Multichannel Spatial Auditory Display for Speech Communications*

DURAND R. BEGAULT, *AES Member*

Human Interface Branch, Aerospace Human Factors Research Division, NASA-Ames Research Center, Moffett Field, CA 94035-1000, USA

AND

TOM ERBE, *AES Member*

School of Music, California Institute for the Arts, Valencia, CA 91355, USA

A spatial auditory display for multiple speech communications was developed at NASA/Ames Research Center. Input is spatialized by the use of simplified head-related transfer functions, adapted for FIR filtering on Motorola 56001 digital signal processors. Hardware and firmware design implementations are overviewed for the initial prototype developed for NASA-Kennedy Space Center. An adaptive staircase method was used to determine intelligibility levels of four-letter call signs used by launch personnel at NASA against diotic speech babble. Spatial positions at 30° azimuth increments were evaluated. The results from eight subjects showed a maximum intelligibility improvement of about 6–7 dB when the signal was spatialized to 60 or 90° azimuth positions.

0 INTRODUCTION

At NASA-Ames Research Center a four-channel spatial auditory display was recently developed for application to multiple-channel speech communication systems in use at NASA John F. Kennedy Space Center (KSC). The KSC communications handbook [1] indicates a list of over 3000 call signs, most of which are spoken as four individual letters, such as NTOC. Communication personnel who monitor multiple radio frequencies during shuttle launches must be able to hear these four letters clearly against manifold layers of speech. This is done currently by using single-earpiece headsets, connected to a radio device capable of monitoring either four or eight separate channels.

A previously specified design ([2], [3]; patent pending) was used to fabricate a prototype design to improve intelligibility within KSC and similar communications contexts. This prototype, referred to as the ASAD (Ames Spatial Auditory Display; see Fig. 1), places four different communication channels in virtual auditory positions about the listener by digitally filtering each input channel with binaural head-related transfer function (HRTF) data. Listening over headphones, one has a spatial sense of each channel originating from a unique position outside the head, namely, as if four people were standing about you, speaking from different directions.

Input channels to the spatial auditory display can be assigned to any position because the ASAD uses four removable EPROMs,[1] with each EPROM corresponding to a particular target position. The EPROMs themselves can contain a binaural HRTF for any given position and measured ear. Hence an important research question is to determine which four positions would be optimal for speech intelligibility of multiple sound sources.

To begin to answer this question, a psychoacoustic investigation focused on what single spatialized azimuth position yielded maximum intelligibility against noise.

* Presented at the 95th Convention of the Audio Engineering Society, New York, 1993 October 7–10; revised 1994 July 6.

[1] EPROM—erasable programmable read-only memory chip.

This was accomplished by measuring intelligibility thresholds at 30° azimuth increments, using a laboratory system that simulated the essential features of the prototype. Intelligibility is defined here as correct identification of a spatialized call sign (signal) against diotic[2] speech babble (noise). The study concludes with results from investigations that suggest what HRTF positions are best used in the filter EPROMs within the ASAD prototype.

1 OVERVIEW OF THE ASAD

1.1 Hardware

The ASAD's hardware is composed of four digital signal processing boards and an analog mixing and pre-filtering board (Figs. 2 and 3). On each digital board there is a signal processing chip (Motorola DSP56001, running at 27 MHz), a codec (Crystal Semiconductor CS4215), three 32K × 8 static memory chips, and a 8K × 8 removable EPROM. On the analog board there are four low-pass filter modules (nominally set at 4 kHz, but interchangeable for other applications), followed by four input gain trimpots. Also on the analog board are two five-channel mixers that matrix the left and right channel outputs. The analog board also contains trimpots for sidetone gain, and headphone and line outputs with individual gain controls are on the front panel.

The ASAD provides FIR filtering for four input signals. The signal flow is as follows. Each signal passes through a 4-kHz low-pass filter with a Butterworth response curve and a dropoff of 36 dB per octave. The signal is then digitized by the ADC section of a Crystal Semiconductor CS4215 codec at a 48-kHz sampling rate and a 16-bit resolution. This PCM encoded signal is received by a Motorola DSP56001 signal processor which performs the FIR filtering. The filtered PCM encoded signal is sent to the DAC section of the CS4215. The resultant analog signal is sent to two five-channel mixers to be combined with the signals from the three other channels as well as the microphone side tone. The mixer output is sent to a headphone amplifier as well as to a line buffer.

[2] Diotic playback is defined as a single audio channel presented to both ears.

