Speech perception in noise with a two-sensor frequency-domain minimum-variance (FMV) beamforming algorithm

Jeffery B. Larsen, Charissa R. Lansing, Robert C. Bilger,[2] Bruce C. Wheeler, Sandeep A. Phatak, Michael E. Lockwood, William O'Brien Jr., and Albert S. Feng

Beckman Institute of Advanced Science and Technology,
University of Illinois, Urbana, 61821, [2] Deceased 26 December, 2002
cfjbl@eiu.edu, crl@uiuc.edu, bcw@uiuc.edu, phatak@uiuc.edu, melockwo@uiuc.edu,
wdo@uiuc.edu, afeng1@uiuc.edu

Abstract: The performance of a two-sensor based, frequency-domain minimum-variance beamforming algorithm (FMV) to extract a signal in the presence of multiple interferers was evaluated. Speech reception thresholds (SRT) and speech intelligibility measures were obtained from listeners with normal hearing or with mild-to-moderate sensorineural hearing loss. Word and sentence length stimuli were processed through the FMV algorithm, directional microphones alone, or with a simple delay-and-sum beamformer. Listener's ratings of speech intelligibility, percent of words repeated correctly, and threshold for words and sentences in the presence of four competing signals showed the FMV to provide significant performance benefits across different listening environments.

PACS numbers: 43.60.Fg, 43.66.Ts [J.C.]
Date Received: June 18, 2003 **Date Accepted:** March 25, 2004

1. Introduction

A binaural adaptive beamforming algorithm that functions in the frequency domain was developed as a front-end for hearing aids [Elledge *et al*., 1999; Lockwood *et al*., 2003]. The frequency-domain minimum-variance beamforming algorithm (FMV) is based on the principle of minimum-variance beamforming operating in the frequency domain [Capon, 1969]. FMV involves minimization of the output energy with the constraint that the target be passed through the system with unity gain. The FMV algorithm differs from other two-sensor beamformers that work in the time domain [e.g., Griffiths and Jim, 1982; Zurek *et al*., 1996; Fischer and Simmer, 1996] by its simplified computational approach to correlating the input of the two sensors and its fast adaptation time (10 to 20 ms). This fast adaptation allows cancellation of multiple interfering sources so long as they do not overlap exactly in frequency and time. A comparison with other two-sensor adaptive beamformers has shown that the FMV outperforms (up to 5 to 6 dB) other adaptive beamformers in computer simulation when more than one noise source is present [Yang *et al*., 2000; Lockwood *et al*., 2003].

The FMV was previously tested by presenting sentences in simulated environments with four interfering signals (jammers) at three signal-to-noise ratios (SNR) (-3, 0, & +3 dB) to listeners with and without hearing loss [Larsen *et al*., 2001]. Listeners rated sentences processed by the FMV more intelligible than the unprocessed sentences. The current study extends the evaluation of the FMV to real-world environments, multiple and varied jammers, and different listening tasks. We evaluated the FMV's ability to extract a target signal from two listening environments of four jammers in the front horizontal plane, located in close proximity to the target signal. The performance of the FMV was compared to that of directional microphones alone (DIR) and of a simple delay-and-sum binaural beamformer (DAS).

2. Method

2.1 Experiment 1

The present experiment, motivated by the need to evaluate the FMV performance in real world environments, replicated the locations of sound sources in the simulated environments [Larsen *et al*., 2001], and used a new set of speech materials as listening targets and jammers. Performance for FMV, DIR, and DAS was measured with fixed and variable SNR tasks. Groups of five listeners with normal hearing (ages 20-40 years) and of nine listeners with bilaterally symmetrical, sensorineural hearing loss (ages 53-81 years) participated. Seven participants with hearing loss had essentially mild to moderate sloping sensorineural hearing loss and two had normal hearing to 1000 Hz with precipitously sloping hearing loss above 1000 Hz. Only two listeners were experienced hearing aid users. All listeners learned American English as a first language.

