Microphone-Array Systems for Speech Recognition Input

Harvey F. Silverman
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, 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 simulataneously. We are using our version 2 system to gather real data for both online and offline experimentation with beamforming, location, tracking, and "talker elimination" algorithms. At the same time, new one- and two-dimensional arrays are in design, allowing us to gather data from larger numbers of microphones. Through the use of a sound-field measurement robot, which is now operational, the difficult acoustics properties of the small room are being measured and appropriate real models are being formulated. Ultimately, these models will lead to algorithms which, implemented in real-time, will give us a robust and useful speech recognition input system.

Recent Accomplishments

The last eight-month period has seen a 100% upgrade in our experimental facility. New (reliable) hardware for an eight microphone linear array system has been completed and installed. Using this, long intervals – say about 10 seconds (10 x 8 mikes x 40,000 bytes/sec = 3.2Mbytes) – may now be recorded for offline algorithm experimentation. In addition, a room-size robot has been completed which allows the accurate placement of sources, as well as the taking of acoustic measurements. We are now in the process of measuring and understanding our environment. Also, an algorithm for the separation of talkers – elimination of an intrusive talker – has been proposed and investigated. To date, the algorithm works perfectly on real speech in a simulated environment, although it fails pretty badly for real speech taken from our real environment. We are on the learning curve and should have some explanations for this soon.

Plans for the Coming Interval

- Continue to measure the properties of the room, transducers, and electronics using the sound-field robot so that models for various algorithms may be designed.
- Continue the implementation of new electronics for a 128-microphone multiple 1-D or 2-D system. This will allow isolation of a volume in space as well as give insight into the reverberation problem for the 2-D system.
- Continue the development of better algorithms for location, tracking, beamforming, and talker elimination.
- Develop simple mechanisms, (filtering etc.) for the current real-time array for cleaning the signal for recognition. Use these to take data for training and testing our talker-independent, HMM alphadigit recognizer, comparing performance to the head-mounted microphone input case.
- Continue to use and gain understanding into the SRC nonlinear minimization algorithm.