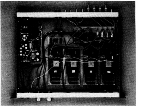

Fig. 2. Internal hardware of ASAD. The four digital boards (bottom right and center) each contain a removable EPROM mounted on ZIF socket, codec (Crystal CS4215), and DSP chip (Motorola 56001). The analog board (upper right) contains the input filtering and mixing circuitry. Power supply is at left.

Fig. 1. Ames Spatial Auditory Display (ASAD). The prototype pictured, fabricated for demonstration purposes, works in conjunction with KSC radio communication equipment. It includes an extra switch for a bypass mode that deactivates the spatial filtering, and an extra knob for adjusting left and right line level test outputs (located on the back of the unit). The EPROMs are accessed by opening the top lid of the unit.

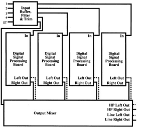

Fig. 3. Schematic of ASAD internal circuitry. Each digital board functions to spatialize sound to a particular position (e.g., left 90, left 30, right 30, and right 90° azimuth). Sidetone is mixed equally to both channels without processing so that it appears in center of head, similar to spatial audition of one's own voice.

1.2 Software

The software for setting up the signal processing chips uses a time domain convolution algorithm for a mono in-stereo out FIR filter. There is also a provision for interchannel delay within the software (0–128 samples). The ADP software is contained in the removable EPROM on each digital board, along with the filter coefficients and delay values. Each EPROM-DSP chip combination utilizes up to 224 coefficients (112 for each channel). This posed a problem for the measured HRTFs that we had available (obtained from F. Wightman and D. Kistler at the University of Wisconsin-Madison), which contained 512 coefficients per channel (that is, an impulse response length of 0.01024 s at a 50-kHz sampling rate). Fewer coefficients could be obtained via windowing or by using minimum-phase reconstructions of the results of principal-components analysis, but these still exceeded the upper limit. However, a previous investigation showed that a filter design approach to simulating the transfer function of the measured HRTFs could be used to obtain a length of 65 coefficients, with minimal difference in the perceived location of speech [4]. Fig. 4 shows an example synthetic HRTF derived from a measured HRTF. The interchannel delay is used to simulate the interaural difference at each position.

1.3 Human Interface Design Philosophy

The goal from the beginning was to make this particular virtual auditory display as "user friendly" as possible. While it is not necessary, different EPROM sets can be installed via the ZIF sockets by a particular user who needs a special filtering set (such as the ones based on their own HRTFs, or to compensate for a particular set of headphones). One could also customize the displayed positions, although some applications would require arrangements known to optimize intelligibility.

Because each EPROM is associated with a particular HRTF and position, there is no need for a user interface (such as selection switches, midi device, or dials) or a host computer. A user therefore only needs to power the device up, and adjust the volume. This is in contrast to

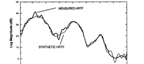

Fig. 4. Magnitude transfer function for measured (512-coefficient) and synthetic (65-coefficient) HRTF for single position. Perceptual studies have found only slight differences in headphone localization of speech for users of "nonindividualized" HRTFs [4].

other "general-purpose" spatial auditory displays that require a computer or a front-panel interface for operation. Most commercially available virtual acoustic displays include many more binaural HRTF sets than are actually needed in a single application, forcing the user to interact selectively with the software and the hardware. This is a type of flexibility that, while desirable in a recording studio, can be a nuisance or even a danger within a high-stress human interface such as space launch radio communications.

2 PSYCHOACOUSTIC VERIFICATION

2.1 Background: Binaural Advantages and Speech Intelligibility

The relationship between binaural hearing and the development of improved communication systems has been understood for over 45 years ([5]; see reviews in [6]; [7]). As opposed to monotic (one-ear) listening—the typical situation in communications operations—binaural listening allows a listener to use head-shadow and binaural interaction advantages simultaneously [7]. The head-shadow advantage is an acoustic phenomenon caused by the interaural level differences that occur when a sound moves closer to one ear relative to the other. Because of the diffraction of lower frequencies around the head from the near ear to the far ear, only frequencies above approximately 1.5 kHz are shadowed in this way. The binaural interaction advantage is a psychoacoustic phenomenon due to the auditory system's comparison of binaurally received signals [7], [8].

Many studies have focused on binaural advantages both for detecting a signal against noise [binaural masking level difference (BMLD)] and for improving speech intelligibility [binaural intelligibility level difference (BILD)]. Studies of BMLDs and BILDs involve manipulation of signal processing variables affecting either signal, noise, or both. The manipulation can involve phase inversion, time delay, and filtering. Speech babble has been used as a noise source in several studies investigating binaural hearing for communication systems contexts (for example, [9]).

