ON FORMAL IDENTIFICATION OF SOME PROBLEMS IN NON-NATIVE FRENCH PRONUNCIATION

*J. F. Malet * and N. Vigouroux***
**California State University, Sacramento, USA*
***Laboratoire IRIT, URA-CNRS n° 1399, Toulouse, France*

ABSTRACT

Description of a method to identify non-native learners' phrase-level pronunciation problems. Devised to inform efforts at automatic detection through HMM modelizing, the method is based on a threefold concern with: 1) preparation of relevant phonetic identifiers, through 2) robust acoustic evidence, applied to 3) training data that is also suggested by traditional didactic practices Examples discussed in light of evidence seen in preparing units for modelization.

INTRODUCTION

In a previous publication [1] a set of informally identified pronunciation problems —commonly encountered by English-speaking learners of French— were discussed from the point of view of potential automatic detection and correction.

Formal identification and reliable detection of such problems —as they occur within speech data supplied by learning subjects— require that they be defined and classified according to a method that interacts with the traditional practices used in non-native language acquisition; as opposed to a mere off-shoot application of phonetic science with a distinct, not easily related didactic module (e.g., SPELL Project [2]).

As it turns out, whereas some mispronunciation items are easily identifiable from the dual point of view of their respective theoretical inception levels (e.g., articulatory, phonotactic, morpho-syntactic, etc.) and of their typical acoustic manifestations, some other items are much more difficult to pinpoint and, even more so, to characterize either theoretically or datawise. A threefold investigative method, catering to the above considerations, is therefore attempted.

From a corpus of 56 French phrases (offering wide phonemic and intonative variety) read (under uniform technical recording conditions) by 42 American English-speaking learners of French, a number of apparent pronunciation problems are selected for a multi-level confirmation of their likely manifestation within the acoustic signal. For each problem taken up, data-analytic strategies are sought, aiming at 1)HMM modelization of phonetic units and 2)at relevant phonotypical mapping into an adapted organic version of the evaluation grids, traditionally used in testing for oral proficiency —e.g., accuracy, delivery rate, stress, phonemic quality, etc.

Phrase-level examples are discussed in light of initial observations, made in preparing for modelization of units that are formal phonetic identifier candidates.

METHODOLOGY

Ultimately, identification of the specifics that constitute a language learner's problem-realization, requires a reliable decisional process. And such a process has better chances of succeeding, of course, if it is highly predictive of variability within the test-data submitted to it.

To achieve this with our method, training-data is prepared with two main considerations in mind: 1)selecting this training data (from a set of learners's realizations) for its efficient problem-oriented value and 2)ascertaining physical evidence of apparent pronunciation problems.

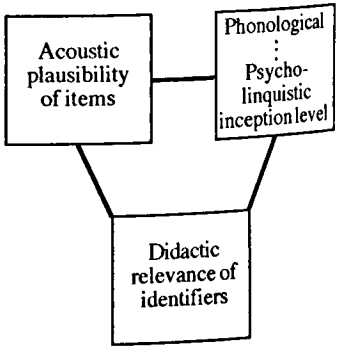

Figure 1. Three-pronged approach to selection of training-data items and formalization of phonetic identifiers.

Training-data Selection

This is a process that can begin with informal observation of non-native accent. However, although common stereotypification of "foreign accents" can initially be somewhat useful, it is preferable to turn to more informed ways, such as long language-lab or classroom experience.

The procedure for choosing training-data is strictly governed by a triad of major concerns that lies at the heart of our method (Figure 1).

In choosing the items, which are to be acoustically modelized in order to acknowledge and/or create useful, formal phonetic identifiers, all observed problem units of speech —and relevant supra-segmental phenomena associated to them— have to be eventually subjected to a decisional process embodying:

- a pertinence strategy guided by didactic needs,
- an acoustic robustness clause excluding speculative observation, and
- a proposal for formal classification of candidate units within a larger, ontological system .

1.The didactic area of concern involves:

a) the preparation of a relevant corpus directly dictated by the content and purpose of a given lesson,

b)once the speech specimens are collected, the preparation of competence groups among the learners —recognized by human expertise,

c) the selection of common learning problems (e.g., in reading aloud, deciphering efforts vs. easy and intelligent reading).

