IMPACT OF PROSODIC TRAINING ON SPEECH RHYTHM IN L2 FRENCH

Lucie Drouillet, University Toulouse Jean Jaurès, Laboratory of NeuroPsychoLinguistics
Charlotte Alazard, University Toulouse Jean Jaurès, Laboratory of NeuroPsychoLinguistics
Corine Astésano, University Toulouse Jean Jaurès, Laboratory of NeuroPsychoLinguistics

Several second language studies show the benefits of prosodic training on fluency, rhythm, comprehensibility, and listening skills over oral comprehension and production activities. However, few approach the acquisition of French as a Foreign Language (FFL). This study tests the impact of FFL prosodic training on learners' speech rhythm. Speech rhythm is understood as a multifaceted construct that includes temporal variables, accentuation patterns, fluency and rhythm metrics. Eleven L1-English speaking learners of French were split into two groups. One group did oral production and comprehension activities and the other trained on French prosody (1.5 hours twice a week for four weeks, 12 hours total). Acoustic measures of speech rhythm were extracted using a free speech task and compared within-speaker (pretest and posttest) and between groups. Preliminary results indicate benefits from the prosodic training over oral production and comprehension activities on fluency and cohesion within Inter Pausal Units.

INTRODUCTION

Teaching pronunciation remains a marginal practice in foreign language classes due to a lack of teacher training and the courses' designs which usually allocate a small part to pronunciation (Billières, 2008; Darcy, 2018; Detey & Durand, 2021). Moreover, segmentals tend to be the main focus, while prosody is often left aside. Teaching pronunciation, especially in English as a Second Language (ESL) contexts, has benefited from a great interest from linguists and didacticians over the past 30 years. Studies have shown the fundamental role of prosody in both speech production (intelligibility and comprehensibility) and perception (speech segmentation and listening comprehension; Derwing & Munro, 2005; McAndrews, 2023). The benefits of prosodic training on fluency, rhythm, comprehensibility, and listening skills over oral comprehension and production activities have also been demonstrated (Saito & Plonsky, 2019). Pronunciation teaching materials and resources are now available for ESL (Celce-Murcia et al., 2010; Grant, 2016), and research in this area is dynamic. Conversely, few studies have focused on the acquisition of French as a Foreign Language (FFL) prosody, and teaching material is scarce (Abel, 2018).

The purpose of this study is to look at the effect of French prosodic training on learners' speech rhythm (amongst other aspects; Drouillet, in preparation). In the next section, we elaborate on our understanding of speech rhythm.

Drouillet, L., Alazard, C., & Astésano, C. (2024). Impact of Prosodic Training on Speech Rhythm in L2 French. In D. J. Olson, J. L. Sturm, O. Dmitrieva, & J. M. Levis (eds), *Proceedings of 14th Pronunciation in Second Language Learning and Teaching Conference*, (pp. 1-13). Purdue University, September 2023. https://doi.org/10.31274/psllt.17100

Speech Rhythm

Rhythm acts as the fundamental structure onto which speech is organized (Di Cristo, 2002). However, definitions of speech rhythm depend on the theoretical angle adopted. In the metrical approaches to prosody (Liberman & Prince, 1977), rhythm refers to the temporal organization of prominences and the alternance of weak and strong syllables (Astésano, 2001). Rhythm can also be understood as emerging from the phonological level through the alternation of vocalic and consonantal intervals linked to syllables' complexity and vowel reduction phenomenon. This came from the *Rhythm Class Hypothesis* (Abercrombie, 1967; Pike, 1945), for which several automatic measures were developed (Grabe & Low, 2002; Ramus et al., 1999). Rhythm can also be seen as the organization of speech into chunks and pauses. These *temporal variables* have been shown to be language-specific (Grosjean & Deschamps, 1975; Judkins et al., 2022). Finally, in studies regarding L2 speech specifically, measures of fluency are frequently used. These include speech rate, phonation time, quantity of disfluencies etc. (Kormos & Dénes, 2004), and also contribute to the speakers' speech rhythm.

