

## A computer-assisted learning of English prosody for French students

Anne BONNEAU    Matthieu CAMUS    Yves LAPRIE    Vincent COLOTTE

Speech Group  
LORIA/CNRS and INRIA  
Campus scientifique BP239  
Vandoeuvre-lès-Nancy 54506, France  
Anne.Bonneau@loria.fr, Matthieu.Camus@loria.fr, Yves.Laprie@loria.fr, Vincent.Colotte@loria.fr

### Abstract

This paper presents a computer-assisted learning system of English prosody for French students. Our main goal is to provide learners with relevant feedback, based upon speech signal transformations and guided by a comparison between French and English prosodies. Considering that teachers should be given the possibility of customizing their courses, the main functions of WinSnoori for editing and modifying the signals has been ported in the form of active X controls. We present here these functions, a set of exercises conceived for the acquisition of English prosody, and give a concrete example illustrating our strategy.

### 1 Introduction

We present a software devoted to the improvement of the production of English prosody by French learners, based on speech signal modifications and phonetic knowledge. Whereas visual feedback on supra-segmentals has proved its efficiency in the domain of language learning [1], speech visualization tools are not yet widely used in schools. One main explanation for this relative failure comes from the very poor feedback much of speech softwares give to learners [2]. As learners build upon their native language, our tutor will provide users with comparative knowledge on the prosody of French and English languages. This will enable learners to be aware of the specificities of each language and student's main production problems. Audio-visual feedbacks, exploiting speech signal transformations and guided by L1 and L2 prosodic knowledge, have been developed in order to correct student's own production so as to make him/her aware of the acoustic correlates expected [3]. An example of our strategy is given in section 4.

In the vein of Philippe Martin's Winpitch [4], our software is intended to be used, at least during the first steps of learning, with the help of teachers.

Our toolkit is made up of two main components.

- a set of speech tools, including speech signal edition and transformation, now available from any MS Word or PowerPoint files.
- An interactive course, comprising a set of progressive exercises, and exploiting all the functionalities of our speech software. The set of exercises, conceived by teachers of English, is intended to make learners aware of prosody in general and French and English specificities in particular. A database of sentences uttered by native speakers of English, is presently recorded. It will serve as a support for the lessons and as references for the correction of students' productions.

In this paper, we present our speech tools for language learning (section 2), the interactive course (section 3), and give a concrete example (section 4).

### 2 Speech tools for language learning

#### 2.1 Signal edition and transformation tools

Our signal editor, WinSnoori [5], includes a set of functions for the edition and the transformation of signal waveforms. Among the main editing functions, we can mention the display of speech signals and spectrograms, F0 contours and intensity curves, as well as speech play back and labelling. Each segment of the signal can be edited, played back, zoomed and modified.

Signal transformations have been developed using an improved version of TD-PSOLA. It exploits a pitch marking algorithm together with a

dynamic resampling algorithm that allows high quality stimuli to be generated [6]. The user can modify the prosodic cues such as duration, fundamental frequency and intensity (independently or at the same time) of every part of the signal. These modifications can be designed by hand in a very simple way: the user monitors curves with the mouse directly on the spectrogram. After this operation, the signal is resynthesized and can be saved.

During exercises, the comparison with a reference followed by the modification of the learner's own voice shows him/her what is expected. Since this modification can be done either by the teacher (at least during the first steps of learning) or the student, we'll use the term "user" to design the teacher or the student indifferently. The user corrects the prosodic parameters that have been wrongly realized, on a segment where the error is the most interesting. As examples, the user can modify the intensity and/or the duration of an accentuated vowel, or change the final melodic contour of a question. Since other acoustic cues (segmental or supra-segmental), are not affected by the manipulation, the student can compare his/her original realization with the locally corrected one, and better apprehend what drives it apart from the target.

Among other useful modification functions, the user can cut copy and filter a segment and listen to the modified signal.

## 2.2 Snorri Active X components.

Software must be user friendly and simple enough so that teachers and students can adapt themselves to it easily. For that purpose we have ported the main functions of our speech analysis software in the form of ActiveX controls that can be easily inserted in any Microsoft Office document (such as Word or PowerPoint), or tutoring tools such as Authorware (Macromedia), as well as web pages. In addition to the labelling tools, notes linked to the signal or spectrogram can be used to focus the attention of the reader and explain some salient phenomenon about various domains (acoustics, prosody, ...). Then it is very simple to prepare speech tutorials for students. Snorri ActiveX Components will be available from our website at <http://www.loria.fr/equipements/parole>

## 2.3 Automatic alignment

It is of great importance to have at one's disposal a good annotation and alignment since a precise

detection of segments is necessary to compare the learner and the reference sentences and to modify speech prosody.

