Секція 5. Лінгвістичні технології інформаційної діяльності

Mispronunciation Detection for CALL Systems

Mariana Holubinska

Applied Linguistics Department, Lviv Polytechnic National University, 12 S. Bandera street, Lviv, UKRAINE, E-mail: m.holubinsk@gmail.com

Abstract – the problem of control and synchronization of video- and audioinfromation tempo stays topical for Computer-Assisted Language Learning systems if the problem is to improve pronunciation or listening comprehension. The paper deals with two methods used for pronunciation enhancement.

Keywords – CALL, pronunciation controller, mispronunciation detection.

I. Introduction

It has been proved that multimedia environment helps foreign language learners to develop both oral and written skills. To perceive contextual information is possible only when sounds are produced clearly, what depends on the human speech production system movements.

Systems based on automatic speech recognition (ASR) technology provide important functionality in CALL applications. This is young but growing area of research motivated by the large number of students studying foreign languages.

II.Pronunciation Controller

While learning a foreign language, a very important constituent of the process stays the learner's pronunciation. Here below is an algorithm suggested in [2]:

1. the problem becomes easier to be solved if translate a given algebraic problem into a geometric one, which can be visualized better. Having solved the geometric problem, it is translated into its previous algebraic domain;

2. the translation from one domain into other is done by using Fast Fourier Transform and its inverse, which can be computed both in time **O(n log n)**, where **n** – the number of points.

In the case it is necessary to compare a sentence pronounced by the teacher with the same sentence but repeated by a learner. Thus, the problem deals with speech matching. That is why each sentence that has to be pronounced by the teacher is stored in a database that consists the following entries: a key to identify the sentence in the program, the sentence in digital (voice) format and the same sentence in text format. The pronunciation controller program works as follows:

▪ when the program requires a sentence, it retrieves it from the database **D** by using the search key;

▪ the program sends the digital version of the sentence to the speaker and stores the sentence in text format into the array **P**, where two consecutive words are delimited by a single blank;

▪ the student repeats the sentence in front of a microphone, and simultaneously the sentence is converted into text (via voice/speech recognition program) that is stored into the array **T**;

▪ then the longest common subsequences of both arrays **P** and **T** are computed to identify the words which have been pronounced correctly;

The algorithm work results in an obtained list of positions in which two arrays (i.e. two sentences) differ. The next step is to store words mispronounced into a dictionary **M** and associate to each word a weight (a kind of *hit/miss ratio*), which is related to the number of times that the word has been pronounced *correctly/incorrectly*.

III.Hidden Markov Model to detect mispronunciation

Mispronunciation detection shares much in common with automatic speech recognition technology and majority of mispronunciation detection systems use statistical models such as Hidden Markov Models (HMMs) in order to detect mispronunciation. Most recent works have attempted to improve the detection accuracy by using extended recognition networks that incorporate cross-language phonological rules, which are rules that dictate how a learner's first language affects his/her pronunciation of a second language. The problem of mispronunciation detection has gained significant interest in the last two decades.

The most robust systems are the Mel-frequency cepstral coefficients (MFCC) which are based on Mel-scale [3]. Though Adaptive Frequency Cepstral Coefficients system has better performance than MFCC has with various commonly used frequency scales. A fundamental criterion for comparing the performance between the Mel-scale and the word adaptive frequency scale is the group separation between the native speakers (with correct pronunciation) and the non-native speakers (with incorrect pronunciation) with respect to their distribution HMM scores.

Conclusion

Software tools for mispronunciation detection provide learners with visual aids such as waveform displays, plots of pitch contours, and spectrograms. The design of multimedia materials stays a significant area of research in CALL.

References

1. Holubinska M., "Computer Assisted Language Learning", Proceedings of the VIth International Scientific and Technical Conference CSIT 2011, pp.276-277.
2. Gentile E., Plantamura P., "Multimedia linguistic learning", IEEE International Conference Proceedings, vol.1, pp.7-10, 2000.
3. Zhenhao Ge., Sudhendu R. Sharma, Mark J.T.Smith, "Adaptive Frequency Cepstral Coefficients for Word Mispronunciation Detection", 4th IEEE International Congress on Image and Signal Processing, pp.2388-2391, 2011.