A bank of Optimus loudspeakers (XTS 40, Cat. No. 400-1991) spaced 20˚ apart and hidden in a wooden case covered in acoustically transparent black foam (Modified Seneimetric Array speaker system) was positioned in a corner of the room. The room measured 3.6 X 3.0 X 3.05 m^3 and was lined with 7.6 cm of acoustically absorbent foam (SONEX Valueline, Illbruck Corp.). The reverberation time for the room averaged just over 0.1 s from 100 Hz to 10 kHz. Two Sennheiser MKE II cardioid microphones spaced 15 cm apart were placed on a tripod. The center of the microphone array was equidistant (0.75 m) from the loudspeakers. All loudspeakers were directed to the center of the microphone array. The listener sat on one side of the room and the signals were presented to the participant via Sennheiser headphones (HDA-200). The listener adjusted the amplitude of an unprocessed signal to a comfortable level to start, and then no more level adjustments were allowed to ensure consistency across signal processing settings of the real-time system. Listeners were tested in the same room as the signal and noise presentation to more closely approximate listening with hearing aids. No attempt was made to shape the frequency of the signals to compensate for the hearing loss of the participants

Fixed SNR task. A compact disk (CD) recording of the Speech Intelligibility Rating (SIR) passages [Cox and McDaniel, 1989] was played at 78 dB SPL from the target loudspeaker at 0˚ relative to a broadside array of the two microphones 15 cm apart. Four competing speech jammers, presented from speakers at $\pm$20˚ and $\pm$ 40˚ relative to the target loudspeaker, consisted of recordings of two male and two female talkers reading sentences from the Revised Speech in Noise Test (R-SPIN) [Bilger *et al*., 1984]. Each talker's speech was reversed to remove linguistic information but preserve the spectro-temporal pattern of the speech, and each was placed in a separate track with no silence between sentences. These competing reversed-speech jammers were presented, one track per loudspeaker, by means of a multichannel playback device (Aark24, Aardvark Audio) and a multichannel amplifier (Knoll MR1250). The RMS amplitude of each of the reversed-speech jammers was adjusted to 76 dB SPL at the center of the microphone array. The level of the reversed speech was constant and started before the target signals. When the four equally intense jammers were presented simultaneously from the loudspeakers the overall level was 82 dB SPL, resulting in a –4 dB SNR at the microphones for the fixed SNR task. SNR measures refer to the SNR as measured at these microphones and not in the ears of the listeners.

Variable SNR tasks. Two variable SNR measures used CID W-1 spondee words spoken by a female talker (BYU Speech Audiometry Materials CD) and Hearing in Noise Test (HINT) sentences [Nilsson *et al*., 1994] as target materials. The jammers were identical to those in the fixed SNR task. To create the variable SNR, the target materials presented from the target loudspeaker at 0° were varied by means of an attenuator (HP AE6 attenuator) with a range of 0 to -80 dB. All signal levels were measured throughout the study by means of a sound level meter set to linear weighting (Bruel & Kjaer 2260).

Processing of signals received by the microphones was accomplished by means of a real-time DSP system based on a Texas Instruments C62x evaluation board. The real-time system processed the output of the two microphones in three ways. One processing scheme (DIR setting) passed the signals through the system with no processing, similar to listening with hearing aids with directional microphones only, with no frequency shaping of the output. A second scheme (DAS setting) summed the signals from the two directional microphones to create a simple delay-and-sum beamformer with a diotic output. The DAS beamformer was included to provide a baseline for simple beamforming against which the performance of the FMV could be compared. Finally, the third scheme (the FMV setting) sent the signals from the two microphones through the FMV algorithm and produced a diotic signal. The FMV is capable of adaptive steering but the beam was fixed perpendicular to the microphone array towards the target speaker to provide a cleaner comparison with the fixed beam of the DIR and DAS settings. The group delay of the system when implementing the FMV was ~32 ms. Effects of group delay on perception were not evaluated.

For the fixed SNR task, the participants listened to 30 s of an SIR passage in the presence of four competing jammers and were asked to rate the intelligibility of the target passage on a scale of 0 to 100 for each setting of the real-time system (i.e., DIR, DAS, FMV). The order of the passages was counterbalanced across settings of the real-time system to compensate for passage-specific effects.

For the variable SNR measures a one-up, one-down procedure was used to obtain ten reversals from which the last seven were used to calculate an SRT for each setting of the real-time system. The FMV produced an elevated noise floor when no competition was present as a result of looking for jammers to cancel. A concern was that these elevated noise levels might potentially bias the variable SNR measures. Therefore, each threshold measure was performed without competing sources for each of the three settings of the real-time system as well as in the competing noise. This allowed performance of the participants to be compared across the three settings of the real-time system by comparing the difference between scores obtained in quiet, and those obtained in competing noise for a particular signal processing method. The order of the SNR measures and the order of the settings for the real-time system during the experiment were randomized to distribute potential order effects in the data.