Recently speech intelligibility studies by Bronkhorst and Plomp [10], [11] have used a mannequin head to impose the filtering effects of the HRTF on both signal and noise sources. The HRTFs were used in an unaltered condition or with either time or amplitude components removed. Their results, summarized in Fig. 5, show a 6–10-dB advantage with the signal at 0° azimuth and speech-spectrum noise moved off axis, compared to the condition where speech and noise originated from the same position. Fig. 5 also shows lower BILDs when either interaural time or amplitude differences are removed from the stimuli. This suggested the inclusion of HRTF filtering within a binaural display for speech communication systems [2], [12]. According to a model proposed by Zurek [7], based on averaged HRTFs specified in Shaw and Vaillancourt [13], the *average* binaural advantage (speech signal fixed at 0°, noise uniformly distributed across all azimuths, head free to move) is around 5 dB, with head shadowing contributing about 3 dB and binaural interaction about 2 dB.

Another advantage for binaural speech reception relates to the ability to switch voluntarily between multiple channels, or streams, of information [14], [15]. The improvement in the detection of a desired speech signal against multiple speakers commonly referred to as the "cocktail party effect" [16], [17] is explained by Bregman [14] as a form of auditory stream segregation. This situation was found to parallel the multiple-channel listening requirements of communication personnel, such as test directors (NTDs) at KSC.

2.2 Experiment Stimuli

The signal portion of the stimulus was drawn from a list of 130 four-letter call signs, selected from the KSC communications handbook [1]. The 130 call signs used in the experiment were selected randomly so that groups of five began with a unique letter of the alphabet. A single male voice was used, with each letter of the call sign spoken discontinuously over a duration of about 2 s. Recordings took place in a soundproof booth, using an AKG C451-EB microphone at a distance of 6 in (152 mm). Once digitized, each call sign combination was normalized in amplitude and then scaled to have equal long-term rms measurement values.

The speech babble used for the noise portion of the stimulus consisted of multiple layers of voices. Two layers were from different airport control tower frequencies, containing both female and male voices, with silent intervals of more than 0.2 s deleted; and two recordings were of different male voices reading technical repair manuals, one played backward, the other's pitch shifted upward 4 semitones. The result was a dense speech layer in which words could occasionally be distinguished, but semantic content was lost.

The noise and speech were stored digitally as separate channels of stereo sound files (see Fig. 6), using an Apple Macintosh II fx and Digidesign's ProTools hardware and software. The duration of each sound file used in each stimulus presentation was adjusted to 5 s, with the noise channel faded in and out over the first and last 0.5 s. The signal was always presented 1.5 s into the sound file, allowing subjects to predict its onset.

Each of the 130 separate noise-signal sound files was played through signal processing software and hardware, using a Crystal River Engineering Convolvotron that also served as the experimental software host computer. (See Wenzel [18] for additional information on the hardware.) The Convolvotron was used instead of the ASAD itself because of the ability to use experimental software control. Upon playback, the Convolvotron passed the speech babble channel unaltered to both ears. Mixed in with this noise was the two-channel signal, after software intensity scaling and HRTF-based spatialization to azimuths at 30° increments between 30 and 330° (all at 0° elevation). A diotic control condition was also used for the signal, where the spatialization was bypassed and only intensity scaling was used.

The minimum-phase HRTFs used for the spatialization were reconstructed from actual HRTF measurements, as described in Kistler and Wightman [19]. The original measurements used were of one subject (SDO in Wightman and Kistler [20]), with the headphone frequency response (Sennheiser HD-430) divided out of the HRTF. The HRTFs used in the ASAD are very similar to these in that 1) they were also based on subject SDO, 2) they use a technique whereby interaural delay is implemented as a constant, and 3) the data suggest little difference in localization for nonindividualized HRTFs between original and synthetic measurements. Although the same model of headphone was used for the subjects in this experiment as in Kistler and Wightman [19], nonlinearities in reproducing the HRTF were introduced as a result of the interaction between different pinnae and the headphone chambers. Data on localization error of speech with nonindividualized HRTFs can be found in Begault and Wenzel [21] and Begault [22].

