

ACOUSTICS2008/754

A reverberation-robust automatic speech recognition system based on temporal masking

Guy Brown^a and Kalle Palomäki^b

^aUniversity of Sheffield, Dept. of Computer Science, Regent Court, 211 Portobello Street, S1 4DP Sheffield, UK

^bAdaptive Informatics Research Centre, Helsinki University of Technology, P.O. Box 5400, 02015 Espoo, Finland

Previously, we have proposed a reverberation-robust system for automatic speech recognition (ASR) based on a temporal masking principle. In the first pathway of this system, speech is analysed by a bank of auditory filters, in order to provide acoustic features for the recogniser. In the second pathway, a bandpass modulation filter (1.5 Hz - 8.2 Hz) detects regions of the envelope in each filter channel that contain strong speech energy. Regions of the modulation filter output that exceed a threshold are labelled as reliable evidence for the speech in a time-frequency mask; regions that fall below the threshold are dominated by reverberation and labelled as unreliable. The acoustic features and time-frequency mask are then decoded by a 'missing data' ASR system. Here we describe modifications of this system that bring it into closer agreement with purported mechanisms of human perceptual compensation for reverberation, as determined by psychophysical studies [Watkins & Makin, JASA 121, 257-266]. Specifically, we use a metric for identifying regions of reliable speech that gives greater emphasis to reverberation 'tails', and is computed independently for each frequency band. The performance of the system is evaluated on a corpus of spoken digits convolved with recorded room impulse responses.