The above are different ways of looking at speech rhythm, yet they are not mutually exclusive. As Arvaniti (2012) pointed out, focusing on segments' intervals can only render part of the picture. In essence, rhythm emerges from both timing and pattern and is strongly underlined by motor aspects (Baills, 2021; Fraisse, 1974; Liberman & Mattingly, 1985). As such, speech rhythm should be understood as a multifaceted construct, combining parameters belonging to different domains and levels of analysis that complement one another.

French and English also have contrasting rhythmic patterns. French is a trailer-timed language where primary accents occur in word and rhythmic group final position, marked mainly through lengthening. English is a leader-time language with accents occurring at the beginning of content words, marked by an increase in intensity and pitch and a potential lengthening (Wenk, 1985). Lastly, it has been shown that chunks - Inter Pausal Units (IPU) - tend to be longer and less numerous in French, and voiced pauses more numerous but less long than in English (Grosjean & Deschamps, 1975; Judkins et al., 2022). These contrasts were points of focus in the course created to teach French prosody to L1-English learners in this study.

Designing a Prosody Course

The course built for this study was a first attempt at creating a curriculum covering mainly rhythmical aspects. The literature regarding L2 prosody acquisition mentions two facilitating means, gesture (e.g., embodiment) and music. Gesturing while talking is an automatic phenomenon that helps complete speech and underlines speech structure (Iverson & Goldin-Meadow, 1998). Krahmer and Swerts (2007) show that prominence-aligned gesture helps prominence perception. Studies that have compared pronunciation training with or without beat-gesture show better posttest performance in the beat-gesture condition (Gluhareva & Prieto, 2017; Yuan et al., 2019). As such, gesture was used in the prosodic training in the current study to help the perception and production of the French syllabic rhythm (e.g., tapping or walking on each syllable; Berdoulat et al., 2018; inspired from the Verbo-Tonal Method, Intravaia, 2000), and also the word/rhythmic group final accent (e.g., stretching a rubber-band between hands on

the final accent to embody lengthening of the syllable [Baills, 2021], from *La Prononciation en Classe* [Briet et al., 2014]).

Furthermore, the link between music and language, and especially prosody, has been shown at the cognitive level (Besson & Schön, 2001; Patel, 2010), and there is empirical evidence on the link between musical and pronunciation abilities and training (Cason et al., 2020; Milovanov & Tervaniemi, 2011).

Study Goal

The aim of the study was to create a French prosody course for L1-English learners and test it in the classroom. With a pretest–training–posttest design and two distinct groups of learners, we compared the impacts of speaking and listening activities commonly used in FFL classes (Oral Group) with the prosody course (Prosody Group) on speech rhythm. The complete study also includes analyses of comprehensibility and accentedness, and speech segmentation abilities, but these will not be presented in this article (see Drouillet, in preparation).

Our research questions were the following:

RQ1: To what extent do the trainings impact learners' speech rhythm?
RQ2: Are there differences between the Prosody Group and the Oral Group?

METHOD

L1-English speakers (n = 11), learning French as an L2 with self-assessed levels of A2-B1 were recruited from universities located in Toulouse and on social media. Three participants were excluded from the study as one was not a native speaker of English and two others did not complete the whole training. The eight remaining participants (M_{age} = 37 years; range = 21–66; 6 females) spoke different varieties of L1 English (1 South-African, 1 American, 3 Scottish, 2 English). Linguistic background information was collected, as well as exposure to French in both formal and informal settings. Participants were split into two groups and randomly assigned to either the Oral Group (n = 4), who received meaning-focused instruction with general speaking and listening activities, or the Prosody Group (n = 4), who received form-focused instruction with rhythmical exercises mainly.

The pretest and posttest were undertaken by the participants with full autonomy, in the comfort of their homes. They received detailed instructions on the tasks to complete and were provided with an online voice recording app (https://online-voice-recorder.com/) to record themselves and save the resulting sound files. The decision to conduct the tests remotely was motivated by two primary factors. Firstly, to spare participants from traveling to the research facility. Secondly, to enhance participants' comfort during the tasks to ensure that speech samples were captured in a naturalistic setting. The pre- and posttests consisted of five tasks in the following order: reading aloud and free speech in L1, free speech and reading aloud in L2, and listen and repeat task (speech segmentation task). The texts for the reading tasks, the topics for the free speech, and the items in the listening task differed in the pre- and posttest.

