Manipulation of foreign-accented speech: improving English-accented French

Colleen C. Martin
Laboratoire de Phonétique et Phonologie (UMR 7018)
CNRS/Sorbonne Nouvelle
19, rue des Bernardins
75006 Paris, France

ABSTRACT

This paper examines the effect of manipulation on a natural voice in the segmental and prosodic domains. Sentences and CV syllables produced respectively by two different American English speakers were modified to match homologous French productions by a single native French speaker. The fundamental frequency (F^O) patterns of the sentences were adjusted to match the French model, and the vowels and voiceless plosive consonants of the CV syllables were adjusted to reduce diphthongs and consequently length, or to eliminate consonant aspiration. The results of a perception test by native French speakers showed that the modifications of the F^O pattern over the entire sentence and the modifications of vowels for diphthongal quality and duration significantly remedied the foreign accent of the non-native French productions. On the other hand, consonant aspiration modification showed no significant improvement. Manipulated stimuli were preferred overall which shows that manipulation *can* improve foreign-accented speech.

1. INTRODUCTION

A foreign accent not only marks the second language (L2) learner as a non-native (favorably or not as the case may be) but it can also inhibit communication between the L2 learner and the native speaker. Pronunciation learning at both the segmental and suprasegmental levels is, therefore, very important for the non-native speaker. It is hypothesized that modification of the natural voice signal can help to reduce the perception of a foreign accent, although acoustic signal manipulation is, at this stage, too primitive to completely eliminate the foreign accent .

Previous studies [4, 5, 8] have shown that manipulation of the acoustic signal of natural speech can improve intelligibility. The present study similarly compares the acceptability of distorted speech before and after manipulation of certain prosodic and segmental elements. Here American English-accented French is manipulated to modify F^O pattern, vowel length and diphthongization, and stop consonant aspiration.

The eventual refinement of speech manipulation methods may lead to a whole new field of "speech airbrushing". Not only may it prove a helpful tool in the media, but it may prove as dangerous as digital photography and filmography in its potential to cover or change (photographic or aural) truth. This study was conducted in the hope of finding a new educational tool for learners of a foreign language.

The field of phonetics is traditionally divided into two fields: segmental phonetics (consonants and vowels) and suprasegmental phonetics (prosody). In order to attain a high level of oral ability in a foreign language, it is essential to master not only the phonetic and phonological rules, but also the intonation and rhythm of the language.

Pitch differences between languages are the result of the different priority given to different prosodic parameters. For example, even though there exists a universal syllable lengthening phenomenon at the end of intonation groups, the relative importance of this phenomenon varies. In French, this end lengthening is dominant. However in English lexical stress, which may not occur on the very last syllable of an intonation group, may take precedence [9].

Traditionally, French is known for its stable monophthongal vowels whereas English vowels tend towards length and diphthongization [1, 7]. It has been well established that during the production of voiced stops in French, the vocal cords vibrate during the entire length of the stop and that this trait alone is sufficient to differentiate between French /p, t, k/ and /b, d,g/. The so called "voiced" consonants of English, /b, d,g/ are often devoiced, especially when they occur non inter-vocalically. English voiceless plosive consonants in initial position are differentiated from their "voiced" counterparts by aspiration [2, 3].

2. METHOD

The goal of this study is to show the effect of modification of the acoustic signal on foreign accent perception. Sentence F^O was adjusted using the computer

ISBN 1-876346-48-5

program *Praat* [6] to match pitch patterns. CV syllable vowel duration and diphthongization were adjusted by cutting out part of the vowel. Consonant aspiration was adjusted by simply cutting aspiration to match the length of the French model.

2.1 SPEAKERS AND CORPUS

Two corpora were used. To produce the first, one native French speaker (Paris, 26 years) and a native English speaker (Upstate New York, 24 years) read seven neutral declarative sentences (five to twelve syllables) in French. The native English speaker had studied French for four years in high school (age 14-18) but had never spent time in a French-speaking country.

