# Application of Acoustic Analysis in English Phonetics Teaching

**Rastislav Sustarsic**
University of Ljubljana, Slovenia
E-mail: rasto.sustarsic@ff.uni-lj.si

## ABSTRACT

Taking into account the prevailing articulatory and auditory approach in phonetics and phonology classes (even at tertiary level), the aim of the paper is to underline the importance of instrumental (acoustic) data in supplementing and enriching the traditional educational approaches in this area.

After briefly discussing applicability of sound synthesis and sound recognition tools in pedagogic environment, I focus on speech analysis, in particular on those computer tools that enable analysis and visual display of both segmental and prosodic features of speech.

Considering my specific situation of teaching English Phonetics to Slovene students of English, the focus of the paper is on contrastive analysis of segmental and prosodic features of English and Slovene. Some of the main differences between the two languages, which often lead to mispronunciations by Slovene students, are illustrated on the basis of instrumental analysis and spectral display.

## 1. INTRODUCTION

It is well-known that articulatory phonetics is often conveniently applied for the description and teaching of consonants, while auditory phonetics is often claimed to be more appropriate for the description and teaching of vowels. Both fields of phonetic study and their applications, however, can and should in my view today be supplemented by making use of the various possibilities offered by the information technology.

I have tried elsewhere (Sustarsic, 2001) to demonstrate some of the applications of speech synthesis and speech recognition computer software for teaching/learning English phonetics and carrying out research in this area. While briefly summarizing the ideas proposed at international meetings devoted to phonetics pedagogy and research, the present paper focuses on a third area of speech technology, more widely and directly relating to phonetics study and research, namely acoustic analysis (rather than text-to-speech or speech-to-text conversion). There are nowadays so many user-friendly, low-cost and even freely distributed programs that allow reliable and precise measurements/analyses of speech signals that there is simply no excuse for those teachers and researchers in the field of phonetic study who persist in restricting themselves to the 'good old' articulation-and-perception approach.

## 2. SPEECH SYNTHESIS

Speech sound synthesis programs seem to be an appropriate point of departure for showing students the straightforward results of changing basic parameters of sound, i.e. duration (length), amplitude (loudness) and frequency (pitch) of individual sounds or sounds in sequence, i.e. longer speech segments. One of the programs that I have used for this purpose is MBROLA (developed at the Faculté Polytechnique de Mons, Belgium). Once the students have become familiar with the set of phonetics symbols, they can experiment freely by changing the numeric values determining the duration and the frequency flow of individual segments, along with the overall pitch/loudness/tempo of the utterance they have composed by way of sequential ordering of sounds. The changing of any numerical data can immediately be assessed by listening either to a sound in isolation or in combination with the preceding or following sounds. One can also insert pauses between parts of the utterance, thus creating more 'realistic' word groups reminiscent of actual speech rather than 'prepared' reading. Figure 1 shows a short synthesized speech segment (3 words), with numbers to the right of each sound, defining the duration of consonants and duration and frequency of each vowel.

```
E 40 0 102
m 50
b 50
r 30
@U 80 5 119 35 126 70 140
l 50
@ 50 50 173
w 100 75 133
V 30 85 114
z 60 75 101
d 60
@ 40
v 40 85 105
E 60 75 121
l 50 70 121
@ 60 60 150
p 50
t 70
```

Figure 1 (above): A short segment of synthesized speech, saying 'MBROLA was developed'


ISBN 1-876346-48-5 

Although I have done more research in the other two areas of instrumental phonetic applications (speech recognition and speech analysis), I believe that speech synthesis programs can be widely used in introductory teaching of some of the most important prosodic features of speech, and that a number of very creative tasks can be developed for students by relatively simple manipulation of numerical data, before proceeding to the more complex analysis of sound waves and spectral displays.

## 3. SPEECH RECOGNITION

While the applications of speech synthesis will mostly be restricted to producing or studying different aspects of 'artificial' speech, speech recognition will usually be based on a certain segment of 'real' speech, i.e. actual oral expression in a particular language. This means that students can be encouraged to record their own speech (in our case English speech produced by Slovene speakers) and then apply a speech recognition program (such as the well-known Via Voice or Dragon Naturally Speaking systems) to convert their utterances to text.

The preliminary results I obtained when testing Via Voice by reading a set of short texts (for details, seeSustarsic, 2001) show that speech recognition can be applied in phonetics (or more precisely in pronunciation) teaching, and that a number of aspects of articulatory and auditory phonetic principles can be observed in the way that speech recognition programs transfer (or fail to transfer) the received speech signals into written form.

It is interesting, however, that (perhaps due to the noise picked up by the microphone and interpreted as sound) in addition to the expected omissions of the so-called 'weak' (unaccented) forms, we can observe a number of unwanted additions to what has actually been articulated by the speaker. As one might expect, it is almost exclusively the grammatical (parts of) words which are added or omitted, i.e. affixes, articles, prepositions, auxiliaries and pronouns, in which the schwa vowel occurs instead of a full vowel. The message presented in the form of text thus often preserves full intelligibility, which can be compared to the actual human speech production and reception, characterized by focusing on lexical (i.e. content) words and backgrounding the grammatical (functional) words which can often be predicted from the different types of context.

