Analog Circuits for Real-Time Spatiotemporal Feature Extraction of Acoustical Signals

Athanasios Mouchtaris

Advisors: Jan Van der Spiegel, Paul Mueller

University of Pennsylvania

ABSTRACT

Current state-of-the-art automatic speech recognition (ASR) systems perform reasonably well on clean speech. Their performance, however, deteriorates considerably in presence of noise or any mismatch between the testing and the training environments. On the other hand, the human system is capable of a much more robust performance even in the presence of significant noise. One of the reasons behind this deficiency in ASR systems is the limited knowledge about the human auditory system, speech characteristics and perception.

In the first part of our work, we investigated the problems of auditory-based processing and feature-based recognition, so as to improve the robustness of speech recognition systems. A new auditory-based front-end speech processing system was developed, called the Average Localized Synchrony Detector (ALSD). We also investigated the acoustic-phonetic characteristics of the obstruents in the framework of a front-end feature-based speaker-independent phoneme-based continuous speech recognition. We studied several acoustic features for their information content and their possible role in the recognition. The features that proved to be vital and rich in their information were extracted and new rule-based algorithms were developed for manipulating these information-rich features for ASR.

In the second part of our work, our focus was to extend the aforementioned research by the addition of a front-end, which will offer the advantage of enhancing the capabilities of that system in cases of performance degradation. We focus on two particular problems of deterioration of ASR, (a) due to inter-speaker variability, and (b) due to noisy background conditions with very low Signal-to-Noise (SNR) ratios. For addressing case (a), we proposed applying voice conversion methods that have been popular within the area of Text-To-Speech (TTS) synthesis. For this project, we were interested in conversion techniques that offer the possibility of feature *transformation,* which can be useful when the initially extracted features are far from those that the system has been designed for. We proposed a new conversion algorithm that addressed previous shortcomings of existing algorithms in this area, with great success. For case (b) we followed the novel approach that the problem of speech conversion has similarities with that of speech enhancement, when the source speech becomes the noisy speech, and the target speech becomes the clean speech. Our results have shown that our feature conversion techniques can efficiently estimate the clean speech features from the noisy speech features, resulting in a significant improvement in performance especially in very noisy conditions.

Finally, another direction where we concentrated on, was towards implementing the initially proposed rule-based system. We performed a great amount of work towards the design and successful implementation of a real-time low-cost front-end for speech recognition that realizes the rule-based algorithm that was proposed in the initial stages of this research.

PUBLICATIONS

[1] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "An Acoustic-Phonetic Feature-based System for the Automatic Recognition of Fricative Consonants", in Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-98)*, vol. II, pp. 961-964, 1998.

[2] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Acoustic-phonetic features for the automatic recognition of stop consonants", *Journal of the Acoustical Society of America*, pp. 2777-2778, 103 (5), 1998.

[3] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Acoustic-Phonetic Features for the Automatic Recognition of Stop Consonants", in Proc. *16th International Congress on Acoustics (ICA) and 135th Meeting of the Acoustical Society of America* (*ASA*), pp. 275-276, 1998.

[4] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Automatic Detection and Classification of Stop Consonants using an Acoustic-Phonetic Feature-Based System", *XIVth International Congress of Phonetic Sciences (ICPhS'99)*, pp. 1709-1712, 1999.

[5] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "A GUI System for Speech Synthesis through Graphical Manipulation of Spectrograms", in Proc. *IEEE International Symposium on Circuits and Systems (ISCAS-99)*, pp. III-106 - III-109, 1999.

[6] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "An Acoustic-Phonetic Feature-Based System for Automatic Phoneme Recognition in Continuous Speech", in Proc. *IEEE International Symposium on Circuits and Systems (ISCAS-99)*, III-118 - III-121, 1999.

[7] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Auditory-Based Acoustic-Phonetic Feature Extraction for the Segmentation and Recognition of Continuous Speech", *The 33rd Annual Conference on Information Sciences and systems, CISS'99*, March 17-19, 1999, The Johns Hopkins University, Baltimore, Maryland.

[8] A. M. Abdelatty Ali, "Auditory-based acoustic-phonetic signal processing for robust continuous speech processing", Ph.D. Thesis, University of Pennsylvania, December, 1999.

[9] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Auditory-based speech processing based on the average localized synchrony detection", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2000)*, Vol. 3. pp. 1623-1626, 2000.

[10] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Speech processing using the average localized synchrony detection ", *Journal of the Acoustical Society of America*, pp. 2908, 107, 2000.

[11] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller, "Robust Classification of Stop Consonants using Auditory-based Speech Processing", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2001)*, vol.1, pp. 81-84, 2001.

[12] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller , "Auditory-based signal processing for robust speech recognition", *35th Annual Conference on Information Sciences and Systems (CISS) - Neuromorphic Engineering and MEMS Sensory Systems,* March 2001, Baltimore.

[13] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller , "Acoustic-Phonetic Features for the Automatic Classification of Fricatives**",** *Journal of the Acoustical Soc. of America (JASA)*, 109 (5), pp. 2217-2235, May 2001.

[14] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller , "Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants**",** *IEEE Trans. Speech and Audio Processing*; Vol. 9, no. 8, pp 833-841, Nov. 2001.

[15] A. M. Abdelatty Ali, Jan Van der Spiegel and Paul Mueller , "Robust Auditory-Based Speech Processing Using the Average Localized Synchrony Detection", *IEEE Trans. on Speech and Audio Processing*, Vol. 10, pp. 279-292, July 2002.

[16] Athanasios Mouchtaris, Jan Van der Spiegel, and Paul Mueller, "Non-Parallel Training for Voice Conversion by Maximum Likelihood Constrained Adaptation", in Proc. *IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP-2004)*, vol. 1, pp. 1-4, May 2004.

[17] Athanasios Mouchtaris, Jan Van der Spiegel, and Paul Mueller, "A Spectral Conversion Approach to the Iterative Wiener Filter for Speech Enhancement", in Proc. *IEEE International Conference on Multimedia and Expo (ICME-2004)*, June 2004.

[18] Athanasios Mouchtaris, Jan Van der Spiegel, and Paul Mueller, "Non-Parallel Training for Voice Conversion based on a Parameter Adaptation Approach", submitted *IEEE Trans. Speech and Audio Processing,* May 2004.

[19] Athanasios Mouchtaris, Jan Van der Spiegel, and Paul Mueller, "A Low-Cost Real-Time Implementation of a Bandpass Filtering Front-End for Speech Recognition," to be submitted *IEEE International Symposium on Circuits and Systems (ISCAS-2005)*.