Annotating, which can be done in sounds and in words under WinSnorri, localizes the segments on their different visual representations. The user can annotate the signal or visualize an annotation file. When this file doesn't exist, and this is generally the case for student's realizations, the sentence can be annotated by automatic alignment software our team has developed. First, some probable phonetic transcriptions are generated from the orthographic transcription, then the software try to determine the exact emplacement of each sound in the signal. This last task is performed with an automatic speech recognition system (ASR), and is facilitated by the knowledge of the segments to be found and of their phonetic context. Since ASR systems are trained with native speakers, we had to adapt our system to French learners speaking English. We are presently collecting sentences from the Timit corpus uttered by young French speakers in one secondary school and two universities of Nancy for that purpose. We have already at our disposal more than two thousands sentences. This corpus will serve to train our HMM models with non-native speakers' pronunciations.

## 3 Acquisition of English prosody

A set of exercises has been designed by teachers of English as a foreign language. This set is intended to make learners aware of prosody in general (lexical stress, rhythm and intonation) and of French and English prosodies in particular. The exercises exploit all the possibilities offered by our signal editor, including modifications of prosodic cues and filtering of speech signals, as well as the facilities of Snorri Active X components. The possibility of copying and modifying exercises will be offered to the users through the generation of scripts.

The initiation stage includes exercises with delexicalised sentences such as "mamama" sentences, or filtered sentences, from various languages, in which only prosodic cues are perceptible. This step is followed by general notions about French and English prosodic specificities. The aim of these lessons is to enable students to distinguish French and English languages without any lexical cues.

After the initiation stage, exercises with natural sentences focus on specific tasks. We are presently collecting sentences from native speakers of English, that will serve both as a support for lessons and as references for comparisons with

students' productions. A very simple example illustrating our strategy is given in section 4.

#### 4 Example

In France, teaching of English prosody for non-specialists insists on the place and the strength of English lexical accent as well as on the main intonation patterns. One of its goals is also to make learners aware of French prosody. In this chapter, we show how our toolkit improves prosody learning, taking the acquisition of the English lexical accent as an example.

English lexical accent is very different from the French one, first because its place is free, whereas the French one is fixed, but also because it is very well marked on an acoustical point of view. Indeed, the stressed syllable is more intense, higher and longer than the unstressed ones. Furthermore, unstressed syllables are sometimes reduced: its vocalic timbre comes close to that of a neutral vowel. The French accent is essentially characterized by a lengthening of the last syllable of a group of words.

In order to show how our toolkit put in light mispronunciations, we will consider the example of the word « important », uttered by a 12 years old French girl, with two years of English lessons. The acoustical characteristics of the stressed syllable “por”, uttered by an English speaker, its high pitch, long duration and great intensity, are very well marked and visible (Fig. 1, centre). We also note that the stop /p/ is aspirated, -there is a weak noise just before the vowel /o/. If the user select (by a simple mouse click) the last syllable “tant” and listen to it separately many times, he could realize that this syllable is just a nasal murmur. The French student is not aware of this reduction and pronounces the syllable with its full timbre (Fig. 1, top). The French learner realisation is characterized by an attenuation of the English accent (the syllable “por” is less long, less high and less intense, with respect to other syllables, than in the reference) and the persistence of a French accent (lengthening of the last syllable, the stressed syllable in French). We have corrected the pitch and the durations on the learner's realisation with Snorri Active X components (Fig.1, bottom).

