Australian Government
Department of Defence
Defence Science and
Technology Organisation

Interpolation of Head-Related Transfer Functions

Russell Martin and Ken McAnally

Air Operations Division
Defence Science and Technology Organisation

DSTO-RR-0323

ABSTRACT

Using current techniques it is usually impractical to measure head-related transfer functions (HRTFs) at a spatial resolution that does not exceed the minimum audible angle, i.e. 1-2° for a source directly in front, by a considerable amount. As a result, measured HRTFs must be interpolated to generate a display in which auditory space is rendered smoothly. The spatial resolution at which it is necessary to measure HRTFs for the display to be of high spatial fidelity will depend on the quality of the interpolation technique. This report describes an interpolation technique that involves the application of a novel, inverse-distance-weighted averaging algorithm to HRTFs represented in the frequency domain. The quality of this technique was evaluated by comparing four listeners' abilities to localise virtual sound sources generated using measured or interpolated HRTFs. The measured HRTFs were shown to be of sufficiently high fidelity to allow virtual sources to be localised as accurately as real sources. Localisation error measures, i.e. lateral errors, polar errors and proportions of front/back confusions, for HRTFs interpolated across up to 30° of either lateral or polar angle, or 20° of both lateral and polar angle, did not differ noticeably from those for measured HRTFs. On the basis of this finding we recommend that HRTFs be measured at a 20° lateral- and polar-angle resolution.

RELEASE LIMITATION

Approved for public release

APPROVED FOR PUBLIC RELEASE

Interpolation of Head-Related Transfer Functions

Executive Summary

The potential benefits of spatial audio displays in military environments, which include quicker visual acquisition of threats and enhanced speech intelligibility where there are multiple talkers, have been demonstrated in a large number of studies. Central to the generation of a high-fidelity spatial audio display is the measurement of the listener's head-related transfer function (HRTF), which describes the way his or her torso, head and ears filter sounds from different directions. Using current techniques it is usually impractical to measure HRTFs at a spatial resolution fine enough to match the abilities of humans to localise sound. (Humans can discriminate sound source locations separated by as little as 1-2°). As a result, an optimal spatial audio display can only be generated by interpolating HRTFs. The spatial resolution at which HRTFs must be measured for the display to be of high spatial fidelity will depend on the quality of the interpolation technique. This report describes a novel HRTF interpolation technique and a study in which its quality was evaluated by comparing four listeners' abilities to localise virtual sound sources generated using measured or interpolated HRTFs.

An evaluation of a HRTF interpolation technique can be misleading if the measured HRTFs used in the evaluation are not of high fidelity. The study began, therefore, by assessing the fidelity of the four listeners' measured HRTFs. It was found that the measured HRTFs were of sufficiently high fidelity to allow the listeners to localise virtual sound sources as accurately as real sound sources. HRTFs were interpolated across lateral angles only, i.e. across locations at different positions along only the left/right dimension, across polar angles only, i.e. across locations at different positions along only the up/down and/or the front/back dimensions, or across lateral and polar angles combined. Localisation accuracy for HRTFs interpolated across up to 30° of either lateral or polar angle, or 20° of *both* lateral and polar angle, was found to not differ noticeably from that for measured HRTFs. The HRTF interpolation technique described in this report, therefore, can be applied to HRTFs measured at a lateral- and polar-angle resolution as coarse as 20° to generate a high-fidelity spatial audio display of arbitrarily high resolution.

The availability of a HRTF interpolation technique that can be applied to HRTFs measured at a 20° lateral- and polar-angle resolution, a resolution that is currently associated with a measurement time of only 5-10 minutes, to generate a display of such high spatial fidelity should facilitate the implementation of spatial audio displays in military and other environments.

Authors

Russell Martin
Air Operations Division

Russell Martin is a Senior Research Scientist in Human Factors in Air Operations Division. He received a Ph.D. in Psychology from Monash University in 1988 and worked at the University of Queensland, Oxford University, the University of Melbourne and Deakin University prior to joining DSTO in 1995.

_____   _____

Ken McAnally
Air Operations Division

Ken McAnally is a Senior Research Scientist in Human Factors in Air Operations Division. He received a Ph.D. in Physiology and Pharmacology from the University of Queensland in 1990 and worked at the University of Melbourne, the University of Bordeaux and Oxford University prior to joining DSTO in 1996.

_____   _____

Contents

1. Introduction

This report describes work conducted as part of the work program of Project Arrangement 10 (PA10). PA10 was a six-year collaborative research and development program in aircraft electronic warfare self-protection systems between the Australian Government and the United States Army, initiated as Project Arrangement A-97-0010. The Australian activities for PA10 were conducted under Project AIR 5406 and included technology and technique development, modelling and simulation, and laboratory and field demonstrations. Within PA10, ten research tasks were created to target specific areas of interest. One of the research tasks, Task 5.1, was directed in part towards the improvement of tactical situational awareness through the development of advanced display concepts that included spatial audio displays. The work described here is concerned with the technical issue of enhancing the resolution and fidelity of spatial audio displays and is part of a broader DSTO research program concerned with the application of those displays in military aviation environments.

The potential benefits of spatial audio displays in military aviation environments have been demonstrated in several studies. These benefits include more rapid visual acquisition of threats/targets by aircraft operators [Begault 1993; Begault & Pittman 1996; Bronkhorst, Veltman & van Breda 1996; Perrott, *et al.* 1996; Flanagan *et al.* 1998; Bolia, D'Angelo & McKinley 1999; Parker *et al.* 2004] and improved speech intelligibility by operators performing multitalker communications tasks [Begault & Erbe 1994; McKinley, Erikson & D'Angelo 1994; Ricard & Meirs 1994; Crispien & Ehrenberg 1995; Drullman & Bronkhorst 2000; Ericson, Brungart & Simpson 2004].

Audio displays of high spatial fidelity can be created by synthesising at a listener's eardrums the signals that would be produced by natural, free-field presentation of sound. This can be achieved by measuring the direction-dependent filtering properties of the listener's torso, head and ears, constructing a set of digital filters having those properties, and filtering sounds with appropriate pairs of filters, i.e. one filter for each ear, before presenting them to the listener via headphones. Recent implementations of this technique have produced displays that allow virtual audio sources to be localised as accurately as are free-field, i.e. real, sources [Martin, McAnally & Senova 2001].

The direction-dependent filtering properties of a listener's torso, head and ears are described by the listener's head-related transfer function (HRTF). HRTFs are typically measured by placing small microphones in a listener's ear canals, or coupling microphones to the ear canals via probe tubes, and recording the microphones' responses to a test signal presented from a range of locations about the listener [Wightman and Kistler 1989; Bronkhorst 1995; Martin, McAnally & Senova 2001]. As listeners are required to remain very still throughout the HRTF measurement procedure and a non-trivial amount of time is required to present each test signal, the spatial resolution of locations is rarely greater than about 10° in azimuth or elevation whenever a broad region of space is sampled. In order to produce a display that renders auditory space smoothly, it is necessary to interpolate the measured HRTFs.

A number of HRTF interpolation techniques have been proposed. These techniques differ in several regards, one of these being the nature of the HRTF representation on which the interpolation algorithm operates. A HRTF describes a set of filters in the frequency domain. Any filter, however, can also be described in the time domain, where the time-domain representation is the inverse Fourier transform of the frequency-domain representation (and is referred to as the filter's *impulse response*). In addition, both frequency- and time-domain representations of a HRTF can be modelled using methods such as principle components analysis [Martens 1987; Kistler & Wightman 1992] or Karhunen-Loève expansion [Chen, van Veen & Hecox 1995; Wu *et al.* 1997]. Proposed interpolation techniques differ also with regard to the specific interpolation algorithm they incorporate. Some algorithms involve calculating a weighted average of the nearest measured filters (in whatever form they are represented), where the weights are the inverses of the linear distances between the target location and the locations associated with those filters [Wenzel & Foster 1993; Hartung, Braasch & Sterbing 1999; Langendijk & Bronkhorst 2000]. Other algorithms involve fitting spherical splines to the entire measured filter set, then solving the splines for the target location [Chen, van Veen & Hecox 1995; Hartung, Braasch & Sterbing 1999].

