ISAPh 2016 International Symposium on Applied Phonetics
25-28 March 2016, Nagoya, Japan

Optimizing Pronunciation and Prosody Teaching in Second Language Learning

Li-chiung Yang

Faculty of Arts, Tunghai University, Taiwan
Yang_lc@thu.edu.tw, lichiung.yang@gmail.com

Abstract

There has been a growing realization of the importance of authentic human interactions attendant to both first and second language learning. By contrast to L1 learners, the older L2 learner faces a greatly changed learning environment and more adult-oriented communicative opportunities, and must overcome loss of aural sensitivity and limited exposure to the prosodic systems of expressiveness appropriate to the target language. Research has shown that exposure to the suprasegmental levels of authentic speech has significant positive effects on learners' fluency in production. The current paper reports preliminary results on the experience of adopting a 2nd language teaching method for medium to advanced learners of English that focuses on improving phonetic pronunciation and prosodic competence towards native-like competency. The method adopted prioritizes shadowing and training in aural sensitivity to authentic language at the suprasegmental level.

Keywords: prosody, pronunciation, spoken communication, aural sensitivity, L2 learning.

1. Introduction

Recent investigations into second language learning have highlighted the critical importance of suprasegmental aspects of language in achieving fluency and proficiency in L2 interpretation and pronunciation, and have suggested that learners who had greater exposure to prosodic aspects of language progressed to spontaneous production better than those who learned only segmental content. [1, 2, 3, 4, 5, 6, 7]. Interactive and emotional involvement is a critical component of L1 language development [8], and suprasegmental language features of L1 are deeply ingrained at an early age [9, 10]. L2 learning has also been found to be significantly and negatively affected by inappropriate cross-language transfer of suprasegmental prosodic features of L1, leading to errors in L2 production and difficulties in perception and intelligibility [2, 3, 4, 5, 6, 7].

The importance of suprasegmental features in both L1 and L2 learning arises from the fundamental function of language as a means for the communicative and expressive sharing of meaning, emotions, and cognitive information. These different elements of communication are core aspects of our basic human instinct to share our responses to our environments and to each other. Thus, the meaning that is communicated in language includes not only simple semantic meaning, but equally importantly, the communication of emotional responses and cognitive evaluations of subject matter, as well as interactive signals to ensure the success of the communication and a satisfactory topic development. Because these extended levels of meaning and pragmatic interactive signals are communicated simultaneously with semantic information, prosody and suprasegmental elements of language have a primary role in communicating the multi-dimensional aspects of meaning in natural conversation [11, 12]. For L1 learners, from the earliest age, language learning occurs in the setting of the family and social environment, and this setting of emotionally and expressively rich human interactions is inextricably tied to how language is learned. While 2nd language acquisition typically occurs at a later developmental stage than L1, the continued presence of emotionally and cognitively rich expressiveness throughout natural language suggests that by increased attention to how communicative goals are expressed in the target language, greater efficiency in reaching fluency and communicative competence may be possible for L2 learners, as well.

Early efforts for L2 learning to take advantage of the characteristics of natural language and the communicative functions provided by supra-segmentals stem from the late 19th century and early 20th century Direct Method and Naturalistic teaching methods, which emphasized imitation of native speech in more natural contexts, and extensive exposure to model native speech to acquire native-like speech [7]. However, the popularity of such methods had been limited in actual teaching practice, partly because the theoretical bases and justification underlying such methods were still not sufficiently known.

10.21437/ISAPh.2016-14

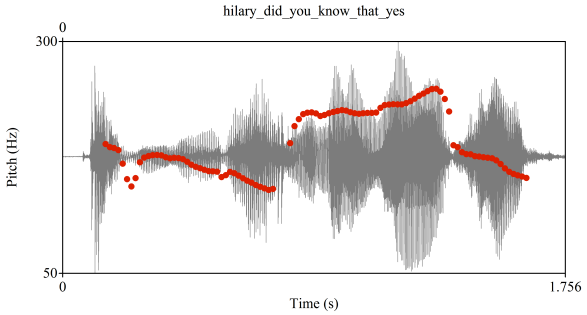
More recent investigations into second language learning have therefore focused on both providing a theoretical foundation for inclusion of supra-segmental elements of language in both 1st and 2nd language learning, and on demonstrating the benefits of adopting such an approach [4, 13]. [2] found that conversational prosody is especially important to language learning because prosody highlights the key important points in a conversation and therefore enhances comprehensibility. Exposure to and mimicry or shadowing of native model speech with native suprasegmental features has been found to be especially beneficial in language learning [2]. Although it is difficult to provide an absolute concordance between specific acoustic variations and specific suprasegmental expressions of meaning [10], consideration of the local interactive context may be necessary to disambiguate the functional value of specific acoustic forms. Although research studies have established the benefits to L2 learning of exposure to native suprasegmental, the methods are still rarely used because of the greater difficulty of teaching suprasegmental speech.

