Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER
UNCLASSIFIED	516241

TITLE

The relative impact of generic head-related transfer functions on auditory speech thresholds: Implications for the design of three-dimensi

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to: CL

This page is left blank

This page is left blank

ORIGINAL RESEARCH

The Relative Impact of Generic Head-Related Transfer Functions on Auditory Speech Thresholds: Implications for the Design of Three-Dimensional Audio Displays

G. R. ARRABITO, M.Sc., S. M. MCFADDEN, M.A., AND R. B. CRABTREE, P.ENG.

ARRABITO GR, MCFADDEN SM, CRABTREE RB. *The relative impact of generic head-related transfer functions on auditory speech thresholds: implications for the design of three-dimensional audio displays.* Aviat Space Environ Med 2001; 72:624–31.

Background: Auditory speech thresholds were measured in this study. ***Methods:*** Subjects were required to discriminate a female voice recording of three-digit numbers in the presence of diotic speech babble The voice stimulus was spatialized at 11 static azimuth positions on the horizontal plane using three different head-related transfer functions (HRTFs) measured on individuals who did not participate in this study. The diotic presentation of the voice stimulus served as the control condition. ***Results:*** The results showed that two of the HRTFS performed similarly and had significantly lower auditory speech thresholds than the third HRTF. All three HRTFs yielded significantly lower auditory speech thresholds compared with the diotic presentation of the voice stimulus, with the largest difference at 60° azimuth ***Conclusion:*** The practical implications of these results suggest that lower headphone levels of the communication system in military aircraft can be achieved without sacrificing intelligibility, thereby lessening the risk of hearing loss.
Keywords: 3-D audio display, head-related transfer functions, auditory speech thresholds.

THE ENVIRONMENT in a modern military aircraft cockpit is highly dynamic and complex. Aircrews are often subjected to high workload and have to maintain situational awareness, while making quick decisions and prompt responses. A relatively new technology, the three-dimensional (3-D) audio display, is being explored for improving aircrew performance. Applications include auditory warnings (10,11,15), air traffic control displays (31), head-up auditory displays for traffic collision avoidance (3,5), enhancing visual target detection and identification (8,13,24), and speech intelligibility (4,16,26). It has been proposed that a 3-D auditory display can support situational awareness and spatial orientation by providing veridical spatial cues to the positions of targets, threats, and beacons (15,18,29). During an inflight study, pilots reported that a 3-D audio display decreased target acquisition time and visual workload while increasing communication capability and situational awareness (21).

The effectiveness of a 3-D audio display depends on the listener's ability to discriminate and localize various sources of information in auditory space. Spatialization of an auditory signal over headphones is accomplished by digitally filtering the signal with head-related transfer functions (HRTFs). HRTFs, which encode the binaural and spectral cues used in sound discrimination and localization, are derived from a series of impulse measurements performed at the ears of an observer in response to a sound source placed at various locations in the vicinity of the head (e.g., 33). It has been argued that a listener's ability to localize a virtual sound is more accurate when using HRTFs measured from his/her own head (*personal*) compared with HRTFs measured from a different head (*generic*) (12,32,34). These studies have shown that generic HRTFs significantly contribute to reversals (i.e., perceiving the mirror image of the presented sound source). However, a recent study reported that the variable of personal vs. generic HRTFs had no significant effect on reversals (7). Perceiving the mirror image of the sound source may lead to inaccurate perceptual cues causing a detrimental effect on situational awareness.

If virtual sources are to be used in a general-purpose 3-D audio display under critical conditions, such as those encountered by military aircrew, then the HRTFs should be optimized for the targeted application. However, it is not presently practical or affordable to measure HRTFs for each potential listener. Furthermore, Bronkhorst's (7) findings suggest that personal HRTFs may not be required. There is a requirement to study the attributes of generic HRTFs, which would contribute to the listener's ability to accurately discriminate and localize the virtual audio signal.

The objective of the present study was to investigate

From the Defence and Civil Institute of Environmental Medicine, Toronto, Ontario, Canada.

This manuscript was received for review in February 2000. It was revised in October 2000. It was accepted for publication in December 2000.

Address reprint requests to: Robert Arrabito, who is Defence Scientist, Defence and Civil Institute of Environmental Medicine, 1133 Sheppard Avenue W., P.O. Box 2000, Toronto, Ontario, M3 M 3B9, Canada; robbie@dciem.dnd.ca.

whether different generic HRTFs yielded significantly different auditory speech thresholds. The paradigm for the experiment reported in this paper closely models the auditory speech threshold study described in Begault and Erbe (4). In the present study, subjects were required to discriminate a female voice recording of three-digit numbers in the presence of diotic (the same sound presented to both ears) speech babble. As in Begault and Erbe (4), the voice stimulus was spatialized at 11 static azimuth positions on the horizontal plane. Three generic HRTFs were used for the spatialization. The diotic presentation of the voice stimulus served as the control condition. Auditory speech thresholds expressed as signal-to-noise ratios for discrimination were calculated for the 12 conditions. The stimuli were intended to emulate an operational listening task in a specific noise background. The sole motivation of the investigators was to rank the relative performance of three generic HRTFs through the measurement of auditory speech thresholds. There was no intent to explore absolute speech intelligibility thresholds using procedures such as diagnostic rhyme test (17) or modified rhyme test (20).

