à la fluence, sauf mention contraire.

Enfin, notre travail portant sur les locuteurs et apprenants du français et de l'anglais, nous avons comparé ces deux langues du point de vue de leur structure rythmique, à tous les niveaux d'analyse.

Le chapitre suivant s'intéresse aux spécificités de la parole en L2 et à l'acquisition du rythme de la parole.

Chapitre II

Nous avons commencé ce chapitre en présentant des modèles de production linguistique en L1 et L2. Nous avons vu que chaque étape du processus (conceptualisation, formulation, articulation) est sujette à des difficultés pour le locuteur L2. Des disfluences peuvent apparaître à chaque étape du processus, car la composante de *monitoring* détecte les erreurs ou les divergences par rapport à l'intention initiale du locuteur. Les problèmes rencontrés lors de la phase de conceptualisation peuvent être liés à une inadéquation entre le concept visé et son expression en L2. La phase de formulation en L2 peut nécessiter beaucoup d'attention et une récupération consciente des différents aspects linguistiques impliqués, ce qui peut ralentir le processus de production et générer des ruptures et des disfluences dans le signal, et il en va de même pour la phase d'articulation où les habitudes de la L1 restent tenaces. Selon le point de vue de Kormos (2006), l'encodage prosodique intervient également à chaque étape majeure du processus.

L'aperçu des différents modèles d'acquisition de la L2 a montré que les auteurs, pour la plupart, considèrent le rôle de processus d'acquisition universels et du transfert de la L1. Il est clair que l'acquisition de la phonologie en L2 n'est pas un processus monochrome, et que l'effet des différents facteurs s'entrecroise de manière dynamique. La similarité L1-L2 est également un aspect crucial sur lequel la plupart des modèles s'appuient pour déterminer les domaines de difficulté. Ces modèles permettent d'élaborer des hypothèses concernant l'origine du phénomène de la parole en L2, et il existe un nombre croissant de recherches sur l'acquisition des suprasegmentaux en particulier (par exemple, Li & Post, 2014 ; Rasier & Hiligsmann, 2007 ; Sánchez-Alvarado, 2022). Cependant, jusqu'à présent, les études se sont concentrées sur un, voire deux niveaux d'analyse, mais l'acquisition du rythme de la parole n'a pas encore été abordée d'une manière globale qui inclurait les niveaux micro, méso et macro ainsi que la fluence.

La seconde moitié de ce chapitre est consacrée à une revue de littérature des études expérimentales portant sur le rythme de la parole en L2. Les recherches sur le rythme au niveau micro ont montré que les mesures normalisées semblent plus robustes car elles éliminent le risque d'une influence du débit de la parole. En particulier, le nPVI s'est avéré particulièrement adapté pour distinguer les niveaux de compétence (Ordin et al., 2011 ; Li & Post, 2014). C'est pourquoi nous avons choisi d'inclure une mesure du nPVI dans l'étude présentée aux chapitres V et VI.

Par ailleurs, plusieurs études pointent vers un processus universel concernant la variabilité de durée des intervalles vocaliques et syllabiques sous la forme d'une trajectoire ascendante, en accord avec les résultats des études sur l'acquisition du rythme en L1. Cependant, ces études sont trop souvent axées sur l'acquisition d'une langue *stress-timed*, et seules quelques-unes d'entre elles incluent des locuteurs à des stades précoces de l'acquisition d'une L2. Par conséquent, il n'existe pas d'éléments probants provenant d'études portant sur l'acquisition d'une langue de type *syllable-timed* et sur des locuteurs de L2 à un niveau de compétence élémentaire. L'étude présentée aux chapitres V et VI de cette thèse, bien qu'elle ne soit pas uniquement axée sur le niveau micro, comble en partie cette lacune.

Les recherches axées sur le rythme au niveau méso mettent en évidence ce qui semble être un phénomène universel dans la tendance à sur-distribuer les accents en L2, ainsi qu'un phénomène de sur-articulation qui réduit la variabilité temporelle entre les syllabes accentuées et non accentuées (Barry, 2007 ; Li & Post, 2014 ; Rasier & Hiligsmann, 2007 ; Verdugo, 2003). Il semble que les détails de la réalisation acoustique des proéminences soient particulièrement problématiques et dépendent des niveaux de compétence en L2 (Frost & O'donnell, 2018 ; Ueyama, 2003, 2016).

En ce qui concerne le rythme au niveau macro et la fluence, nous avons vu que les mesures de fluence sont des indicateurs pertinents de la fluence perçue. Cependant, l'impact de chaque mesure individuelle varie selon les paires de langues, les tâches et les choix méthodologiques (Suzuki et al., 2021), et la perception de la fluidité par les auditeurs peut être influencée par d'autres facteurs que les variables temporelles, comme les erreurs d'intonation (Trofimovich et al., 2017). Les mesures de fluence les plus fortement corrélées aux niveaux de compétence sont les mesures composites, les mesures de vitesse et les mesures de rupture (Baker-Smemoe et al., 2014 ; Saito et al., 2018 ; Tavakoli et al., 2020). Les résultats sont plus contrastés sur les mesures de réparation. Il apparaît également que le développement de la fluidité dans une paire de langues peut différer d'une autre paire, et qu'il est donc difficile de dégager des traits communs entre les langues (Baker-Smemoe et al., 2014 ; Préfontaine et al., 2016 pour le français L2).

Enfin, nous avons abordé la relation entre le macro-niveau L1 et L2 et les schémas de fluence. Grâce aux études examinées et à la présentation d'une étude que nous avons nous-mêmes menée sur le français et l'anglais L1 et L2, qui comprenait une analyse intra-sujet et inter-sujet, nous avons mis en évidence le fait que ces aspects sont fortement influencés par les patterns de L1 et les spécificités de la langue cible. Il est apparu que, globalement, les mesures de pause sont fortement corrélées à la L1, les mesures de vitesse modérément, et les mesures de réparation faiblement.

Nous avons également soulevé le fait que la variabilité entre deux temps de test en L1 n'est jamais examinée par rapport à la même variabilité en L2. Par conséquent, outre le rôle que joue la L1 dans les modèles de la L2, nous pourrions obtenir des mesures encore plus précises en utilisant un design longitudinal, si elles sont également débarrassées de la variabilité entre deux temps de la L1. Ce point sera abordé dans l'analyse des résultats de l'étude présentée aux chapitres V et VI.

Dans le troisième chapitre, nous aborderons la perception du discours de la L2 par les locuteurs natifs et les capacités de segmentation des locuteurs de la L2 vis-à-vis de la langue cible.

Chapitre III

Ce chapitre est focalisé sur la perception de la parole en L2, d'abord du point de vue des locuteurs natifs, puis du point de vue des apprenants.