On average, listeners with hearing loss rated SIR passages (fixed SNR task) as processed by the FMV to be 30% more intelligible than processing with the DAS and approximately 50% more intelligible than with DIR. Normal-hearing listeners rated FMV-processed SIR passages as 20% more intelligible than DAS and 30% more intelligible than DIR processed SIR passages. For the variable SNR task with listeners who had hearing loss, the mean speech reception thresholds (SRTs) in noise were 8 and 11 dB better with FMV than with DAS and DIR processing, respectively. Normal-hearing listeners also showed significant improvement when listening to speech materials processed by the FMV as compared to DAS or to DIR (i.e., mean improvement was 6 or 8 dB, respectively).

2.2. Experiment 2

The superior FMV performance in experiment 1 prompted questions about whether performance with the FMV would be maintained across various environments and jammer configurations. Experiment 2 was designed as a preliminary investigation of FMV performance with a new set of jammers and a new configuration different from those used in all previous evaluations of the algorithm. Cafeteria noise with reduced modulation [Ricketts and Dhar, 1999] was chosen as a jammer because it has been used in previous investigations of hearing aid performance. The jammers used in experiment 2 were unequally spaced and played at unequal RMS amplitudes to represent a more realistic acoustic environment. Six adult listeners with essentially mild-to-moderate, bilaterally symmetrical, sloping sensorineural hearing loss participated. Four had participated in experiment 1. Two listeners were experienced hearing aid users.

A target at 0° and four jammers were also used for experiment 2. The levels, as measured at the center of the microphone array, were 67 dB SPL for the primary jammer at +20° and 60.3 dB SPL for the summed output of three secondary jammers located at –80°, -40°, and +60°. The target level was fixed at 65 dB SPL to achieve an overall SNR of –2.8 dB for the fixed SNR task. Two variable SNR tasks were also performed with the jammer levels remaining constant, and the target level being varied as described in experiment 1.

For the fixed SNR task, the target speech was 50 keywords from two Connected Speech Test passages [Cox and McDaniel, 1989] presented one at a time. For the variable SNR tasks, the target speech was HINT sentences [Nilsson *et al.*, 1994]. Four independent tracks of cafeteria noise, shaped to the long-term spectrum of HINT sentences and altered to reduce its modulation [Ricketts and Dhar, 1999], were used as jammers for the first variable SNR task. The reversed speech used in experiment 1 was used in this new configuration to provide a comparison of performance across jammer configurations. Procedures for the fixed and variable SNR tasks were similar to those in experiment 1.

Approximately one year separated the two experiments. The six listeners with sensorineural hearing loss correctly identified an average of 23% and 36% more CST keywords when listening with the FMV than with DAS and DIR, respectively. Thresholds for FMV processing were on average 7.5 dB lower than those for DAS, and 13 dB lower than for DIR, in the presence of the cafeteria noise. On average, thresholds for FMV processing were 9 dB lower than for DAS and 13 dB lower than for DIR in the presence of the reversed speech jammers. The results show that performance with the FMV in the new environment of experiment 2 was consistent to those of experiment 1.

3. Overall results

Figure 1 shows results for all listeners with hearing loss (N= 9 for experiment 1, and $N = 6$ for experiment 2). Data in Fig. 1 were analyzed individually with paired *t*-tests and Bonferroni corrections. Higher ratings of intelligibility and lower thresholds were achieved with FMV than with DIR setting in all cases ($p < 0.01$). Also, significantly higher intelligibility ratings and lower thresholds were achieved with FMV than with DAS ($p < 0.05$) with one exception. For the fixed SNR results of experiment 2 (top right box) where FMV mean intelligibility scores were 23% better than the DAS condition, the difference was not statistically significant. To analyze the data conservatively, the data from the four listeners who participated in both experiments were analyzed with a repeated measures analysis of variance (ANOVA). The fixed SNR percent-correct data were submitted to an arcsine transformation to stabilize the error variance before being included in the ANOVA [Weiner, 1962]. The untransformed data are reported for simplicity, because the results for the transformed data were similar. Listener performance with the FMV, DAS, and DIR differed significantly [$F(2,6) = 6.22$, MSE =151.18, $p < 0.05$]. Differences in the performance of individual listeners and how listeners performed with a particular speech measure were not significant. No interaction between factors reached statistical significance. Further investigation of the signal processing condition factor through paired comparisons between the three processing conditions showed significant differences between the DIR and DAS conditions [$F(2,6) = 4.94$, $MSE = 8.69$, $p < 0.05$], the FMV and the DIR conditions [$F(2,6) = 26.54$, $MSE = 8.69$, $p < 0.01$], and the FMV and DAS conditions [$F(2,6) = 21.60$, $MSE = 8.69$, $p < 0.01$].