METHODS

Subjects

Six female and six male paid subjects voluntarily participated in this study. The subjects ranged in age from 20–32 yr with a mean age of 24.7. A Békésy audiometric test was administered to each subject. All had less than a 20–dB bilateral hearing loss at frequencies between 125 Hz and 8 KHz, and reported no history of hearing abnormalities. Subjects were fluent in English. Five of the subjects were in-house employees

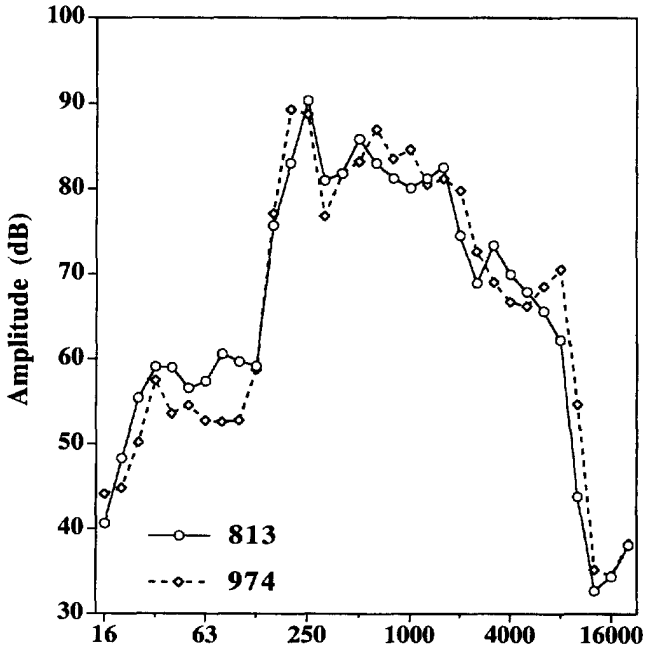

Fig. 1. Frequency spectrum for the numbers 813 and 974 (female voice stimulus).

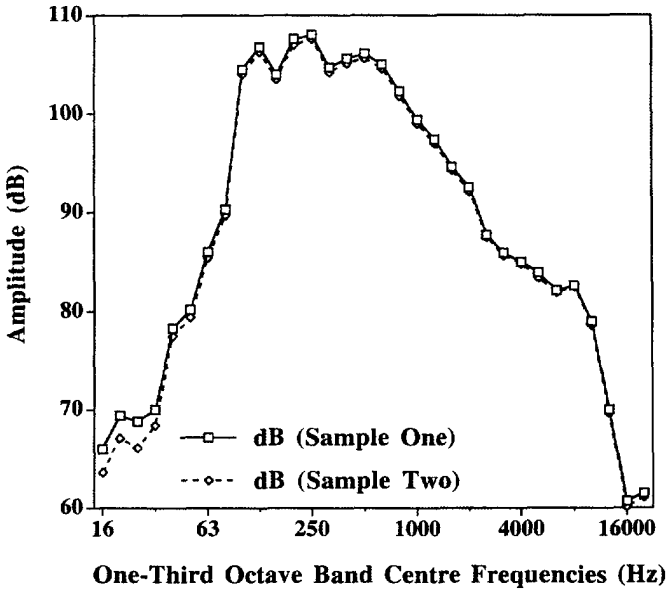

Fig. 2. Two different 4-min frequency spectra for speech babble.

while the others were university students. Two of the subjects had previously participated in psychoacoustic studies. Three of the subjects participated in all of the HRTF conditions and the remaining nine participated in only one of the HRTF conditions. The DCIEM human ethics committee approved the experimental protocol and informed consent was obtained from the subjects.

Stimuli and Apparatus

The signal portion of the stimulus consisted of a female voice recording of 100 three-digit numbers obtained from Gagnon and Cupples (19). Each number was spoken as individual digits. For example, the number "123" was spoken as "1," "2," "3." The shortest amount of time taken to speak a number was 739 ms, while the longest was 1163 ms. The recorded numbers were digitally stored as separate, single-channel sound files on the hard disk of a NeXT Turbo Cube that served as the host computer in this study. The frequency spectrum for the recordings of the numbers 813 and 974 is shown in **Fig. 1**. Each sound file was scaled to have equal long-term RMS amplitude via "STITST" which is the routine of STIDAS (Speech Transmission Index Device using Artificial Signals) for measuring the amplitude of speech signals (28).

The noise portion of the stimulus consisted of diotic speech babble acquired from Abel (1). The speech babble was a continuous, looped, 30-min recording of several men and women speaking simultaneously. The occasional word could be distinguished but semantic content was lost. The speech babble was stored on a digital audio-tape (DAT) and played on a Panasonic SV3500 DAT player. Two different 4-min frequency spectra of the speech babble are shown in **Fig. 2**.

