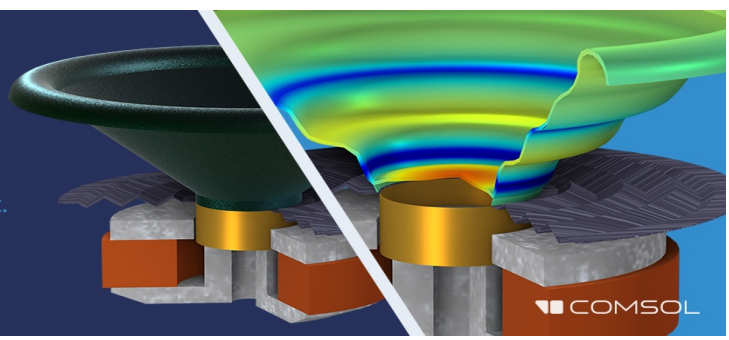
Matched transaural synthesis with probe microphones for psychoacoustical experiments

Aimee Shore, Anthony J. Tropiano and William M. Hartmann

Citation: The Journal of the Acoustical Society of America **145**, 1313 (2019); doi: 10.1121/1.5092203
View online: https://doi.org/10.1121/1.5092203
View Table of Contents: https://asa.scitation.org/toc/jas/145/3
Published by the Acoustical Society of America



Matched transaural synthesis with probe microphones for psychoacoustical experiments

Aimee Shore, Anthony J. Tropiano,[a] and William M. Hartmann[b]
Department of Physics and Astronomy, Michigan State University, 567 Wilson Road, East Lansing, Michigan 48824, USA

(Received 14 April 2018; revised 27 January 2019; accepted 7 February 2019; published online 11 March 2019)

Transaural synthesis using loudspeaker signals determined through contemporaneous ear canal calibration is proposed as an alternative to headphone presentation for critical psychoacoustical experiments. The proposed technique can afford greater accuracy, improved reproducibility, and continuous signal monitoring. It allows the experimenter to compare listener responses to real and virtual presentations. In this article, the advantages of transaural (three or four loudspeakers) compared to crosstalk cancellation (two loudspeakers) are shown through computer modeling and manikin measurements in a moderately reverberant room. Measurements employ binaurally challenging signals and speech from a distant source. Transaural synthesis is shown to be a better solution to the essential inverse problem resulting in reduced average synthesis amplitudes, fewer large-amplitude outliers, improved amplitude and phase accuracy for real and imagined sources, and improved noise immunity. Immunity to inadvertent listener head rotation depends sensitively on loudspeaker placement and is not an advantage in general. Appendixes review the relevant mathematical foundation and extend it to the relationship between ear canal signals and eardrum signals.
 https://doi.org/10.1121/1.5092203

[JB] Pages: 1313–1330

I. INTRODUCTION

Psychoacoustical experiments depend on precise and reproducible delivery of signals to the ear canals of human subjects. This requirement is especially important in studies of the binaural system where the experimenter must control interaural differences. The usual approach to precise stimulus delivery has been to use headphones with important features: The headphones themselves modestly attenuate unwanted background noise. For all practical purposes, left and right ear signals are independently controllable. Headphones can provide a constant signal in the ear canals even if the subject changes orientation or listening environment. Headphone delivery is inexpensive with minimal requirements on the experimental environment and no loudspeaker arrays. Similarly, anatomically local signal delivery is normally used in physiological experiments on other animals (e.g., Yin *et al.*, 1984). Nevertheless, techniques that deliver signals using loudspeakers can afford advantages as described in detail in this article. These sound synthesis techniques are based on methods established in the audio industry with the additional critical feature that signals are monitored in the ear canals. The ear canal measurements, in turn, determine the signals sent to the loudspeakers. As noted by Akeroyd *et al.* (2007), the calibration and the synthesis need to be *matched* for listener, environment, and instrumentation.

Loudspeaker delivery of experimental signals has characterized some recent psychoacoustical work with particular attention to accuracy and comparison with real-world sound sources (Akeroyd *et al.*, 2007; Moore *et al.*, 2010; Zhang and Hartmann, 2010; Majdak *et al.*, 2013; Hartmann *et al.*, 2016). The present article reviews loudspeaker techniques and points out advantages while observing technical problems. It then shows how transaural synthesis, employing more than two synthesis loudspeakers, can remediate some of the problems and lead to more accurate synthesis.

A. Crosstalk cancellation

In binaural (two-channel) headphone presentation, the signal intended for the left ear is sent to the left headphone, and the signal intended for the right ear is sent to the right headphone. For normal hearing listeners, there is essentially no crosstalk between the channels. However, when a binaural signal is presented by loudspeakers, some of the sound intended for the left ear arrives at the right ear and vice versa. These are cases of crosstalk.

Bauer (1961) proposed an audio method for replacing two-channel headphone listening by two loudspeakers. The crosstalk was reduced by adding filtered versions of the right and left channels to the left and right channels, respectively, in order to cancel the crosstalk at the ears. Schroeder and Atal (1963) and Schroeder (1975) implemented computational versions of crosstalk cancellation in terms of anatomical transfer functions. Here, synthesis through loudspeakers is described in terms of signals measured in the ear canals. This reformulation is the variation necessary to apply the method to individual listeners in psychoacoustical or physiological experiments.