Nous avons commencé par définir le concept d'accent étranger et avons noté la définition de Rasier & Hiligsmann (2007), que nous avons trouvé la plus complète. L'accent étranger est avant tout un phénomène perçu, lié à la prononciation et influencé par la L1 du locuteur. Des données empiriques confirment que les déviations segmentales et suprasegmentales ont un impact sur la perception de l'accent étranger (par exemple : Anderson-Hsieh et al., 1992 ; Sereno, Lammers, & Jongmann, 2014). Cependant, le poids respectif et la manière dont ces deux niveaux interagissent restent assez flous, et varient très certainement en fonction des paires de langues, des contextes de parole et de la sensibilité des auditeurs (Ulbrich & Mennen, 2016 ; Polyanskaya et al., 2017).

Trois concepts essentiels dans la recherche impliquant les jugements perceptuels des auditeurs natifs sur la parole en L2 ont été définis à la suite du travail pionnier de Munro & Derwing (1995a) :

Accentedness correspond au degré d'accent étranger perçu et est évalué à l'aide d'une échelle de Lickert allant de 1 à 9.

La compréhensibilité correspond à l'effort nécessaire à l'auditeur pour parvenir à la compréhension. Elle est liée à la facilité ou à la difficulté rencontrée par l'auditeur dans le processus de décodage. Elle est généralement mesurée de la même manière que l'accentuation, sur une échelle de Lickert en 9 points.

L'intelligibilité est définie comme le degré de compréhension du message par l'auditeur. Elle est généralement évaluée au moyen d'une tâche de transcription.

Des études ont exploré la relation entre les trois concepts et il en ressort qu'ils semblent à la fois se chevaucher et être partiellement indépendants. L'intelligibilité semble être la moins influencée par l'accentuation et la compréhensibilité, étant donné qu'une parole fortement accentuée et de faibles scores de compréhensibilité peuvent être associés à des transcriptions réussies du message. Cela signifie qu'un fort accent n'empêche pas de facto la compréhension, mais qu'il peut ralentir le processus. C'est pourquoi la compréhensibilité donne une image plus précise du traitement de la parole en L2.

Comme nous nous intéressons particulièrement à l'accentuation et à la compréhensibilité (mesures que nous utilisons dans notre étude), nous avons compilé des données sur les corrélats linguistiques de ces deux concepts à partir de la littérature. Les aspects segmentaux ont tendance à être associés à l'accentedness, mais moins à la compréhensibilité. Les caractéristiques suprasegmentales telles que l'accent, le rythme et l'intonation sont associées aux deux concepts. Les mesures de fluence ont tendance à être plus souvent associées à la compréhensibilité qu'à l'*accentuation*. Enfin, les variables non liées à la prononciation sont davantage associées à la compréhensibilité qu'à l'accentedness.

D'un point de vue pédagogique, cela implique que la prosodie et la fluence devraient être au cœur des cours de prononciation en L2, afin d'améliorer la compréhensibilité des apprenants. Inversement, dans la recherche d'une prononciation de type natif, le travail sur le segmental devrait être ajouté puisqu'ils semblent participer autant que le suprasegmental à la perception de l'accent étranger. Dans une classe où, de nos jours, le principe d'intelligibilité prévaut (Lévis, 2005), les activités de prononciation devraient mettre l'accent sur les aspects suprasegmentaux et de fluence, peut-être avant les aspects segmentaux.

Ceci constitue la base des choix que nous avons faits lors de la conception de notre étude présentée dans les chapitres V & VI.

Dans la seconde moitié de ce chapitre, nous avons donné un aperçu des processus en jeu dans la segmentation de la parole et souligné les défis qu'ils représentent pour un locuteur L2 écoutant la langue cible parlée par des natifs. La reconnaissance et la distinction des phonèmes sont entravées par le filtre phonologique de la L1, l'activation lexicale est soumise à la taille du vocabulaire qui est beaucoup plus réduit en L2 qu'en L1, et les indices phonotactiques et prosodiques des différences de segmentation conduisent à l'utilisation inefficace de stratégies probabilistes qui fonctionnent dans la L1 mais sont inadéquates dans la L2.

Nous avons ensuite présenté l'hypothèse de la segmentation rythmique, qui insiste sur l'importance de la structure rythmique des langues dans la segmentation de la parole (Cutler, 2012), ainsi qu'un examen des études portant sur les capacités de segmentation de la parole des apprenants de L2. Les résultats suggèrent qu'il est tout à fait possible d'apprendre à utiliser de nouveaux indices prosodiques, que la facilité ou la difficulté à le faire dépend de la similarité ou de la différence entre la L1 et la L2, et de la dimension de la *L2 Intonation Learning Theory* (Mennen, 2015) qu'ils concernent. Il semble que la dimension systémique soit particulièrement susceptible de prédire des difficultés d'apprentissage.

Dans la dernière section, nous avons posé la question de savoir si le fait d'enseigner aux apprenants de L2 les indices prosodiques pouvait les aider à développer des compétences de segmentation de la parole en L2. Cette question constitue l'une des questions de recherche abordées dans notre étude (chapitres V et VI).

L'enseignement de la prononciation et son efficacité sur les performances des locuteurs L2 en matière de production et de perception est le sujet central du chapitre suivant.

Chapitre IV

L'objectif de ce chapitre était de présenter un état des lieux des pratiques actuelles dans les salles de classe en ce qui concerne le développement des compétences orales des apprenants L2, et en particulier l'enseignement de la prononciation. Même s'il existe un consensus général sur la nécessité d'inclure l'enseignement de la prononciation dans les cours et les classes de langues étrangères, dans la réalité, cet enseignement a toujours tendance à être à la traîne par rapport à d'autres aspects linguistiques considérés comme plus prioritaires (Alazard, 2013 ; Billières, 2014 ; Darcy, 2018 ; Detey & Durand, 2021).

En outre, les formations des enseignants ne comprennent souvent pas suffisamment de connaissances et de techniques pédagogiques spécifiques aux aspects phonologiques et à l'enseignement de la prononciation. Les professeurs de langues ont signalé à plusieurs reprises leur manque de compétence et donc de confiance pour enseigner la phonétique, ainsi que le manque de ressources disponibles sur lesquelles s'appuyer (Breitkreutz et al., 2001 ; Foote et al., 2011, 2016 ; Henderson et al., 2012).

Néanmoins, une variété de méthodes et d'outils ont été développés au cours des 50 dernières années. L'approche articulatoire constitue généralement une base dans les manuels de prononciation, mêlant exercices de perception et de production, et explications explicites. La méthode verbo-tonale (MVT, Guberina, 1956, 1975), en revanche, utilise une approche axée sur la forme et le corps, centrée sur l'erreur de l'apprenant, en mettant l'accent sur les caractéristiques prosodiques de la langue cible.