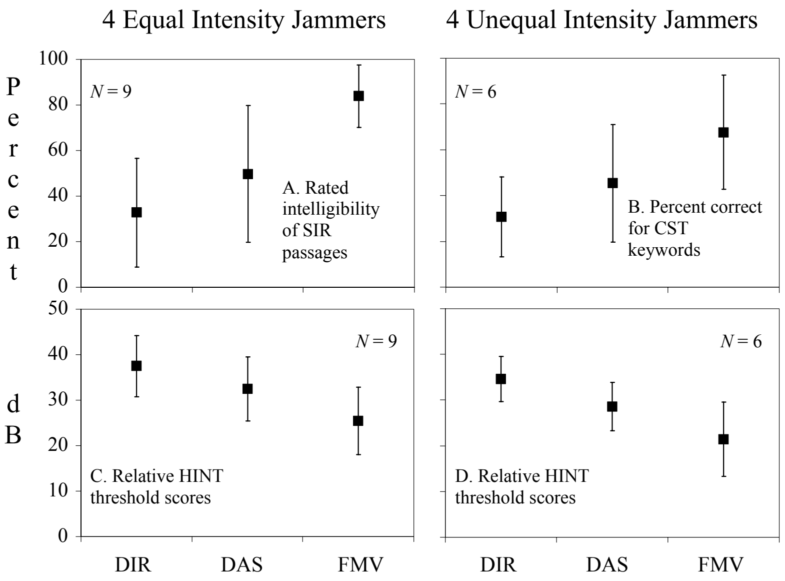

Fig. 1. Group data from listeners with SNHL for two listening environments with four jammers of equal intensity (left panels) and of unequal intensity (right panels). The two environments differed in the configurations of the jammers. Filled square markers represent group means and bars represent the standard error of the mean. For the top two panels, higher scores are better and for the two bottom panels, lower scores are better. Threshold scores in the lower panels represent the difference between the scores in quiet and in competing noise (see text). DIR = directional microphone, DAS = delay-and-sum, FMV = Frequency domain minimum-variance beamformer.

4. Discussion

Results for listeners with normal hearing mirrored those of the listeners with hearing loss. The FMV resulted in statistically superior performance on all measures as compared to DAS or DIR, although the benefit to the listeners with normal hearing was not as large as that observed for the listeners with hearing loss.

The FMV conferred significant benefits for listeners in environments with more sources of noise than sensors when noise sources are in close proximity to the target signal in the front hemisphere. These environments represent real world listening situations that are problematic for other beamforming techniques. The FMV takes advantage of amplitude and temporal modulation characteristics of the speech of individual talkers in a multi-talker environment to accurately and rapidly switch between off-target signals to cancel them. This gives the FMV a performance edge in the presence of competing speech. However, this strategy penalizes the FMV when the interfering sources comprise multi-talker babble and broadband signals with small temporal gaps. Theoretically, in the presence of a completely diffuse environment with no temporal gaps in competing signals, the performance of the FMV would be no worse than the DAS condition.

Ongoing investigations address group delay, efficient signal delivery, and performance in reverberant environments. Listeners with normal hearing object to group delays of ≥10 ms [Agnew and Thornton, 2000]. Group delay may interfere with speech reading and auditory perception for listeners with hearing aids. The current implementation of the FMV has a group delay of approximately 32 ms and therefore must be evaluated under auditory-visual conditions. Further, because the FMV is envisioned as the front-end of a hearing instrument, issues of power consumption, a method to mix signals from the two sensors in a wireless fashion, and a small packaging size must be resolved. Finally, the effect of reverberation on FMV processing must be studied. Preliminary assessment of the current real-time implementation of the FMV in reverberant environments (T60 = 1.2 s) has been encouraging. In spite of these open issues the current results are encouraging.