The synthesis technique is described in the frequency domain where linear filtering is mathematically represented

[a]Present address: Department of Physics, The Ohio State University, 191 Woodruff Ave., Columbus, OH 43210, USA.
[b]Electronic mail: hartmann@pa.msu.edu

by multiplication. Let **x** be a column vector with two elements representing the signal in the left and right ear canals. Let y be another column vector with two elements representing the signal sent to left and right synthesis loudspeakers. Both **x** and y are complex functions of frequency with amplitude and phase information.

The signals in the ear canals are related to the loudspeaker signals by a transfer function matrix, **H**,

$$\mathbf{x} = \mathbf{H}y, \tag{1}$$

where the off-diagonal elements of **H** represent the crosstalk. This transfer function matrix includes the responses of the loudspeakers, the environment, the listener's anatomy, and the microphones. Equation (1) is representative of all the equations in this article. All quantities are functions of frequency—none are functions of time. Vectors are given by lower case symbols and matrices by upper case symbols. Further, quantities that occur physically are in boldface, while quantities that are computed or invented by the user are in plain italic text. The boldface convention extends also to "measured" quantities in computational experiments.

The crosstalk cancellation method begins with a two-element signal vector x' intended to be exactly the spectra in left and right ear canals for which headphone signals would otherwise be a first approximation. Vector x' can be realized in the ear canals using the loudspeaker system described by transfer functions **H** in Eq. (1) by inverting matrix **H**,

$$y' = H^{-1}x'. \tag{2}$$

Vector y' now describes the signals that must be sent to the left and right synthesis speakers in order to obtain x' in the left and right ear canals. When vector y' is then processed by transfer functions **H**, the result in the listener's ear canals is **x**,

$$\mathbf{x} = \mathbf{H}y' = \mathbf{H}H^{-1}x' = Ix', \tag{3}$$

where I is the identity matrix. Therefore, $\mathbf{x} = x'$, which satisfies the goal. Because **H** includes all the elements of the signal path, multiplying by its inverse guarantees that the signal in the ear canals will be exactly the desired signal.

Since its introduction in the early 1960s, the crosstalk cancellation method has been studied and applied a number of times. Damaske (1971) used dummy-head recordings as the binaural stimuli to be presented through an empirical crosstalk cancellation network to a human listener whose task was to localize the sound source in the horizontal or vertical planes. Morimoto and Ando (1980) applied the matrix inversion technique given in Eq. (2) to loudspeaker presentation of head-related transfer functions (HRTFs) measured in a different context. Hartmann *et al.* (2016) applied the technique to make controlled modifications of real signals.

B. Transaural synthesis

Transaural synthesis is a term used by Cooper and Bauck (1989) to generalize the concept of crosstalk cancellation to a context with multiple listening points in space and multiple loudspeakers. The concepts are mathematically straightforward if the number of listening points is equal to the number of speakers. Then matrix **H** is square, and its inverse is formally well defined, though it may be subject to the practical difficulties endemic to inverse problems. If the number of loudspeakers is less than the number of potential listening points (typical in multichannel presentation with unconstrained listening locations), the matrix is nonsquare and the inverse problem is overdetermined – there is no solution. If the number of loudspeakers is greater than the number of listening points, the matrix is again nonsquare, and the inverse problem is underdetermined—there is more than one solution. Indeed, there is an infinite number of solutions because there is an infinite number of different signals that could be sent to the loudspeakers while still obtaining the desired signals at the listening points. In a remarkable article, Bauck and Cooper (1996) showed how the Moore-Penrose pseudoinverse matrix can be used to obtain an effective inverse for the nonsquare matrix. Further, the pseudoinverse leads to an optimum solution in the sense that the signals sent to the loudspeakers have the least norm (smallest power; Moore, 1920; Penrose, 1955a,b).

C. Practical transaural synthesis (PTS)

Following the work of Bauck and Cooper (1996), other investigations critically analyzed crosstalk cancellation and suggested extensions. Kirkeby *et al.* (1998a,b) introduced the "stereo dipole" with synthesis loudspeakers close together to enlarge the area of controlled synthesis. Ward and Elko (1999) did calculations based on the geometry of loudspeakers and receiving points to show how loudspeaker placement could be optimized to maximize robustness of crosstalk cancellation filters. Takeuchi and Nelson (2002) proposed the optimal source distribution (OSD) with increasing angular separation between loudspeakers with decreasing frequency.

Kirkeby *et al.* (1998c) and Kirkeby and Nelson (1999) calculated crosstalk cancellation filters using matrix regularization, an approximation that limits the maximum gain allowed in the crosstalk filters, mitigating the practical difficulties endemic to inverse problems. Thereafter, regularization became a standard method to deal with unwanted large gains in crosstalk cancellation filters.

