Practical issues in the use of a frequency-domain delay estimator for microphone-array applications

John Adcock, Joe DiBiase, Michael Brandstein, et al.

Citation: The Journal of the Acoustical Society of America **96**, 3244 (1994); doi: 10.1121/1.411095
View online: https://doi.org/10.1121/1.411095
View Table of Contents: https://asa.scitation.org/toc/jas/96/5
Published by the Acoustical Society of America

8:55

2aSP2. Practical issues in the use of a frequency-domain delay estimator for microphone-array applications. John Adcock, Joe DiBiase, Michael Brandstein (Box D, Brown Univ., Providence, RI 02912), and Harvey F. Silverman (Brown University, Providence, RI 02912)

A frequency-domain delay estimator has been used as the basis of a microphone-array talker location and beamforming system [M. S. Brandstein and H. F. Silverman, Techn. Rep. LEMS-116 (1993)]. While the estimator has advantages over previously employed correlation-based delay estimation methods [H. F. Silverman and S. E. Kirtman, Comput. Speech Lang. **6**, 129–152 (1990)], including a shorter analysis window and greater accuracy at lower computational cost, it has the disadvantage that since delays between microphone pairs are estimated independently of one another, there is nothing to ensure that a set of estimated delays corresponds to a single location. This not only introduces errors in talker location but degrades the performance of the beamformer. A method for delay estimation and talker location with a microphone array is described that preserves the low computational complexity and rapid tracking ability of the frequency-domain delay estimator, while improving the coherence and stability of the estimated delays and derived source locations. Experimental results using data from a real 16-element array are presented to demonstrate the performance of the algorithms. [Early work principally funded by DARPA/NSF Grant IRI-8901882, and current work by NSF Grant No. 9314625.]

9:15

2aSP3. A constant-directivity beamforming microphone array. Gary W. Elko, Thomas C. Chou, Robert J. Lustberg, and Michael M. Goodwin (Acoust. Res. Dept., AT&T Bell Labs., 600 Mountain Ave., Murray Hill, NJ 07974)

The quality of audio teleconferencing in large rooms and noisy environments can be increased with the use of steerable directional microphone arrays. A minimum bandwidth of 4 oct is required to faithfully transmit the speech signal. In a typical teleconferencing arrangement, only discrete angular directions are of interest and therefore the microphone steering directions are quantized. A standard delay-sum beamformer can result in noticeable frequency response changes as the talker moves between these steering locations. In an effort to mitigate this problem, a broadband constant-directivity beamformer has been designed and constructed. A few of the algorithms developed in this work will be discussed and compared to existing techniques. Basically, the solution revolves around the design of FIR filters that are inserted in the delay-sum beamformer after each element. A constant-beamwidth 4 oct steerable linear array microphone using directional elements will be described. A real-time implementation utilizing multiple AT&T DSP3210 digital signal processors is also described.

9:35

2aSP4. Hands-free mobile telephony by means of an adaptive microphone array. Sven Nordholm, Ingvar Claesson, Sven Nordebo, and Mattias Dahl (Dept. of Signal Process., Univ. of Karlskrona/Ronneby S-372 25 Ronneby, Sweden)

By employing speech generation models and new algorithms more and more *a priori* information about speech signals is utilized in speech recognition and speech coding. A fair signal-to-noise ratio is therefore required to ensure that the *a priori* information is correct. This implies a need for noise reduction under adverse conditions, such as hands-free operation of telephones in the car compartment or speech recognition in cars [S. Nordholm *et al.*, "Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars," IEEE Trans. Veh. Tech. **42**, 514–518 (1993)]. The paper presents two adaptive microphone array schemes, aimed for this situation. The first, denoted spatial filtering generalized sidelobe canceller (SFGSC), gives good noise suppression with little distortion of the speech but requires careful calibration. The second, denoted adaptive microphone array employing calibration signals recorded on-site (AMAEC), facilitates a simple built-in calibration. It is beneficial from a user point of view to use a calibration signal recorded on site eliminating amplifier tuning and microphone selection. The calibration can be done within 60 s. The AMAEC calibrates the array to the speakers' location, microphone positions and lobe gains, amplifiers, and to the acoustic environment in the car. No *a priori* information about signal statistics or array geometry is utilized. [Work supported by Nutek.]

9:55–10:10 Break

10:10

2aSP5. Robust hands-free speech recognition. Qiguang Lin, Chi Wei Che, and James Flanagan (CAIP Ctr., Rutgers Univ., Piscataway, NJ 08855-1390)

When speech recognition technology moves from the laboratory to real-world applications, there is increasing need for robustness. This paper describes a system of microphone arrays and neural networks (MANN) for robust hands-free speech recognition. MANN has the advantage that existing speech recognition systems can directly be deployed in practical adverse environments where distant-talking sound pickup is required. No retraining nor modification of the recognizers is necessary. MANN consists of two synergistic components: (1) signal enhancement by microphone arrays and (2) feature adaptation by neural network computing. High-quality sound capture by the microphone array enables successful feature adaptation by the neural network to mitigate environmental interference. Through neural network computation, a matched training and testing condition is approximated which typically elevates performance of speech recognition. Both computer-simulated and real-room speech input are used to evaluate the capability of MANN. Measurements of isolated-word recognition in noisy, reverberant, and distant-talking conditions show that MANN leads to a word recognition accuracy which is within 4%–6% of that obtained under a close-talking condition in quiet.

10:30

2aSP6. Wideband microphone array for hearing aid preprocessing. Kung Yao (Elect. Eng. Dept., Eng. IV, 68-113, UCLA, Los Angeles, CA 91403-1594), Sigfrid D. Soli (House Ear Inst., Los Angeles, CA 90057), and Dan Korompis (UCLA, Los Angeles, CA 91403-1594)

Speech communication in environments with low signal/noise ratios (SNRs) is a primary complaint of the hearing impaired. Microphone beam formation techniques provide an effective approach to improving SNR in these environments. A novel, fixed