Techniques for interpolating HRTFs can be evaluated in two different ways. Measured and interpolated HRTFs can be compared numerically or psychophysically where the perceptual attributes of sounds filtered with measured and interpolated HRTFs are compared. The second of these approaches is preferable, as it is unlikely that the perceptual consequences of any identified numerical differences between measured and interpolated HRTFs could be accurately predicted. In the study described in this report, a HRTF interpolation technique was evaluated by comparing listeners' abilities to localise virtual sound sources generated using measured or interpolated HRTFs. HRTF measurements were made for each listener, i.e. HRTFs were individualised. Frequency-domain representations of measured HRTFs were interpolated by a novel algorithm that involved calculating an inverse-distance-weighted average of the four filters nearest to the target location.

2. Methods

2.1 Participants

Two female and two male volunteers, of ages ranging from 26 to 45 years, participated in this study. All were employees of the Defence Science and Technology Organisation and all reported having normal hearing.

Each participant had considerable prior experience localising real and virtual sound sources under the conditions of the present study.

2.2 HRTF Measurement

The HRTF of each participant was measured using a "blocked ear canal" technique, [e.g. Møller et al. 1995]. Miniature microphones (Sennheiser, KE4-211-2) encased in swimmer's ear putty were placed in the participant's left and right ear canals (see Figure 1). Care was taken to ensure that the microphones were positionally stable and that their diaphragms were at least 1 mm inside the ear canal entrances.
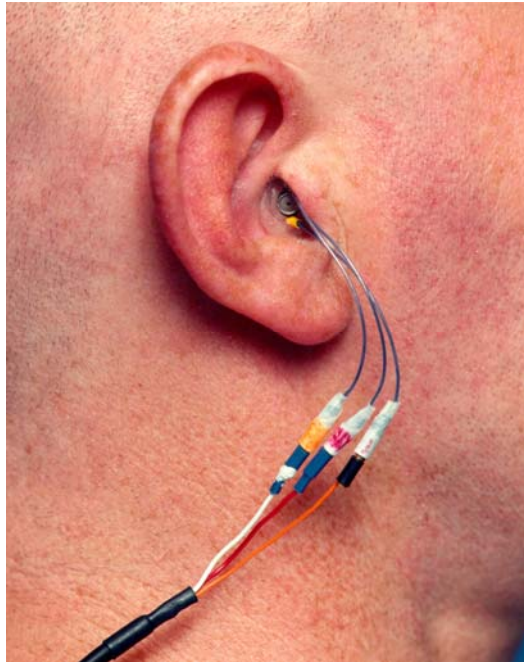

Figure 1.   Miniature microphone inserted in a participant's right ear canal

The participant was seated in a 3 x 3 m, sound-attenuated, anechoic chamber at the centre of a 1 m radius hoop on which a loudspeaker (Bose, FreeSpace tweeter) was mounted (see Figure 2). The hoop could be rotated by programmable stepping motors to position the loudspeaker with a resolution of 0.1° at any azimuth and from -50 to +80° of elevation. A convention of describing elevations in the hoop's lower hemisphere as negative was followed.

The participant placed his/her chin on a rest that helped position his/her head at the centre of the hoop and orient it toward 0° azimuth and elevation. Head position and orientation were tracked magnetically via a receiver (Polhemus, 3Space Fastrak) attached to a plastic headband that was worn by the participant. The position and orientation of the participant's head were displayed on a bank of light emitting diodes (LEDs) mounted within the participant's field of view. HRTF measurement did not proceed unless the participant's head was stationary, i.e. its x, y and z coordinates did not vary by more than 2 mm and its azimuth, elevation and roll did not vary by more than 0.2° over three successive readings of the head tracker made at 20-ms intervals, no more than 3 mm from the hoop centre in the x, y and z directions, and oriented within 1° of straight and level.

HRTFs were measured for attainable sound-source locations, i.e. locations where the loudspeaker could be positioned, at lateral angles, i.e. angles subtended at the centre of the head between the location and the nominal median-vertical plane, ranging from -90 to +90° in steps of 10°, and polar angles, i.e. angles of rotation around the nominal interaural axis, ranging from 0 to 359.9° in steps of 360° (for +/- 90° lateral angles), 30° (for +/- 80° lateral angles), 20° (for +/- 70 and 60° lateral angles) and 10° (for all other lateral angles). For each location, two 8192-point Golay codes [Golay 1961] were generated at a rate of 50 kHz (Tucker-Davis Technologies, System II), amplified and played at 75 dB SPL (A-weighted) through the hoop-mounted loudspeaker. The signal from each microphone was low-pass filtered at 20 kHz and sampled at 50 kHz (Tucker-Davis Technologies, System II) for 327.7 ms following initiation of the Golay codes. An impulse response was derived from each sampled signal [Zhou, 1992], truncated to 128 points and stored.
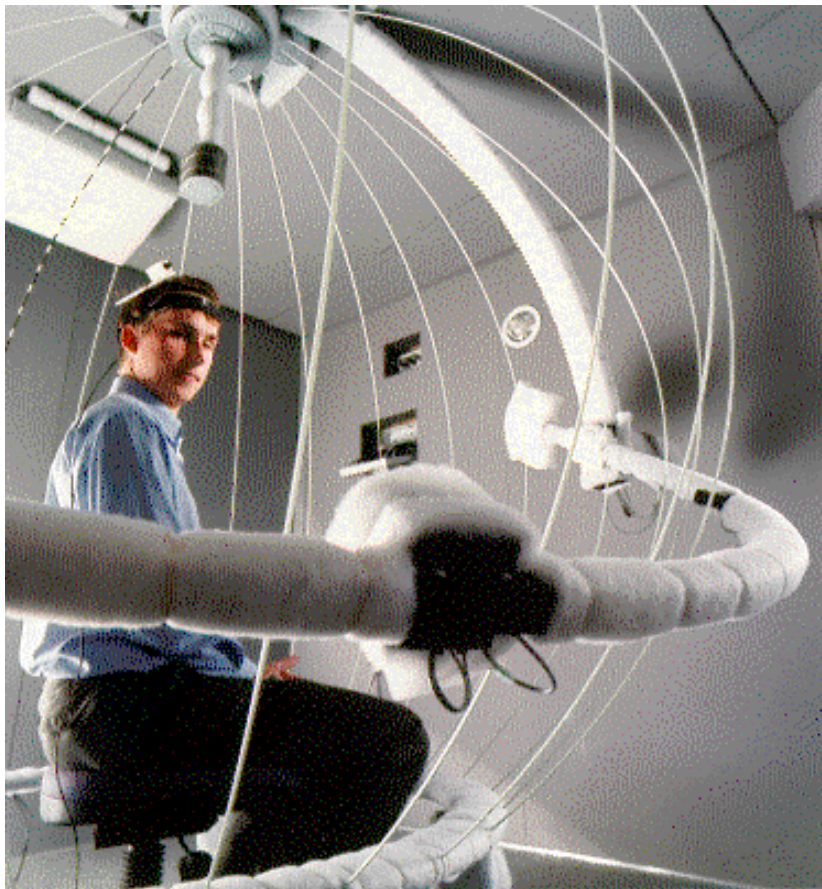

Figure 2. A participant seated at the centre of the hoop. (The acoustically transparent cloth that normally covers the fibreglass rods has been removed for clarity. Note that this picture was taken before the anechoic treatment of the chamber was completed.)

The transfer functions of the two miniature microphones were then measured together with those of the headphones that were subsequently used to present stimuli during localisation trials (Sennheiser, HD520 II). The headphones were carefully placed on the participant's head and Golay codes were played through them while the signals from the microphones were

sampled. An impulse response was derived from each sampled signal, truncated to 128 points, zero-padded to 370 points, inverted in the complex frequency domain and stored. The transfer function of the system through which the Golay codes were presented, i.e. the hoop-mounted loudspeaker, etc, had been measured previously using a microphone with a flat frequency response (Brüel and Kjær, 4003). The impulse response of this system, which had been truncated to 128 points, was deconvolved from each HRTF by division in the complex frequency domain.