2. The importance of prosody and suprasegmental forms in natural speech

In natural speech, the key element for expression of cognitive and emotional states, and negotiation of interactive relationships is prosody. Prosodic signals provide information on speaker intentions and emphasis, and communicate how ideas in the stream of conversation are related [14, 15]. Because of their role in communicating cognitive and emotional states, the ongoing variations in the elements of prosody of pitch, speech rate, amplitude, and rhythm provide an ever-changing simultaneous underlying commentary on the linear semantic elements of speech.

The signaling of emotional and syntactical information through pitch may share universal elements across many languages [20, 21], but large differences in 1st and 2nd language phonemic or syntactic structure, such as between tone and non-tone languages, also result in significant differences in the prosody of a language. The melody and rhythm of a speaker's first language become well-ingrained from an early age, and make it more difficult for listeners to interpret the specific prosodic signals of the 2nd language and achieve adequate comprehension. We illustrate some of the prosodic features that L2 learners face below:

Example 1
B: Did you know that your brother-in-law and sister-in-law are my neighbor?
*A: I **do** know that"*
B: D'you know that?"
A: Yes!



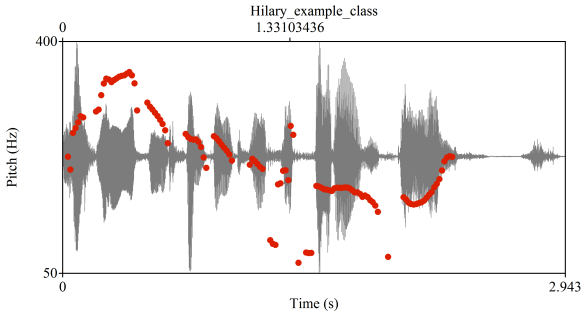
I do know that D'you know that Yes

Figure 1: Pitch contours showing emphasis, emotion, and interactions

Prosodically, a high pitch level is generally associated with increased emphasis or focus, frequently occurring on the key informational elements of an intonational phrase [9]. A rising pitch level usually points to continuation, questioning, or prompting, and a falling pitch with certainty or completeness. In the example of Figure 1, A is responding to a question from B: "I **do** know that", with emphasis on "**do**", signaling certainty. B responds with rising pitch, probing for additional confirmation: "Did you know that?" A answers affirmatively, "Yes!", again signaling certainty with an elevated but falling pitch level.