5. Conclusions

Results show that, when compared to the performance of directional microphones alone or of delay-and-sum beamforming algorithm, the FMV allowed for lower thresholds, higher ratings of intelligibility, and a higher percentage of words repeated correctly for both listeners with and without hearing loss. Also, the benefit of the FMV algorithm was observed for two different environments and two different types of interfering noise. Further evaluation of the FMV is necessary to address implementation of the algorithm in a hearing instrument.

Acknowledgments

This work was supported by Grant No. R21 DC04840 to the University of Illinois at Urbana-Champaign from the NIDCD at the National Institutes of Health, and the Beckman Institute. We acknowledge the House Ear Institute permission to use the HINT test. Portions of this manuscript were presented at the 141st Acoustical Society of America Meeting in Chicago, IL and at the 2002 American Academy of Audiology Convention in Philadelphia, PA.

References and links

Agnew, J. and Thornton, J.M. (**2000**). “Just noticeable and objectionable group delays in digital hearing aids,” J. Am. Acad. Audio. **11**, 330–336.

Bilger, R.C., Neutzel, J.M., and Rabinowitz, W.M. (**1984**). “Standardization of a test of speech perception in noise,” J. Speech Hear. Res. **21**, 5–36.

Capon , J. (**1969**). “High-resolution frequency-wavenumber spectrum analysis,” Proc. IEEE **57**(8), 1408–1419.

Cox, R.M. and McDaniel, D.M. (**1989**). “Development of the Speech Intelligibility Rating (SIR) Test for hearing aid comparisons,” J. Speech Hear. Res. **32**, 347–352.

Elledge, M.E., Lockwood, M.E., Bilger, R.C., Feng, A.S., Goueygou, M., Jones, D.L., Lansing, C.R., Liu, C., O'Brien, W.D. Jr., and Wheeler, B.C., (**1999**). “A real-time dual-microphone signal-processing system for hearing aids,” J. Acous. Soc. Am. **106** (Pt. 2), 2279A.

Fischer, S. and Simmer, K.U. (**1996**). “Beamforming microphone arrays for speech acquisition in noisy environments,” Speech Comm. **20**, 215–227.

Griffiths, L.J. and Jim, C.W. (**1982**). “An alternative approach to linearly constrained adaptive beamforming,” IEEE Trans. Antennas Propog, **AP-30**(1), 27–34

Larsen, J.B., Lockwood, M.E., Lansing, C.R., Bilger, R.C., Wheeler, B.C., O’Brien, W.D., Jones, D.L., and Feng, A.S. (**2001**). “Human performance in a multisource environment with a frequency-banded minimum-variance beamforming algorithm,” J. Acoust. Soc. Am. **109**(5), 2494.

Lockwood, M.E., Jones, D., Bilger, R.C., Lansing, C.R., O’Brien, W.D., Wheeler, B.C., and Feng, A.S. (**2004**). “Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms,” J. Acoust. Soc. Am. **115**(1), 379–391.

Nilsson, M.J., Soli, S.D., and Sullivan, J. (**1994**). “Development of a hearing in noise test for the measurement of speech reception threshold,” J. Acoust. Soc. Am. **95**, 1985–1999.

Ricketts, T.A. and Dahr, S. (**1999**). “Aided benefit across directional and omni-directional hearing aid microphones for behind-the-ear hearing aids,” J. Am. Acad. Audio. **10**, 180–189.

Weiner, B.J. (**1962**). *Statistical Principles in Experimental Desig,* (McGraw-Hill, New York), pp. 221–222.

Yang, K.L., Lockwood, M.E., Elledge, M.E. and Jones, D.L. (**2000**) “A comparison of beamforming algorithms for binaural acoustic processing,” Proc. 9th IEEE Digital Signal Processing Workshop, Hunt, TX, 15-18 October.

Zurek, P.M., Greenberg, J.E., and Rabinowitz, W.M. (**1996**). “Prospects and limitations of microphone-array hearing aids,” in *Psychoacoustics, Speech and Hearing Aids,* edited by B. Kollmeier (World Scientific, Singapore).