In my experience, both additions and omissions are outnumbered by the more intriguing 'substitutions' or 'misinterpretations'. Table 1 shows a sample of general vocabulary (rather than technical vocabulary, abbreviations or proper names, which seem to give different results), with the 'recognized' items on the left and the actually spoken words on the right.

| General Vocabulary |
|---|
| 21st year – *20 1st year* |
| a – *of* |
| a – *the* |
| a cult – *adult* |
| a power – *at our* |
| a tart depart – *at our department* |
| a the – *at the* |
| a very – *three* |
| about – *adult* |
| deferred – *differ* |

Table 1: Some general vocabulary mis-recognitions (on the left) of the words and phrases on the right

While 'unknown' technical terms, abbreviations and proper names are often a total mystery for a speech recognition program, resulting in amusing 'word salad', the errors in the case of general vocabulary are usually less outrageous and can frequently be explained in terms of phonetic similarity and in terms of misinterpretation of word boundaries.

The very first example is most interesting: a human language-user would never notice any ambiguity in the phrase *twenty first-year students,* simply because *twenty-first year students* makes no sense in any education system, but for a computer program it is perfectly acceptable.

We also find a number of examples of homophony here, such as neutralisation of weak forms 'of, a, are' etc., and amusing results of deletion of unaccented syllables, such as 'a tart depart' for 'at our department'.

In order to draw some conclusions on the applicability of the ViaVoice program in English pronunciation teaching and learning, I decided to focus on the general vocabulary, where the most interesting phonetic misinterpretations seem to occur, involving the above mentioned homophony, as well as a number of misinterpretations (and possibly also some mispronunciations) of individual sounds occurring in connected speech. Among the consonants, the weak obstruents (in particular /b, d, ð, and v/) and the sonorants (in particular the nasals and /l/) seem to be confused the most often, but with students who, due to the influence of Slovene, tend to neutralise several pairs of 'similar' English vowels, one would expect more of the following cases of mismatch: even – *given,* fanatics – *phonetics*, drummer – *drama,* sturdy – *study,* etc. These are some of the mistakes which clearly indicate the possibility of using the program in pronunciation teaching. Others concern the above-mentioned word-boundary mismatch, e.g.:

- ◊ use ability - *usability*
- ◊ yields beach - *yield speech*
- ◊ die a chronic – *diachronic*

ISBN 1-876346-48-5  

There are also examples of confusion of lenis/fortis consonants, e.g.:

- ◊ lags – *lax*
- ◊ your needs – *units*

It should be pointed out that using any speech recognition program with English pronunciation students has several other justifications. Firstly, the program needs to be trained to one's voice, which requires a great deal of loud reading. The basic rule is: the more you train the program, the higher will be the accuracy of recognition, and thus the usefulness of the program for any practical task (such as reading out an e-mail message, for example). Once the program has attained the level of accuracy around 95%, the student will be even more motivated to switch from the keyboard to speaking. And this is undoubtedly something that all teachers of phonetics would be only too pleased to see.

A word of caution though: speech recognition programs deal better with relatively slow, clear and 'careful' (non-assimilated) style of speech, which might encourage students to train the 'foreign speaker' word-by-word pronunciation - precisely the opposite of what we want them to achieve. On the other hand, this may not be such a serious problem if we instruct students to apply the more natural colloquial style already in the process of training the computer to their voices, following the simple philosophy that the computer is the tool of the user and not the other way round.

The fact that these programs have not been developed for our particular purpose is by no means a disadvantage, since we can use them to fulfil the main purpose of foreign-language pronunciation training, namely that of making one's speech understood – admittedly, in this case, by a machine rather than another human being.

## 4. SPEECH ANALYSIS

A range of free and low-cost computer tools are available nowadays for the analysis and visual display of both segmental and prosodic features of speech. Considering my specific situation of teaching English Phonetics to Slovene students of English, my basic application of such tools lies in the contrastive analysis of phonetic and phonological features of English and Slovene. I have selected some of the main differences between the two sound systems which often lead to mispronunciations by Slovene students, and performed instrumental analysis to see how they reflect on waveform and/or spectral display. The selected features which I have analysed contrastively for the two languages concern the following features:

- ♦ aspiration of voiceless plosives
- ♦ devoicing of sonorants by voiceless obstruents
- ♦ release masking of plosives
- ♦ pre-fortis clipping of vowels
- ♦ neutralization of English vowels
- ♦ places of articulation in English/Slovene

Two of the features are illustrated contrastively for the two languages in Figures 2 and 3 below, using the PRAAT program.