The ecological validity of the data collected was privileged. Most L2 studies with similar designs use read speech and test the participants immediately after the training (Saito & Plonsky, 2019). Therefore, the findings are confined to the controlled setting and specific speech style that is reading aloud, and conclusions cannot be extended to real-life communication situations. Here, we made the choice to get as close as possible to the reality of our speakers' language use by focusing on free speech (read speech will serve as a comparison), administrating the tasks autonomously and at home, and delaying the posttest. Participants were given the instruction to send their posttests recording in the seven days following the end of the training and most were received between the fifth and seventh day. The course contents are detailed below.

Oral Group Course Content

Each session followed the same format where the teacher would introduce a topic to the participants and encourage an informal conversation to get started. This was followed by a listening comprehension exercise, which introduced relevant vocabulary. Finally, participants were asked to prepare a short presentation - by themselves or in pairs - and share it with the class. The session usually ended with a broader discussion on the subject. A new topic was introduced and developed in each session. During class, the teacher gave minimal pronunciation feedback only when necessary and did not provide any explicit rules or exercises related to pronunciation. The course was solely focused on oral practice and listening comprehension. Authentic material was used, such as extracts from radio shows and video clips, as well as activities from TV5 Monde (https://enseigner.tv5monde.com/) and French as Foreign Language textbook *L'Atelier B1* (Cocton et al., 2020).

Prosody Group Course Content

The prosody course started with a musical training to rhythm. Activities adapted from the Dalcroze Method – a holistic method using embodiment developed by a composer in the 1900s (Juntunen, 2016) – included synchronizing movements (clap or walk) to the rhythm of a tune, match a movement to the perceived accent in the melody, anticipate the coming accent in the melody. To help transition between musical rhythm and language rhythm, we used French songs with lyrics where the rhythmic pattern emphasized the final accent. As in the Verbo-Tonal Method (Intravaia, 2000), logatom sequences (the same repeated syllable such as “dadada”) were used to isolate prosody from meaning. All sessions included perception and production activities and followed a progression from controlled exercises toward an application in spontaneous speech. Table 1 below summarizes the sessions’ content with examples of activities.

Table 1

Prosody Course Content Summary and Sources

Aspects in focus	**Activity examples**
Musical entrainment (2 X 1.5hour)	Dalcroze method exercises - Synchronising walk / clap to the rhythm of a tune - Spot accents in the melody + synchronise a movement - Anticipate and jump on the final accent of a sequence
Syllabic rhythm & final accent (2 X 1.5hour)	- Matching logatom sequences with speech sequences - Repeat dialogue and tap each syllables - Spot the accentuated syllable in songs/speech samples - Reproduce final accent on words with a matching movement
Rhythmic group (RG) (2 X 1.5hour)	- Identifying RG in speech samples - Liaisons & linking within the RG - Production exercises
Intonation & pauses (2 X 1.5hour)	- Match a movement onto perceived intonation contours - Spot pauses placement & insert pauses in a dialogue
Sources: Dalcroze Method exercises kindly provided by F. Baills (Zhang et al., in press); *La prononciation en classe* (Briet et al., 2014); *Les 500 exercices de Phonétique* (Abry-Deffayet & Chalaron, 2009); online platform Fonetix.org (Berdoulat et al., 2018).	

The length and volume of the training were decided based on feasibility constraints. The training had to be substantial enough to cover the desired material but not too long to minimize participant loss during the process. We settled on a 12-hour training, divided into 8 sessions of 90 minutes twice a week, spread over 4 consecutive weeks. All sessions took place on the university premises, and both groups attended their sessions during the same 4 weeks. The experimenter taught the sessions for both groups and spoke only French in class to ensure equivalent input.