Two different hardware configurations were used in the course of this study. For both configurations, an IBM compatible 486 personal computer (PC) was slaved to the NeXT host computer. The first configuration is based on a commercial 3-D audio product. This product

consists of software routines compiled in a library that run solely on a PC in conjunction with a Turtle Beach Multisound Tahiti sound card resident in the PC. The diotic presentation of the voice stimulus was achieved by routing the audio output from the NeXT computer to two separate channels on a Niche Audio Control Module (ACM) MIDI programmable attenuator/mixer. The ACM served as a mixer for the voice stimulus and speech babble inputs and as an amplitude control device for the voice stimulus. A spatial presentation was achieved by routing the NeXT's audio output to the input of the Turtle Beach Multisound Tahiti sound card. One of the software library routines spatialized the inputted signal. The stereo output from the Tahiti card and the monaural audio output of the Panasonic SV3500 DAT player (speech babble) were each routed to two separate channels on the Niche ACM. The mixed output signal from the Niche ACM was amplified using an Oxmoor 4 × 4 Buffer Amplifier which in turn drove a Stax type SRM-T1 headphone amplifier. The voice stimulus and speech babble were simultaneously presented over Stax electrostatic headphones (model SR-Λ Signature). The HRTF (denoted herein as HRTF1) was integrated in the software library. The measurement and psychoacoustic validation techniques for HRTF1 were not available.

The second configuration is based on the Tucker-Davis Technologies (TDT) equipment. The audio output of the NeXT computer was routed to a Kemo type VBF/23 low-pass elliptic filter set to 10 KHz and then routed to the TDT PD1, which spatialized the voice stimulus. The output from the TDT PD1 was routed to two TDT PA4 programmable attenuators, which varied the sound level of the voice stimulus. The left and right channels of the attenuated voice stimulus were mixed with the speech babble playing on the Panasonic DAT player, using a TDT SM3 stereo signal mixer. The TDT SM3 in turn drove the Stax SRM-T1 headphone amplifier. The voice stimulus and speech babble were simultaneously presented over the Stax electrostatic headphones. Two HRTFs (denoted herein as HRTF2 and HRTF3) were integrated into the TDT system. HRTF2 was obtained from the University of Wisconsin-Madison, while HRTF3 was obtained from the Aeronautical and Maritime Research Laboratory, Melbourne, Australia. The measurement and psychoacoustic validation techniques for HRTF2 and HRTF3 are reported in previous studies (12,25,33,34).

The three different generic HRTFs were used in this study to spatialize the voice stimulus in real-time. Given that the HRTFs were platform specific, it was not possible to implement the three HRTFs on the same hardware configuration. In spite of the two different hardware configurations used in this study, all of the HRTFs were calibrated to have a maximum headphone output level of 70 dB(A) ± 2 dB for the voice stimulus. The output of the voice stimulus in the diotic presentation was 70 dB(A). The headphone output of the speech babble was approximately 60 dB(A). There were no elements in the signal paths to "color" or modify the signals. The HRTFs were treated as "black boxes" in the sense that the parameters were not altered in any way. The calibration was performed with a Brüel and Kjær B&K 4144 1-in condenser microphone mounted within a shock-mounted flat-plate coupler. The microphone recorded the signal output from the headphone earcup that was pressed against the coupler plate. The signal from the microphone was fed to a B&K 2133 frequency analyzer.

The study took place in an Industrial Acoustics Company (IAC) double-wall sound attenuation booth. The ambient level in the booth was less for all frequencies than the maximum allowed for open-ear headphone testing (2). The booth contained a chair, NeXT console, and keyboard.

Task

The subject's task was to listen to the voice stimulus in the presence of diotic speech babble and then to enter, via the NeXT computer keyboard, the number believed to have been spoken.

Experimental Design

The voice stimulus was spatialized at 11 static azimuth positions on the horizontal plane ranging from 30–330° azimuth in 30° increments. As reported in this paper, azimuth increases clockwise on the horizontal plane with 0° positioned directly in front of the listener. The diotic presentation of the voice stimulus served as the control condition. The spatial and diotic presentation was treated as a within-subject factor and was counter-balanced across subjects. The three HRTFs were treated as a between-subject factor for the nine subjects that participated in only one HRTF condition and as a within-subject factor for the three subjects who participated in all three HRTF conditions.

Procedure

Subjects were individually tested in the IAC sound-attenuation listening booth. Each subject was seated facing the NeXT console and keyboard. Testing began by familiarizing the subject with the computer set up, voice stimulus, and experimental protocol, which took approximately 15 min. Subsequently, data collection began. For each trial a 100 ms, 400 Hz tone was used to cue the subject to attend to the spoken number. The subject was then prompted on the NeXT console to enter his/her response. A response was scored as "correct" only if all three digits were entered in the sequence presented by the voice stimulus. If less than three digits were entered the subject was repeatedly prompted, without rehearing the number, until a 3-digit number was entered.

On completing a condition the subject proceeded to the next. There were 12 conditions in total. Response time was unlimited and feedback was provided after each trial. The software on the host computer varied the amplitude of the voice stimulus against the speech babble using the PEST adaptive psychophysical procedure in MOUSE mode (22). Testing began with a maximum step size of 4 dB and decreased to a minimum step size of 0.5 dB. The auditory speech threshold was determined for an 80% probability of detection. It was cal-

culated for every condition as defined by the level of the voice stimulus on the last trial. This constituted the subject's "raw" auditory speech threshold.