2.3 Fidelity of Measured HRTFs

The fidelity of measured HRTFs was assessed by comparing the accuracies with which participants could localise real and virtual sound sources. Each participant completed four sessions each containing 42 trials in which real sources were localised and four sessions each containing 42 trials in which virtual sources were localised. The order of presentation of conditions was counterbalanced within and across participants following a randomised-blocks design.

For all sessions the participant was seated on a swivelling chair at the centre of the loudspeaker hoop in the same anechoic chamber in which his/her HRTFs had been measured. The participant's view of the hoop and loudspeaker was obscured by an acoustically transparent, 99-cm radius cloth sphere supported by thin fibreglass rods. The inside of this sphere was dimly lit to allow visual orientation. Participants wore a headband on which a magnetic-tracker receiver and a laser pointer were rigidly mounted. When localising virtual sound sources they also wore the headphones for which transfer functions had been measured.

At the beginning of each trial the participant placed his/her chin on the rest and fixated on an LED at 0° azimuth and elevation. When ready, he/she pressed a hand-held button. An acoustic stimulus was then presented, provided the participant's head was stationary, no more than 10 mm from the hoop centre in the x, y and z directions, and oriented within 3° of straight and level. Participants were instructed to keep their heads stationary during stimulus presentation.

For both real and virtual sound sources, each stimulus consisted of an independent sample of Gaussian noise generated at a sampling rate of 50 kHz (Tucker-Davis Technologies AP2). Each sample was 328 ms in duration and incorporated 20-ms cosine-shaped rises and falls. For real sources, the Gaussian noise sample was filtered to compensate for the transfer function of the stimulus presentation system, i.e. the hoop-mounted loudspeaker, etc, converted to an analogue signal (Tucker-Davis Technologies PD1), low-pass filtered at 20 kHz (Tucker-Davies Technologies FT5), amplified (Hafler Pro 1200) and presented via the hoop-mounted loudspeaker at 60 dB SPL (A-weighted). For virtual sources, the Gaussian noise sample was filtered with the participant's location-appropriate HRTF, filtered to compensate for the transfer function of the stimulus presentation system, i.e. the headphones, etc, converted to an analogue signal (Tucker-Davis Technologies PD1), low-pass filtered at 20 kHz (Tucker-Davies Technologies FT5) and presented via the headphones at 60 dB SPL (A-weighted).

Following stimulus presentation, the head-mounted laser pointer was turned on and the participant turned his/her head (and body, if necessary) to orient the laser pointer's beam toward the point on the cloth sphere from which he/she perceived the stimulus to come. The location and orientation of the laser pointer were measured using the magnetic tracker, and the point where the beam intersected the sphere was calculated geometrically. Prior calibration had established that the absolute error associated with this procedure did not exceed 2.5° for any of 354 locations spread across the part sphere extending from 0 to 359.9° in azimuth and from -40 to +70° in elevation and was less than 1° when averaged across these locations.

The sound-source location for each trial was chosen pseudorandomly from 428 of the 448 locations for which HRTFs had been measured. (This was the case for real as well as virtual sources.) The part-sphere extending from 0 to 359.9° in azimuth and from -47.6 to +79.9° in elevation was divided into 42 sectors of equal area. Each sector contained from 7 to 15 locations for which HRTFs had been measured. To ensure a reasonably even spread of source locations in each session, one sector was selected randomly without replacement on each trial and a location within it was then selected randomly. For real sources, the loudspeaker was moved to the new source location before each trial began. Loudspeaker movement occurred in two steps to reduce the likelihood of participants discerning the source location from the duration of movement. During the first step, the loudspeaker was moved to a randomly-chosen location at least 30° in azimuth and elevation away from both the previous and the new locations. During the second step, it was moved to the new location.

Stimuli presented via the hoop-mounted loudspeaker were calibrated using a microphone (Brüel and Kjær, 4003) and a sound-level meter (Brüel and Kjær, 2209). Stimuli presented via headphones were calibrated using an acoustic manikin incorporating a sound level meter (Head Acoustics, HMS II.3). The manikin was placed inside the anechoic chamber such that its head was centred with respect to the hoop and oriented straight and level. The hoop-mounted loudspeaker was positioned at 270° azimuth/0° elevation and Gaussian noise that had been low-pass filtered at 20 kHz was presented via it at a level that produced a 60 dB SPL (A-weighted) signal at the centre of the hoop. The sound level at the manikin's left ear was recorded. The headphones that were used to present stimuli during localisation trials were then placed on the manikin's head and Gaussian noise that had been filtered with the manikin's HRTFs for 270° azimuth/0° elevation and low-pass filtered at 20 kHz was presented via them. The level of the noise was adjusted until the sound level at the manikin's left ear was equivalent to that associated with presentation of noise from the hoop-mounted loudspeaker.

2.4 Fidelity of Interpolated HRTFs

The fidelity of interpolated HRTFs was assessed for interpolation across lateral angles only, polar angles only, and lateral and polar angles combined. For interpolation across lateral angles only, the spacing of measured HRTFs made available to the interpolation algorithm was 10, 20, 30, 60 or 90° with respect to lateral angle and *as measured* (see section 2.3) with respect to polar angle. Each participant completed four 50-trial sessions for each condition of HRTF lateral-angle spacing, i.e. 10, 20, 30, 60 or 90°, and for a non-interpolated HRTF condition. Conditions were presented in an order that was counterbalanced within and across

participants following a randomised-blocks design. For interpolation across polar angles only, the spacing of available HRTFs was *as measured*, i.e. 10°, with respect to lateral angle and at least 10, 20, 30, 60 or 90° with respect to polar angle. Each participant completed four 50-trial sessions for each condition of HRTF polar-angle spacing and for a non-interpolated HRTF condition. Conditions were presented in an order that was counterbalanced within and across participants following a randomised-blocks design. For interpolation across lateral and polar angles combined, the spacing of available HRTFs was 20° with respect to lateral angle and at least 20° with respect to polar angle. Each participant completed four 50-trial sessions for the interpolated HRTF condition and for a non-interpolated HRTF condition. Conditions were presented in an order that was counterbalanced within and across participants following an ABBA design.

Where interpolation was across lateral angles only, the sound-source lateral angle for each trial was selected randomly from the range extending from -90 to +90°. From the lateral angles associated with available HRTFs, the two nearest to the selected lateral angle were identified. The sound-source polar angle was then selected randomly from those at which HRTFs were measured for the lateral angle of larger absolute value. (Two exceptions were where the lateral angle of larger absolute value was -90 or + 90°, in which case the sound-source polar angle was selected randomly from those at which HRTFs were measured for the lateral angle of *smaller* absolute value.) For example, if the lateral-angle spacing of available HRTFs was 20° and a sound-source lateral angle of -63.26° was selected, then the sound-source polar angle would have been selected randomly from the range extending from 0 to 340° in steps of 20°, i.e. the polar angles at which HRTFs were measured for a lateral angle of -70°. Where interpolation was across polar angles only, the sound-source lateral angle for each trial was selected randomly from those associated with measured HRTFs. The sound-source polar angle was then selected randomly from the range extending from 0 to 360°, with the constraint that the sound-source elevation was within the range from -50 to +80°. Where interpolation was across lateral and polar angles combined, the sound-source lateral angle was selected randomly from the range extending from -90 to +90°, then the sound-source polar angle was selected randomly from the range extending from 0 to 360°, with the constraint that the sound-source elevation was within the range from -50 to +80°.

In all other respects, the procedures followed were identical to those described in the previous section in relation to the assessment of participants' abilities to localise virtual sound sources.

2.5 HRTF Interpolation

HRTFs were interpolated in the frequency domain. From the locations associated with available HRTFs, the four nearest to the sound-source location were identified. The HRTFs for those locations were then split into log-magnitude and phase components.

