Robust Recognition of Vowels In Speech Impediment Therapy Systems

Dénes Paczolay, András Bánhalmi, and András Kocsor

Learning to speak with damaged or missing auditory feedback is very dificult. It would be a real help for the hearing impaired if a computer were able to provide a real-time visual feedback of the quality of the uttered vowels. For speech impediment therapy and teaching reading the SpeechMaster software package was developed in 2004 [7] and was financially supported by the Hungarian Ministry of Education. The members of the Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged were the coordinators of the software development, and SpeechMaster is freely available from our Internet site: http://www.inf.u-szeged.hu/beszedmester.

The software package has been downloaded numerous times and many people use it successfully at many schools and at home. The speech impediment therapy part of SpeechMaster was tested at the School for the Hearing Impaired in Kaposvár. The results of the experiments show that the project achieved its ambitions [5]. Following therapy each young subject pronounced vowels more intelligibly, and the number of mistakes became noticeably fewer. Hence the use of the SpeechMaster can indeed speed up the learning of the utterance of vowels to a significant degree. The speech therapy usually starts in early childhood with introductory exercises like loudness and pitch control drills. The success of our package among children is due to its simple user interface and the playful sound formation exercises. Often children open up more quickly and easily and start uttering sounds using a computer than with a therapist who they do not know. The eficiency of the software can be attributed to two factors. First, the therapy is based on a series of varied and customizable drills. Second, the software produces effective real-time vowel recognition with machine learning [2, 9, 12] and gives a clear visual feedback. Although the project came to an end in 2004, the development of software is still ongoing. We fixed several bugs and added some new features. Currently we are working on improving the accuracy of the real-time vowel recogniser based on experiments and test results.

The goal of the research is twofold. The first goal is to make the software package more user independent. This is needed because the recogniser has to provide on objective rating of the quality of the uttered vowels for all sorts of speakers. To achieve this goal we tried applying real-time speaker normalization [3] and classifier combinations [4]. We found these techniques to be valuable for improving the recognition accuracy [6, 8]. In the current system we divide our vowel database into male, female and child speaker types. We trained three distinct recognisers on these data sets, and trained another to separate these three groups and detect which one is active. In this article we examine what happens if we create categories automatically using a machine learning algorithm. We performed our tests with 3-8 categories using the K-Means method and the Unweighted Pair Group Method with Arithmetic Mean [10] with several configurations.

The second goal is the research to improve the accuracy and stability of the recogniser near the boundary of vowels in the transient period. This is required when the students practice the pronunciation of vowels in words. In the current realisation we trained three-three recognisers separately on isolated vowels and non-isolated vowels e.g. which were uttered in words (separately pronounced vowels and ones uttered in a word). Then we get the scores of the quality of the uttered vowels as follows. First we select the active group (man, woman or child), next we evaluate the two active classfiers (isolated and not-isolated ones), and lastly we get a result using a classifier combination. To improve the stability we tried out new schemes, like training a machine learner to decide whether the actual sound is a vowel or a nasal or if it is a border. Earlier the SpeechMaster package applied three-layer Artificial Neural Networks [1], but in this paper we describe the effects of using a Core Vector Machine [11] classifier as well.

References

[1] C.M. Bishop. Neural Networks for Pattern Recognition. *Oxford Univ. Press*, 1995.

[2] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. *John Wiley and Son*, New York, 2001.

[3] E. Eide and H. Gish. A parametric approach to vocal tract length normalization. In *ICASSP*, pages 1039-1042, Munich, 1997.

[4] T.K. Ho. Data complexity analysis for classifier combination. *Lecture Notes in Computer Science*, 2096:53-68, 2001.

[5] A. Kocsor. Acoustic technologies in the speechmaster software package. VI(2):3-8, 2005.

[6] D. Paczolay, L. Felföldi, and A. Kocsor. Classifier combination schemes in speech impediment therapy systems. *Acta Cybernetica*, 17:385-399, 2005.

[7] D. Paczolay, A. Kocsor, Gy. Sejtes, and G. Hégely. A 'beszédmester' csomag bemutatása, informatikai és nyelvi aspektusok. IV(1):57-79, 2004.

[8] D. Paczolay, A. Kocsor, and L. Tóth. Real-time vocal tract length normalization in a phonological awareness teaching system. *In Text Speech and Dialogue*, volume 2807, pages 4-37, Czech Republic, 2003. Springer.

[9] L.R. Rabiner and B.H. Juang. Fundamentals of Speech Recognition. *Englewood Cliffs, NJ, Prentice Hall*, 1993.

[10] P.H.A. Sneath and R.R. Snokal. Numerical Taxonomy: The Principles and Practice of Numerical Classification. *W.H. Freeman and Company*, 1973.

[11] I.W. Tsang, J.T. Kwok, and P. Cheung. Core vector machines: Fast svm training on very large data sets. 6:363-392.

[12] V.N. Vapnik. Statistical Learning Theory. *John Wiley and Son*, 1998.