The 100 numbers in the voice stimulus "vocabulary" were presented randomly without replacement. If the number of trials exceeded 100, then the list was reshuffled. The maximum allowable number of trials in any condition was 230, in order to minimize the effect of fatigue. In this rare occurrence the condition was terminated.

A session consisted of 12 conditions (11 spatial positions and the diotic condition) and took approximately 70 min to complete. Relative auditory speech thresholds were calculated for each session as defined by the algebraic difference between the diotic and spatial raw speech auditory thresholds. Subjects completed four sessions, each on separate days, using the same HRTF. Six subjects participated in each HRTF condition. Of the 12 subjects who took part in this study, 9 participated in only one of the three HRTF conditions. The other 3 subjects (denoted herein as 2, 3, and 5) participated in all three HRTF conditions. Due to lack of subject and hardware availability, the presentation of HRTF conditions was not randomized across these subjects. Presentation was selected as follows: subject 2 was administered HRTF1, HRTF3 and HRTF2; subject 3 was administered HRTF1, HRTF2 and HRTF3; subject 5 was administered HRTF2, HRTF3 and HRTF1. Approximately 6 mo elapsed between the presentation of HRTF conditions for these three subjects.

RESULTS

As observed in **Fig. 3**, the three generic HRTFs yielded different auditory speech thresholds. In general, performance was best with HRTF3 and poorest for HRTF1. However, the pattern of the results for the different azimuth positions were similar across the different HRTF conditions as illustrated in **Table I**. For all three HRTF conditions, performance was best at 60° azimuth. However, azimuth positions of 90° and 300° were similar to those at 60° azimuth. As well, performance tended to be better when the voice stimulus was positioned in front of the subject (30°, 60°, 300°, 330° azimuth) and to the subject's right (30°, 60°, 90°, 120°, 150° azimuth).

Prior to the main data analysis, the relative auditory speech thresholds of the subjects who participated in all HRTF conditions (2, 3, and 5) were compared with the thresholds of those subjects who participated in only one HRTF condition. As observed in Fig. 3, the performance of these two groups of subjects was similar. This similarity was confirmed by an analysis of variance (ANOVA), $F < 1$.

A repeated measures ANOVA was performed on all 12 subjects. There was a significant effect of HRTF, $F(2,4) = 15.1$, $p < 0.02$, and azimuth position, $F(10,110) = 226.0$, $p < 0.01$. Similar results were also obtained when separate analyses were performed on the two groups of subjects. A Scheffé post hoc analysis at the 0.05 alpha level revealed that: a) there was a significant difference between HRTF1 and both HRTF2 and HRTF3; and b) there was no significant difference between HRTF2 and HRTF3. There were no significant differences among 60°, 90°, and 300° azimuth across all HRTFs. An ANOVA comparing front/back and right/left differences showed a significant difference as a function of front/back azimuth positions $F(1,11) =$

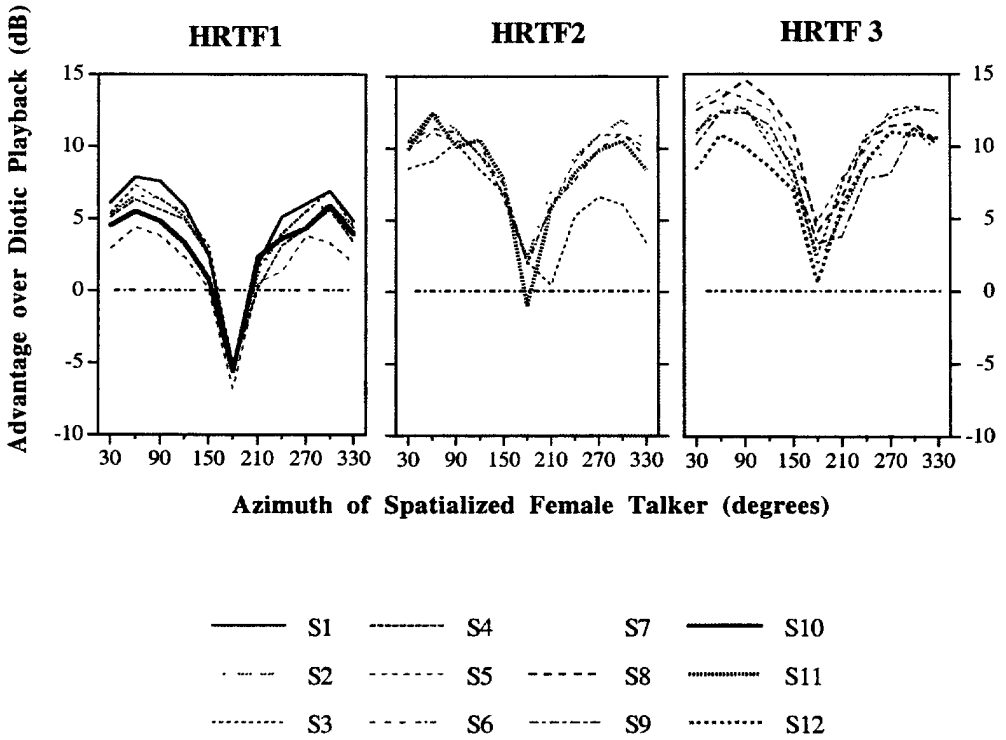

Fig. 3. Relative auditory speech thresholds for each HRTF condition. Data are averaged over four sessions for each subject.