The log-magnitude components of the four HRTFs were summed after each was multiplied by the following weight:

$$((LatStep - |Lat_{HRTF} - Lat_{SS}|) * (PolStep - |Pol_{HRTF} - Pol_{SS}|)) / (LatStep * PolStep)$$

where,
- o *LatStep* is the lateral-angle spacing of available HRTFs
- o *Lat_{HRTF}* is the lateral angle of the HRTF
- o *Lat_{SS}* is the lateral angle of the sound source
- o *PolStep* is the polar-angle spacing of available HRTFs for the HRTF lateral angle
- o *Pol_{HRTF}* is the polar angle of the HRTF
- o *Pol_{SS}* is the polar angle of the sound source.

Phase components were interpolated in two steps. For each of the two lateral angles associated with the four HRTFs, the phase components of the two HRTFs for that lateral angle were adjusted on a frequency-by-frequency basis by adding 360° to one or the other until the unsigned difference between the two was less than or equal to 180°. The phase components of the two HRTFs were summed after each was multiplied by the following weight:

$$(PolStep - |Pol_{HRTF} - Pol_{SS}|) / PolStep$$

The two summed phase components were adjusted on a frequency-by-frequency basis by adding 360° to one or the other until the unsigned difference between the two was less than or equal to 180°. They were then summed after each was multiplied by the following weight:

$$(LatStep - |Lat_{Sum} - Lat_{SS}|) / LatStep$$

where,
- o *Lat_{Sum}* is the lateral angle of the two HRTFs from which the summed component was generated.

2.6 Data Analysis

Localisation accuracy was described in terms of two errors: lateral error and elevation error. Lateral error was defined as the unsigned difference between the true and perceived sound-source lateral angles. Elevation error was defined as the unsigned difference between the true and perceived sound-source elevations. The true location of virtual sources was calculated taking the position and orientation of the participant's head at the time of stimulus presentation into account.

For each participant, a median lateral error and a median elevation error were calculated for each condition after removal of data for those trials on which a front/back confusion was made. (Medians were preferred to means because the distributions of these errors tended to be skewed. Data from trials on which a front/back confusion was made were removed because front/back confusions appear to be qualitatively different from other localisation errors.) A front/back confusion was deemed to have been made if two conditions were met. The first was that neither the true nor the perceived sound-source location fell within a narrow exclusion zone symmetrical about the vertical plane dividing the front and back hemispheres of the hoop. The width of this exclusion zone, in degrees of azimuth, was 15 divided by the cosine of the elevation. (Note that the arc length associated with 1° of azimuth is greatest at 0° of elevation and becomes progressively smaller as either vertical pole is approached.) The second condition was that the true and perceived sound-source locations

were in different front-versus-back hemispheres. The proportion of front/back confusions was calculated for each participant and condition by dividing the number of trials on which a front/back confusion was made by the number of trials on which neither the true nor the perceived sound-source location fell within the exclusion zone.

3. Results

3.1 Fidelity of Measured HRTFs

Median lateral and elevation errors averaged across participants (by calculating arithmetic means) are shown in Figure 3 for localisation of real sound sources and virtual sound sources generated from measured, i.e. non-interpolated, HRTFs. Neither error measure differed substantially across sound sources. Median lateral errors for individual participants ranged from 4.6 to 7.0° for real sources and from 4.7 to 7.5° for virtual sources. Median elevation errors for individual participants ranged from 5.5 to 8.3° for real sources and from 5.8 to 7.1° for virtual sources.
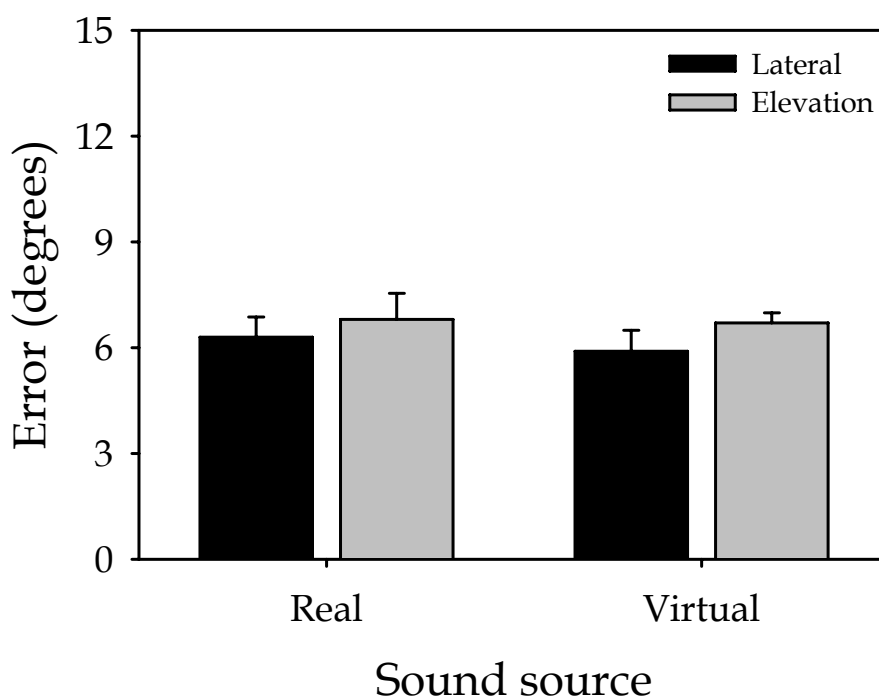

Figure 3:   Median lateral and elevation errors averaged across participants for real and virtual sound sources. Each error bar shows one standard error of the average.

Proportions of front/back confusions averaged across participants are shown in Figure 4 for localisation of real and virtual sound sources. The average proportion of front/back confusions for virtual sources was twice that for real sources, but neither was particularly high. Proportions of front/back confusions for individual participants ranged from 0.01 to 0.06 for real sources and from 0.05 to 0.06 for virtual sources.
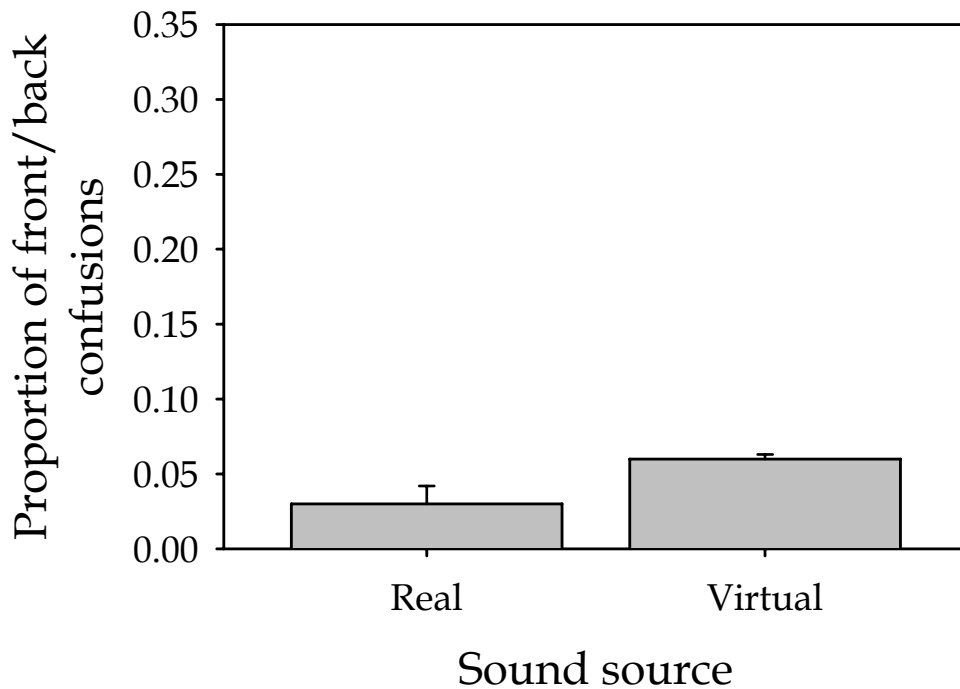
*Figure 4:   Proportions of front/back confusions averaged across participants for real and virtual sound
sources. Each error bar shows one standard error of the average*

The similarity across sound sources of all error measures indicates that the fidelity of
measured HRTFs was high.