En fait, plusieurs méthodes d'enseignement de la prononciation font appel à la gestuelle et au mouvement (par exemple, The Silent Way, Gattegno, 1972, 1976, 2010 ; et The Essential Haptic-Integrated English Pronunciation framework, Acton, 2012). Les théories de la cognition et de l'apprentissage incarnés ont été à la base des recherches montrant l'effet amplificateur de l'incarnation (*embodiment*) sur l'apprentissage. L'utilisation du geste est particulièrement appropriée pour travailler sur les sons qui peuvent sembler assez abstraits pour les apprenants (Kontra et al., 2012 ; McCafferty, 2006).

D'une manière générale, l'enseignement multimodal de la prononciation a beaucoup de sens puisqu'il reflète la nature multimodale de la parole elle-même. En outre, les liens entre la musique et le langage, et en particulier la relation étroite entre les aptitudes musicales et les compétences phonologiques en L2, ont motivé l'utilisation de la musique dans l'enseignement de la prononciation, avec des impacts positifs sur les compétences phonologiques productives et perceptives en L2 (Baills et al., 2021 ; Chobert & Besson, 2013 ; François et al., 2013).

Bien que la recherche empirique sur l'enseignement de la prononciation comparant l'efficacité de différentes méthodes soit encore naissante, nous disposons déjà de données solides montrant les avantages de l'enseignement explicite, de l'incarnation et des activités musicales. Nous nous sommes inspirés de ces techniques pour construire un cours multimodal de prosodie en français L2 et avons testé ses effets sur le rythme de la parole, la compréhensibilité, l'accentuation et les compétences de segmentation en L2 des apprenants d'anglais L1 de niveau A2-B1. Les deux chapitres suivants de cette thèse présentent la conception et les résultats de notre étude.

Chapitre V

Ce chapitre présente la méthodologie employée pour la mise en place de l'étude menée. Ce travail a été motivé par les points suivant :

- la volonté de contribuer à combler une lacune dans la littérature concernant l'acquisition et l'enseignement de la prosodie en français L2
- le désir de construire un prototype de cours de prosodie en français L2, qui pourrait servir de base à de futures recherches et au développement de ressources pédagogiques
- la nécessité de comparer les effets de l'enseignement de la prosodie avec ce qui se fait habituellement dans les cours de FLE, c'est-à-dire les activités communicatives (expression orale et compréhension orale)
- la nécessité de combiner différents types de mesures pour examiner les effets du type d'enseignement sur plusieurs aspects de la performance des apprenants (mesures acoustiques du rythme de la parole, jugements des auditeurs natifs, capacités de segmentation des apprenants), ce qui permet à son tour de rechercher des corrélations potentielles entre ces mesures et ces aspects.

La conception de l'étude a impliqué la création de deux cours correspondant aux deux modalités d'enseignement : un cours sur la prosodie et un cours sur la production/compréhension orale. Elle comportait également deux phases distinctes de collecte de données.

La phase 1 consistait en des sessions de pré-test et de post-test (ci-après T1 et T2) qui ont eu lieu la semaine précédant et suivant l'enseignement. Les participants ont été invités à s'enregistrer lors d'une tâche de lecture à haute voix en L1 et en L2, ainsi qu'une tâche d'expression libre en L1 et en L2. Ces enregistrements constituent nos données de parole. Ils ont également effectué une tâche d'écoute et de répétition, dans le but de tester leurs compétences en matière de segmentation de la parole (ci-après appelée tâche de segmentation).

La phase 2 a été réalisée sept mois plus tard et consistait en une tâche d'évaluation de la compréhensibilité et d'accentedness par des auditeurs de langue maternelle française.

Plusieurs questions de recherche ont guidé ce travail :

QR1 : quel est l'effet d'un entraînement prosodique sur le rythme de la parole en L2, et comment se compare-t-il à celui d'un entraînement à l'expression et à la compréhension orales ?

QR2a : quel est l'effet d'un entraînement prosodique sur la compréhensibilité et l'accentuation, et comment se compare-t-il à celui d'un entraînement à l'expression et à la compréhension orales ?
QR2b : les scores de compréhensibilité et d'accentedness sont-ils corrélés ?

QR3 : quel est l'effet d'un entraînement prosodique sur les capacités de segmentation, et comment se compare-t-il à celui d'un entraînement à l'expression orale et à la compréhension ?

L'opérationnalisation du rythme de la parole suit notre vision du rythme de la parole comme un concept à multiples facettes (voir chapitre I). Les mesures comprennent des aspects des quatre niveaux d'analyse : micro-niveau, méso-niveau, macro-niveau, fluence.

Chapitre VI

Ce dernier chapitre présente les résultats et discussions s'y rapportant.

Résultats sur le rythme de la parole :

Afin de saisir les tendances générales au sein de chaque groupe, nous avons résumé les résultats ci-dessous. Pour chaque mesure, nous indiquons la tendance montrée par la majorité des participants au sein de chaque groupe, basée sur les valeurs moyennes et exprimée dans l'unité d'origine de chaque mesure. Entre parenthèses, nous indiquons le nombre de participants qui suivent la tendance principale par rapport au total des participants de ce groupe. A titre d'exemple, dans le groupe Prosodie, la tendance principale pour le taux d'articulation est une augmentation de 0,16 syllabe par seconde en moyenne, pour trois participants sur quatre. Lorsqu'aucune tendance ne peut être identifiée parce que chaque participant suit une tendance différente, nous utilisons le terme dispersé.

MEASURE	ORAL GROUP	PROSODY GROUP
nPVI	Dispersé	Baisse de 3.1 (3/4)
Ratio Durée Syllable NAC/FA	Augmentation de 0.25 (2/3)	Baisse de 0.15 (4/4)
IPU quantity	Augmentation de 3.1 (2/3)	Augmentation de 3.6 (3/4)
IPU duration	Augmentation de 0.262 (2/3)	Baisse de 0.189 (4/4)
External pauses duration	Baisse de 0.105 (3/3)	Augmentation de 0.093 (3/4)
Articulation rate	Augmentation de 0.08 (3/3)	Augmentation de 0.16 (3/4)
Disfluencies quantity	Augmentation de 8.6 (2/3)	Baisse de 6.9 (3/4)
Voiced Pauses duration	Augmentation de 0.035 (2/3)	Baisse de 0.03 (3/4)
Filled external pauses proportion	Dispersé	Augmentation de 6.76% (3/4)

Nous allons maintenant discuter des résultats présentés par rapport à notre première question de recherche :

QR1 : Quel est l'effet d'un entraînement prosodique sur le rythme de la parole en L2, et comment se compare-t-il à celui d'un entraînement à l'expression et à la compréhension orales ?

Evolution T1-T2 du micro-niveau de rythme

Nous avons vu que les scores nPVI en L2 se situent généralement entre les valeurs de la L1 et de la langue maternelle cible (Carter, 2005 ; Ordin et al., 2011 ; Ordin & Polyanskaya, 2015). Comme les scores nPVI de l'anglais L1 sont systématiquement plus élevés que ceux du français L1, nous nous attendions à ce que les participants diminuent progressivement leurs scores nPVI au fur et à mesure qu'ils se rapprochaient du rythme syllabique du français.