Data Collection and Extraction

After the pre- and post-test, participants uploaded their recorded speech samples onto a secured online platform provided by the university. In order to extract speech rhythm measures, the samples were first semi-automatically segmented into syllables using the Praat Plug-in EasyAlign (Goldman, 2011). The alignment was manually corrected, and files were manually annotated in Praat (Boersma & Weenink, 2022). Annotations included Inter Pausal Units (IPU; uninterrupted run of speech delimited by silent pauses of at least 250ms, Kormos & Dénes, 2004) and disfluencies (voiced pauses, false starts, lengthenings, code-switching). For accentuation-related measures, the three authors individually annotated the perceived prominences for initial accents, final accents, and accents transferred from English. The first author centralized the three annotations, and where they diverged, the three authors met, listened again, and discussed to find a consensus.

RESULTS

Table 2 below lists the selection of measures presented in this section, with a brief definition.

Table 2

Selection of Measures Presented and their Definitions

Measure	**Definition**
IPU quantity	IPU total*100/total time IPU (Derwing et al., 2009)
IPU duration	Duration of IPU in seconds
External Pauses duration	IPU-external pauses of at least 250ms of silence
External Voiced Pauses quantity	IPU-external voiced pauses between silences >250ms
Disfluency quantity	IPU-internal voiced pauses, false starts, lengthenings
nPVI	Normalized Pairwise Variability Index (Grabe & Low, 2002)

The speech samples of one participant in the Oral Group had to be excluded from the analysis as it appeared that they were reading instead of talking spontaneously in the free speech task. As such, the following results show data from 3 participants in the Oral Group and 4 in the Prosody Group. In Figures 1–5, the data is presented by group (unless mentioned otherwise), and the vertical lines show the distribution range. Given the small sample size, our analysis relies on descriptive statistics only.

IPU and External Pauses

Figures 1 and 2 show the IPU quantity and duration for each group at T1 (pretest) and T2 (posttest).

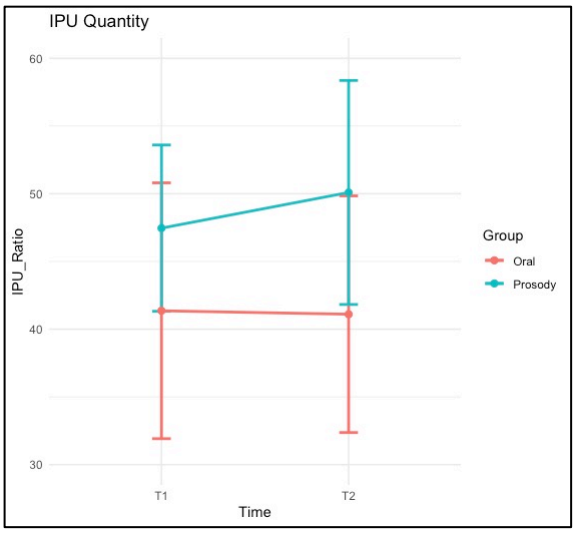

Figure 1. IPU quantity per Time and Group.

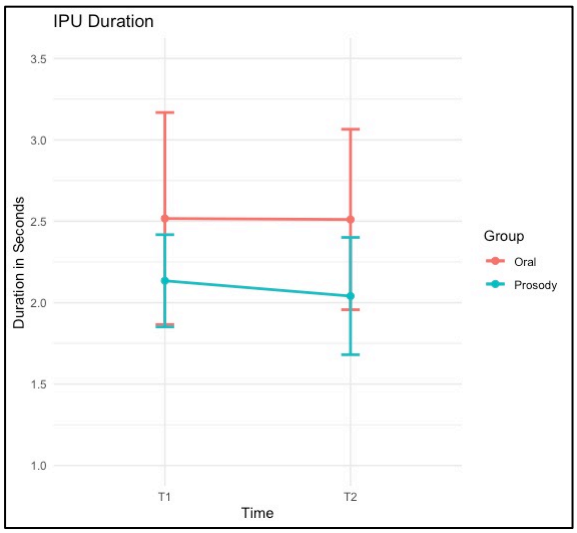

Figure 2. IPU duration in seconds per Time and Group.