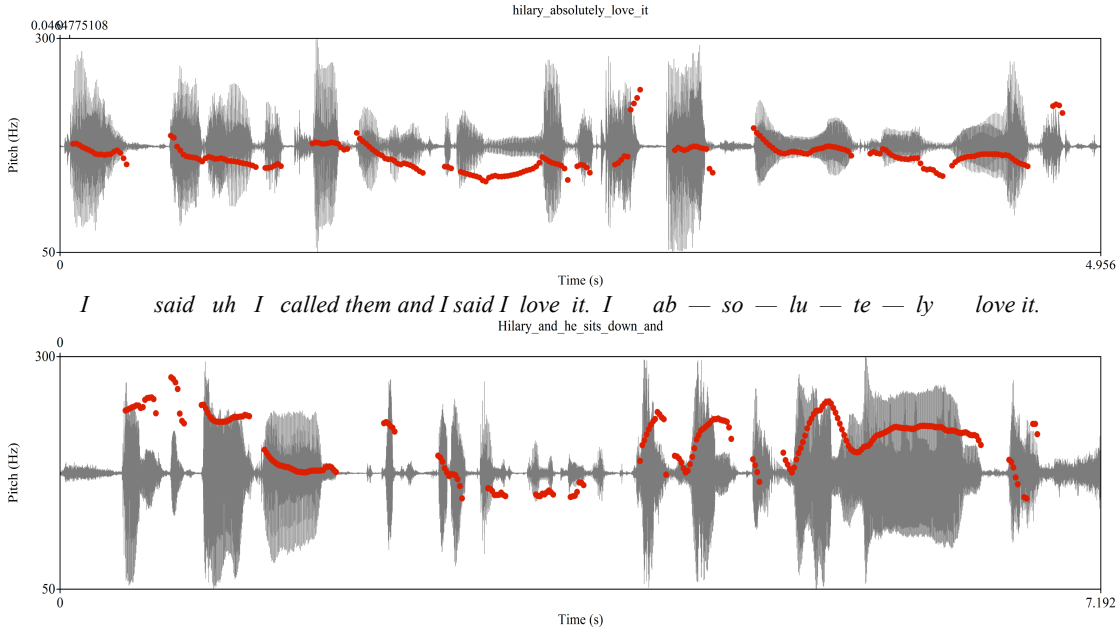
Differential focus on lexical items frequently induces coarticulation and syllable reduction. In Figure 1, B's contraction of "Did you know" to "D'you know", gives relatively greater focus to "know" with much longer duration. Syllable reduction is tied to low information elements and de-emphasis, and is a common feature of connected speech [9].

Example 2
A: just when you think something can't get better, and it does? (rising pitch)
B: Yeah (minimally).



just when you think something can't get better, and it does?

Figure 2: Different expressions of intensity



Figures 3-4: Emphasis and phrasal unity (top); Prominence and iconicity in prosody (bottom)

Variations in pitch are often associated with different degrees of emotional intensity. High pitch is generally associated with increased emotional intensity. In the fragment of Figure 2, A is highly involved in a more narrative-like section, resulting in an overall pitch level that is much higher than in Example 1. The rising pitch on A's "*does*" at phrase end acts to prompt for agreement from Speaker B, and its pragmatic force is signaled both through the strong rising pitch and lengthened duration (.556s). Conversely, a low and falling pitch often signals understanding and agreement. In this fragment, B's pause and short subdued low pitch "*Yeah*" expresses full understanding and agreement, in sharp contrast to the emphatic "Yes" answer of A in the first example. This interchange also shows syllable reduction: A's "*and it*" is sharply reduced in duration and acoustic content to "*n't*".

Example 3

A: I said uh I called them and I said I love it. I ABSOLUTELY love it. (see Figure 3)

Language typically proceeds through a sequence of ideas, each presented as a single information unit or intonational phrase. Each information unit has a single idea of focus. Syntax and prosody work together to delineate and integrate the flow of intonational phrase to develop a coherent stream of ideas. Thus, suprasegmental features are critical for L2 learners to achieve intelligibility in both perception and production. Repetition, voice quality, iconic emphasis, prosodic expression of pragmatic intentions all work to signal phrasal organization, and intelligibility of the information.

The specific variations in prosody and timing of the fragment of Figure 3 show emphasis and enthusiasm, and also provide unity to the ideas expressed in the sentence. The initial high speech rate is accompanied by pauses and delay markers (*uh, and*) that indicate the speaker's efforts in cognitive planning efforts. The strong emphasis on "*love it*" occurs through both repetition and through lengthening. The key word, "*absolutely*" has a duration that is three times as long as the initial "*love it*", and the latter "*love it*" is also lengthened to .707s. This lengthening gives primary emphasis to "*absolutely*" with progressively stronger stress on the two instances of "*love it*".

Example 4

A: And he sits (pause) down and (pause) he puts his foot up on the (pause) on the table, one-one foot, one—long—foot. (see Figure 4)

The fragment of Figure 4 is clearly delineated into a sequence of short intonational phrases by pauses and repetition. This fragment also illustrates a universal property of prosody, iconicity. Prosody is fundamentally iconic [16], in the sense that movements of acoustic features can often be mapped systematically to different degrees of intensity of speaker state. Syllable duration is a key suprasegmental feature that frequently exhibits iconicity. In the example of Figure 4, the 1st "*and*" has typical duration of .128s, but the 2nd "*and*" has a markedly extended duration (.583s) to delay and for pragmatic effect, strikingly longer than the reduced "*and*" of Figure 3. The durations on "*one foot*" and "*one—long—foot*" also reflect the significance of

this idea to the speaker. The first instance (.723s) is already relatively long, and this is further emphasized at "*one—long—foot*", with "*long*" iconically lengthened to over 1 second in duration.