Nos résultats montrent qu'ils suivent en effet la tendance attendue de réduction de leurs scores nPVI - c'est-à-dire la variabilité temporelle globale des syllabes. Cependant, la différence entre T1 et T2 est très faible et pourrait ne pas être significative. A l'inverse, les scores du groupe Oral sont très inconstants dans le sens de l'évolution entre T1 et T2. On pourrait s'arrêter là et conclure que l'entraînement prosodique a conduit les participants de ce groupe à suivre la tendance attendue associée à l'amélioration, alors que le groupe Oral est trop inconsistant pour conclure quoi que ce soit.

Mais en regardant les scores des deux groupes à T2, une convergence apparaît. Tous les participants à T2 ont obtenu des scores centrés autour de 50 (+/-3) quels que soient leurs scores à T1, et quelle que soit la distance séparant leur score à T1 de 50. Cela suggère que les deux types d'enseignement ont eu pour effet d'aligner les participants sur un score d'interlangue commun, qui se situe à mi-chemin entre la L1 anglaise (autour de 60) et la L1 française (autour de 40). Étant donné que le type de formation ne semble pas faire de différence, il se peut que l'exposition et la pratique dont les deux groupes ont bénéficié au cours de la formation - bien que sous des formes différentes - puissent expliquer ce résultat. Il est également possible que la familiarité avec la tâche du post-test à T2 ait eu un effet d'homogénéisation, mais cet effet serait limité au nPVI puisqu'il n'est pas observé pour d'autres mesures.

Néanmoins, nous devons interpréter les résultats de cette mesure avec prudence car le nPVI n'a pas été utilisé souvent dans des études expérimentales similaires, et surtout, pour des locuteurs passant d'une langue accentuelle à une langue syllabique. De plus, un certain nombre d'études ont montré la sensibilité de cette mesure à des facteurs tels que la variation interindividuelle, le matériel vocal et la fréquence des mots (Arvaniti, 2012 ; Harris & Gries, 2011 ; Wiget et al., 2010).

Evolution T1-T2 du niveau méso du rythme

En ce qui concerne la durée des syllabes non accentuées et accentuées, nos résultats montrent que si les syllabes non accentuées restent stables dans le temps, la durée des syllabes accentuées diminue dans le groupe Prosodie alors qu'elle a tendance à augmenter dans le groupe Oral. Ceci est en accord avec Pasdeloup (2004) qui a également trouvé que les syllabes accentuées sont plus susceptibles de varier alors que les syllabes non accentuées restent stables.

Par conséquent, le rapport temporel entre les syllabes accentuées et les syllabes non accentuées diminue dans le groupe Prosodie et augmente dans le groupe Oral. Cela contredit les résultats d'Alazard (2013) qui a constaté qu'après une formation à la MVT, le rapport entre les apprenants débutants en anglais L1 et en français L2 augmentait (mais sur un style de parole différent et avec un post-test immédiat).

Il y a très peu de données disponibles dans la littérature concernant ce ratio en anglais et en français L1. Delattre (1966) a trouvé un ratio de 1,6 pour l'anglais L1 et un ratio de 1,78 pour le français L1, Astésano (2001) a également trouvé un ratio de 1,7 pour le français L1. D'après ces valeurs, on s'attendrait plutôt à une tendance à l'augmentation de T1 à T2 comme signe d'amélioration. Cependant, nous avons constaté que le ratio à T1 dans les deux groupes était déjà de 1,8 en moyenne, ce qui est légèrement supérieur au ratio de Delattre et Astésano pour le français L1.

Cela pourrait être dû à la tendance à la sur-articulation signalée dans la littérature sur les L2 (Barry, 2007 ; Gut, 2003), mais cette littérature concerne l'acquisition d'un rythme accentuel, et à notre connaissance, nous ne disposons pas de données empiriques sur les locuteurs d'une langue accentuelle apprenant une langue syllabique sur cet aspect. Néanmoins, si la sur-articulation peut expliquer le ratio élevé à T1, la diminution du ratio à T2 dans le groupe Prosodie peut être considérée comme une amélioration, puisque tous les participants atteignent un ratio L1-français similaire de 1,7 en moyenne après la formation. A l'inverse, le groupe Oral augmente encore plus son ratio à T2. Cela suggère que le type de formation pourrait avoir un impact différent sur ce ratio, et que seule la formation prosodique a des résultats positifs.

De plus, nous avons vu que l'évolution de ce ratio dans le temps est partiellement indépendante de l'effet du taux d'articulation (voir Pasdeloup 2004). Enfin, les résultats du modèle mixte indiquent que la différence entre les groupes Prosodie et Oral concernant la durée des syllabes accentuées à T2 est significative (voir Annexe).

Les résultats sur le nPVI et le ratio de syllabes dans le groupe Prosodie sont pour la plupart cohérents, puisqu'ils indiquent tous deux une réduction de la variabilité de durée syllabique. Cette tendance est conforme à une variété interlangue évoluant vers des modèles rythmiques du français L1, que le groupe Oral ne suit pas.

Evolution T1-T2 du niveau macro du rythme et de la fluence

Les résultats concernant le niveau macro des mesures de rythme et de fluidité étant interdépendants, nous les examinerons ensemble. En ce qui concerne ces deux aspects de la parole, le groupe Prosodie et le groupe Oral présentent des tendances divergentes. Dans le groupe Prosodie, la plupart des participants produisent des IPU plus nombreuses et plus courtes, des pauses externes plus longues et plus souvent remplies d'hésitations, une augmentation du taux d'articulation, moins de disfluences et des pauses remplies plus courtes après l'entraînement. Des études antérieures ont indiqué que l'amélioration de la fluidité/de la compétence tend à être corrélée à des IPU moins nombreuses et plus longues, et à un taux d'articulation plus rapide (Judkins et al., 2022 ; Kormos & Dénes, 2004 ; Lennon, 1990 ; Préfontaine et al., 2016 ; Saito et al., 2018 ; Tavakoli et al., 2020).

Par conséquent, nous pourrions interpréter l'augmentation de la quantité d'IPU et la réduction de la durée de l'IPU comme allant à l'encontre d'une amélioration. Cependant, l'augmentation du taux d'articulation ainsi que la diminution des disfluences et des pauses remplies plus courtes au sein de l'IPU expliquent la diminution de la durée de l'IPU de manière positive. En fait, une diminution du nombre de disfluences a été associée à une amélioration en français L2 (Trofimovich et al., 2017). Cela contraste avec les résultats d'études portant sur d'autres langues cibles, qui ont rapporté des résultats mitigés sur la relation entre l'amélioration de la fluence/de la compétence et le nombre de disfluences (Kormos & Dénes, 2004 ; Saito et al., 2018). Cependant, Baker-Smemoe et al. (2014) ont montré que les résultats concernant le développement de la fluence dans une paire de langues ne peuvent pas être généralisés à d'autres paires de langues.