2.The concern with acoustic robustness aims at establishing the physical basis of problems perceived in speech. This involves:

a. the preparation of acoustico-phonetic (AP) data: choice of subjects, collection of speech specimens, manual segmentation and labelling of relevant signal data (a step based on observation of wave-form, spectral display, frequency analysis and listening[3]),

b. a classification of AP data files in competence groups, in general with respect to perceived quality of pronunciation, but also in particular with respect to mastery of chosen difficulties (See examples),

c. a further classification of AP data-files with a view to retaining data that is reliably observable, and suggesting the use of well mastered parameters to bring out potential candidate units or clusters to formal phonetic status.

3.The third area of concern involves the detection of the (most likely) inception level of problem units of realization; whether articulatory, phonotactic, ..., lexical, syntactic, supra-segmental (prosodic), or of higher order (i.e., linguistic and psycho-linguistic levels sharing their competence with didactic concerns —e.g., deciphering versus intelligent reading, as in First Example, below). Determination of the inception level of problem-realizations is necessary to clinch the formalization process of identification. It calls on the joint expertise of teachers, phoneticians, linguists, psycho-linguists, etc..

Summary Of Purpose

Basically, we are looking for types of corpus realization that can be HMM-modelized and confirmed as formal identifiers of a non-native kind of speech.

However, given the high degree of variability, this search cannot be a random one and is better served by a didactic line of progress.

An identifier does not necessarily have to be a special interlanguage unit so long as it can serve an interim learning strategy [4]; it can be a known phoneme (or an infra-phonemic unit, or a phonemic cluster, or again a supra-segmental phenomenon) that is realized in a specific location where it should not occur (e.g., /t/ in First Example, below) or again that is not realized, or hardly so, when it should be (/p/ or [m] in Second Example).

EXAMPLES

Two sample sets of AP datafiles are now briefly examined and commented.

First Example

Our database contains the phrase (actually a sentence): Et c'est sain. (Eng.: "And it's healthy too.") with barebone phonetic transcription /eses$\tilde{\epsilon}$/. This phrase has been selected on the following grounds:

1) it involves twice the same consonant, /s/, which is easy to realize,

2) it uses the same vowel, /e/ twice in a row; inducing assimilation of the third vowel /ɛ̃/ to [ẽ], especially with American Anglophones as such speakers are not used to an [ ]-producing buccal aperture concomitantly with essential nasalization,
3) its metric structure is somewhat anapestic and it can possibly be stressed for rhythm on the last syllable.
As a result, it should not be difficult to choose a didactically useful phonotypical transcription leading to comparison of non-native realizations. Such a theoretical templet can entail:
• no pause (minimum phrasal continuity, reasonably demanded from learners),
• a strictly unvoiced fricative (/s/),
• an easily perceptible and measurable time structure with variants quite directly reflecting nuances in the meaning of the phrase,
• a strict exclusion of any diphthong,
• some fading on the median vowel is preferable to avoid a staccato effect.

A set of 41 learners supplied the voice data. These were divided into 3 groups according to the perceived resemblance of their speech to either French or (American) English pronunciation: Grp.F (15 speakers) was perceived as fairly francophone, Grp.A (17 speakers) as dominantly anglophone. Speakers in the third group were difficult to assess.

Not all speakers realized a sixth phoneme /n/ but, in Table 1, figures are adjusted on the virtual presence of this sixth phoneme.

Mean duration:	Fr.-type 780 ms	Ang.-type 1 963 ms	
% phrase time		Ratio MeanEn./time	
Franco.	Anglo.	Franco	Anglo
e 16.7	e.. 14.9	5.77	4.92
s 19.4	s 16.8	5.71	4.95
e 13.2	ei 17.0	7.15	4.60
s 20.3	s 15.6	5.60	4.95
ɛ̃ 28.4	e 26.8	7.60	2.64
ŋ 10.5	ŋ 14.8	6.10	4.10

Table 1.

As can be observed, both Grp.F and Grp.A realized a somewhat anapestic metric structure, devoting nearly half of phrase-time to the last syllable. However, a wide difference is to be noted between the two groups in that a slight fading of the median vowel occurs in Grp.F, relieving the staccato effect perceived in Grp.A. The monotone ratio figures of mean energy to duration of units might also account for the perception of such an effect. The phenomenon might in fact be owed to strenuous efforts at deciphering each syllable as it comes up in the reading, as opposed to a more competent, flowing rendering of the written corpus.