3.2 Fidelity of Interpolated HRTFs

Median lateral errors, median elevation errors and proportions of front/back confusions
averaged across participants are shown in Figures 5, 6 and 7, respectively, for localisation of
virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across
10, 20, 30, 60 and 90° of lateral angle.

Average lateral errors for HRTFs interpolated across 10, 20 and 30°of lateral angle were no
greater than the average lateral error for non-interpolated HRTFs. Median lateral errors for
individual participants ranged from 5.0 to 8.7° for non-interpolated HRTFs and from 5.3 to
7.5° for HRTFs interpolated across 30° of lateral angle. Average lateral errors for HRTFs
interpolated across 60 and 90° of lateral angle were, respectively, 2.2 and 4.1° greater than the
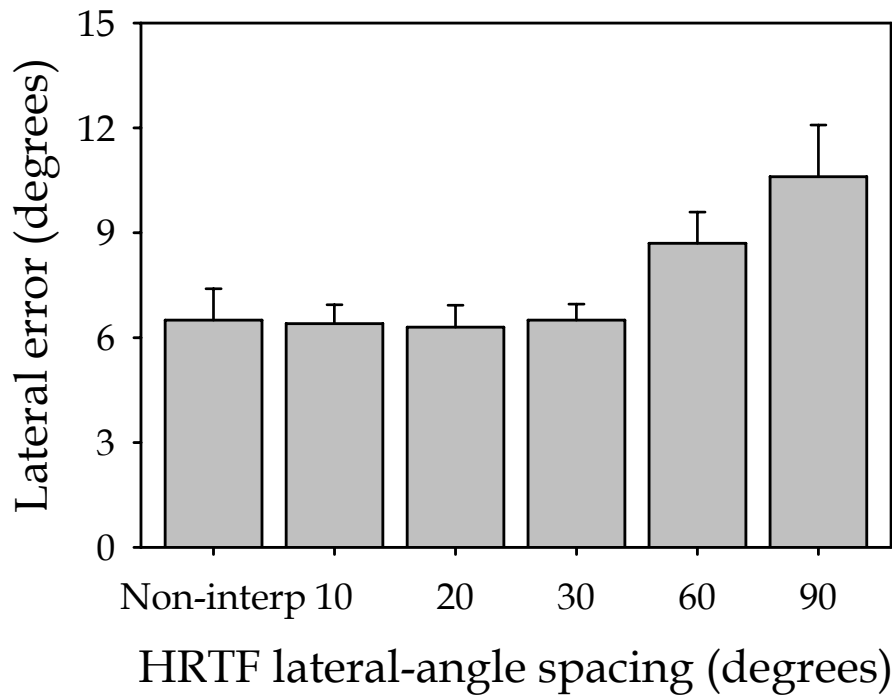average lateral error for non-interpolated HRTFs.

Figure 5: *Median lateral errors averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of lateral angle. Each error bar shows one standard error of the average.*

Figure 6: *Median elevation errors averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of lateral angle. Each error bar shows one standard error of the average.*
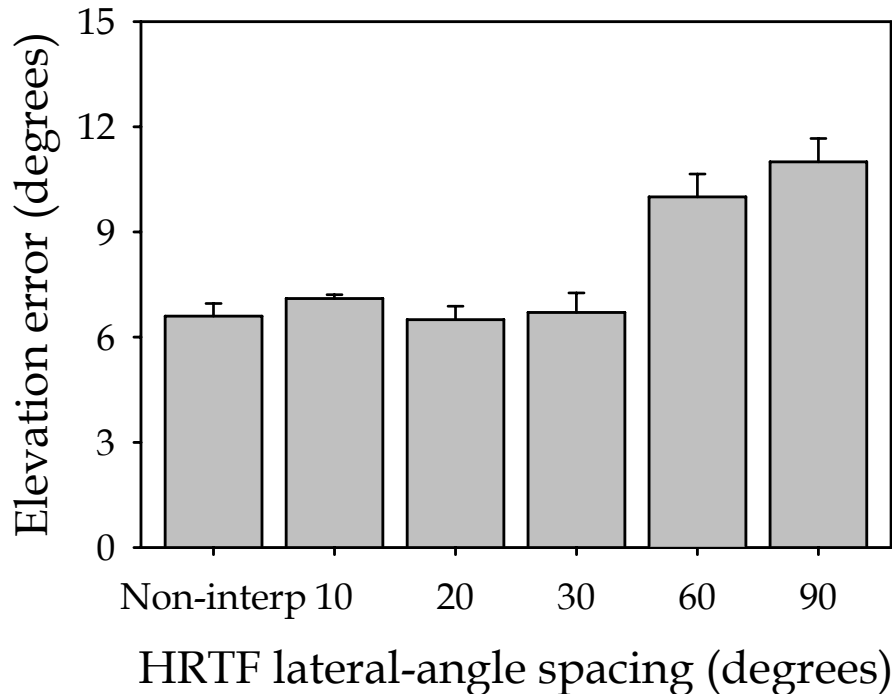
Figure 7: *Proportions of front/back confusions averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of lateral angle. Each error bar shows one standard error of the average.*
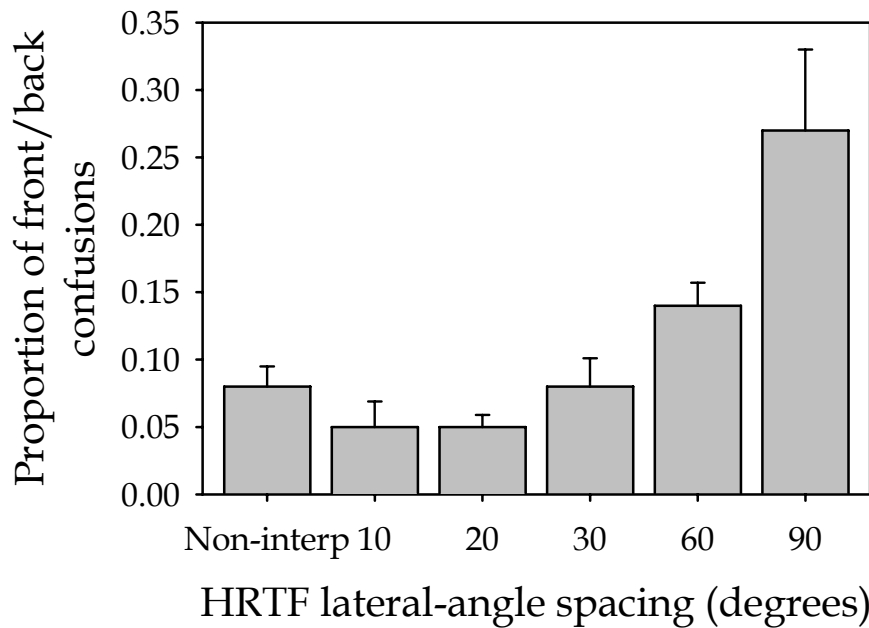
Average elevation errors for HRTFs interpolated across 10, 20 and 30° of lateral angle were similar to the average elevation error for non-interpolated HRTFs. Median elevation errors for individual participants ranged from 5.9 to 7.6° for non-interpolated HRTFs and from 5.5 to 7.9° for HRTFs interpolated across 30° of lateral angle. Average elevation errors for HRTFs interpolated across 60 and 90° of lateral angle were, respectively, 3.4 and 4.4° greater than the average elevation error for non-interpolated HRTFs.

Average proportions of front/back confusions for HRTFs interpolated across 10, 20 and 30° of lateral angle were no greater than the average proportion of front/back confusions for non-interpolated HRTFs. Proportions of front/back confusions for individual participants ranged from 0.05 to 0.12 for non-interpolated HRTFs and from 0.02 to 0.12 for HRTFs interpolated across 30° of lateral angle. Average proportions of front/back confusions for HRTFs interpolated across 60 and 90° of lateral angle were, respectively, 0.06 and 0.19 greater than the average proportion of front/back confusions for non-interpolated HRTFs.