In most studies looking at the evolution of IPU and pauses across L2 proficiency levels, speakers tend to make fewer and longer IPU as they become more proficient. However, here, the two groups behave differently: the Oral Group stays constant both in terms of quantity ($M_{quantT1}$ = 41.3, *SD* = 9.4; $M_{quantT2}$ = 41.1 *SD* = 8.7) and duration (M_{durT1} = 2.56, *SD* = 1.85; M_{durT2} = 2.41, *SD* = 1.48), whereas the Prosody Group tends to make more ($M_{quantT1}$=47.4, *SD* = 6.1; $M_{quantT2}$ = 50.1, *SD* = 8.2) and shorter IPU in T2 (M_{durT1} = 2.11, *SD* =1.34; M_{durT2} = 2.02, *SD* = 1.33). Figure 3 below shows the evolution of external pauses duration (pauses in between IPU).

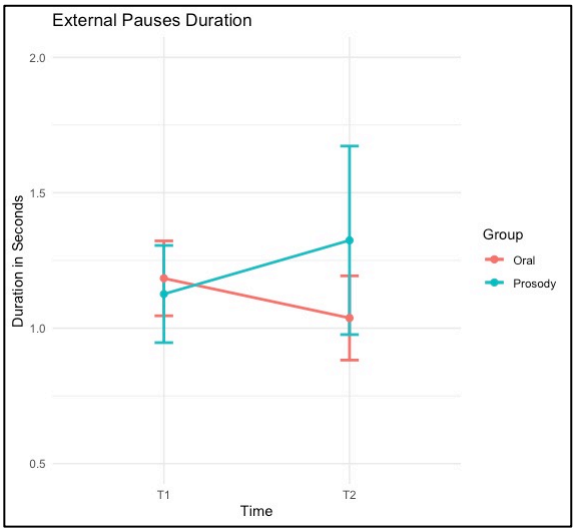

Figure 3. External pauses duration in seconds per Time and Group.

Here too, the two groups are going in the opposite direction. External pauses get shorter in the Oral Group (M_{durT1} = 1.18, *SD* = 0.98; M_{durT2} = 1.03, *SD* = 0.91) and longer in the Prosody Group (M_{durT1} = 1.13, *SD* =1.05; M_{durT2} = 1.28, *SD* = 1.23).

Data on disfluencies, presented in Figures 4 and 5 below, help to explain these tendencies.

Disfluency and External Voiced Pauses

Disfluencies include voiced pauses, false starts, and lengthenings, all occurring within IPU. Figure 4 shows more disfluencies in T2 for the Oral Group (M_{T1} = 58.9, *SD* = 9.3; M_{T2} = 61.8, *SD* = 1.8), whereas the Prosody Group shows a decrease (M_{T1} = 53.3, *SD* 11.1; M_{T2} = 50, *SD* = 8.7). This helps explain why the IPUs are shorter in T2 for the Prosody Group. Figure 5 shows the quantity of external voiced pauses (EVP, outside IPU). While it appears that the Prosody Group tend to produce slightly more EVP in T2 (M_{T1} = 5.8, *SD* = 2.6; M_{T2} = 6.2, *SD* = 3.6) and the Oral group less (M_{T1} = 7.9, *SD* = 4; M_{T2} = 6.3, *SD* = 1.9), the important and overlapping distributions of the two groups prevent from drawing any firm conclusion.