Il est intéressant de noter que la durée des pauses externes a augmenté, en relation avec la proportion accrue de pauses externes remplies d'hésitations. Alors que plusieurs études associent des pauses externes plus courtes à une amélioration de la fluence ou de la compétence (Bosker et al., 2013 ; Cucchiarini et al., 2002 ; Lennon, 1990 ; Suzuki & Kormos, 2020 ; Tavakoli et al., 2020), Préfontaine et al. (2016) ont constaté que plus les pauses étaient longues, plus le score de fluence était élevé dans le cas du français L2. Cela est logique à la lumière des résultats concernant le modèle de pause en français L1 (Grosjean & Deschamps, 1975 ; Judkins et al., 2022), qui ont montré que les locuteurs natifs français font moins de pauses externes, mais plus longues, que les locuteurs anglais L1.

En tenant compte des spécificités du modèle du français L1 en ce qui concerne le macro-niveau du rythme et de la fluidité perçue, les résultats globaux du groupe Prosodie confirment une amélioration puisque leurs productions se rapprochent des schémas du français L1. De plus, l'amélioration du groupe

Prosodie semble indiquer que leur formation a contribué à améliorer leurs processus de programmation de la parole, comme le suggèrent les théories sur le lien entre la fluence de l'énoncé et la fluence cognitive (Goldman-Eisler, 1968 ; Segalowitz, 2010).

A l'inverse, le groupe Oral présente des tendances globalement opposées, à l'exception du taux d'articulation qui augmente également à T2, bien que dans une proportion moindre que le groupe Prosodie. Le groupe oral a tendance à produire des IPU plus nombreuses et plus longues, mais il augmente également le nombre de disfluences à l'intérieur de celles-ci, ainsi que la durée des pauses remplies, ce qui explique probablement l'augmentation de la durée de l'IPU, plutôt que la construction de syntagme plus élaborés. En outre, la durée des pauses externes diminue, mais la proportion de pauses externes avec hésitations diffère d'un participant à l'autre.

Par conséquent, compte tenu des schémas du français L1 exposés ci-dessus, le groupe Oral ne semble pas s'améliorer globalement après l'entraînement en ce qui concerne le rythme au niveau macro et la fluence.

Variabilité T1-T2 en L1:

L'objectif de cette section était de compléter le tableau de nos résultats en L2 en examinant les données en L1 et d'aborder une question méthodologique concernant la variabilité induite par le temps, intrinsèque aux études longitudinales. La simple exploration de nos données met clairement en évidence la variation non négligeable entre les temps en L1. Elle soulève la question de la fiabilité avec laquelle nous pouvons attribuer le changement T1-T2 à un traitement, compte tenu de la variabilité nécessaire induite par un tel modèle expérimental.

Comme De Jong et al. (2015) qui proposent deux façons différentes d'étudier la fluence en L2 tout en excluant l'influence des schémas de fluence en L1 sur les données, il serait intéressant de trouver un moyen d'éliminer la variabilité T1-T2 des facteurs indépendants de l'entraînement, afin d'isoler l'effet de l'entraînement uniquement. La tâche est complexe car la variabilité T1-T2 en L1 et L2 peut partager certains facteurs, mais aussi des facteurs différents spécifiques à chaque langue.

À notre connaissance, cette question n'a pas été étudiée et des recherches dans ce domaine seraient les bienvenues pour acquérir des connaissances sur la variabilité temporelle au sein d'un même sujet et pour affiner la méthodologie et les résultats de l'approche longitudinale.

Une façon d'aborder cette question à l'avenir serait d'emprunter la voie des tests statistiques, mais cela n'est pas nécessairement compatible avec les designs expérimentaux axés sur l'écologie tels que le nôtre. Néanmoins, une autre façon de découvrir si les changements T1-T2 en L2 sont pertinents et peuvent être liés au type de traitement consiste à utiliser des mesures perceptives telles que les évaluations de la compréhensibilité et de l'accentedness. La section suivante présente nos résultats sur les mesures perceptives et discute des liens avec nos résultats sur les mesures acoustiques.

Compréhensibilité & accentedness :

Nos résultats sont en accord avec les études précédentes qui ont montré l'importance des aspects suprasegmentaux sur la compréhensibilité et d' accentedness. Plusieurs méta-analyses soutiennent l'effet positif de l'enseignement de la prononciation sur les évaluations globales de la compréhensibilité et d'accentedness (Lee et al., 2015 ; Saito, 2012 ; Saito & Plonsky, 2019).

En outre, des études antérieures ont également souligné que l'enseignement suprasegmental aidait les apprenants à améliorer leur accentedness et leur compréhensibilité plus que l'absence d'enseignement explicite de la prononciation - c'est-à-dire des exercices d'expression orale et d'écoute tels que l'entraînement suivi par notre groupe Oral - en particulier sur la parole spontanée (Derwing et al., 1998 ; Gordon & Darcy, 2016 ; R. Zhang & Yuan, 2020).

Il est toutefois surprenant que le groupe Oral ne se soit pas amélioré du tout, et qu'il ait même empiré après la formation, en ce qui concerne les deux concepts. Pourtant, des résultats similaires ont été constatés par le passé (Derwing et al., 1998 ; Gordon & Darcy, 2016).

Néanmoins, nous devons reconnaître que les deux types d'enseignement ne diffèrent pas seulement dans l'orientation de l'enseignement, mais aussi dans les outils et les techniques utilisés en classe. Nous avons vu au chapitre IV que l'utilisation de gestes et d'exercices musicaux renforce les compétences phonologiques des apprenants de L2 (Baills, Santiago, et al., 2022 ; Gluhareva & Prieto, 2017 ; Good et al., 2015 ; Y. Zhang et al., 2020, 2023, à paraître). Dans notre étude, seul l'entraînement prosodique comprenait l'utilisation d'activités gestuelles et musicales, nous ne sommes donc pas en mesure de déterminer les effets différentiels de l'orientation de l'enseignement et de la nature multimodale de l'entraînement.

Résultats sur la tâche de segmentation :

Nos résultats indiquent que la formation prosodique n'a pas aidé les participants à s'améliorer dans la tâche de segmentation. Cela va à l'encontre de notre prédiction et des études précédentes qui ont soutenu l'effet positif d'une formation sur le suprasegmental sur les compétences d'écoute (Kissling, 2018 ; Luu et al., 2021 ; McAndrews, 2023 ; Yenkimaleki et al., 2023). La plupart de ces études incluaient un groupe contrôlé qui recevait des exercices de compréhension orale, et ces groupes se sont également améliorés après la formation, bien que dans une proportion moindre que les groupes formés sur les éléments suprasegmentaux. Par conséquent, l'amélioration - même si elle n'est pas significative - montrée par le groupe Oral est cohérente avec cela au moins.