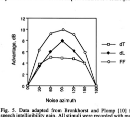

Fig. 5. Data adapted from Bronkhorst and Plomp [10] for speech intelligibility gain. All stimuli were recorded with mannequin head. Speech signal fixed at 0°; noise moved along azimuth at 0° elevation. FF—data including effects of HRTF; dT—same data with binaural amplitude differences removed; dL—same data with binaural time differences removed.

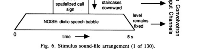

Fig. 6. Stimulus sound-file arrangement (1 of 130).

2.3 Experimental Subjects

Five paid subjects (four males, one female) participated in the study over two 3-h sessions. This was the "naive subjects" group in that they had no exposure to the call sign list. Another group of three lab personnel (three males) who had previous exposure to the call sign list constituted the "experienced subjects" group; their data are analyzed separately from the naive subject group's. This group included a subject whose voice was used in the signal.

All subjects were evaluated for normal hearing from 0.1 to 8 kHz in a pure-tone audiometer test. Subjects were given a training session before starting the experiment to familiarize themselves with the computer, the time when to expect the signal in relation to the noise, and the procedure for entering responses. This training session consisted of a dummy block where the level of the signal was clearly audible against the noise and was never scaled. The formal blocks were begun after approximately 20 trials.

2.4 Experimental Procedure

Software was developed for presenting stimuli and gathering data from subjects using an interleaved, transformed up–down "staircase" method [23]. The software varied the level of the signal against the noise, starting with a maximum step-size interval of 6 dB and decreasing to a minimum step size of 1 dB. The response sequences were evaluated in such a way as to determine the 50% speech intelligibility level.

The decibel level between the diotic stimuli and the spatialized stimuli was considered to be equal with reference to the long-term rms value of speech-spectrum noise filtered by a left-ear 0° HRTF (obtained from the same HRTF set used for the other spatialized positions). The playback level was around 55 dB SPL, when the noise and 0° HRTF-filtered calibration signals were played simultaneously.

Six blocks were administered to each subject over three or four days, with each block containing four staircases randomly chosen from the 11 possible spatial positions and the one diotic signal condition. The four staircases within each block were presented randomly, as were the 130 call sign–speech babble sound files used for a particular stimulus block. The staircases within the blocks were arranged so that 10 threshold values were obtained from each subject for each spatial condition and the diotic condition. No block contained two simultaneous staircases for the same spatial condition of the signal.

Upon hearing the stimulus, the subject typed the four letters he or she thought to have heard onto a computer keyboard, and then after a short pause the software would present the next trial. The duration to complete each block of four staircases was about 15–20 min. Testing was administered in a soundproof booth. No feedback was given as to the correct identification of the call signs. The subjects were only notified when the 20 staircases within a particular block (four spatial conditions times five staircases) were completed.

2.5 Experimental Results

Figs. 7 and 8 summarize the data for the six naive subjects and the three experienced subjects, respectively. The mean values for each position were obtained before grouping the data by first subtracting each individual subject's threshold for the diotic signal versus the diotic speech babble condition. As expected, the results in Figs. 7 and 8 show a greater intelligibility advantage as the signal is moved to either side of the head; the advantage is maximum between 60 and 90° and 270 and 300°. These are locations where both head shadowing is maximized and the binaural interaction advantage mechanism is given maximum time differences.

Fig. 9 summarizes the data from Figs. 7 and 8, with the mean values for symmetric left–right positions about the head. This suggests that, without reference to the side where a sound is spatialized, the preferred order for HRTF processing for maximum intelligibility is 60

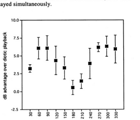

Fig. 7. Data for naive subject group (4 males, 1 female). Mean value for diotic signal condition was subtracted from each spatialized signal value. Standard deviation bars were based on 10 staircase solutions obtained for each condition.

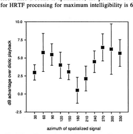

Fig. 8. Data for experienced subject group (ref. Fig. 7).

or 90°, then 120°, then 30°, then 150°, and finally 180°. The latter is hardly better than performance with the diotic stimuli. Fig. 9 also shows that the three experienced subjects achieved about a 1-dB additional intelligibility advantage over the five naive subjects. However, an analysis of variance revealed that no significant difference existed between these two subject categories; $F(1, 6) = 2.90$, $p = 0.14$.