TABLE I. MEANS OF THE RELATIVE AUDITORY SPEECH THRESHOLDS FOR EACH HRTF AND HRTF AVERAGE.

	HRTF1		HRTF2		HRTF3		Average	
Azimuth	Mean	Scheffé Grouping	Mean	Scheffé Grouping	Mean	Scheffé Grouping	Mean	Scheffé Grouping
60	6.3	A	11.3	A	12.6	A	10 1	A
90	5.7	B A	10.6	B A	12.5	A	9.6	B A
300	5.7	B A	10.5	B A	11.7	B A	9.3	B A C
30	4.9	B C	9.8	B A C	11.1	B A	8.5	B C D
270	4.9	B C	9.7	B A C	11.0	B A C	8.5	B C D
120	4.5	B C	9.6	B A C	10.9	B A C	8.4	C D
330	3.8	C	8.4	B A C	10.8	B A C	7.7	D E
240	3.5	C	8.1	B C	9.5	B C	7.0	E F
150	1.9	D	7.2	D C	8 6	C	5.9	F
210	1.2	D	5 1	D	6.1	D	4.1	G
180	−5.6	E	1.6	E	2 8	E	−0.4	H

Note: Means which contain the same letter were not significantly different ($p < 0.05$).

496.8, $p < 0.01$ but no effect of right/left azimuth positions.

To further assess the similarities and differences in the relative auditory speech thresholds for the three HRTF conditions, the effect of azimuth position was analyzed separately for each HRTF condition; HRTF1 F (10,50) = 38.9, $p < 0.01$; HRTF2 F (10,50) = 223.1, $p < 0.01$; HRTF3 F (10,50) = 57.2, $p < 0.01$. A Scheffé post hoc analysis at the 0.05 alpha level revealed a similar pattern of differences in relative auditory speech thresholds across the 11 azimuth positions for each HRTF. There were some exceptions as shown in Table I. The ordering for 30°, 120°, 270°, and 330° azimuth varied slightly as a function of HRTF. However, there were no significant differences in performance across these azimuths.

The improvement in auditory speech thresholds relative to the diotic condition is shown in Fig. 3 (the dashed line at 0 on the y-axis). As observed, performance was almost uniformly better when the voice stimulus was spatialized relative to the speech babble. The one exception was the 180° azimuth position with HRTF1. A repeated measures ANOVA was performed on the normalized raw data averaged across all HRTFs and as a function of HRTF condition in order to assess the significance of the differences between the diotic and the spatial presentation of the voice stimulus. The procedure for normalizing the data consisted of first calculating the mean auditory speech threshold across the diotic condition and all spatial positions for each HRTF and across all HRTFs. The mean for each HRTF was then subtracted from each auditory speech threshold value contributing to that mean. Finally, the overall mean was added to the resulting differences.

The analysis of the normalized data showed a significant difference as a function of acoustic condition (the 11 azimuth positions and the diotic condition) F (11,121) = 137.6, $p < 0.01$, and across session F (3,33) = 8.1, $p < 0.01$. A Scheffé post hoc analysis at the 0.05 alpha level indicated that the diotic condition was significantly poorer than all spatial positions, with the exception of 180° azimuth, $p < 0.05$. There was also a small but significant improvement between session one and session three, $p < 0.05$.

A significant effect of acoustic condition was also found for each of the HRTF conditions: HRTF1 F (11,55) = 185.6, $p < 0.01$; HRTF2 F (11,55) = 63.4, $p < 0.01$; and for HRTF3 F (11,55) 102.6, $p < 0.01$). However, a significant effect of session was found only for HRTF3 F (3,15) = 5.6, $p < 0.01$. The Scheffé post hoc analysis showed a significant difference between the diotic condition and each of the spatial conditions for each of the HRTFs except for 210° azimuth with HRTF1 and 180° azimuth for HRTF2. The 180° azimuth condition was significantly poorer than the diotic condition for HRTF1.

DISCUSSION

Of the three generic HRTFs used in this study, HRTF1 yielded a significantly smaller advantage in auditory speech thresholds than HRTF2 and HRTF3. The absence of published information on the generation of HRTF1 makes an explanation of the performance difference highly speculative. However, several factors are known to affect the spatial synthesis of HRTFs. For example, the measurement techniques of HRTFs differ across laboratories and are motivated by the different goals of the investigators (see references 2, 3, 5–27 of Moller et al.; 23). Some of the parameters that vary significantly in the measurement of HRTFs are type of test stimulus (e.g., sinusoidal tones or noise bursts), the point in relation to the ear canal where the measurement is made (e.g., at the blocked ear canal or a point somewhere along the ear canal), and the number of source positions.