Aside from these global observations, a number of unwanted pauses were detected. Two cases of extreme fading of the median vowel with a resulting centralizing to /ə/ and an English type of accentuation on the last syllable. On the other hand, ten speakers realized this median vowel taking anywhere between 150 and 320 ms, with five of these realizing a perceptible dipthong [ei]. Six speakers realized a [t] with five of them supplying a phonetic profile [etses] and one [etsetsə].

All such unwanted units are potential candidates for potential modelization and identifier status.

Second Example

The French phrase *un pseudo voeux* (Eng.: "a pseudo wish") was chosen for three didactic *a priori* 's:
1) the gender distinction of the singular indefinite adjective,
2) the requirement to delete an English phonological rule (initial occl.+ /s/ —> /s/),
3) the mastery of vocalic timbres for /ø/ and final /o/.
Out of 37 learners, who read the phrase, two different types of realizations turned out for the first three syllables: respectively, along a fairly francophone phonetic scheme /œ̃psødo.../ (12 speakers) and along a quite definitely anglophone profile [œ̃nsyIdʰəo...] (15 speakers) where [yI] symbolizes a vowel that can be perceived either as some French /y/ or as an English /I/ but is neither. This [yI] can be considered a good candidate to become an interlanguage vocalic identifier.
In Table 2, it can be seen that the anglophone type of subjects took, on average, some 45 % more time to realize the first three syllables of the phrase. While a more detailed examination of the data points to, at least, three areas of discrepancy in the duration of the phenomena looked at:
1) francophone-type [p] and (in this specific context only) its *de facto* substitute [n], with a possibility of treating [œp+ s] and [øn+s] as clusters, for /œ̃[m]sødo/ is

Mean duration:	Fr.-type 1 034 ms	Ang.-type 1 499 ms	
% phrase time		Ratio MeanEn./time	
Franco.	Anglo.	Franco	Anglo
œ̃ 12.2	œ̃ 6.1	5.93	6.41
p 9.2	n 4.5	4.86	7.62
s 16.6	s 12.6	5.26	3.42
ø 8.2	yI 4.2	8.52	8.59
d 6.4	dʰ 4.2	7.60	7.36
o 17.0	əo 49.9	3.89	

Table 2.

also a native francophone type of realization
2) francophone-type [ø] and interlanguage candidate identifier [yI],
3) francophone-type [o] and American English-type [əo] or [əɔ].
Whereas [p] and [n],in this context, reveal a very different ratio of mean energy to duration —4.8 vs 7.62— vowel realization is not as clearly differented by this ratio and calls for frequential parameter definition (through formant tracking or, more recently, noise in certain frequency bands [5]).

CONCLUSION

The method, presented in this article, offers seemingly endless investigative possibilities, while it remains *de facto* contained by the practical and realistic demands of assisted non-native language acquisition.

At the same time, whereas a certain tendency to sprawling investigation is inherent to a thorough observation of bulk AP data, such a tendency is checked by the necessity to justify the creation of ever more formal identifiers —for these must eventually be fitted in the larger context of a language science.

However, as an empirical tool serving didactic purposes whence it is in great part derived, this method appears promising.

Obviously,the training-database required to achieve automatic detection of mispronunciation in test-data supplied by learners, has to be considerably enlarged so has to enable HMM modelizing of potential formal phonetic identifiers.

REFERENCES

[1][Malet 1992] J.F. Malet, G. Pérennou & N.Vigouroux, "Repérage automatique d'unités acoustico-phonétiques pour l'enseignement de la prononciation française," *Journal de Physique*, Suppl. IV, 1992.

[2][Lefèvre 1992] J.P. Lefèvre, M. Jack, C. Maggio, M. Recife, M. Savino & L. Santagelo, "An Interactive System For Automated Pronunciation Improvement" in *Proceedings Of ICLSP-92*, Banff, Canada.

[3] *La parole et son traitement automatique*, CALLIOPE, Eds. J.P. Tubach, L.J. Boe, P. MArtin, J. Caelen, J.M. Pierrel R. Descout, C. Sorin, J.J. Mari-ani (Massons: Paris, 1989).

[4] N. Yamada, "Japanese Accentuation Of Foreign Learners And Its Interlanguage," ICSLP'94, Yokohama, pp. 1227-1230. IV, 1992.