To produce the second corpus, the same native French speaker read a list of framing sentences, "J'ai dit ___, pas ___" containing French CV syllables composed of the six French plosive consonants /p, t k, b, d, g/ each followed by the vowels, / i, ɛ, a, o, y/. A native English speaker (Southern California, 22 years) similarly read a list of framing sentences, "I said ___, not ___" containing English CV syllables composed of each of the six English plosive consonants /p, t, k, b, d, g/ followed by the English vowels / i, e, æ, o, u/. Note that the symbols used here are the traditional phonemic symbols used to describe English and French. These particular vowels were chosen after an initial study comparing the formants of the vowels in English and French. The vowels presented here are those which resemble each other the most. In fact, the English /e/ and /o/ are usually diphthongized and the Californian /u/ is quite anterior though less rounded than the French /y/. French /a/ is quite anterior and hence closer to English /æ/ than /ɑ/. The English monophthong /ɛ/ could not be compared to its French counterpart because it does not occur in open syllables. Each speaker pronounced 30 syllables: the six stops each followed by the five vowels.

2.2 SPEECH MODIFICATION

Measurements were taken of the duration of each phoneme and the first four formants of the vowels (at each 1/3 of the vowel) in the CV syllables, and of the F^O (every 8 milliseconds) in the sentences.

The modified English F^O contours were modeled after the native French F^O contours. Because this work concentrated on F^O *patterns*, the modification was a modification of the *slope* or *shape* of the F^O contour rather than a replacement of pure numbers. This ensured that the F^O slopes were distributed correctly over the phonemes which sometimes varied widely in duration between the two speakers. It also minimized changes in the English speaker's natural pitch.

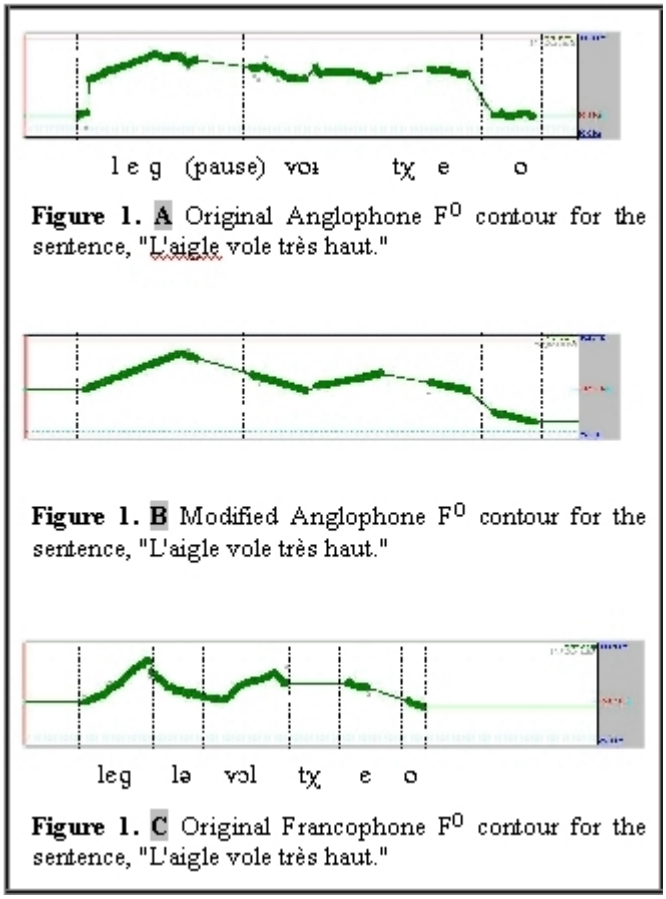

Figure 1. A Original Anglophone F^O contour for the sentence, "L'aigle vole très haut."

Figure 1. B Modified Anglophone F^O contour for the sentence, "L'aigle vole très haut."

Figure 1. C Original Francophone F^O contour for the sentence, "L'aigle vole très haut."

The goal of the F^O manipulations was to copy the global contour of the French model. Micro-melodies caused by consonants were not taken into account. The occasional cases of creaky voice produced by the English speaker were conserved.