#### 5 Conclusion

Our strategy is based upon a comparison between a sentence uttered by a native speaker and the learner's production. The association of visual

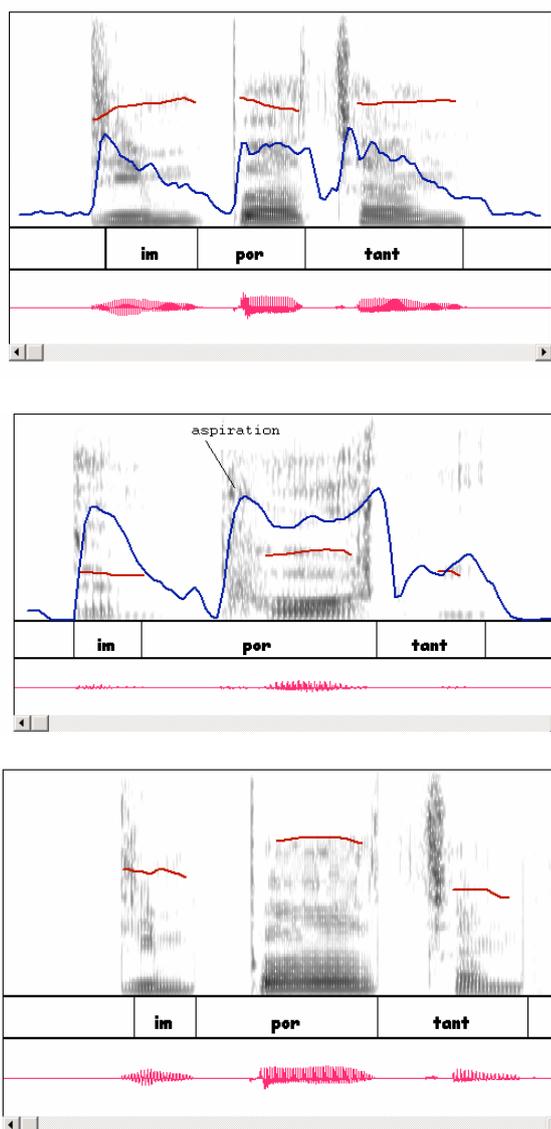
and auditory feedbacks makes the learner aware of the differences between the correct pronunciation and his/her own, which improves his perception of the second language. Then the user can try to bring his realization closer to the target and to improve his pronunciation.

Although these kinds of modification can be implemented automatically to a certain extent (see [7] for instance), a crucial issue is how to make learners aware of the deviations with respect to the expected acoustic correlates. This issue is all the more difficult since signal transformations presented before must be used carefully to prevent any acoustic and/or prosodic artefact. The work of Makarova [8] can be used to predict the strength of modifications to guarantee that learners perceive exaggerated correct or incorrect acoustic correlates in the case of Russian and Japanese. We have the project of conducting the same kind of perception experiment in the case of French and English.

It is important to note that, if visual feedback gives objective measures to compare the learner's utterances with that of the references, the interpretation of these measures is far from trivial. The help of the teacher is necessary to decide what is pertinent or not and guide the beginner in his/her initiation. This raises the issue of developing a simplified model of the target prosody that could give some information about the expected acoustic cues. Even if it is hard to imagine a general model it should be possible within the framework of single words or simple sentences. This is an objective for our future work.

#### 6 Acknowledgements

Our thanks go to teachers of English as a foreign language of the Lorraine region.



**Figure 1: Word “important” uttered by a French speaker (top) and by a native speaker of English (centre). Correction of the French speaker pronunciation (bottom), realized with Snorri Active X components. The non-continuous line, superimposed onto the spectrogram, represents the F0 contour, the continuous line the energy.**

## References

- [1] J. Anderson-Hsieh, “Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction”, *CALICO*, Vol 11.4, 1994.
- [2] D.M. Chun, “Signal analysis software for teaching discourse intonation” *Language Learning and Technology*, Vol. 2, 1. 1998.
- [3] A. Bonneau, K. Balci, Y. Laprie and V. Colotte (2003) « Design and development of a computer-assisted learning of prosody” Proceedings of the *International Congress of Phonetic Sciences*, Barcelona, Spain.
- [4] A. Germain and P. Martin, “Presentation d’un logiciel de visualisation pour l’apprentissage de l’oral d’une langue seconde” , *ALSIC*, Vol. 3. 1., 2000
- [5] Y. Laprie “Snorri, a software for speech sciences”, *MATISSE*, 1999.
- [6] V. Colotte and Y Laprie, “Higher pitch marking precision for TD-PSOLA”, Proceedings of XI European Signal Processing Conference (EUSIPCO), Toulouse, 2002
- [7] Y. Meron and H. Keikichi, “Language training system utilizing speech modification”, Proceedings of the fourth International Conference on Spoken Language Processing, pp 1449-1452, Philadelphia, 1996
- [8] V. Makarova, “Perceptual correlates of sentence-type intonation in Russian and Japanese”, *Journal of Phonetics*, Vol. 29.2, 2001.