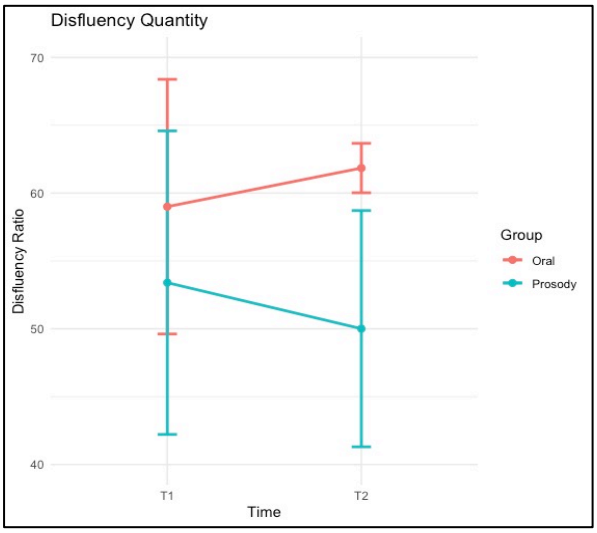

Figure 4. Disfluency quantity per Time and Group (nb*100/total time IPU+external pauses; Bosker et al., 2013).

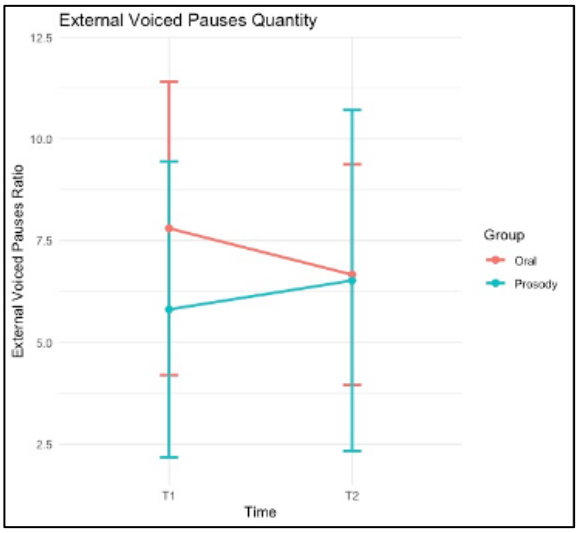

Figure 5. External voiced pauses quantity per Time and Group (nb*100/total time IPU+external pauses; Bosker et al., 2013).

Normalized Pairwise Variability Index

We computed the normalized Pairwise Variability Index (nPVI) using the Grabe and Low (2002) equation. The nPVI corresponds to the durational difference between two successive syllables, divided by the mean duration of the pair (to factor in the local speech rate), and finally combines the results of all pairs into an average multiplied by 100. The lower the score, the lower the variability, which is usually associated with a syllable-timed rhythm; the higher the scores, the greater variability, associated with a stress-timed rhythm pattern (Grabe & Low, 2002). The nPVI score therefore reflects the rhythmical difference between French - considered syllable-timed - and English - considered stress-timed. As a reference, previous studies find that nPVI in L1-French is around 43 and 57 in L1-English (Grabe & Low, 2002; Mairano & Romano, 2011). Acquisition of French rhythm by L1-English speakers would be illustrated by a lower nPVI in L2-French than in L1-English. Since the Prosody Group received specific training on the syllabic structure and rhythm of French, we expect the nPVI scores to decrease in T2, in greater

proportion than in the Oral Group. Figure 6 below shows that there is a stronger decreasing tendency in the Prosody Group, although differences are quite small for three participants and one shows a higher score in T2.

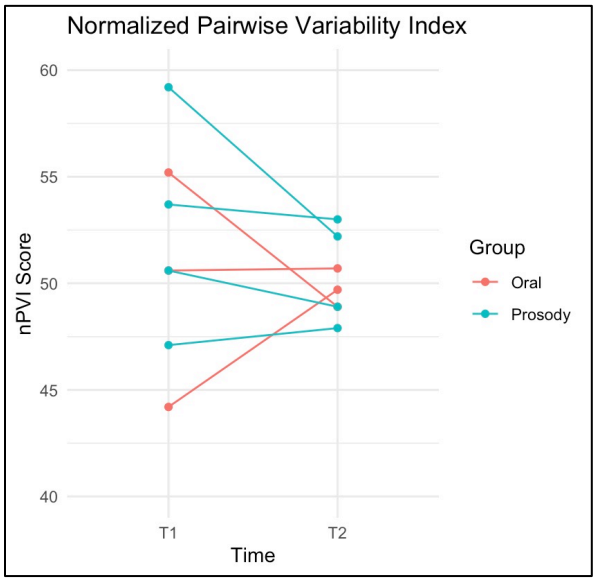

Figure 6. Normalized Pairwise Variability Index per Time for each participant.

