Feedback design philosophy in the Computer Assisted Language Learning systems

Degtyarev M. A., Krinov S. N., Marchuk Yu. N.,
Moscow State Regional University

misha_degtyarev@mail.ru, skrinov@mail.ru

Abstract

Most of commercially available Computer Assisted Language Learning (CALL) systems teaching phonetics require a student to judge on his own whether or not his pronunciation is correct. However he is not always capable of evaluating his own speech in an adequate way without help from the other party. In this article we suggest an approach to teaching foreign language pronunciation with the help of CALL system in which the process of evaluation of pronunciation and generation of suggestions for its correction is fully carried out by a computer program. We will present a general model of the effective Computer Assisted Pronunciation Training (CAPT) system as we see it. We also will consider the matter of feedback in CAPT systems, e.g. the response of the computer program on the student’s attempts to reproduce presented language material. We will outline some basic requirements to the effective feedback and discuss pros and cons of the feedback used in modern CAPT systems. At the end we will present a form of feedback that we consider to be effective.

1. The general model of the CAPT system

Basically any effective CAPT system has to fulfill all the functions of a tutor giving private lessons to a student at home. In general its work can be described as follows: it has to provide the student with some input (it can be words, phrases, short texts), apprehend his/her output (reaction) and provide some sort of feedback (it can be mere grading remark or a sort of advice that can help the student to correct his mistake). In Fig. 1 this model is presented in a more formal way.

![Input Teacher \(\rightarrow\) Output Student \(\rightarrow\) Feedback Teacher](image)

Fig. 1 General model of the language learning process

These are three main components of any language learning process [1]. Let us consider every one of these items more closely.

1.1 Input

According to observations of scholars working in the field of second language acquisition students studying foreign language pronunciation have to be trained on phrases. Only when learning a foreign language by phrases the student can learn to sound natural. Any efficient CAPT system has to have a rich input. Its form is up to developers. It can be phrases, proverbs, poems, dialogues, short stories, radio interviews and so on.

1.2 Output

The way the learners react to the input, e.g. student’s output, also has SM\textsubscript{L2} to be considered. Since the goal of the pronunciation training is to teach the learners to speak correctly most of the output should be oral. The student can be asked to repeat the phrase or read it from the screen. Also it has to be decided whether the students simply do “listen-repeat” exercises or carry a conversation with the use of a predefined set of phrases.

1.3 Feedback

The form of feedback may vary, whoever the effective feedback should match certain definite criteria [1]. Before we look at these criteria lets consider the forms of feedback that are widely used in modern CAPT systems. One of the most popular forms of feedback is a waveform of the students attempt to reproduce the model phrase. The student is to compare the waveform of his attempt to reproduce the model phrase with the waveform of this model phrase and then adjust his pronunciation in one way or the other. This form of feedback can be successfully used in training supra-segmental elements of speech like rhythm and intonation. However, it is hard to prove its effectiveness in teaching pronunciation of particular sounds. Some scholars argue that the only benefit this kind of feedback might have is the increase in amount of time spent on particular phrases. Other systems present feedback in a less confusing and more interactive way. Instead of (or with) presenting the waveforms they also grade students pronunciation. Seemingly this kind of feedback is more effective because it creates more of a classroom like learning environment, when the teacher grades the student.
Fig. 2 General model of the CAPT system
But simply knowing that the attempt was not successful or even knowing what particular segment of the phrase was mispronounced is not sufficient, at least for beginners. It was proven that students learning foreign language develop the skill of correcting their own mistakes only after at least 100 hours of training under professional foreign language teacher.

Taking these drawbacks of the feedback used in modern CAPT systems into account lets outline some criteria of effective feedback. Scholars argue that effective feedback in CALL in general and in CAPT systems in particular has to be (1) provided in a real-time, (2) pertinent and (3) easy-to-interpret [1]. Second and third criteria are fully related to pedagogical aspect of language teaching whereas first criterion is related to the implementation of Automatic Speech Recognition (ASR) technologies in CAPT systems.

In a sense the computer program has to carry out the functions of a tutor who instructs the students during their learning process. Just like the tutor it has to identify the mistakes made by learner and present a real-time, pertinent, easy-to-interpret feedback. Neither waveforms nor grading match all the criteria of the effective feedback. Waveforms can hardly be easy-to-interpret. Also it is doubtful that they are pertinent at least for earning segmental units of speech. Grading is much easier to interpret then the waveforms but is not sufficient for correcting mispronunciation.

2. Model of the CAPT system

At the end we will present a form of feedback that we consider to be more effective but now let us look at the model of the CAPT system. It is presented in the Fig. 2. It consists of three main blocks that correspond to three items of the learning process discussed above. In this article we present very general idea of what elements can be included in each of three blocks. The ultimate selection and organization of all the items in the model is up to developers.

2.1 Input block

In the input block of the CAPT system all input shall be stored. Basically input is all the language material that is to be presented to the students in their study course. It has to be stored in this database. The language material can be anything from isolated sounds, syllables and simple words to phrases, poems, short texts, radio interviews and etc. Since it is a pronunciation training system all the input shall be presented both in written and oral form.

It shall be decided how the input will be presented. Will it be exercises, drilling aimed at learning to pronounce some particular sounds or just a set of common phrases that a student will study one by one? It is purely pedagogical aspect of CAPT system development.

2.2 Output block

The student has to react to the input in a certain way. Since we are dealing with pronunciation training in most cases he will produce some kind of speech signal. Usually it will be mere repetitions of the input. Student’s speech signal shall be recorded and stored for the future references (for instance playback). Since we aim at developing a system that will grade the students and correct their pronunciation absolutely automatically, recorded speech signal has to be segmented and recognized. The analysis of spectrum, rhythm and intonation has to be done. For this some common speech analysis methods can be used.

