9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

# MIAPARLE: Online training for discrimination and production of stress contrasts

*Sandra Schwab*[1,2]*; Jean-Philippe Goldman*[1]

[1]Université de Genève, Switzerland
[2]Universität Zürich, Switzerland
sandra.schwab@unige.ch, jean-philippe.goldman@unige.ch

## Abstract

MIAPARLE is a web application for the general public that is designed to offer an innovative range of CAPT (computer-aided pronunciation teaching) tools for second language (L2) learners. In this paper, we describe MIAPARLE, a web application that focuses on stress perception and production in L2. Such a tool is particularly useful for speakers whose L1 is a fixed-stress language, such as French. These speakers have difficulties perceiving and discriminating stress contrasts, and therefore struggle when producing them. To help them with this so-called stress 'deafness', we present a methodology for perception training which is based on successive questions in which a visual pattern is associated with the stress pattern of a lexical item. After successively completing pre-tests, training and post-tests in the perception part, the participant is given their improvement score, i.e., the difference between pre- to post-test scores. In the production part, the participant is shown pictures and has to pronounce the word for the item they see in Spanish. The participant's production is acoustically analyzed and visually represented with the stressed syllable highlighted. A comparison between the participant's production and the Spanish word's canonical production serves as feedback.

**Index Terms**: CAPT, pronunciation training, prosody, stress deafness

## 1. Introduction

Second language learners tend to imprint the prosody of their native language onto the second language (L2) (e.g., [1]; [2]). This prosodic cross-language transfer is often combined with segmental transfers, which can lead to the presence of a foreign accent. A foreign accent can not only hamper communication between learners and native speakers (e.g., [3]), but it can also affect the credibility of learners and how they are evaluated by others; this can sometimes lead to social discrimination (e.g., [4]; [5]; [6]). Moreover, despite the crucial role of prosody in speech flow segmentation (e.g., [7]; [8] for English; [9] for French), it is rarely taught in language courses, even in foreign-language pronunciation courses.

In the framework of *computer-assisted pronunciation teaching* (CAPT), the goal of this project is to develop a methodology which improves the prosodic skills of L2 learners in order to reduce cues of foreign accent. Although there has recently been a growing interest in CAPT applications (e.g., for French[1], Spanish[2], English[3], and Norwegian[4]), there have been few applications that focus specifically on prosodic features. Among those that do, some have developed procedures for automatic assessment of non-native prosody (e.g., [10] for English L2; [11], [12] for Spanish L2), while others have examined the impact of feedback in learning L2 prosody (e.g., [13], [14] for German L2; [15] for English L2). Although some existing systems were specifically built for the automatic detection of stress errors (e.g., [16] for English L2; [17] for German L2, [18] for Hungarian L2), and hence designed to improve the learners' production of stress in L2, none to our knowledge have focused on learners' perception of stress in a second language.

This paper focuses on the development of a tool designed to improve the discrimination and production of stress contrasts in a second language. Although this method is intended for the general public, and will address various other aspects of pronunciation in the near future, training in stress contrasts is mostly useful for language-learners with a fixed lexical stress native language (such as French, Czech, or Hungarian). For example, French-speaking listeners tend to experience difficulties in perceiving stress contrasts in a second language such as Spanish, and thus experience difficulties when producing such contrasts. These difficulties are the basis of the *stress deafness* hypothesis (e.g., [19]). According to this hypothesis, the degree of *stress deafness* is related to the stress properties of the L1, more specifically to the nature of lexical stress (free or fixed). In a free-stress language such as Spanish or English, lexical stress has a distinctive function, since it distinguishes segmentally identical words; such as in Spanish *<u>ú</u>mero* (['numeɾo], English *(the) number*) and *nu<u>me</u>ro* ([nu'meɾo], in English *I number*)[5]. As a consequence, speakers of a free-stress language need to encode stress position in their mental representation of the words. On the other hand, the position of stress in a fixed-stress language such as French is not variable, and is therefore not contrastive. Consequently, the stress information does not need to be stored in the lexical representation for speakers of these languages. The *stress deafness* hypothesis claims that speakers of fixed-stress languages have difficulties in perceiving stress contrasts in free-stress L2s since they are not able, or at least not trained, to encode stress information in their mental lexicon (e.g., [19]). Nevertheless, as shown by our recent research ([20]), French learners are able to improve their stress detection ability after a 4-hour training session. Based on this methodology and its promising results, our aim is that our tool for training of stress contrast discrimination and production will give L2 learners the bootstrapping required to start encoding stress information in

