

ACOUSTICS2008/3436
**”Polyaural” array processing for robust automatic speech
recognition in noisy and reverberant environments**

Richard Stern, Evandro Gouvea and Kshitiz Kumar
Carnegie Mellon University, Department of Electrical and Computer Engineering and Language
Technologies Institute, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

It is well known that human binaural processing is very useful for separating incoming sound sources as well as for improving the intelligibility of speech in reverberant environments. In this paper we present a new method of signal processing for robust speech recognition using multiple microphones. The method, loosely based on the human binaural hearing system, consists of passing the speech signals detected by multiple microphones through bandpass filtering and nonlinear halfwave rectification operations, and then cross-correlating the outputs from each channel within each frequency band. These operations provide rejection of off-axis interfering signals. These operations are repeated (in a non-physiological fashion) for the negative of the signal, and an estimate of the desired signal is obtained by combining the positive and negative outputs. We demonstrate that the use of this approach provides substantially better recognition accuracy than delay-and-sum beamforming using the same sensors for target signals in the presence of additive broadband and speech maskers, and it provides substantial improvements in specific reverberant environments as well. [Supported by NSF and DARPA]