The modified contours did not, of course, change length but were applied over the phonemes in exactly the same manner as in the French speaker's productions. For example, in one sentence, " A l'aube, la fille jette mille roses vers la ville", the French speaker has a rising F^O slope during the first 42% of the vowel /o/ in the word "aube", followed by a falling slope for the rest of the vowel. The 42% mark was calculated for the same phoneme produced by the English speaker, according to the duration of her production in order to faithfully imitate the French speaker's slope pattern.

The English word stimuli were modified in two ways. One set of stimuli underwent modification of vowel length and diphthongization. In the second set, the aspiration of the initial consonant of the English speaker's productions was reduced to match the native French speaker's corresponding stimuli (The release was kept). The aspiration of the stop consonant was measured from the burst to the onset of the F2 of the following vowel. The English aspiration was 63 ms longer on average.

ISBN 1-876346-48-5

It was hypothesized that manipulation of vowels would be easier working from an English production towards a French production rather than in the opposite direction because the diphthongal quality of English allows for a choice of formant patterns. Especially when working with diphthongized phonemes such as /e/ [eɪ] and /o/ [ow], it is often possible that at least a part of the English production resembles the French monophthongal vowel while the rendering of a French monophthong into a diphthong would be much more complicated. The same goes for consonant aspiration which is easier to synthetically cut out than to add in.

2.3 TESTING PERCEPTION

Untrained native French speakers from Paris and the surrounding region were asked to take a perception test, the results of which were used to determine the relative importance of the modifications made to the English speakers' original productions. Their previous knowledge of English or other languages was not taken into account. The French sentences were presented to ten adult Parisians. The French and English CV syllables were presented to eighteen adult Parisians. The original and manipulated stimuli were presented in pairs to the native francophone subjects who were then asked to judge which sentence in each pair seemed the most "French".

3. RESULTS

Conforming with the basic hypothesis of this study, while modified stimuli tended to be preferred over the original English-accented stimuli, the French productions were overwhelmingly preferred. As expected, *according to the perception test results, French listeners preferred the native French stimuli, both for CV syllables and phrases, over the English accented stimuli 100% of the time.*

When the original sentences in English and those modified for the fundamental frequency were compared, the modified stimuli were chosen 67% of the time. The difference in score was found to be significant at the $p < .01$ level (two-tailed t test; $t = 4.042$, $df = 15$).

When the original English CV syllables and the English CV syllables modified to reduce vowel segment duration and diphthongization were compared, the modified stimuli were chosen 76% of the time. The difference in score was found to be significant at the $p < .0001$ level (two-tailed t test; $t = 5.799$, $df = 15$).

When the original English CV syllables and the English CV syllables modified for consonant aspiration reduction (using only voiceless stops) were compared, the modified stimuli were chosen only 55% of the time. The difference in scores was found not to be significant, $p > .10$ (two-tailed t test; $t = .631$, $df = 8$).

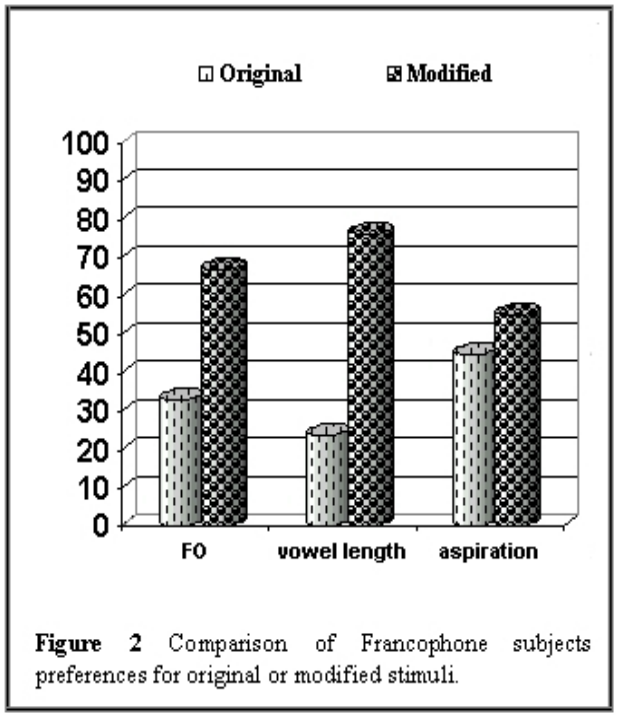

Figure 2 Comparison of Francophone subjects preferences for original or modified stimuli.

