Microphone-Array Systems for Speech Recognition Input

Harvey F. Silverman, Principal Investigator

Laboratory for Engineering Man/Machine Systems (LEMS)
Division of Engineering
Brown University
Providence, RI 02912

Objective

An understanding of algorithmic and engineering techniques for the construction of a high-quality microphone array system for speech input to machines is the goal of this project; it implicitly assumes that wearing a head-mounted microphone, or sitting at a fixed location in front of a table microphone is most often an unacceptable imposition for a user of a speech recognizer. Rather, the array system should electronically track a particular talker in a small room having other talkers and noise sources and should provide a signal of quality comparable to that of the head-mounted microphone. In particular, the project is focused on the measurement of small-room acoustic properties and the derivation of underlying mathematical algorithms for the acoustic field, for talker tracking and characterization, for array layout, for correlated and uncorrelated noise elimination, and for beamforming. Simultaneously, digital array systems are being designed, built, and used.

Approach

The current approach is one in which many of the issues for these systems are being investigated simultaneously. We are using a linear (1D) array to gather real data for both online and offline experimentation with beamforming, location, tracking, and "talker elimination" algorithms. At the same time, new 1D, 2D, and 3D arrays and associated all-digital data processors are being completed, allowing the facile acquisition of data from various transducers and acoustic conditions. Acoustic repeatability is deemed essential to understanding; thus, a sound-field robot is used to automatically move transducers. Careful experimental technique is being used to understand the properties of such systems, as well as develop algorithms and systems.

Recent Accomplishments

During the last year, the current system has been further upgraded for more flexible conversion of existing, stored digital data. The properties of our high-performance output transducer have been measured and inverse-filter mechanisms now are used to acoustically reproduce a speech output without coloration. Automatic systems have been written to convert large, stored speech databases to another acoustical situation. For example, the array may be used to collect speech from a target talker and an interfering talker, both with known locations and output levels. New results for talker location are about to be published. The method works about 95% of the time the talker is speaking, and incorporates a two-stage search and our Stochastic Region Contraction algorithm for nonlinear optimization to find the source location in real-time on a typical, fixed-point DSP chip. Our 128-microphone, all-digital system should be operational before this summer. The architecture is both simple and elegant and there is some interest in it on the part of NUWC for ultrasonic data acquisition. This system allows the formation of up to 128 total microphones in virtually any spatial arrangement. This array should allow for very flexible experimentation with spatial filtering of all sorts. A recent experiment has set a baseline for the performance of our speech recognizer relative to its close-talking performance. A 12% performance degradation resulted from using a simple beamformer system with no noise compensation as input to the recognizer. Now, all new algorithms can be quantitatively tested for effect.

Plans for the Coming Interval

- Continue the modeling of the properties of the room, transducers, and electronics, so that effective "deconvolving" algorithms may be designed.

- Finish the 128-microphone system and apply it in many spatial variants to the problem of filtering out reverberations.

- Continue the development of better algorithms for location, tracking, beamforming, and talker elimination.

- Incorporate noise-reduction mechanisms into the current real-time array for cleaning the signal for recognition. Compare these results to the current baseline.

- Continue to use and gain understanding of the SRC nonlinear minimization algorithm.