Example 5: An illustrative example 'oh'
Figure 5 provides an illustration of how meaning is effectively encoded in even very short feedback utterances such as 'oh' for spoken communication.

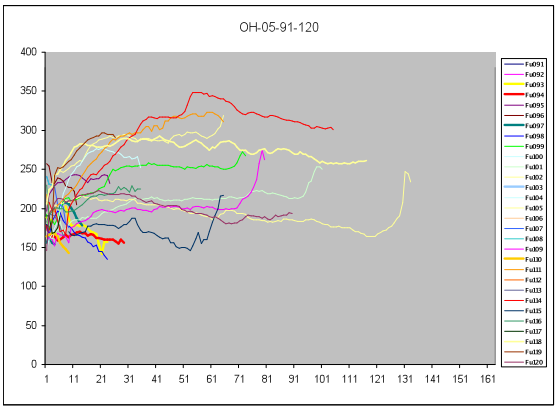


Figure 5: A selection of 30 instances of a female's feedback marker *oh*s in a conversation segment, showing varying forms relating to varying states and functions

3. Suprasegmental forms and L2 learning

The usefulness of suprasegmental exposure to second language learners arises both from the functions of prosody in natural language and from the distinct learning environment that learners encounter. Prosodic signals are critical in communicating focus, speaker intention, and the relationships among ideas and in signaling the important elements of the conversation.

Mature learners typically have limited opportunities for extended one-on-one phonetic guidance, and have characteristically different communicative and interactive goals. Most prominently for pronunciation acquisition, L2 learners must overcome the loss of sensitivity to non-native sound distinctions, and also must overcome their limited exposure to the systems of expressiveness of the target language.

Increased aural sensitivity is best accomplished through model speech that encompasses all levels of language structure. This process parallels the process for 1st language learning, as it places language learning in an emotionally rich context.

The benefits to L2 learning of adequate exposure to suprasegmental and prosodic elements pointed to by the above research may stem from the higher bilateral brain activation that occurs with prosodically rich language perception [17], which ties together the perception and production of target language phonetic, pragmatic, and cognitive elements. The current study investigated how L2 language learning can be enhanced through a promising technique of near-simultaneous mimicry of native speech [18], *shadowing*, in which learners mimic the key acoustic features of a target language model nearly simultaneously.

4. Integrating prosodic features in 2nd language teaching: A classroom study

The current paper reports on the adoption of a 2nd language teaching method for intermediate to advanced learners of English that focuses on improving phonetic pronunciation and prosodic competence towards native-like competency. Our goal was to integrate the insights from linguistic theory and language acquisition research to improve English proficiency, particularly in pronunciation and prosodic perception and production in L2 learners.

Participants were university and graduate level students who had 9-10 years of English language education starting from an early age. The study includes both advanced level classes with 10-12 students per class, and intermediate level classes with 20-25 students per class. Classes met for 3 hours per week, with about one hour devoted to the method adopted, including preparatory activities, for 15 weeks per semester.

Different languages have different rhythms and different ways of expressiveness within the grammatical and phonetic systems of each language, and this points to the critical importance of developing sensitivity to how things are expressed in the target language. In the imitation/shadowing method adopted, we utilize several principles of language learning to help students overcome acclimatization to their first language and loss of sensitivity to sound distinctions in other languages. We prioritize hearing and encountering authentic language in real context. With the ability to access unlimited video and oral language from "real life" currently, model language on topics of high interest to 2nd language learners can readily be found.

For both intermediate and advanced English classes, the video series "Connect with English", BBC's Learning English, and selected TED talks were adopted as model language corpora. Although acted, "Connect with English" provides an extended integrated exposure to everyday English in a context that is very appropriate for college level students. TED talks are also chosen based on topics that are of interest to students, and are especially suitable for advanced learners, as they are more professional and at a higher level, covering a wide range of topics.