Moreover, performance difference might be attributed to certain individuals being "better localizers" than others due to physiological differences (34). Good localizers are subjects whose free-field localization performance is better than average and whose headphone localization performance in virtual auditory space closely matches his/her free-field localization performance. The authors have no information on the subjects used to measure the HRTFs. However, F. L. Wightman (Personal communication, March 2, 1997) indicated that his laboratory has been unsuccessful in documenting any relation between HRTF characteristics and localization performance despite suggestions made in an earlier study (34). While it is clear that some subjects may have

less spectral detail to work with because their pinnae are smooth, it is not clear that this translates into poor performance. With several cues to work with, some individuals seem simply to emphasize one or more cues depending on their own physical characteristics.

The investigators concede that HRTF1 is a commercial product, perhaps not as sophisticated as the others. Since cost may ultimately be the determining factor in future applications under study, its inclusion is highly relevant and as far as the investigators are concerned, represents no loss in generality. Given that it is not presently practical nor affordable to measure the HRTFs for each potential listener, it is probable that a variety of generic HRTFs will be made commercially available and that their measurement and psychoacoustic validation techniques may not be disclosed.

Despite a similarity in the performance of HRTF2 and HRTF3, the measurement techniques employed by the investigators were different. Wightman and Kistler (33) measured impulse responses in an anechoic chamber (a room without reverberation cues down to a specified cutoff frequency) which had a higher cutoff frequency compared with the one used by Pralong and Carlile (25). This resulted in measured reflections that had to be windowed out of the final HRTFs. Although using a similar measurement approach to Wightman and Kistler (33), Pralong and Carlile (25) developed a miniature "in-ear" recording system to firmly hold the measurement microphone in the subject's ear canal. This minimized internal ear canal reflections that may have been caused by the earmold shell used by Wightman and Kistler (33). Indeed, Wightman indicated (Personal communication, March 2, 1997) deficiencies in the earmold shell described by Wightman and Kistler (33), as it changed the resonant frequency of the ear canal. When Pralong and Carlile (25) compared the localization performance of their HRTFs to those of Wightman and Kistler (34), they reported a reduction in the number of front-back reversals in the perception of azimuth and an increase in the accuracy in the judgement of elevation. "This suggests that some particularly fine detail of the HRTFs may not have been properly captured in their recordings" (25). These "fine details," which Carlile and Pralong (12) suggest may have been missed by Wightman's measurement technique, may have played a key role in the slightly improved performance of the Pralong and Carlile (25) HRTFs.

All three generic HRTFs used in this study yielded a significantly lower auditory speech threshold value compared with the diotic presentation of the voice stimulus. The maximum improvement values of 11.3 dB and 12.6 dB for HRTF2 and HRTF3, respectively, result in an approximate 75% reduction in the required headphone voltage. HRTF1 yielded a maximum improvement of 6.3 dB which would result in an approximate 50% reduction. Any of the HRTFs would allow a significant reduction in the communication system volume level without sacrificing intelligibility. Presently, headphone levels in the cockpit of military aircrafts are too loud (27). Lower listening levels over headphones could lessen the risk of hearing loss.

A greater advantage over the diotic presentation of the voice stimulus than that reported in this study might have been obtained if personal HRTFs had been used. However, personal HRTFs are traditionally derived from binaural measurements in the ears of the end-listener seated in an anechoic chamber. This requires a substantial investment in infrastructure and equipment, and is presently impractical in most applications.

The maximum gain in release from masking for all three generic HRTFs was at the 60° azimuth position. However, there was no significant difference between performance at 60° and 90° azimuth. The azimuth position in this study that yielded the lowest auditory threshold was similar to that found by Begault and Erbe (4). This is not surprising, as the 90° azimuth position corresponds to the largest interaural time difference (ITD). As illustrated in Fig. 1, the numbers used in this study have most of the spectral energy below 1.5 KHz. The spatial location and discrimination of speech in noise are partially cued by ITD below 1.5 KHz, while frequencies above 1.5 KHz are cued by interaural level difference (ILD). Dirks and Wilson (14) showed that the greatest gain in intelligibility corresponded to the presence of ITDs. Furthermore, Bronkhorst and Plomp (9) showed that ILD decreases the effectiveness of binaural unmasking of speech due to ITD.

The 6.3 dB lower auditory speech threshold at 60° azimuth for HRTF1 closely agrees with the 6–7 dB reduction at 60–90° azimuth reported by Begault and Erbe (4). In that study, the Convolvotron (Crystal River Engineering) was used to spatialize a male recording of four-letter call signs on the horizontal plane at the same azimuth positions used in the present study. The male recording was presented in the presence of diotic speech babble. As observed in **Fig. 4**, Begault and Erbe