[1] http://www.coelang.tufs.ac.jp/mt/fr-swiss/dmod/index_en.html
[2] http://www.enterate.unam.mx/Articulos/2007/abril/sarahi.html
[3] http://www.carnegiespeech.com/products/nativeaccent.php
[4] https://orgesuniversitetet.no/prosjekt/computer-assisted-listening-and-speaking
[5] The underlined syllable in these examples, and in the rest of the paper, corresponds to the stressed syllable.


10.21437/SpeechProsody.2018-116

their mental lexicon; this will, in turn, enable them to correctly reproduce lexical stress when speaking the second language.

## 2. MIAPARLE

MIAPARLE is a web application that can host various tools and activities dedicated to pronunciation training. It is based on PyBossa, a micro-tasking crowdsourcing platform. PyBossa is developed and maintained by SciFabric; it runs on the flask micro-framework and is well designed to be responsive on smaller screens such as those in tablets and smartphones.

In this paper, we describe a training prototype for learning stress contrasts in Spanish. The training is divided into a perception and a production part. In the perception part, the participants' initial performance is evaluated with two different pre-tests, they then follow a full perceptual training program, and eventually take post-tests in order to quantify their improvement in the perception of lexical stress. In the production part, the participants are shown pictures and have to name the item they see in Spanish. The participants' production is acoustically analyzed and visually represented as bars which highlight the syllable that the participants stressed. A comparison between the participants' production and the Spanish word's canonical production serves as feedback.

### 2.1. Perception

#### *2.1.1. Pre- and post-tests*

The pre- and post-tests are used to evaluate the training and comprise similar tasks. Two types of exercises are available: Localization and Odd-One-Out, as described below. Each exercise has 15 items in the pre-tests (the participants go through 30 items in total during two pre-tests). During the post-test, each activity shows the same 15 items and 15 additional items for both exercises, in order to test the generalization of the training. Thus, the learner is shown 60 items during the post-tests (Table 1).

Table 1. Number of items and duration (Dur) for pre-tests, training and post-tests.

| | Pre-tests | Training | Post-tests |
|---|---|---|---|
| | Localization (N = 15) Dur = 3 min | A or B | Localization (N = 15+15) Dur = 6 min |
| | Odd-One-Out (N = 15) Dur = 3 min | | Odd-One-Out (N = 15+15) Dur = 6 min |
| Total | N = 30 Dur = 6 min | 104 or 90 | N = 60 Dur = 12 min |

*Localization of a given stress pattern*

In the *Localization* exercise, learners have to localize a given stress pattern. They hear trials of different words produced by a native Spanish female speaker. Each word is associated with a representative drawing. They have to answer to the following question: "Which word has stress on the {antepenultimate, penultimate, final} syllable?". For example, as shown in Figure 1, the learners hear the words "río" (river), "mesa" (table) and "champú" (shampoo) and have to indicate which word has stress on the final syllable (i.e., "dernière syllabe" in French). The learners indicate their response by clicking on the drawing corresponding to the word with the given stress pattern. Before answering, they have the opportunity to listen to the words as many times as they want. They do not receive any feedback after each trial, but their score is displayed at the end of the exercise.

The words the participants have to localize in the trials can have stress on the final, the penultimate or the antepenultimate syllable. The difficulty of the task is increased during the course of the exercise when words produced by another (male) speaker are introduced, and by increasing the number of words presented in the trials (2, 3, 4 words) or/and the number of syllables in the words (2, 3, and 4 syllables).