Given these contextually rich language models, to increase sensitivity to both sounds and to the

rhythmic and prosodic patterns of goal-oriented and emotional expressiveness in the target language, students are trained to follow along, sentence-by-sentence, and paragraph-by-paragraph with the speakers, and mimic/shadow the speaker nearly instantaneously, trying to imitate both the sound patterns and intonation patterns of the speaker exactly. In class, explaining the transcripts ahead of time takes the stress off of understanding, and allows emphasis on training the ear to be sensitive to the sound patterns. By mimicking first, students gain confidence quickly, and rapidly progress to innovation and their own production. As class time is limited, and must also be devoted to other critical aspects of language learning such as understanding and self-motivated production, in-class training and correction of students is oriented towards showing the students how to listen carefully and mimic well, so that home study can be utilized to using the online talks, and modeling their speech on native speech patterns. Evaluation is done every class, and also at the mid-semester and final oral proficiency exams, where students' proficiency level was being rated by their teacher, based on 10-point rubrics.

The results of using this method have been remarkable. Students have demonstrated greatly improved language confidence in a short time, and significant improvements in pronunciation, presentation, and expressiveness of their own ideas and topics are achieved within as little as two months, and greater over the course of a semester. When mimicking is done at a high level, and involves ideas, rather than just rote repetition, students gain insight into the thought processes of the speaker, since linkages between ideas and emotions attached to ideas are represented in the prosody and rhythms which the speaker uses; in reproducing these sound patterns, students inevitably reproduce the idea linkages in their own minds. Thus, this method both activates the valuable motivational context for language learning, and trains student sensitivity to meaningful sound distinctions of the target language, providing a highly effective and efficient means to achieve comprehensibility and intelligibility.

There are a number of important advantages found for this method of language learning. First, mimicking real high-level language is highly beneficial for phonetic and prosodic pronunciation. Using natural databases leads to exposure to a very rich language corpus and exposure to a wide variety of speech styles and accents. This also allows students to intuitively absorb aspects of language that go beyond the current explicitly known rules for the language. This is especially true for prosodic aspects of language. Because of the high level model speech, students gain satisfaction and confidence in handling the target language "on its own terms" and a raised awareness of language features. The technique is also particularly suitable for computer assisted software and home assignments. For more motivated students, free speech software such as *Praat* and *Wavesurfer* can be utilized to allow learners to self-evaluate their own production, and practice can be adjusted to each student's own level. When enhanced with other class activities such as discussion, role play, and spontaneous production, students achieve a very high degree of proficiency in just a couple of months or a semester.

5. Cognitive and neural basis for suprasegmental approach to language learning

Suprasegmental and prosodic acoustic features communicate cues to speaker characteristics such as age or sex, as well as information on speaker intention, cognitive evaluation, and emotions. The presence of prosodic cues in conversation also increases the intelligibility of the communicated information. Conversely, disruption to paralinguistic aspects of voice perception and production often causes serious problems in effective psychosocial functioning [19].

For L2 learning, a prosodically rich and cognitively relevant approach parallels the path followed for 1st language learning. The effectiveness of exposure to prosodically rich language for language learning has been corroborated through experimental studies in neuroscience. Yune-Sang Lee, et al. have demonstrated that an acoustic rich speech signal results in a stronger bilateral activation of specific brain regions, especially the right hemisphere, in comparison to acoustically less-rich signals, suggesting that "the neural systems involved in speech perception are finely attuned to the type of information available" [17].

Near simultaneous mimicry, or *shadowing* of a model speaker's natural intonation provides an especially effective technique for L2 learning. Simulation of authentic model speech activates the innate neural ability to mirror speech sounds [8]. Recent research on "neural mirroring systems" in the brain also provides promising suggestions that imitation of speech may have powerful effects because of activation of parallel emotional and cognitive neural activity called up by mirroring. Recent research of [20] which used simultaneous acoustic, articulatory and MRI neural measurements has found that imitative speech bilaterally activates important language areas of the brain in L2 learners. Research has also suggested that imitative speech is causally correlated with social likeability.

The information provided by suprasegmental signals appears to be independent of the phonetic/semantic information [21]. The researchers found that speakers engaged in shadowing "tend to subconsciously echo phonologically irrelevant acoustic details", including speech rate, prosody, and voice-onset time. In their account, the shadowing process involves the transfer from perception of perceptual signals into production, and the degree of success of shadowing correlates with greater activation of specific regions of the brain consistent with paralinguistic and phonetic localization found by other researchers. In their work, greater activation of bilateral areas of the brain occurred when shadowing was done on multiple speakers, which could be attributed to greater effort by shadowers to normalize the different acoustic characteristics of different speakers [21]. They further pointed out that this greater activation occurred in neural regions associated with greater attention, thus, multiple speakers may require greater neural effort to normalize the acoustic signals.