Median lateral errors, median elevation errors and proportions of front/back confusions averaged across participants are shown in Figures 8, 9 and 10, respectively, for localisation of virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of polar angle.

Average lateral errors for HRTFs interpolated across 10, 20, 30, 60 and 90° of polar angle were similar to the average lateral error for non-interpolated HRTFs. Median lateral errors for individual participants ranged from 5.5 to 6.9° for non-interpolated HRTFs and from 4.8 to 7.5° for HRTFs interpolated across 90° of lateral angle.
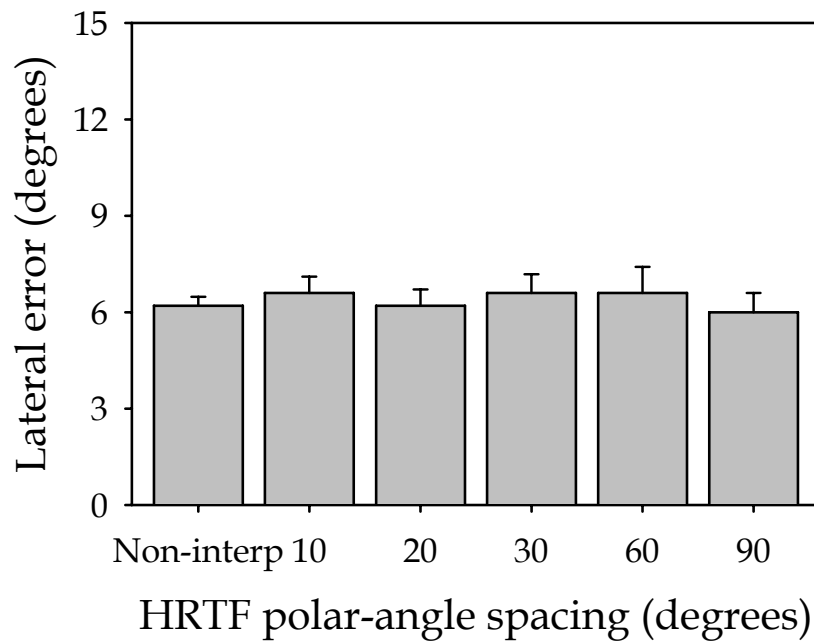
Figure 8: *Median lateral errors averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of polar angle. Each error bar shows one standard error of the average.*

Figure 9: *Median elevation errors averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of polar angle. Each error bar shows one standard error of the average.*

Figure 10: *Proportions of front/back confusions averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 10, 20, 30, 60 and 90° of polar angle. Each error bar shows one standard error of the average.*
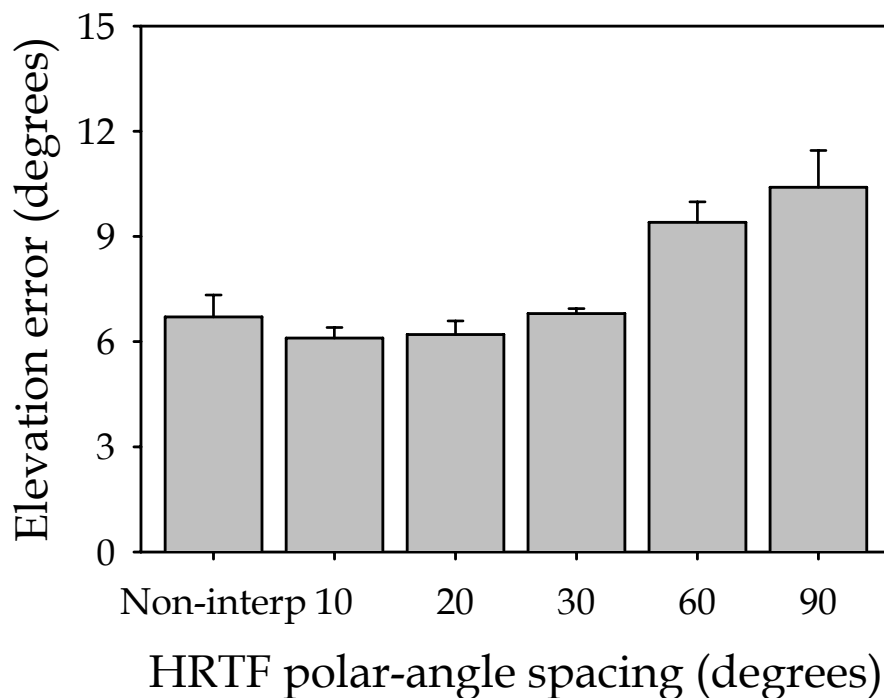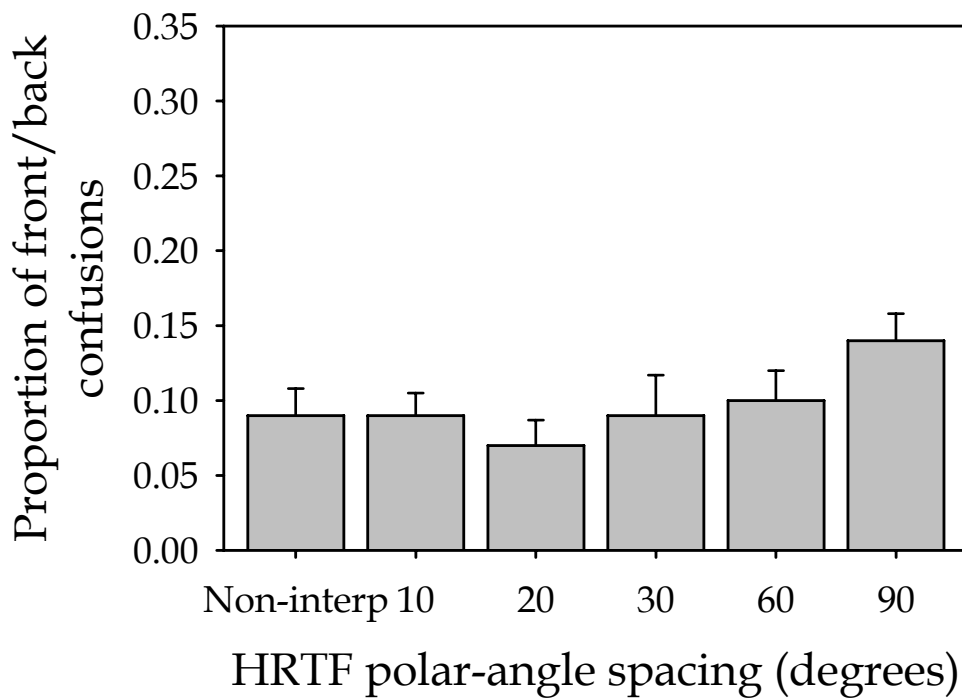
Average elevation errors for HRTFs interpolated across 10, 20 and 30° of polar angle were similar to the average elevation error for non-interpolated HRTFs. Median elevation errors for individual participants ranged from 5.5 to 8.3° for non-interpolated HRTFs and from 6.5 to 7.1° for HRTFs interpolated across 30° of polar angle. Average elevation errors for HRTFs interpolated across 60 and 90° of polar angle were, respectively, 2.7 and 3.7° greater than the average elevation error for non-interpolated HRTFs.

Average proportions of front/back confusions for HRTFs interpolated across 10, 20, 30 and 60° of polar angle were similar to the average proportion of front/back confusions for non-interpolated HRTFs. Proportions of front/back confusions for individual participants ranged from 0.05 to 0.14 for non-interpolated HRTFs and from 0.07 to 0.13 for HRTFs interpolated across 60° of polar angle. The average proportion of front/back confusions for HRTFs interpolated across 90° of polar angle was 0.05 greater than that for non-interpolated HRTFs.