Fifteen trials were used in the pre-test. In addition to these 15 trials, we used 15 new trials in the post-test in order to evaluate the generalization of the training method to new items.


Figure 1: Localization of a given stress pattern. Question "Click on the word which has stress on the final syllable"

*Odd-One-Out*

In the *Odd-One-Out* exercise, learners perform an Odd-One-Out task ([21]). They hear a trial of three segmentally identical stimuli (e.g., /numeɾo/). Among them, two stimuli present the same stress pattern (e.g., stress on the penultimate syllable) and one (i.e., the *odd*) presents a different stress pattern (e.g., stress on the final syllable). After hearing each trial, learners have to indicate which of the three elements is the *odd* one (i.e., the different one), by clicking on the corresponding option on the screen (Mot 1, Mot 2, Mot 3; in English Word 1, Word 2, Word 3).

We used triplets of trisyllabic Spanish words that differ with respect to position of the stressed syllable. Each triplet is composed of a word with stress on the first syllable (e.g., número), a word with stress on the penultimate syllable (e.g., numero) and a word with stress on the final syllable (e.g., numeró). Two native speakers of Spanish (one female and one male) produced the words.

All combinations of stress contrasts are tested. The position of the odd word within the trial is assigned randomly. The difficulty of the task increases in the second part of the exercise by introducing words produced by the male speaker. Fifteen trials were used in the pre-test. In addition to these 15 trials, we used 15 new trials in the post-test, as in the *Localization* exercise described above.

#### *2.1.2. Training*

The training *per se* is the principal, and therefore the most time-consuming activity in the whole process. The method we used

in the training is based on the empirical results of our previous research [20]. In that study, we compared the effectiveness of two training methods – the one explicit and the other implicit – for learning stress contrasts in Spanish L2. The explicit training approximated the learning environment of a L2 phonetics classroom, in which learners received metalinguistic explanations about the Spanish accentual system and performed various exercises commonly used in corrective phonetics. In the implicit training, the learners did not receive any metalinguistic explanations and always performed the same task, namely a shape/word matching task (see below for a description of the task). The results were as follows: 1) both training methods significantly improved the ability of French and German learners to perceive Spanish lexical stress; 2) the differences observed between the progressions with explicit and implicit training methods were subtle; 3) learners with no training did not improve their performance. Given these findings, and the easier implementation of implicit training, the method used for the current implementation of MIAPARLE is based on the implicit training developed in [20].

In the next sections, we present two versions of the training, both based on the methodology presented in [20]. Participants in Version A learn the association between shapes and Spanish words, while in Version B they learn the association between shapes and stress patterns. In other words, Version A focuses on learning some specific Spanish words and Version B on learning Spanish stress patterns.

*Version A*

In Version A, learners perform a shape/word matching task. They hear a word and 4 of 6 possible shapes appear on the screen (see Figure 2). Learners have to click on the shape which they think corresponds to the word they hear. After giving their response, they receive feedback: they hear the word again and the correct shape is indicated with a green frame. If they had clicked on an incorrect shape, it is indicated with a red frame. The feedback enables the learners to learn the correspondence between the words and the shapes.

Two triplets of trisyllabic Spanish words are used in this training: cáscara (shell), cascara (that he cracked), cascará (he will crack) and módulo (module), modulo (I modulate), moduló (he modulated); these are associated to 6 shapes, as shown in Figure 2. As can be observed, there is no logical/iconic relationship between the form of the shapes and the stress pattern of the words.

Each triplet is composed of a first-syllable stressed word (i.e., cáscara and módulo), a second-syllable stressed word (i.e., cascara and modulo) and a third-syllable stressed word (i.e., cascará and moculo). The six words are produced by another female native Spanish speaker (i.e., different to the one who recorded the stimuli in the pre-tests).

The training is divided into three blocks. Each word (e.g., cáscara) is presented six times per block (6 words x 6 times = 36 times per block). Among the four shapes that appear on the screen, only one corresponds to the target word and three others are distractors. Among the three distractors, one corresponds to a word with a different stress pattern (e.g., cascara) and the two other shapes correspond to two words from the other triplet (e.g., módulo and modulo). The position of the shapes on the screen is assigned randomly. The order of presentation of the trials is determined semi-randomly in such a way that no more than two same stress patterns and no more than two members of each triplet follow each other.