2.6 Discussion

Overall, a 6–7-dB advantage for left and right 60 and 90° positions was found in the present study, which exceeds the binaural advantage cited in Zurek's model [7] by 1–2 dB. This means that headphone listening with static spatial positions through the hardware prototype is at least as good as a normal-hearing binaural listener who is free to move his or her head. Although Bronkhorst and Plomp [10] found a 10-dB advantage for a signal at 0° azimuth and speech-spectrum noise at 90°, their results are not directly comparable to those found here since both signal and noise were HRTF filtered by their mannequin head, and in the present study the noise portion of the stimulus was diotic. The additional release from masking they found may have been attained through either HRTF filtering of both signal and noise, the use of noise rather than speech babble, or both.

The results found here are limited by the fact that only one male speaker was used for the signal portion of the stimulus. In spite of the care taken in preparing the stimulus through digital editing, there is the potential that extraneous variation was introduced into the results because of the variability of spoken intelligibility (ANSI [24]). Furthermore, the average spectrum of this particular speaker might have interacted differently with the HRTF filtering than that of another speaker (for example, a female voice). We have subsequently found the binaural advantage to be lowered by 1–2 dB, depending on azimuth, when using a modified rhyme test and multiple speakers, according to ANSI [24]. Finally the variability in HRTF measurements due to different persons

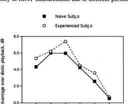

Fig. 9. Mean values from Figs. 7 and 8 collapsed about symmetrical left–right positions.

or reconstruction techniques could influence the results of any experiment that uses only one set of HRTFs. This is one reason why the ASAD was designed to allow interchangeable EPROMs—individuals could tailor systems to their best advantage by using a preferred set of HRTFs.

3 CONCLUSION

The advantage of the ASAD for multiple communication channels has been demonstrated through a case study of a single signal at incremented 30° azimuth positions against a diotic speech babble noise source. The 6–7-dB advantage for 60 and 90° HRTF-filtered speech represents a halving of the intensity (acoustic power) necessary for correctly identifying a four-letter call sign typical of those used in communication systems at KSC. This reduction in the likelihood of misinterpreting call signs over communication systems is an important safety improvement for "high-stress" human–machine interface contexts. The binaural advantage could also benefit communications personnel because the overall intensity of communications hardware could be reduced without sacrificing intelligibility. Lower listening levels over headphones could possibly reduce the risk of threshold shifts, the Lombard reflex (raising the intensity of one's own voice; see Junqua [25]), and overall fatigue, thereby making additional contributions to safety.

Overall, the findings here suggest that the use of a spatial auditory display such as the ASAD could enhance both occupational and operational safety and efficiency of multiple-communication-channel contexts. Further studies are under way at Ames Research Center to determine the additional benefits, if any, of spatial audio communications displays.

4 ACKNOWLEDGMENT

The assistance of Dr. Elizabeth M. Wenzel in all aspects of this work is gratefully acknowledged. Further acknowledgment is due to Drs. Cynthia Null and Key Dismukes for comments on the manuscript; to Bill Williams and Jim Devault at KSC for their support of this project; and to Phil Stone and Rick Shrum for programming, laboratory management, and overall assistance.

5 REFERENCES

[1] NASA-KSC. *KSC Operational Intercommunications System Call Sign/Word Handbook*, no. DRD 016, rev. 18, John F. Kennedy Space Center (1991).

[2] D. R. Begault and E. M. Wenzel, "Technical Aspects of a Demonstration Tape for Three-Dimensional Auditory Displays," Tech. Memo. TM 102286, NASA-Ames Research Center (1990).

[3] D. R. Begault, "Audio Spatialization Device for Radio Communications," patent disclosure ARC 12013-1CU, NASA-Ames Research Center (1992).

[4] D. R. Begault, "Perceptual Similarity of Measured and Synthetic HRTF Filtered Speech Stimuli," *J.*

Acoust. Soc. Am., vol. 92, p. 2334 (1992).

[5] J. C. R. Licklider, "The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise," *J. Acoust. Soc. Am.*, vol. 20, pp. 150–159 (1948).

[6] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, J. Allen, Transl. (MIT Press, Cambridge, MA, 1983).

[7] P. M. Zurek, "Binaural Advantages and Directional Effects in Speech Intelligibility," in G. A. Studebaker and I. Hochberg (Eds.), *Acoustical Factors Affecting Hearing Aid Performance* (Allyn and Bacon, Needham Heights, MA, 1993).

[8] H. Levitt and L. R. Rabiner, "Predicting Binaural Gain in Intelligibility and Release from Masking of Speech," *J. Acoust. Soc. Am.*, vol. 42, pp. 820–829 (1967).