Median lateral and elevation errors averaged across participants are shown in Figure 11 for localisation of virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 20° of lateral angle and at least 20° of polar angle. Neither error measure differed substantially across HRTF type. Median lateral errors for individual participants ranged from 5.0 to 9.1° for non-interpolated HRTFs and from 4.4 to 8.2° for interpolated HRTFs. Median elevation errors for individual participants ranged from 5.0 to 7.5° for non-interpolated HRTFs and from 5.5 to 8.8° for interpolated HRTFs.
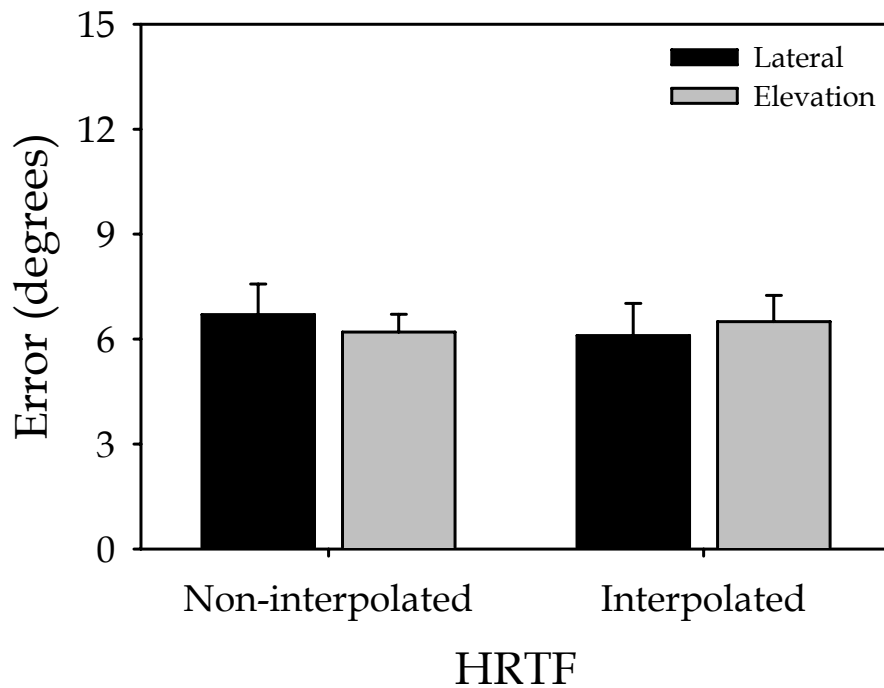
Figure 11: Median lateral and elevation errors averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 20° of lateral angle and at least 20° of polar angle. Each error bar shows one standard error of the average.
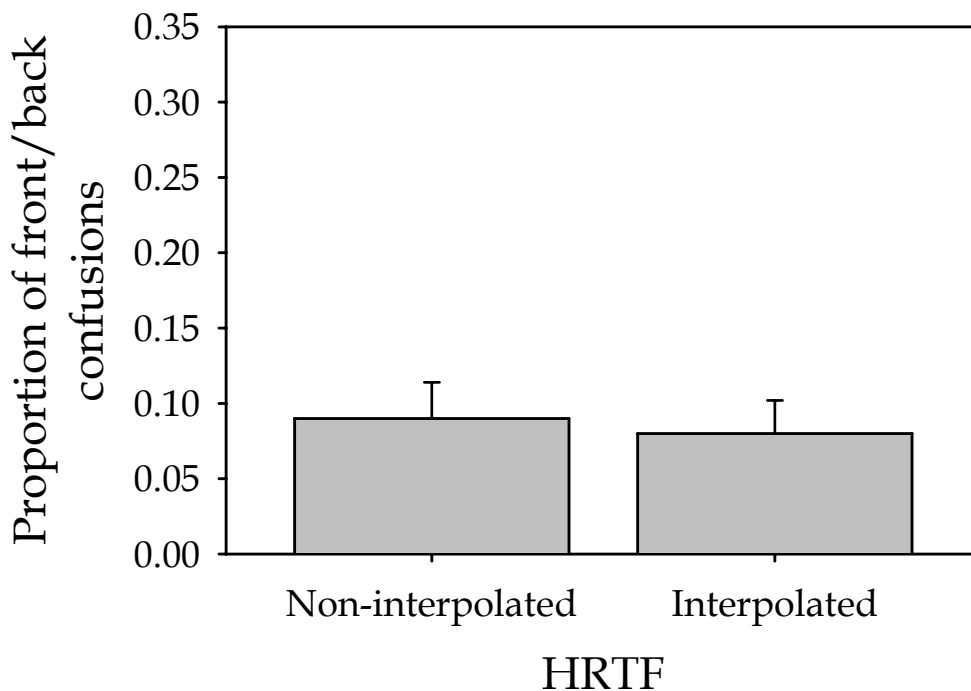
Figure 12: Proportions of front/back confusions averaged across participants for virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 20° of lateral angle and at least 20° of polar angle. Each error bar shows one standard error of the average.

Proportions of front/back confusions averaged across participants for localisation of virtual sound sources generated from non-interpolated HRTFs and HRTFs interpolated across 20° of lateral angle and at least 20° of polar angle are shown in Figure 12. Average proportions of front/back confusions were similar across HRTF type. Proportions of front/back confusions for individual participants ranged from 0.06 to 0.16 for non-interpolated HRTFs and from 0.04 to 0.13 for interpolated HRTFs.

4. Discussion

When evaluating techniques for interpolating HRTFs it is essential to ensure that the measured HRTFs are of high perceptual fidelity. If they are not, the inadequacies of a poor interpolation technique may not be revealed [see, for example, Wenzel & Foster 1993 in which poorly localised, non-individualised HRTFs were interpolated]. The measured HRTFs in the present study were shown to be of sufficiently high fidelity to allow virtual sound sources to be localised as accurately as real sound sources. They therefore provide a demanding standard against which the fidelity of interpolated HRTFs can be judged.

Localisation error measures for HRTFs interpolated across up to 30° of either lateral or polar angle were observed in this study to not differ noticeably from those for measured HRTFs. Furthermore, this was also the case for HRTFs interpolated across 20° of *both* lateral and polar angle. On the basis of these findings we recommend that HRTF measurements be made at 20° lateral and polar angle steps. Increasing lateral and polar angle steps from 10° (the size that has routinely been used in our and several other laboratories) to 20° reduces both the number of locations for which HRTFs are measured and the time taken to make the measurements by a factor of approximately four. Given the difficulties people experience remaining very still for extended periods of time, as they must do to avoid an undesirable level of contamination of measured HRTFs, a reduction of this magnitude is particularly significant.

At least two previous studies have examined the effect of varying the spatial resolution of the HRTF set upon which interpolation is based on the fidelity of interpolated HRTFs for measured HRTFs of established high fidelity. Langendijk & Bronkhorst [2000] evaluated a nearest neighbour interpolation algorithm, i.e. an algorithm of the general type evaluated in the present study, applied to frequency-domain representations of individualised HRTFs for HRTF sets having spatial resolutions of 5.6, 11.3 and 22.5°. The fidelity of interpolated HRTFs was assessed psychophysically in a discrimination test. Langendijk & Bronkhorst found that HRTFs interpolated from a set having a spatial resolution of 5.6° could not be discriminated from non-interpolated HRTFs but those interpolated from a set having a spatial resolution of 11.3 or 22.5° could be. As Langendijk & Bronkhorst did not perform an explicit localisation test, it is not clear if the discriminable differences they observed between interpolated and non-interpolated HRTFs would have led to differences in the accuracies with which sound sources synthesised from those HRTFs could be localised.

