

Abstract

Automatic speech recognition (ASR) is a computerized speech-to-text process, in which speech is usually recorded with acoustical microphones by capturing air pressure changes. This kind of air-transmitted speech signal is prone to two kinds of problems related to noise robustness and applicability. The former means the mixing of speech signal and ambient noise usually deteriorates ASR performance. The latter means speech could be overheard easily on the air-transmission channel and this often results in privacy loss or annoyance to other people.

This thesis research solves these two problems by using channels that contact to the human body without air transmission, i.e. by vibrocervigraphic and electromyographic methods. The vibrocervigraphic (VCG) method measures the throat vibration with a ceramic piezoelectric transducer contact to the skin on the neck, and the electromyographic (EMG) method measures the muscular electric potential with a set of electrodes attached to the skin where the articulatory muscles underlie. The VCG and EMG methods are inherently more robust to ambient noise and they make it possible to recognize whispered and silent speech to improve applicability.

The major contribution of this dissertation includes feature design and adaptation for optimizing features, acoustic model adaptation for adapting traditional acoustic models onto different feature spaces, and articulatory feature classification for incorporating articulatory information to improve recognition. For VCG ASR, the combination of feature transformation methods and maximum a posteriori adaptation improves the recognition accuracy even with a very small data set. On top of that, additive performance gain is achieved by applying maximum likelihood linear regression and feature space adaptation with different data granularities in order to adapt to channel variations as well as to speaker variations. For EMG ASR, we propose the Concise EMG feature that extracts representative EMG characteristics. It improves the recognition accuracy and advances the EMG ASR research from isolated word recognition to phone-based continuous speech recognition. Articulatory features are studied on both VCG and EMG ASR to analyze the systems and improve recognition accuracy. These techniques are demonstrated to be effective on both experimental evaluations and prototype applications.