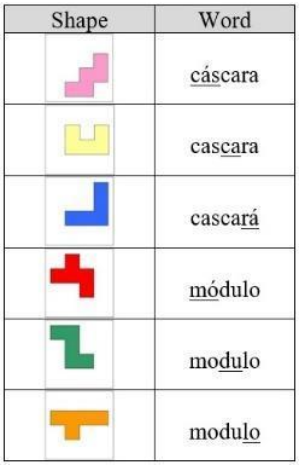

Figure 2: Shapes and words used in Version A.

*Version B*

In Version B of the training phase, learners perform a shape/stress pattern matching task. They hear a word and three shapes appear on the screen (see Figure 3). The shape iconically represents a trisyllabic word with stress on one of the three syllables (see Figure 3). The participants have to click on the shape which they think corresponds to the stress pattern they hear. After giving their response, they receive feedback: they hear the word again and the correct shape is indicated with a green frame. If they had clicked on an incorrect shape, it is indicated with a red frame. The feedback enables the learners to learn the correspondence between the stress patterns and the shapes.

Ten triplets of trisyllabic Spanish words are used in this version of the training. Each triplet is composed of a first-syllable stressed word (e.g., cálculo), a second-syllable stressed word (e.g., calculo) and a third-syllable stressed word (e.g., calculó). The 30 words are produced by a female native Spanish speaker (the same who produced the stimuli in Version A).

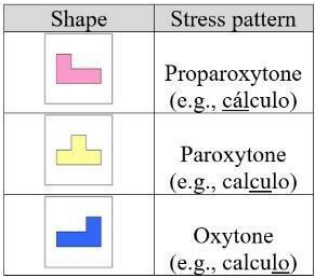

Figure 3: Shapes and stress patterns used in Version B (proparoxytone = first-syllable stressed, paroxytone = second-syllable stressed, oxytone = third-syllable stressed).

Similarly to Version A, the training is divided into three blocks. Each word (e.g., cálculo) is presented once per block. Among the three shapes that appear on the screen, only one corresponds to the target stress pattern and the two others correspond to the other two stress patterns. The position of the shapes on the screen is assigned randomly. The order of presentation of the trials is determined semi-randomly in such a way that no more

than two same stress patterns and no more than two members of each triplet follow each other.

#### *2.1.3. Preliminary results*

The platform was launched in mid-September 2017. Six native French-speaking testers went through the perception part (pre-tests, training and post-tests) and gave us some useful qualitative feedback in respect of user experience.

The platform uses an A/B testing mechanism, so that every other registered participant is led alternatively to Version A or Version B of the training. For Version A, the score (i.e., percent correct) increased from 56% at the pre-test to 72% at the post-test, i.e., an improvement of 16%. For Version B, the score increased from 66% to 74%, i.e., an improvement of 8%. In both versions, we noted that the best improvement was as high as 20%.

We also noticed that the highest score in the post-test was 80%, showing a limited degree of possible improvement for speakers with a higher initial ability level in L2 Spanish. Finally, we estimated that the total time that participants spent on the whole training was 30 minutes.

### 2.2. Production

In the production part (which is still undergoing development) the participant is shown a picture and has to name the item they see in Spanish. For example, a picture of a *bedsheet* is presented and the participant has to produce the word "sábana" with stress on the first syllable. The participant's production is sent to our server, automatically aligned with EasyAlign ([22]) and acoustically analyzed using ProsoProm, a tool for the automatic detection of prominent syllables ([23]). For each syllable, ProsoProm computes a prominence score based on two acoustic parameters (normalized relatively to surrounding syllables): 1) fundamental frequency (F0) (of tones identified by Prosogram [24]), and 2) syllable duration.