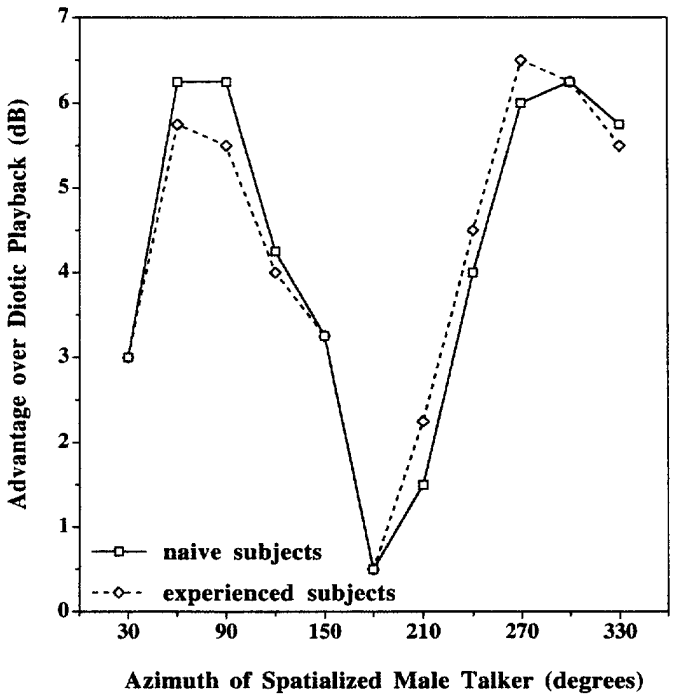

Fig. 4. Relative auditory speech thresholds with a Convolvotron system used to spatialize 4-letter call signs (male voice stimulus) in the presence of diotic speech babble. Data adapted from Begault and Erbe (4).

(4) obtained symmetrical performance centered at 180° azimuth. Performance at this position was not significantly poorer than with the diotic condition. In comparison, the 180° azimuth position was significantly poorer than with the diotic condition.

The choice of stimuli and experimental design employed in the present study necessitates further elaboration in order to assist the reader in interpreting the results. The investigators concede that auditory thresholds obtained with alternate source material may be different than those reported in this paper, but argue that there is no reason to believe that the relative rankings of the HRTFs would be different. This is partially demonstrated by the similarity in threshold levels between HRTF1 and those of Begault and Erbe (4), despite the differences in gender and vocabulary of the stimuli. The fact that only 3 of the 12 subjects participated in all three HRTF conditions does not substantially lower the generality of the results. This can be demonstrated by examining the data of subjects 2, 3, and 5, who participated in all three HRTF conditions. For each HRTF condition, their data are very similar to the data of subjects who participated in only one HRTF condition (see Fig. 3). The pattern of auditory speech threshold values for the different azimuth positions were similar across the three HRTFs (see Table I). Finally, intersubject variability was small. This is possibly due to participation in multiple sessions and the use of the PEST adaptive psychophysical procedure for determining the auditory speech thresholds. These findings suggest that data of the between- and within-subject groups across HRTF conditions were primarily due to differences in HRTFs.

The lower auditory speech thresholds of all HRTFs used in this study, compared with a diotic presentation of the voice stimulus, closely agree with the results of studies conducted in the free-field; see Blauert (6) for a review. For example, Tonning (30) measured the intelligibility of speech in the presence of broadband noise. The speech and masker sources could be independently positioned at 0°, 90°, 180°, and 270° azimuth, thus yielding 16 different combinations. The largest gain in intelligibility was 8.2 dB. This occurred when the speech was positioned at −270° azimuth and the masker was positioned at 180° azimuth. The poorest position occurred when the speech and noise were simultaneously positioned at 0° azimuth. In another study, Dirks and Wilson (14) obtained an approximate 6 dB gain in intelligibility. This was accomplished when the speech was positioned at −270° azimuth and the masker was positioned at 0° azimuth, compared with when the signal and masker were both positioned at 0° azimuth.

CONCLUSIONS AND RECOMMENDATIONS

As found in this study, not all generic HRTFs would yield the same performance if used in a general-purpose 3-D audio display. All three HRTFs yielded significantly lower auditory speech thresholds compared with the diotic presentation of the voice stimulus. Although not every subject participated in all the HRTF conditions, the present results still suggest that lower communication system headphone levels can be achieved without sacrificing intelligibility. Lower headphone levels will lessen the risk of hearing loss in aircrew.

Given that 3-D audio displays may become commonplace in cockpits, it is vital that further research be performed on the accurate modeling and implementation of HRTFs. Methods need to be developed to quickly and accurately select a generic HRTF for the targeted application. Commercial vendors of 3-D audio displays should provide some insight into measurement and psychoacoustic validation techniques for HRTFs. Additional testing is required to determine if lower auditory speech thresholds are specific to the horizontal plane and can be obtained under conditions of military aircraft tactical maneuvers.

ACKNOWLEDGMENTS

Technical assistance was provided by Babak Asadi, Dr. Phiroz Dastoor, Garry Dunn, Raymond Grondin, Stephen King, Patti Odell, Athanasia Pallas, and Andrew Welker Ruth Croxford and Ying Liu assisted with preliminary statistical data analysis