The ties between prosodically rich language and greater bilateral neural activation suggests that shadowing authentic native speech may be an especially powerful technique for learning the appropriate native prosodic patterns used for expressiveness in the target language. In reproducing the natural native sound patterns through shadowing or mimicking, learners inevitably reproduce these idea linkages in their own minds. This enhances both intelligibility and motivation for the target language, and learners exhibit enhanced confidence in discovering their ability to approximate native speech successfully.

6. Conclusions

One of the critical aspects of a contextualized and prosodic focus is the *authenticity* of the language learning experience. Authenticity of learning material allows learners to experience first-hand the full hierarchy of semantic, interactive, and emotional meaning that are developed at all levels of speech. Learners thus experience the target language as native speakers use it, with its full emotional and cognitive expressiveness. Such experience of the target language takes advantage of primal neural processes that provide structure and efficiency to the learning process. Exposure to speech with rich prosodic content gives speech communication significance and meaning, and provides a great motivating power for language learning.

7. References

[1] Anderson-Hsieh, J, Johnson, R, Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*. 42(4):529–555.

[2] Chela-Flores, B. 2004. Optimizing the Teaching of English Suprasegmentals, *BELLS: Barcelona English Language and Literature Studies*.

[3] Edwards, J., Zampini, M. L. eds. 2008. *Phonology and Second Language Acquisition*, John Benjamins.

[4] Derwing, Tracy, Munro, M. 2015. *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. John Benjamins.

[5] Romero-Trillo, J. (ed.) 2012. *Pragmatics and Prosody in English Language Teaching*. Dordrecht & London: Springer.

[6] Saito, Y., Saito, K. 2016. Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*. 1-20.

[7] Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. 2010. *Teaching Pronunciation: A course book and reference guide*. 2nd edition. Cambridge: Cambridge University Press.

[8] Kuhl, P. 2004. Early language acquisition: Cracking the speech code, *Nature Reviews Neuroscience*, 5, 831-843.

[9] Ladefoged, P. & Johnson, K. 2015. *A Course in Phonetics*, 7th edition, Cengage.

[10] Meng, H., Tseng, C., Kondo, M., Harrison, A., Viscelgia, T. 2009. Studying L2 Suprasegmental features in Asian Englishes: A position paper. *Proc. Interspeech 2009*, Brighton, UK. 1715-1718.

[11] Brown, G., Yule, G. 1983. *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge University Press.

[12] Yates, L. 2014. Learning how to speak: Pronunciation, pragmatics and practicalities in the classroom and beyond. *Language Teaching*, 1-20.

[13] Gilbert, J. B. 2008. *Teaching Pronunciation*. Cambridge University Press.

[14] Bolinger, D. 1989. *Intonation and Its Uses -- Melody in Grammar and Discourse*. Stanford, California: Stanford University Press.

[15] Chafe, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing.* Chicago: The University of Chicago Press.

[16] Ohala, J. 1983. Cross-language Use of Pitch: An Ethological view. *Phonetica*, 40, 1-18.

[17] Lee, Y-S., Min, N., Wingfield, A., Murray Grossman, M., Peelle, J. 2016. Acoustic richness modulates the neural networks supporting intelligible speech processing. *Hear Res* 23;333:108-17.

[18] Luo, D., Shimomura, N., Minematsu, N., Yamauchi, Y., Hirose, K. 2008. Automatic pronunciation evaluation of language learners' utterances generated through shadowing. *Proc. Interspeech 2008*, Brisbane, Australia. 2807-2810.

[19] McGettigan, C. 2015. The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. *Front. Hum. Neurosci.*

[20] Carey, D., McGettigan, C. 2015. Magnetic resonance imaging of the brain and vocal tract: applications to the study of speech production and language learning. *Front. Hum. Neurosci.*

[21] Peschke, C., Ziegler, W., Kappes, J., Baumgaertner, A. 2009. Auditory–motor integration during fast repetition: The neuronal correlates of shadowing. *NeuroImage*. 47 (1): 392–402.