It is interesting to note how the participants' scores dispersion gets compressed around the mean in T2 (M_{T2} = 50.6). This seems to indicate an effect of the training in both groups with a convergence around a 50-point score which is about halfway between L1-French and L1-English scores reported in the literature, and therefore is consistent with an interlanguage variety. However, it is difficult to interpret these results without knowing about nPVI intra-individual variation in L1. L1 data has yet to be analyzed to complete this picture.

DISCUSSION AND CONCLUSION

Results on IPU, pauses, and fluency indicate interesting differences between the two groups that we can interpret through the connection between fluency and speech programming processes (Goldman-Eisler, 1968; Segalowitz, 2010). The Prosody Group produces longer external pauses but less disfluency within the IPU. This indicates an improvement as more time is dedicated to preparing the next chunk, which in turn, is realized more fluently and with better internal cohesion. The Oral Group produces shorter external pauses in T2, so less time is dedicated to speech programming; therefore, more fine-tuning appears within the IPU. This observation goes toward our expectations as it indicates benefits from the prosody training over general oral activities.

Results regarding the nPVI show that more participants in the Prosody Group follow the expected tendency of lowering their score in T2, indicating a progression towards the syllable-timed rhythm of French. However, differences between T1 and T2 are for the most part quite small and difficult to interpret in the absence of comparable data in L1. The complete study will include L1 data on all acoustic measures, and measures of comprehensibility and accentedness.

Limitations of this study - aside from the small sample size - are connected to the ecological value of the data collected. A delayed posttest and an at-home testing procedure implies a risk of results not being as reliable as in controlled conditions. It is also possible that more than a 12-hour training is needed to build significant change in pronunciation, although recent studies have demonstrated the effects of shorter training (Schwab & Dellwo, 2022).

This study is the first step toward developing a new course focused on French prosody for French learners, and testing its effect on both acoustic and perceptual measures. We hope the additional measures included in the full study will help get a better understanding of the effects of the trainings.

ACKNOWLEDGMENTS

The authors would like to thank Florence Baills from the University of lleida for her precious help with the Dalcroze musical exercises, and Laure Fesquet and Sébastien Palusci from Fonetix (www.fonetix.com), who gave us access to exercises on rhythm and intonation from their online learning platform for L2 French pronunciation.

ABOUT THE AUTHORS

Lucie Drouillet (formally **Judkins**) is a doctoral candidate in Linguistics at the Laboratory of NeuroPsychoLinguistics (LNPL) at University Jean Jaurès in Toulouse, France. Her interests include L2 prosody acquisition and teaching, English and French prosody, L1 and L2 speech rhythm.
Contact information:
University Toulouse Jean Jaurès
5, allée Antonio Machado
31058 Toulouse Cedex, France
E-mail: lucie.drouillet@univ-tlse2.fr

Charlotte Alazard-Guiu is an Associate Professor in the French as a Foreign Language Department at University Jean Jaurès Toulouse. Her research includes investigating the effects of the Verbo-Tonal Method for phonetic correction, multimodality, and experimental didactics.
Contact information:
University Toulouse Jean Jaurès
5, allée Antonio Machado
31058 Toulouse Cedex, France
E-mail: charlotte.alazard@univ-tlse2.fr

Corine Astésano is a Full Professor in the Linguistics Department at University Jean Jaurès Toulouse. Her work focuses on accentuation in French and speech rhythm, embodied cognition, and rhythm in pathological and L2 speech.
Contact information:
University Toulouse Jean Jaurès
5, allée Antonio Machado

31058 Toulouse Cedex, France
E-mail: corine.astesano@univ-tlse2.fr

REFERENCES

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*, 1-17.

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Abry-Deffayet, D., & Chalaron, M.L. (2009). *Les 500 exercices de phonétique niveau B1-B2*. Hachette Langue Etrangère.

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*. L'Harmattan.