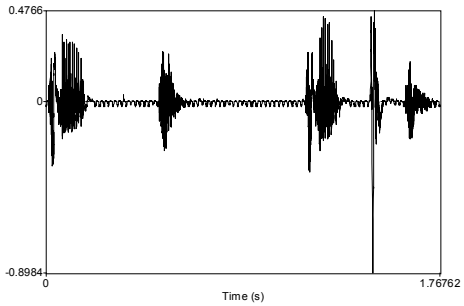


Figure 2: Waveforms showing inaudibly (left) and audibly (right) released /k/ in the word 'act'

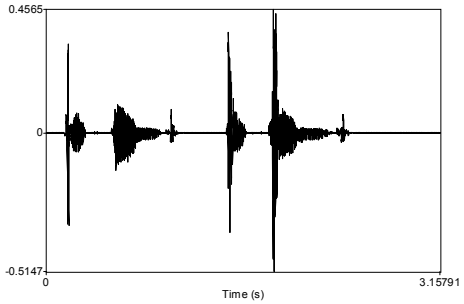


Figure 3: Waveforms showing unaspirated /t/ (left) and aspirated /t/ (right) in the word 'pretend'

Some of the typical errors in the pronunciation of English by Slovene students' are considered to reflect these differences and are ascribed to the transfer of Slovene pronunciation features to English. Typically, Slovene lacks the features of aspiration, devoicing and release masking which are characteristic of Standard British English (as well as a number of regional accents). It is still unclear to what extent this is also true of the reduction of vowels before strong consonants, but again this feature seems to be disregarded by Slovene learners in terms of neutralizing the length of vowels before weak and strong consonants. A different kind of neutralization can be observed in students' simplifying of the English vowel system, again most likely due to the fact that the Slovene vowel system is much less complex. These differences between the two systems can be clearly represented on a vowel chart, provided we have the necessary average data concerning the first two formants (F1 and F2) for speakers of both languages. Finally, the important differences in places of articulation of English and Slovene consonants (e.g. glottal versus velar) can also be observed when comparing the spectra of the sounds in question.


ISBN 1-876346-48-5 

However, similarly to speech synthesis programs, also speech analysis software can be used at a very basic and general level, when students are acquiring the concepts of frequency and amplitude (which can be shown in the form of waveforms), as well as in studying the basic differences between vowels and consonants on the one hand, and sonorant consonants and obstruents on the other. Further differentiation, namely that of individual vowel qualities, can then be studied by introducing the above mentioned concept of formant structure, and from there on by using the knowledge of formant structure in plotting different vowels on the vowel chart.

Of course, segmental analysis is only the first step in the multitude of possible applications of acoustic analysis in phonetics pedagogy. When moving on from realizations of individual sounds to sounds in sequence, various assimilatory processes can be studied and observed in spectral displays. In addition, various prosodic features can be studied, such as word accent and pitch variation, the acoustic analysis of which can either support or change our approaches with regard to syllable prominence and sentence intonation.

Perhaps the most challenging task of students and researchers in the field of acoustic phonetics is the so-called 'reading' of spectrograms, which is much more demanding than even the notorious reading of text from phonetic or phonemic transcription. Similarly to solving a difficult crossword puzzle, reading spectrograms demands a great deal of knowledge about acoustic features of individual sounds and the influence of sounds on those following and preceding them. We often need to provide certain clues to avoid making the task too difficult, for example (the idea I got from Martin Barry, Department of Linguistics, University of Manchester) by telling the students that the spectrogram represents the title of one of Shakespeare's plays, the name of a famous actor/actress, etc. Even when the number of possibilities is relatively small, it may not be easy to work out the meaning of the spectrogram unless we have sufficient knowledge of the display of acoustic differences between vowels and consonants to begin with, then obstruents and sonorants, plosives and fricatives, sibilant and non-sibilant fricatives, nasals and laterals, etc.

## 5. CONCLUSIONS

I have tried briefly to discuss some applications of acoustic phonetics (speech synthesis, speech recognition and speech analysis) in the teaching and learning of English Phonetics, in particular from the point of view of English-Slovene contrastive analysis. It should perhaps be pointed out that at tertiary level the students, once they have become familiar with the main features of the tools used for acoustic analysis, could be assigned a number of different practical and research tasks in the three areas. Ideally, these tasks should lead the students beyond the area of (contrastive or general) phonetics, and show them the usefulness of phonetic research in the frame of their overall linguistic research and use of information technology.

## REFERENCES

[1] R. Sustarsic, "Using a Speech Recognition Program in Teaching English Pronunciations", in *PTLC2001: Proceedings of the Phonetics Teaching and Learning Conference,* J. Maidment, and E. Estebas-Vilaplana, Eds, pp 47-50. University College London, 2001.

[2] MBROLA Speech Synthesizer. Version 3.02a. http://tcts.fpms.ac.be/synthesis/mbrola.html

[3] ViaVoice 98. Version 5.2. IBM Corporation 1998.

[4] PRAAT 4.0.4 by P. Boersma and D. Weenink. www.praat.org.