2.3 Evaluation block

After the student’s speech signal is analyzed it can be processed further and evaluated. Student can make a mistake in pronunciation and it shall be detected. Based on this feedback can be generated.

It is important to define how the mistakes or mispronunciations can be determined automatically. It is quite tricky matter. The computer program has to have some definite guidelines based on which the judgment whether the student’s pronunciation is correct or not can be made.

In most CAPT systems ASR technologies are implemented in the following way: the student’s speech signal is compared with the model speech signal; the program grades the “similarity” of these two signals and based on its value the program goes on to the next phrase or asks the student to repeat the phrase one more time.

Some programs based on “similarity” grade the student. The grade can be presented in different ways. It can be presented as a grading bar showing the degree of similarity or as a encouraging “Good job!” or demanding “Please, try again!”! Some more advanced programs provide some visual presentation of the students speech signal (waveform, spectrogram or transcription) with the erroneous segment indicated.

This approach to grading based on similarity of two signals was borrowed from general speech recognition systems. Surely it makes the fancy CALL systems that are attractive to the students. However, it is questionable that this kind of analysis can be used for generation of the adequate feedback that can be utilized by the student to correct his/her mispronunciations.

We have a few arguments to support this statement. First of all, the system utilizing similarity as a guideline for grading students in a sense does not operate with phonetically meaningful information. The “similarity” is being calculated in a purely mathematical way. This means that even if the particular sound pronounced erroneously is detected correctly, the system will not be able to provide any information on the nature of the mispronunciation. At best this kind of feedback will only help the student to realize what sound he did not pronounce correctly. As
we have said before this is not sufficient for most users of CAPT systems.

Secondly, it is questionable that the system working the way described above can correctly identify the definite segments of the student’s speech signal corresponding to mispronounced sounds. It can also mistake the correctly pronounced sounds for mispronunciations.

There is a great theoretical study underlying these arguments. We cannot present it in full in this article. We will only outline the problem and then describe the solution as we see it.

Every person has his own unique voice. When trying to imitate the teachers voice or prerecorded speech the student even when speaking “perfectly” still speaks in his own unique way. Even when pronounced correctly learners output cannot be identical to the teachers input, thought the phonetic model of foreign language is reproduced correctly. In the acoustic space the representations of source and students sounds also do not have to match because even every other pronunciation of the particular sound produced by one speaker cannot be identical to other pronunciations of the same sound produced by the same speaker even in the same context [2].

However we see a solution to this problem. It is known that there is some definite correspondence between spectral parameters of vowels and articulation. Therefore mistakes in student’s articulation can be detected on basis of the formant tracks and some other parameters. Formally the sound system of any language can be presented as a relation between vowels in an acoustic space (sound model). A native speaker of a certain language has his own unique sound model. For him to learn the pronunciation of the other language means to acquire the sound model of that language. It can be said that the sound model of his native language has to be transformed into the sound model of the foreign language to be learned. This process is illustrated in Fig. 3.

\[
\text{SM}_{L_1} \quad \Rightarrow \quad \text{SM}_{L_1/L_2} \quad \Rightarrow \quad \text{SM}_{L_2}
\]

Fig. 3 Transformation of the sound model of the native language (\(\text{SM}_{L_1}\)) into the model of the foreign language to be learned (\(\text{SM}_{L_2}\)). \(\text{SM}_{L_1/L_2}\) is a transitional model in which the sounds from the native language still have some influence on \(\text{SM}_{L_2}\).

In order to implement this idea, before the training starts the sound model of the student’s native language has to be built. For this he might be asked to read some specially designed short text which will be automatically analyzed. Based on the information extracted during this analysis the sound model of the native language can be built. Then this model shall be automatically transformed into the corresponding model of the foreign language to be learned. This way every student will have his “own” particular \(\text{SM}_{L_2}\). This model will be used in evaluation of the student’s speech and also in generation of feedback during his learning process.

2.4 Feedback block

Since we see the process of acquiring the foreign language pronunciation as transformation of the native language sound model, we suggest the following feedback ideology. This kind of feedback has been successfully implemented in a regular class-room foreign language lessons. It is aimed at correcting some particular sounds pronunciation. It is very simple in essence and it can be formalized as follows:

\[
\text{Sound}_{FL} \text{ in Word}_{FL} \text{ is pronounced as }
\]
\[
\text{Sound}_{NL} \text{ in Word}_{NL} \text{ only with X,}
\]

where \(\text{Sound}_{FL}\) is the sound from the foreign language word (\(\text{Word}_{FL}\)), \(\text{Sound}_{NL}\) is the sound from the native language word (\(\text{Word}_{NL}\)) and \(X\) — is an articulatory hint that will help the student to pronounce the desired sound correctly by changing the articulation of some sound from his native language. Here is the example of the implementation of this formula for Russian speaking student learning English: “The sound [c] in the word car is pronounced just like [a] in Russian word kak only with the mouth almost closed”. “…Only with the mouth almost closed” is the articulatory hint that should help the Russian speaker to produce the corresponding English sound with least effort. This kind of feedback proved to be quite intuitive enough for Russian speaking students just starting to learn English as a foreign language.

Together with this kind of hints the synthesized version of the mispronounced phrase can be presented. Modern methods allow to change all the mispronounced segments of the student’s output so that he will hear the correct phrase spoken by his own voice. This kind of feedback shall increase student’s incentive to learn speak correctly.

3. Conclusion

We have presented the general model of the CAPT system. Our main goal was to show how it is possible to create a CAPT system, a computer tutor, that will help the students to learn the pronunciation of a foreign language. Basically we just outlined some problems. Surely this matter requires further elaboration.

4. References