Bailles, F. (2021). *Using the hands to embody prosody boosts phonological learning in a foreign language*. [Doctoral dissertation, University of Pompeu Fabra Barcelona]. http://hdl.handle.net/10803/673777

Berdoulat, H., Fesquet, L., & Palusci, S. (2018). *Fonetix*. [Online platform]. https://fonetix.org/

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, *930*(1), 232–258.

Billières, M. (2008). Le statut de l'intonation dans l'enseignement/apprentissage du FLE, *Le français dans le monde, série recherches et applications, 43*, 27–37.

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer. [Computer program]. Version 6.2.06, retrieved 23 January 2022 from https://www.praat.org.

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.

Briet, G., Collige, V., & Rassart, E. (2014). *La prononciation en classe de langue.* Presses Universitaires de Grenoble.

Cason, N., Marmursztejn, M., d'Imperio, M., & Schön, D. (2020). Rhythmic abilities correlate with L2 prosody imitation abilities in typologically different languages. *Language and Speech, 63*(1), 149–165.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A course book and reference guide*. Cambridge University Press.

Cocton, M. N., Kohlmann, J., Rabin, M., & Ripaud, D. (2020). *L'atelier B1. Méthode de français.* Paris: Didier.

Darcy, I. (2018). Powerful and effective pronunciation instruction: How can we achieve it? *CATESOL Journal, 30*, 13-45.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397.

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*(4), 533–557.

Detey, S., & Durand, J. (2021). Acquisition de la prononciation en langue étrangère. In P. Leclercq, A. Edmonds & E. Sneed German (Eds). *Introduction à l'acquisition des langues étrangères* (pp. 111–126). De Boeck Superieur.
Di Cristo, A. (2002). De la métrique et du rythme de la parole ordinaire: l'exemple du français. Dans: E. Bordas (dir.). *Le Rythme de la Prose*, PUFC. Paris: Les Belles Lettres, pp. 25–44.
Drouillet, L. (in preparation). L2 prosody: French and English rhythms. [Manuscript in preparation. Doctoral dissertation, University Jean Jaurès Toulouse].
Fraisse, P. (1974). *Psychologie du rythme*. FeniXX.
Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research, 21*(5), 609–631.
Goldman, J.-P. (2011). Easyalign: An automatic phonetic alignment tool under praat. *Proceedings of Interspeech* 2011, 3233–3236.
Goldman-Eisler, F. (1968). *Psycholinguistics*. Academic Press.
Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology, 7*(1982), 515–546.
Grant, L. (2016). *Well said: Pronunciation for clear communication* (4th ed.). Heinle & Heinle.
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.
Intravaia, P. (2000). *Formation des professeurs de langue en phonétique corrective*. Le système verbotonal. Didier Édition et Mons : CIPA.
Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*(6708), 228–228.
Judkins, L., Alazard, C., & Astésano, C. (2022). Chunking and pausing in native vs non-native speech. In *XXXIVe Journées d'Études sur la Parole–JEP 2022* (pp. 910–919).
Juntunen, M.-L. (2016). The Dalcroze Approach. In Carlos R. Abril & Brent M. Gault (Eds.). *Teaching general music: Approaches, issues, and viewpoints* (pp. 141–168). Oxford University Press.
Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145–164.
Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396–414.
Liberman, M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36.
Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *2,* 249–336.
Mairano, P., & Romano, A. (2011). Rhythm metrics for 21 languages. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1318–1321).
McAndrews, M. (2023). The effects of prosody instruction on listening comprehension in an EAP classroom context. *Language Teaching Research*, *27*(6), 1480-1503.
Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology, 2*, 1–6.
Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.
Pike, K. L. (1945). *The intonation of American English*. University of Michigan Press.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292.
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.
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.
Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
Wenk, B. J. (1985). Speech rhythms in second language acquisition. *Language and Speech*, *28*(2), 157–175.
Yuan, C., Gonzalez-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition, 41*(1), 5–32.
Zhang, Y., Baills, F., & Prieto, P. (in press). Embodied music training may help improve speech imitation and pronunciation skills. *Language Teaching*.