[9] I. Pollack and J. M. Pickett, "Stereophonic Listening and Speech Intelligibility," *J. Acoust. Soc. Am.*, vol. 30, pp. 131–133 (1958).

[10] A. W. Bronkhorst and R. Plomp, "The Effect of Head-Induced Interaural Time and Level Differences on Speech Intelligibility in Noise," *J. Acoust. Soc. Am.*, vol. 83, pp. 1508–1516 (1988).

[11] A. W. Bronkhorst and R. Plomp, "Effect of Multiple Speechlike Maskers on Binaural Speech Recognition in Normal and Impaired Hearing," *J. Acoust. Soc. Am.*, vol. 92, pp. 3132–3139 (1992).

[12] D. R. Begault and E. M. Wenzel, "Techniques and Applications for Binaural Sound Manipulation in Human–Machine Interfaces," *Int. J. Aviation Psychol.*, vol. 2, no. 1, pp. 1–22 (1992).

[13] E. A. G. Shaw and M. M. Vaillancourt, "Transformation of Sound-Pressure Level from the Free Field to the Eardrum Presented in Numerical Form," *J. Acoust. Soc. Am.*, vol. 78, pp. 1120–1122 (1985).

[14] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1990).

[15] D. Deutsch, "Auditory Illusions, Handedness, and the Spatial Environment," *J. Audio Eng. Soc.*, vol. 31, pp. 607–622 (1983 Sept).

[16] E. C. Cherry, "Some Experiments on the Recognition of Speech with One and Two Ears," *J. Acoust. Soc. Am.*, vol. 25, pp. 975–979 (1953).

[17] E. C. Cherry and W. K. Taylor, "Some Further Experiments on the Recognition of Speech with One and with Two Ears," *J. Acoust. Soc. Am.*, vol. 26, pp. 549–554 (1954).

[18] E. M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence: Teleoperators and Virtual Environments*, vol. 1, pp. 80–107 (1992).

[19] D. J. Kistler and F. L. Wightman, "A Model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction," *J. Acoust. Soc. Am.*, vol. 91, pp. 1637–1647 (1992).

[20] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *J. Acoust. Soc. Am.*, vol. 85, pp. 858–867 (1989).

[21] D. R. Begault and E. M. Wenzel, "Headphone Localization of Speech," *Human Factors*, vol. 35, no. 2 (1993).

[22] D. R. Begault, "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems," *J. Audio Eng. Soc.*, vol. 40, pp. 895–904 (1992 Nov.).

[23] H. Levitt, "Transformed Up–Down Methods in Psychoacoustics," *J. Acoust. Soc. Am.*, vol. 49, pp. 467–477 (1970).

[24] ANSI S3.2-1989, "American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems," American National Standards Institute, New York (1989).

[25] J. Junqua, "The Lombard Reflex and Its Role on Human Listeners and Automatic Speech Recognizers," *J. Acoust. Soc. Am.*, vol. 4, pp. 510–524 (1993).

THE AUTHORS

D. R. Begault

T. Erbe

Durand R. Begault received the Ph.D. degree in computer music from the University of California, San Diego, in 1987, and an M.F.A. from Mills College in 1981.

Since 1988, he has worked with the Aerospace Human Factors Research Division of NASA Ames Research Center, under the auspices of the National Research Council and the San Jose State University Foundation. His work has included basic research into sound localization, and applications research and development for advanced audio displays in commercial airline cockpits, reverberation modeling, and speech intelligibility. He is the author of "3-D Sound for Virtual Reality and Multimedia" (Academic Press).

Tom Erbe was born in Milwaukee, WI, in 1960. He studied computer science and music at the University of Illinois, Urbana-Champaign, receiving his degree in 1984. For three years he was the technical manager at the Computer Audio Research Lab at the University of California, San Diego. While there, he was involved in the development of an electronic violin, a DSP-based sound processor, and a generalized computer music workstation.

In 1987 Mr. Erbe became the technical director at the Center for Contemporary Music (CCM) at Mills College, where he worked with composers Robert Ashley, David Rosenboom, Larry Polansky, and James Tenney as computer music engineer and recording engineer. His technical work included the program SoundHack, a signal processing and soundfile manipulation program for the Macintosh, and the design of a DSP-based sound processor for use with the computer language HMSL.

In 1993, Mr. Erbe joined the faculty of the California Institute of the Arts in Valencia, CA, where he is the technical director of the computer music and integrated media studios. He also works as an independent consultant in computer music and computer audio applications.