De plus, les études citées ci-dessus rapportent systématiquement l'effet supérieur de l'enseignement basé sur des exercices de perception par rapport aux exercices de production. Même si l'entraînement prosodique comprenait les deux types d'exercices, le groupe Oral a été plus souvent exposé à des documents audios authentiques. Par conséquent, cette différence pourrait jouer un rôle dans la meilleure performance du groupe Oral.

Cependant, le principal problème dans la comparaison entre le groupe Oral et le groupe Prosodie concerne leur différence significative de niveau de performance à T1. Le groupe Prosodie ayant déjà atteint des scores autour de 60 %, sa marge de progression a été réduite par rapport au groupe Oral qui a commencé avec des scores inférieurs à 50 %.

En examinant le profil des participants dans chaque groupe, nous sommes parvenus à la même conclusion que les groupes sont relativement équivalents en termes de durée de résidence et d'utilisation hebdomadaire du français. Le seul cas aberrant est celui de B31 qui déclare utiliser le français 100 % du temps dans son environnement scolaire/professionnel et qui s'est attribué la note la plus élevée en matière de compétence à l'écoute. Ceci est cohérent avec sa performance à T1 qui était la plus forte de tous les participants.

Afin d'éviter un tel déséquilibre entre les groupes, il aurait été idéal de répartir les participants dans les groupes en fonction de leurs performances T1. Cependant nous n'avons pas eu ce luxe.

Quoi qu'il en soit, il est possible que l'absence d'évolution dans le groupe Prosodie soit due à un effet plafond. Cependant, les résultats de Charles et al. (2015) - qui ont inspiré la conception de notre tâche de segmentation - montrent que leurs participants (L1 chinois - L2 anglais intermédiaire-avancé) obtiennent un score de 66% à T1, et atteignent tout de même 80% lors d'un post-test.

Pourtant, nos participants sont à un niveau pré-intermédiaire. Il est possible que la progression des compétences de segmentation varie en fonction du niveau de compétence. Malheureusement, à notre connaissance, cela n'a pas encore été étudié. A ce jour, la segmentation de la parole en L2 reste un domaine sous-étudié, et davantage de données sont nécessaires pour mieux comprendre comment l'enseignement pourrait aider les apprenants à s'améliorer dans ce domaine.

Discussion et conclusion:

L'objectif de notre recherche était d'évaluer l'effet d'une formation prosodique par rapport à une formation l'expression et compréhension orale, sur le rythme de la parole, les évaluations de compréhensibilité et d'accentedness, et les capacités de segmentation des apprenants L1-anglais du français L2 à un niveau de compétence pré-intermédiaire.

L'impact positif d'un enseignement centré sur les aspects suprasegmentaux a déjà été démontré sur les évaluations globales de compréhensibilité et accentedness (e.g., Derwing et al., 1998 ; Gordon & Darcy, 2016), moins sur les mesures acoustiques du rythme de la parole (e.g., Trofimovich et al., 2017), et encore plus rarement sur les capacités de segmentation (mais François et al., 2013 ; Luu et al., 2021 par exemple).

En outre, des méta-analyses d'études sur l'enseignement de la prononciation ont conclu que les modèles devraient plus souvent inclure d'autres langues que l'anglais L2, un échantillon de parole spontanée, un post-test différé, et l'association de notations globales et de mesures acoustiques (Lee et al., 2015 ; Saito, 2012 ; Saito & Plonsky, 2019).

Enfin, il a été démontré que l'utilisation d'activités incarnées et musicales renforçait l'efficacité d'une formation axée sur les aspects prosodiques (Baills, Santiago, et al., 2022 ; Gluhareva & Prieto, 2017 ; Good et al., 2015 ; Y. Zhang et al., 2020, 2023, à paraître).

Compte tenu de tout ce qui précède, nous avons conçu un protocole expérimental, animée par la volonté de mettre en évidence la validité écologique de l'étude. Ainsi, les conditions des phases de test et des formations ont été créées pour être aussi proches conditions naturelles que possible. Les participants ont effectué les phases de test dans le confort de leur domicile, et les sessions de formation se sont déroulées dans les locaux de l'université et ont ressemblé à des cours de français L2 ordinaires, bien qu'avec un nombre limité d'étudiants. Par ailleurs, les échantillons de discours analysés étaient de nature spontanée, et le post-test a été différé (une semaine après la fin de la formation).

Globalement, nos résultats montrent que l'entraînement prosodique (groupe Prosodie) a conduit à une amélioration du rythme de la parole et des scores de compréhensibilité et d' accentedness des apprenants, alors que ce n'est pas le cas pour l'entraînement à l'expression et compréhension orale (groupe Oral). Cependant, en ce qui concerne les capacités de segmentation, seul le groupe Oral s'est amélioré après la formation. Cependant, la différence entre les groupes au pré-test rend l'interprétation de ces résultats difficile

.

Outre les principales questions de recherche, nous avons également montré l'importance de prendre en compte les schémas rythmiques de la langue cible dans l'interprétation de ce qui constitue ou non une amélioration. Étant donné que la grande majorité des études sur le rythme et la fluence en L2 se concentrent sur l'acquisition de l'anglais L2, ce qui est généralement considéré comme une amélioration est en fait principalement pertinent pour l'anglais en tant que langue cible. Dans le cas du français L2, certaines caractéristiques qui ont été associées à de moins bonnes évaluations globales dans la plupart des études sur la fluence (par exemple, des pauses externes plus longues) sont en fait évaluées positivement par les juges du français L1 (Préfontaine et al., 2016).

De plus, nous avons également confronté nos données L2 aux données L1 des participants. Ce faisant, nous nous sommes interrogés sur l'influence de la variabilité T1-T2 intrinsèque à un modèle longitudinal sur l'interprétation de l'évolution T1-T2. Nous avons constaté que la variabilité est souvent plus importante en L1 qu'en L2, en particulier pour les mesures de macro-niveau et de fluence, ce qui met en évidence l'influence de facteurs intra-sujet autres que la langue et le niveau de maîtrise de celle-ci. Cela confirme la nécessité d'associer des mesures perceptives à des mesures acoustiques afin d'évaluer la fiabilité de l'attribution des changements T1-T2 à la formation.

Perspectives

Les perspectives qui se dégagent de ce travail sont aussi passionnantes que nombreuses. D'une part, les données vocales recueillies pour cette étude, ainsi que le corpus B-FREN3 issu de recherches antérieures (Drouillet et al., 2023), n'ont pas encore été explorés en profondeur. D'autre part, des études complémentaires à celle présentée ici permettraient de confronter nos résultats et/ou de tester d'autres conditions d'entraînement et de test.