A visual representation of the participant's production is then given. As shown in Figure 4, each syllable is represented by a bar, and the syllable that has been stressed is represented by the highest bar. The participant's production is compared to the canonical production of the word. If they match, the production is represented in green (Figure 4a), if they don't, it is shown in red (Figure 4b). At the end of the production exercise, a score of the participant's correct productions is calculated and displayed.

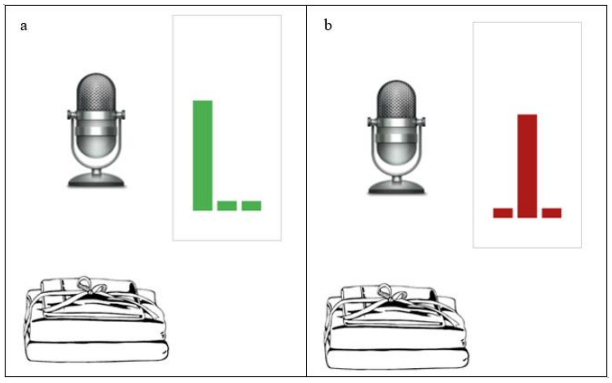


Figure 4: Visual representation of the correct (a) and incorrect (b) pronunciation of the Spanish word "sábana".

The link between perception and production has been extensively studied in the acquisition of segmental aspects (see for example [25], [26], [27]). In contrast, the examination of such a link for prosodic aspects is scarcer. MIAPARLE allows us to collect and correlate, for each participant, perception and production data. Indeed, correlational analyses between the performance at the post-test and the production score can be conducted.

## 3. Conclusion

MIAPARLE is a web platform which offers a perceptual training program that helps learners to discriminate and produce stress contrasts in an L2 such as Spanish. It allows us to test several methodological configurations in parallel (i.e., Version A and B of the perceptual training) and compare their teaching impact.

Our three main short-term tasks are: 1) to continue taking the qualitative feedback of our testers into account in order to refine the didactic instructions and the more general user experience; 2) to deploy the platform to a larger number of participants in order to consolidate our preliminary results about the efficiency of the perceptual training; 3) to develop other sets of exercises focusing on the production of stress contrasts. We also plan to adapt our platform for other L1s, and to expand the L2 material to other free-stress languages such as English and German, as well as tone languages such as Mandarin.

## 4. Acknowledgements

This project is supported by the INNOGAP proof-of-principle fund from UNITEC, University of Geneva.

## 5. References

[1] Barquero, Mª A., Racine, I., Baqué, L. & Schwab, S. (2014). La estructuración acentual: estudio comparativo en la interlengua español-francés. Caso de la lectura. In: Y. Congosto Martín, M. L. Montero Curiel & A. Salvador Plans (Eds.), *Fonética experimental, educación superior e investigación*. (Vol. II, pp. 9-28). Madrid: Arco/Libros.

[2] Muñoz Garcia, M. (2010). *La perception et la production de l'accent lexical de l'espagnol par des francophones: aspects phonétiques et psycholinguistes*. Thèse de doctorat, U. Toulouse 2/Universitat Autònoma de Barcelona.

[3] Anderson-Hsieh, J. & Koehler, K. (1988). The Effect of Foreign Accent and Speaking Rate on Native Speaker Comprehension. *Language Learning*, 38, 561–613.

[4] Dailey, R. M., Giles, H. & Jansma, L. L (2005). Language attitudes in an Anglo-Hispanic context: the role of the linguistic landscape. *Language and Communication*, 25, 27–38.

[5] 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, 1093-1096.

[6] Purnell, T, Idsardi W. & Baugh J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18, 10-30.

[7] Cutler, A. (1984). Stress and accent in language production and understanding. In: D. Gibbon & H. Richter (eds.) *Intonation, Accent and Rhythm: Studies in Discourse Phonology*. Berlin: Walter de Gruyter.

[8] Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.