This result is surprising because the amount of modification was significant (An average 63 ms were cut from the syllables which averaged only 333 ms to start with.). Also in the author's own experience as a native English speaker living in France, voiceless stop aspiration is one of the most recognizable traits of an English accent in French so it would be expected that this kind of modification would be important in reducing a perceivable foreign accent.

4. DISCUSSION

It was hypothesized that the French phrases and the CV syllables pronounced by the native American English speakers would be identified as "more French" by native French speakers once the original stimuli had been corrected by three different kinds of modifications: F^O pattern, vowel length and diphthongization, and voiceless stop aspiration.

When looking at the English productions, the results show that manipulation can correct or at least improve a foreign accent. Especially when comparing between the original and manipulated productions for F^O and vowel length, the modified stimuli were chosen overwhelmingly. The results of this study show that correction is viable for a foreign accent.

Modifications of the English speaker's F^O to match the F^O pattern of the native French speaker as well as the modifications of duration of the vowel segments and the

ISBN 1-876346-48-5 © 2003 UAB

consequent diphthong reduction were statistically important. Only the reduction of initial consonant aspiration seems not to have had an effect on native French speakers' judgments.

5. CONCLUSION

It is impossible to objectively and quantitatively compare the three types of manipulations performed in this experiment. Therefore it would not be prudent to make a global statement comparing the importance of one factor to the other.

However, the fact that the manipulated stimuli were preferred overall shows that manipulation *can* improve foreign-accented speech.

This study was conducted in the hope of finding a new educational tool for learners of a foreign language. The learner may be greatly benefited by not only hearing another's native pronunciation, but by hearing his own original pronunciation compared with his corrected pronunciation. It would also help the learner to concentrate on a single element such as F^O pattern, vowel duration, or consonant aspiration in his own voice according to his needs.

This study has shown, above all, that manipulation of a foreign accent is a viable tool, and that manipulation can correct or at least improve a foreign accent.

REFERENCES

[1] P. Delattre, *Comparing the Phonetic Features of English, French, German and Spanish*, Heidelberg: Julius Groos Verlag, 1965.

[2] J. Flege and W.S. Brown, "The voicing function between English /p/ and /b/ as a function of stress and position-in-utterance," *Journal of Phonetics,* vol. 10, pp. 335-345, 1982.

[3] J. Flege and J. Hillenbrand, "A differential effect of release bursts on the stop voicing judgments of native French and English listeners," *Journal of Phonetics,* vol. 15, pp. 203-208, 1987.

[4] B. Maasen and D.J. Povel, "The effect of correcting fundamental frequency on the intelligibility of deaf speech and its interaction with temporal aspects," *Journal of the Acoustical Society of America*, vol. 76, pp. 1673-1681, 1984.

[5] B. Maasen and D.J. Povel, "The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech," *Journal of the Acoustical Society of America*, vol. 78, pp. 877-886, 1985.

[6] *Praat,* version 4.0.28, P. Boersma and D. Weenink, www.praat.org

[7] D. O'Shaughnessy, "A study of French vowel and consonant durations," *Journal of Phonetics*, vol. 9, pp. 512-521, 1981.

[8] K. Tajima, R. Port and J. Dalby, "Effects of temporal correction on intelligibility of foreign-accented English, "*Journal of Phonetics*, vol. 25, pp. 1-24, 1997.

[9] J. Vaissière, "Language-independent prosodic features," in *Prosody: Models and Measurements*, Cutler and Ladd, Eds., pp. 53-66. Berlin: Springer-Verlag, 1983.