Tout d'abord, les échantillons de parole recueillis pour la présente étude comprennent une tâche de lecture. Il a été démontré que dans des études similaires sur l'enseignement de la prononciation, les tâches contrôlées tendent à produire des résultats plus robustes que les tâches de parole libre (Lee et al., 2015 ; Saito, 2012 ; Saito & Plonsky, 2019). Une analyse similaire combinant des mesures acoustiques et des jugements perceptifs sur les échantillons de parole lue permettrait a) de vérifier si un effet de la formation est constaté sur ce style de parole - ce qui renforcerait les résultats obtenus sur la parole spontanée, et b) d'apporter des connaissances supplémentaires sur la différence entre la parole lue et la parole libre pour l'analyse de l'impact d'une formation à la prononciation. Par exemple, Kennedy et al. (2017) ont constaté que les mesures de l'accentuation et de la compréhensibilité n'étaient pas associées aux mêmes corrélats linguistiques dans leur tâche de lecture par rapport à la tâche narrative. La réalisation d'une comparaison similaire entre les tâches apporterait des connaissances supplémentaires sur ce sujet en français L2.

De plus, dans l'étude présentée ici, nous avons limité notre analyse du niveau méso du rythme à la durée et à la relation entre les syllabes accentuées et non accentuées, en ne considérant que l'accent final primaire du français. L'analyse pourrait être complétée par des données concernant les accents initiaux, les accents nucléaires de phrase et l'utilisation de f0 dans le marquage de la proéminence.

Outre l'effet des entraînements sur ces aspects prosodiques, nous pourrions par exemple étudier la fréquence des accents initiaux et leur réalisation acoustique en français L2, en relation avec nos résultats sur les accents finaux. Ce faisant, nous pourrions tester l'hypothèse de l'interférence prosodique-apprentissage (PLIH) proposée par Trembley et al. (2016) et inspirée par le SLM de Flege (1995) et le PAM-L2 de Best (1995). L'hypothèse postule qu'une caractéristique prosodique sera plus difficile à acquérir lorsqu'elle est presque identique mais différente d'une caractéristique L1, que si elle est complètement différente. L'accent initial français et l'accent de mot anglais ont des caractéristiques communes : ils sont tous deux réalisés sur une syllabe initiale et avec une augmentation de f0. Cependant, l'ampleur de la montée de f0 tend à être modérée en français alors qu'elle est plus prononcée en anglais (Astésano, 2001 ; Jun & Fougeron, 2000 ; Lieberman, 1960). En outre, alors que l'accentuation des mots en anglais est déterminée lexicalement, l'accent initial en français est déterminé rythmiquement, et c'est aussi un marqueur de frontières gauche au niveau du mot prosodique (Astésano, 2017).

Par conséquent, selon la PLIH, l'acquisition de l'accent initial en français devrait être plus difficile que celle de l'accent final, qui diffère grandement des modèles de l'anglais L1. Cette étude contribuerait à mieux comprendre l'acquisition du rythme en L2 de plusieurs façons : elle permettrait de tester la pertinence de la PLIH sur la production de la parole plutôt que sur la perception ; et elle constituerait l'une des rares contributions sur l'accent initial en français L2.

Le corpus B-FREN3 (Drouillet et al., 2023) est également une source précieuse de données puisqu'il comprend des échantillons de parole bidirectionnelle en français, en anglais, en L1 et en L2 - il comprend également de la parole libre et de la parole lue. L'extraction de mesures supplémentaires du rythme de la parole sur ce corpus permettrait une comparaison perspicace de la production L1-L2 de et vers les deux langues, ce qui à son tour rendrait possible l'observation du transfert L1 par rapport aux modèles universels de rythme de la parole L2.

Afin de renforcer la fiabilité de nos résultats, et puisque le matériel expérimental est prêt à être utilisé, il serait intéressant de reproduire l'étude et d'augmenter la taille de l'échantillon. Cela permettrait d'utiliser des tests statistiques sur les mesures acoustiques et d'accroître globalement la robustesse des conclusions. Des données supplémentaires permettraient également de tester les corrélations entre les mesures des différents niveaux de rythme de la parole, et donc de mieux comprendre la relation entre eux.

Nos résultats suggèrent par exemple une relation entre le nPVI et le rapport entre syllabes accentuées et syllabes non accentuées. Ceci n'est pas surprenant étant donné que l'accent primaire en français est de nature durative. Cependant, cette observation illustre un lien étroit entre les deux mesures à prendre en compte pour le français en particulier. Les disfluences et le taux d'articulation ont également un impact sur l'IPU et les pauses externes. Une exploration plus approfondie des liens entre chaque niveau permettrait une cartographie plus précise de ces interdépendances.

Une modalité d'entraînement supplémentaire - entraînement prosodique sans geste/musique - pourrait être ajoutée afin d'isoler l'effet de la nature multimodale de l'entraînement prosodique. Des études antérieures ont soutenu les avantages supérieurs de l'enseignement multimodal (par exemple, Alazard, 2010 ; Baills et al., 2022), mais aussi de l'enseignement explicite sur les suprasegmentaux (par exemple, Derwing et al., 2018 ; Gordon & Darcy, 2016). Cependant, les comparaisons entre les trois modalités et sur une tâche de parole spontanée semblent plus rares. Compte tenu des recherches antérieures, nous nous attendrions à ce que les deux formations prosodiques conduisent à une amélioration plus importante que le groupe de contrôle, et à ce que la formation multimodale produise de meilleurs résultats que la formation non multimodale-prosodique.

En outre, comme nos conclusions sur l'effet des entraînements sur les capacités de segmentation sont limitées par la différence entre les groupes à T1, nous souhaitons repenser une expérience dans ce but. La segmentation de la parole en L2 et la manière d'aider les apprenants à cet égard est encore un sujet très peu étudié (Charles et al., 2015), d'autant plus en français L2.

Les observations faites sur la variabilité intertemporelle entre L1 et L2 ont piqué notre curiosité. Comme il semble que l'on manque d'informations à ce sujet, il s'agit d'un thème que nous aimerions approfondir, idéalement avec un échantillon de taille suffisante pour permettre l'utilisation de statistiques, à l'instar de ce que De Jong et al. (2015) ont proposé. Dans leur étude sur la relation entre les schémas de fluidité L1 et L2, les auteurs ont séparé la variance expliquée par les schémas L1 des mesures L2. De la même manière, nous pourrions imaginer d'éliminer la variabilité temporelle ainsi que l'influence de la L1 pour obtenir des mesures L2 « propres ». Toutefois, la question de l'origine de la variabilité temporelle intra-individuelle en L1 vs L2 doit être explorée plus avant.