[9] Bagou, O., Fougeron, C. & Frauenfelder, U. H. (2002). Contribution of prosody to the segmentation and storage of

"words" in the acquisition of a new mini-language. *Speech Prosody 2002*, Aix-en-Provence, France, 159-162.
[10] Hönig, F., Batliner, A., Weilhammer, K. & Nöth, E. (2010). Automatic assessment of non-native prosody for English as L2, *Speech Prosody 2010*, paper 973.
[11] Cardeñoso-Payo, V., González-Ferreras, C., Escudero-Mancebo, D. (2014). Assessment of Non-native Prosody for Spanish as L2 using quantitative scores and perceptual evaluation. *LREC 2014*.
[12] Escudero-Mancebo, D., González-Ferreras, C. Aguilar, L. & Estebas-Vilaplana, E. (2017) Automatic Assessment of Non-Native Prosody by Measuring Distances on Prosodic Label Sequences. *Interspeech 2017*.
[13] Demenko, G, Wagner, A., Cylwik, N. & Jokisch, O. (2009): An audiovisual feedback system for acquiring L2 pronunciation and L2 prosody, *SLaTE-2009*, 113-116.
[14] Niebuhr, O. (2017), Comparing visualization techniques for learning second language prosody. First results. *International Journal of Learner Corpus Research* 3:2 (2017), 250–277.
[15] Bang, J., Kang, S. & Lee, G.G. (2013). An automatic feedback system for English speaking integrating pronunciation and prosody assessments, *SLaTE-2013*, 83-89.
[16] Hönig, F., Batliner, A., Weilhammer, K. & Nöth, E. (2009). Islands of failure: employing word accent information for pronunciation quality assessment of English L2 learners, *SLaTE-2009*, 41-44.
[17] Vakil, A. S. & Trouvain, J. (2015). Automatic classification of lexical stress errors for German CAPT. *SLaTE 2015*.
[18] Sztahó, D., Kiss, G. Czap, L. & Vicsi, L. (2014). A Computer-Assisted Prosody Pronunciation Teaching System. *Fourth Workshop on Child Computer Interaction (WOCCI 2014)*.
[19] Dupoux, E., Pallier, C., Sebastián, N. & Mehler, J. (1997). A destressing 'deafness' in French?, Journal of Memory and Language, 36(3), 406-421.
[20] Schwab, S. & Dellwo, V. (2018). Explicit and implicit training methods for the learning of stress contrasts in Spanish. In Lahoz-Bengoechea, J. M., Pérez Ramón, R., & Villa Villa, J. (Eds.), *Subsidia. Tools and resources for speech sciences*. Malaga: Universidad de Malaga.
[21] Schwab, S. & Dellwo, V. (2016). The use of the Odd-One-Out task in the study of the perception of lexical stress in Spanish by German-speaking listeners. *Proceedings of the 6th International Conference on Speech Prosody*, Boston, May, 31 – June, 3, 2016.
[22] Goldman, J.-P. & Schwab, S. (2014). EasyAlign Spanish: a (semi-)automatic segmentation tool under Praat. En: Y. Congosto Martín, M. L. Montero Curiel & A. Salvador Plans (eds.), *Fonética experimental, educación superior e investigación* (Vol. I, pp. 629-640). Madrid: Arco/Libros.
[23] Goldman, J.-P, Avanzi, M., Lacheret-Dujour, A., Simon, A.-C. & Auchlin, A. (2007), A Methodology for the Automatic Detection of Perceived Prominent syllables in Spoken French, *Proceedings of Interspeech 2007*, Antwerpen, August 27 - 31, 2007.
[24] Mertens, P. (2004). The prosogram: semi-automatic transcription of prosody based on a tonal perception model, *Proceedings of Speech Prosody 2004*, 549-552.
[25] Llisterri, J., (1995). Relationships between Speech Production and Speech Perception in a Second Language", *Proc. of the 13th Int. Congress of Phonetic Sciences*, Vol. 4, 92-99, 1995.
[26] Flege, J., (1999). The Relation Between L2 Production and Perception., *Proc. of the 14th Int. Congress of Phonetic Sciences, Vol. 2*, 1273–1276, 1999.
[27] Peperkamp, S. & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. *INTERSPEECH 2011,* , Florence, Italy, August 27-31, 2011