Carlile, Jin & van Raad [2000] evaluated a nearest neighbour and a spherical spline interpolation algorithm applied to the results of principal components analyses of frequency-domain representations of individualised HRTFs. The fidelity of interpolated HRTFs

generated by both algorithms was assessed numerically by calculating root-mean-square errors between the magnitude components of interpolated and non-interpolated HRTFs. The fidelity of interpolated HRTFs generated by the spherical spline algorithm was also assessed psychophysically using an explicit localisation test. The spherical spline algorithm was found to perform better than the nearest neighbour algorithm according to the numerical assessment. (A similar advantage of a spherical spline over a nearest neighbour interpolation algorithm was reported by Hartung, Braasch & Sterbing [1999].) For both algorithms, root-mean-square errors increased markedly when the steps between adjacent locations in the HRTF set on which the interpolation was based increased above 15°. Likewise, the psychophysical assessment of the spherical spline algorithm indicated that 15° is the critical step size above which the fidelity of interpolated HRTFs starts to decrease.

To the extent that it can be judged, it appears that the interpolation algorithm evaluated in the present study performed as well as those evaluated in previous studies. This certainly seems to be the case with respect to the algorithms evaluated by Carlile, Jin & van Raad [2000]. The spherical spline algorithm applied by Carlile, Jin & van Raad is arguably more sophisticated than the nearest neighbour algorithm applied by us. For one thing, spherical spline algorithms take the inherent, i.e. spherical, geometry of HRTF data sets into account. It is possible, however, that nearest neighbour algorithms benefit from focussing on the way HRTFs change with location in the region of space local to the target location. Spherical spline algorithms, in contrast, are based on an approximation of the pattern of HRTF changes across the entire sphere (or at least the part of it covered by the HRTF data set).

5. References

Begault, D.R. (1993). Head-up auditory displays for traffic collision avoidance systems advisories: a preliminary investigation. Human Factors, 35, 707-717.

Begault, D., & Erbe, T. (1994). Multichannel spatial auditory display for speech communications. *Journal of the Audio Engineering Society*, **42**, 819-826.

Begault, D.R., & Pittman, M.T. (1996). Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. *The International Journal of Aviation Psychology*, **61**, 79-93.

Bolia, R.S., D'Angelo, W.R., & McKinley, R.L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors*, **41**, 664-669.

Bronkhorst, A.W. (1995). Localization of real and virtual sound sources. *Journal of the Acoustical Society of America*, **98**, 2542-2553.

Bronkhorst, A.W., Veltman, J.A., & van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors*, **38**, 23-33.

Carlile, S., Jin, C., & van Raad, V. (2000). Continuous virtual auditory space using HRTF interpolation: acoustic & psychophysical errors. *International Symposium on Multimedia Information Processing*, 220-223, Sydney.

Chen, J., van Veen, B.D., & Hecox, K.E. (1995). A spatial feature extraction and regularization model for the head-related transfer function. *Journal of the Acoustical Society of America*, **97**, 439-452.

Crispien, K, & Ehrenberg, T. (1995). Evaluation of the "cocktail-party effect" for multiple speech stimuli within a spatial auditory display. *Journal of the Audio Engineering Society*, **11**, 932-941.

Drullman, R., & Bronkhorst, A.W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *Journal of the Acoustical Society of America*, **107**, 2224-2235.

Ericson, M.A., Brungart, D.S., & Simpson, B.D. (2004). Factors that influence intelligibility in multitalker speech displays. *The International Journal of Aviation Psychology*, **14**, 313-334.

Flanagan, P., McAnally, K.I., Martin, R.L., Meehan, J.W., & Oldfield, S.R. (1998). Aurally and visually guided visual search in a virtual environment. *Human Factors*, **40**, 461-468.

Golay, M.J.E. (1961). Complimentary series. *IRE Transactions on Information Theory*, **7**, 82-87.

Hartung, K., Braasch, J., & Sterbing, S.J. (1999). Comparison of different methods for the interpolation of head-related transfer functions. In *The Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, 319-329, Audio Engineering Society.

Kistler, D.J., & Wightman, F.L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, **91**, 1637-1647.

Langendijk, E.H.A., & Bronkhorst, A.W. (2000). Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *Journal of the Acoustical society of America*, **107**, 528-537.

Martens, W.L. (1987). Principal components analysis and resynthesis of spectral cues to perceived direction. In J. Beauchamp, (Ed.) *Proceedings of the International Computer Music Conference*, 274-281, San Francisco, International Computer Music Association.

Martin, R.L., McAnally, K.I., & Senova, M.A. (2001). Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society*, **49**, 14-22.

McKinley, R.L., Erikson, M.A., & D'Angelo, W.R. (1994). 3-Dimensional auditory displays: Development, applications, and performance. *Aviation, Space, and Environmental Medicine*, **65**, A31-38.

Møller, H., Sørenson, M.F., Hammershøi, D., & Jensen, C.B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, **43**, 300-321.

Parker, S.P.A., Smith, S.E., Stephan, K.L., Martin, R.L., & McAnally, K.I. (2004). Effects of supplementing head-down displays with 3-D audio during visual target acquisition. *International Journal of Aviation Psychology*, **14**, 277-295.

Perrott, D.R., Cisneros, J., McKinley, R.L., & D'Angelo, W.R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors*, **38**, 702-715.

Ricard, G.L., & Meirs, S.L. (1994). Intelligibility and localization of speech from virtual directions. *Human Factors*, **36**, 120-128.

Wenzel, E.M., & Foster, S.H. (1993). Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. In *Proceedings of the ASSP (IEEE) 1993 Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, Institute of Electrical and Electronics Engineers.

Wightman, F.L., & Kistler, D.J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America*, **85**, 858-867.

Wu, Z., Chan, F.H.Y., Lam, F.K., & Chan, J.C. (1997). A time domain binaural model based on spatial feature extraction for the head-related transfer function. *Journal of the Acoustical Society of America*, **102**, 2211-2218.

Zhou, B., Green, D.M., & Middlebrooks, J.C. (1992). Characterization of external ear impulse responses using Golay codes. *Journal of the Acoustical Society of America*, **92**, 1169-1171.

<table>
<tr><td colspan="2">
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA
</td><td>
1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)
</td></tr>
</table>

2. TITLE Interpolation of Head-Related Transfer Functions	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)
4. AUTHOR(S) Russell Martin and Ken McAnally	5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation 506 Lorimer St Fishermans Bend, Victoria 3207 Australia

6a. DSTO NUMBER DSTO-RR-0323	6b. AR NUMBER AR-013-842	6c. TYPE OF REPORT Research Report	7. DOCUMENT DATE February 2007

8. FILE NUMBER 2006/1119127	9. TASK NUMBER AIR 04/024	10. TASK SPONSOR HEWSD	11. NO. OF PAGES 20	12. NO. OF REFERENCES 27

13. URL on the World Wide Web http://www.dsto.defence.gov.au/corporate/reports/DSTO-RR-0323.pdf	14. RELEASE AUTHORITY Chief, Air Operations Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for public release*   OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT No Limitations

17. CITATION IN OTHER DOCUMENTS   Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.htm Acoustical signal processing; hearing; transfer functions; head related transfer functions; spatial interpolation; military operations and interpolation

19. ABSTRACT

Using current techniques, it is usually impractical to measure head-related transfer functions (HRTFs) at a spatial resolution that does not exceed the minimum audible angle, i.e. 1-2° for a source directly in front, by a considerable amount. As a result, measured HRTFs must be interpolated to generate a display in which auditory space is rendered smoothly. The spatial resolution at which it is necessary to measure HRTFs for the display to be of high spatial fidelity will depend on the quality of the interpolation technique. This report describes an interpolation technique that involves the application of a novel, inverse-distance-weighted averaging algorithm to HRTFs represented in the frequency domain. The quality of this technique was evaluated by comparing four listeners' abilities to localise virtual sound sources generated using measured or interpolated HRTFs. The measured HRTFs were shown to be of sufficiently high fidelity to allow virtual sources to be localised as accurately as real sources. Localisation error measures, i.e. lateral errors, polar errors and proportions of front/back confusions, for HRTFs interpolated across up to 30° of either lateral or polar angle, or 20° of both lateral and polar angle, did not differ noticeably from those for measured HRTFs. On the basis of this finding we recommend that HRTFs be measured at a 20° lateral- and polar-angle resolution.