REFERENCES

1. Abel SM. The development of speech communication capability tests. Toronto, Ontario: DCIEM, 1993; Final Report for DCIEM Contract W7711-8-7047.
2. American National Standards Institute Maximum permissible ambient noise levels for audiometric test rooms. New York. Author, 1991; ANSI-S3.1.
3. Begault DR. Head-up auditory displays for traffic collision avoidance system advisories: a preliminary investigation. Hum Factors 1993; 35:707–17.
4. Begault DR, Erbe T Multichannel spatial auditory display for speech communications. J Audio Eng Soc 1994; 42:819–26.
5. Begault DR, Pittman MT Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. Int J Aviat Psychol 1996; 6.79–93
6. Blauert J. Spatial hearing. the psychophysics of human sound localization. Allen J, Trans. Cambridge: MIT Press, 1983.
7. Bronkhorst AW Localization of real and virtual sound sources. J Acoust Soc Am 1995, 98:2542–53.
8. Bronkhorst AW, Veltman JAH, van Breda L. Application of a three-dimensional auditory display in a flight task. Hum Factors 1996, 38.23–33.
9. Bronkhorst AW, Plomp R. The effect of head-induced interaural time and level differences on speech intelligibility in noise. J Acoust Soc Am 1988, 83:1508–16.
10. Calhoun GL, Janson WP, Valencia G. Effectiveness of three-dimensional auditory directional cues. In. Proceedings of the Human Factors Society 32nd Annual Meeting. Santa Monica, CA: Human Factors Society, 1988; 68–72.
11. Calhoun GL, Valencia G, Furness TA. Three-dimensional auditory cue simulation for crew station design/evaluation. In: Proceedings of the Human Factors Society 31st Annual Meeting. Santa Monica, CA:Human Factors Society, 1987; 1398–402.
12. Carlile S, Pralong D. The location-dependent nature of perceptually salient features of the human head-related transfer functions. J Acoust Soc Am 1994; 95:3445–59.
13. D'Angelo WR, Bolia RS, McKinley RL, Perrott DR. Auditory guided visual search with visual distracters. J Acoust Soc Am 1997; 101:3107.
14. Dirks DD, Wilson RH The effect of spatially separated sound sources on speech intelligibility. J Speech Hear Res 1969; 12:5–38
15. Doll TJ, Gerth JM, Engelman WR, Folds DJ. Development of simulated directional audio for cockpit applications Wright-Patterson AFB, OH:AAMRL, 1986; Report No. AAMRL-TR-86-014.
16. Ericson MA, McKinley RL. The intelligibility of multiple talkers separated spatially in noise. In: Gilkey H, Anderson TR, eds.

Binaural and spatial hearing in real and virtual environments. Mahwah, NJ: Lawrence Erlbaum Associates, 1997, 701–24.
17. Fairbanks G. Test of phonemic differentiation: the rhyme test. J Acoust Soc Am 1958; 30:596–600
18 Furness TA The super cockpit and its human factors challenges. In: Proceedings of the Human Factors Society's 30th Annual Meeting. Santa Monica: Human Factors Society, 1986, 48–52
19. Gagnon L, Cupples EJ Automatic speech recognition in additive noise II, 1995: B-1110, Brussels.NATO; Technical Report AC/243, Panel 3, TR/17
20. House AS, Williams CE, Hecker MHL, Kryter KD. Articulation-testing methods. consonantal differentiation with a closed-response set J Acoust Soc Am 1965; 37:158–66.
21. McKinley RL, Ericson MA. Flight demonstration of a 3-D auditory display In: Gilkey RH, Anderson TR, eds Binaural and spatial hearing in real and virtual environments. Mahwah, NJ. Lawrence Erlbaum Associates, 1997; 683–99
22. Macmillan NA, Creelman CD. Detection theory: a user's guide. Cambridge Cambridge University Press, 1991, 200.
23 Moller H, Sorensen MF, Hammershoi D, Jensen CB Head-related transfer functions of human subjects. J Audio Eng Soc 1995; 43:300–21
24. Perrott DR, Cisneros J, McKinley RL, D'Angelo WR. Aurally aided visual search under virtual and free-field listening conditions Hum Factors 1996; 38:702–15.
25. Pralong D, Carlile S Measuring the human head-related transfer functions: a novel method for the construction and calibration of a miniature "in-ear" recording system. J Acoust Soc Am 1994, 95:3435–44.
26 Ricard GL, Meirs SL. Intelligibility and localization of speech from virtual directions. Hum Factors 1994; 36:120–8.
27. Rood GM Noise and communication. In. Ernsting J, King P, ed. Aviation medicine, 2nd ed. Toronto: Butterworths, 1988; 353–82
28. Steenekan HJM, Houtgast T. A physical method for measuring speech-transmission quality J Acoust Soc Am 1980; 67:318–26.
29 Stinnett TA Human factors in the supercockpit. In: Jenson RS, ed. Aviation psychology. Brookfield: Gower Publishing, 1989; 1–37.
30. Tonning FM. Directional audiometry: II. The influence of azimuth on the perception of speech. Acta Otolaryngol 1971; 72.352–7.
31. Wenzel EM. Spatial sound and sonification In. Kramer G, ed Auditory display –sonification, audification, and auditory interfaces. Reading Addison-Westley Publishing Co , 1994; 127–50
32. Wenzel EM, Arruda M, Kistler DJ, Wightman FL Localization using non-individualized head-related transfer functions. J Acoust Soc Am 1993; 94.111–23.
33 Wightman FL, Kistler DJ. Headphone simulation of free-field listening. Part I: stimulus synthesis. J Acoust Soc Am 1989; 85:858–67.
34. Wightman FL, Kistler DJ. Headphone simulation of free-field listening. Part II: psychophysical validation J Acoust Soc Am 1989; 85:868–78