Une autre idée a surgi pendant la rédaction de cette thèse. Elle concerne la relation entre le rythme de la parole et le rythme cérébral. Comme nous l'avons exposé au chapitre I, un mécanisme neuronal important est en jeu dans le traitement et la compréhension de la parole. L'entraînement neuronal au rythme de la parole entendue permet un traitement rapide et efficace, et donc l'accès à la compréhension (Peelle & Davis, 2012). Il serait fascinant de tester si un entraînement prosodique a un effet sur la capacité de l'apprenant L2 à se synchroniser sur le rythme de l'interlocuteur.

Dans l'ensemble, les études portant sur l'acquisition d'une langue autre que l'anglais, et plus spécifiquement sur les langues dites à rythme syllabique, sont encore beaucoup moins nombreuses que les études sur l'anglais L2. Afin de tester les théories actuelles sur l'acquisition de la prosodie en L2, telles que celle d'Ordin & Polyanskaya (2015), ou d'en proposer de nouvelles, davantage de données sont nécessaires sur différentes paires de langues, mélangeant les langues à temps de stress et les langues à temps de syllabe de toutes les manières possibles.

Enfin, comme mentionné dans la section précédente, l'un des objectifs de cette étude était de concevoir un cours de prosodie de français L2 en classe, avec l'idée de développer des ressources pédagogiques et des outils pour les enseignants. Nous espérons que nos efforts continus pourront aboutir à la publication d'un manuel de prosodie du français L2, dans l'esprit d'un rapprochement entre la recherche et les pratiques de classe.

Appendix 7: Related publications

Drouillet, L., Astésano, C., & Alazard-Guiu, C. (2023). Another voice for another language? The impact of language on vocal register. *Language, Interaction and Acquisition*, *14*(2), 247–279. https://doi.org/10.1075/lia.00018.dro

Judkins, L., Alazard, C., & Astésano, C. (2022). Phénomènes de groupement et pause en parole native vs non-native. *XXXIVe Journées d'Études Sur La Parole–JEP 2022*, 910–919. https://hal.science/hal-03980725/

Judkins, L., Alazard-Guiu, C., & Astésano, C. (2022). How do we chunk and pause in non-native vs native speech? *Speech Prosody 2022*, 792–796. https://hal.science/hal-03980742/

Drouillet, L., Alazard, C., & Astésano, C. (2024). Impact of Prosodic Training on Speech Rhythm in L2 French. *Pronunciation in Second Language Learning and Teaching Proceedings*, *14*(1).
https://www.iastatedigitalpress.com/psllt/article/id/17100/

Titre : Approche intégrative du rythme de la parole en L2: Enjeux pour les apprenants et impact de l'enseignement de la prosodie en français L2

Mots clés : didactique, rythme, L2, segmentation, compréhensibilité, acquisition

Résumé : Cette thèse porte sur le rythme de la parole dans le contexte de l'acquisition et de l'enseignement des L2. Une approche originale du rythme de la parole est proposée : des arguments sont donnés en faveur d'une vision intégrative réunissant les aspects phonologiques, prosodiques, temporels et de fluence. Les corrélats acoustiques sont présentés et discutés à la lumière des résultats des études sur l'acquisition de la L2. Un examen des modèles théoriques et des études empiriques met en évidence le rôle essentiel de la prosodie dans la production et la perception de la parole en L2. Les méthodes et techniques actuelles utilisées dans l'enseignement de la prononciation sont explorées et l'impact positif de l'enseignement multimodal axé sur la prosodie est souligné. Dans la deuxième partie de cette thèse, nous présentons une étude expérimentale qui vise à tester l'effet d'un enseignement de la prosodie du français L2 sur le rythme de la parole, la compréhensibilité, le degré d'accent étranger et les capacités de segmentation d'apprenants L1-anglais de niveau pré-intermédiaire. 8 participants ont été recrutés et séparés en deux groupes. Un groupe a reçu la formation à la prosodie (n=4), tandis que l'autre a suivi un cours d'expression et compréhension orale (n=4). Les cours consistaient en des sessions d'une heure et demie, deux fois par semaine, sur une période de quatre semaines. Les participants ont été testés avant et après la formation sur une tâche d'expression libre et une tâche de segmentation. Plusieurs mesures du rythme de la parole ont été extraites, et des scores de compréhensibilité et d'accentuation ont été recueillis auprès de locuteurs natifs français (n=9). Les résultats globaux montrent que le groupe Prosodie a produit des schémas rhythmiques plus proches du français L1 que le groupe Oral, et que seul le groupe Prosodie s'est amélioré en termes de compréhensibilité et de degré d'accent étranger après la formation. Cela suggère que l'enseignement multimodal axé sur les aspects prosodiques amène les apprenants à progresser davantage en termes d'expression orale, comparé à un enseignement basé sur des exercices d'expression et compréhension orale. Cependant, dans la tâche de segmentation, seul le groupe oral s'est amélioré, mais une différence importante entre les groupes au pré-test rend l'interprétation de ce résultat difficile. Les implications des résultats en termes de pratique pédagogique, de décisions méthodologiques et de niveau d'analyse du rythme de la parole sont discutées.

Title: AN INTEGRATIVE APPROACH OF L2 SPEECH RHYTHM: LEARNERS CHALLENGES, AND IMPACT OF PROSODIC INSTRUCTION IN L2 FRENCH

Key words: language teaching, Rhythm, L2, segmentation, comprehensibility/accentedness, acquisition

Abstract: This dissertation focuses on speech rhythm in the context of L2 acquisition and teaching. A novel approach of speech rhythm is proposed: arguments are given in favor of an integrative view reuniting phonological, prosodic, temporal, and fluency aspects. Acoustic correlates are presented and discussed in light of findings from L2 acquisition studies. Through a review of theoretical models and empirical studies, the essential role of prosody in L2 speech production and perception is highlighted. Current methods and techniques used in pronunciation instruction are explored and the positive impact of multimodal instruction focused on prosody is underlined. In the second part of this dissertation, we present an experimental study that aims at testing the effect of a training in L2 French prosody on the speech rhythm, comprehensibility, accentedness, and segmentation abilities of L1-English learners at a pre-intermediate level. 8 participants were recruited and separated into two groups. One group received the prosodic training (n=4), while the other followed a course with oral expression and comprehension activities but no explicit instruction on pronunciation aspects (n=4). The courses consisted of 1.5 hour sessions, twice a week, over a four week-period. Participants were tested before and after training on a free speech task and a segmentation task. Several measures of speech rhythm were extracted, and comprehensibility and accentedness scores were collected from French native speakers (n=9). Results overall show that the Prosody group produced patterns closer to L1 French than the Oral group, and only the Prosody group improved in both comprehensibility and accentedness after training. This suggests that multimodal instruction focused on prosodic aspects is more beneficial for the development of learners' speaking skills, as compared to oral expression and comprehension exercises. However, in the segmentation task, only the Oral group improved, but an important between-group difference at pretest makes the interpretation of this outcome difficult. Implications of the findings in terms of pedagogical practice, methodological decisions, and level of analysis of speech rhythm are discussed.