The subsequent decade saw much effort to optimize loudspeaker placement for maximal robustness to head displacements (Takeuchi *et al.*, 2001; Rose *et al.*, 2002; Nelson and Rose, 2005; Bai and Lee, 2006; Parodi and Rubak, 2010). These experiments and simulations were primarily concerned with maximizing the "sweet spot," the region of space over which the illusion of a virtual sound source holds. They were less concerned with precise signal delivery to a listener in a fixed position, possibly with head clamped. In these articles the researchers used regularization to obtain well-behaved crosstalk cancellation filters; thus, the solutions were only approximate.

Some researchers have incorporated more than two loudspeakers into their crosstalk cancellation networks. Takeuchi and Nelson (2001, 2002) and Akeroyd *et al.*

(2007) used two channels that were fed into a crossover network coupled to three loudspeaker pairs spaced at small, mid, and large angles for synthesizing high, mid, and low frequencies in their OSD system. This was essentially an extension of the two loudspeaker system. Bai *et al.* (2005) used six independent loudspeakers to deliver signals to a listener and incorporated multiple control points to gain greater control over the sound field. The goal was to widen the sweet spot. In all cases, researchers used regularization to limit maximum gains in the crosstalk cancellation filters, resulting in approximate solutions.

D. Experimenter's transaural synthesis (ETS)

Whereas PTS makes minimal demands on the listener's condition or environment, the subject of the present article might be called "ETS" because it is primarily useful as a means to deliver signals in controlled experiments, making unusual demands on the listener. It is concerned with enhanced accuracy and begins with the expectation that the intentional errors caused by regularization can be avoided.

In ETS, the listener has probe microphones in the ear canals at all times.[1] The signals x' intended to appear in the ear canals may be signals from real sources as recorded through the probe microphones. Alternatively, they might be intentionally altered versions of those real-source signals or something else invented by the experimenter (e.g., Hartmann and Wittenberg, 1996). With ETS, the experimenter always knows the signals in the listener's ear canals and the stimulus can alternate randomly between virtual signals and signals from real sources.

The control that is available with the ETS method is further described in Sec. VII. This control is not available with other methods of stimulus presentation. However, the ETS method can encounter computational problems as will be described in Sec. II.

E. Plan of the article

This article makes computational and experimental investigations of ETS. It begins with the crosstalk cancellation method with two ears and two synthesis loudspeakers (2×2 system). It then expands to the cases of three synthesis loudspeakers (2×3 system), as shown in Fig. 1, or four synthesis loudspeakers (2×4 system). We emphasize that this report involves almost no new mathematics beyond the article by Bauck and Cooper (1996), but the application is different, and the physical meaning of the variables is somewhat different.

Section II is entirely computational. It begins with a brief description of the pseudoinverse method, and then computes distributions of synthesis amplitudes for random ear canal signals and random transfer function matrices to test the expectation that the synthesis amplitudes should be markedly smaller for a larger number of synthesis speakers. Section III is an experimental test of the techniques from Sec. II, using two or three synthesis loudspeakers and comparing synthesis amplitude distributions with Sec. II. Section IV continues the experiments, testing how well synthesis can reproduce (1) a challenging, invented binaural signal and (2) signals from a real source. Section V computes the sensitivity of signals at the eardrums to small, random changes in the transfer functions, and shows that increasing the number of synthesis loudspeakers increases immunity to this kind of noise. Section VI considers the sensitivity of signals at the eardrums to small, systematic rotations of the listener's head. An analytic spherical head model is used to compute the changes in the transfer functions, and experiments are done with a manikin. Section VII is an interlude describing, in general terms, the benefits of loudspeaker stimulus presentation as guided by measurements in the ear canals. Finally, Sec. VIII reviews the conclusions of our modeling and experiments. Appendixes A and B expand the mathematics of the transaural synthesis for greater clarity.

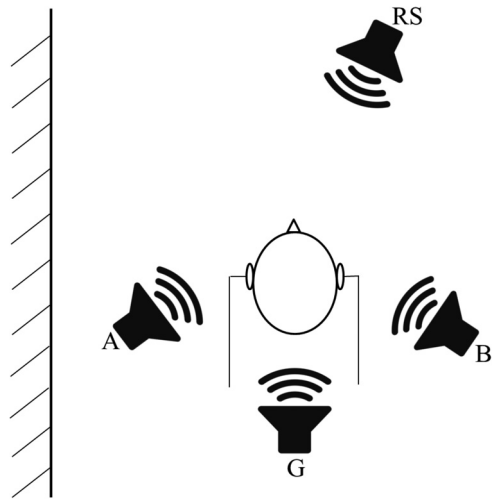


FIG. 1. Head with probe microphones in the ear canals. There is one real-source loudspeaker (RS) 28° to the right of the listener's forward direction, and three synthesis loudspeakers at ±120° and 180°. Synthesis loudspeakers are at about 1 m distance from the head, but the real source is at about 4 m distance. A reflecting wall is 1 m to the left of the listener. Not to scale.

II. SYNTHESIZED AMPLITUDES

Experimental transaural synthesis with probe microphones in the ear canals can be accomplished using two synthesis loudspeakers (2×2 system) and inverting matrix $\mathbf{H}$. As noted in Appendix A, inverting matrix $\mathbf{H}$ leads to a determinant in the denominator, and sometimes this determinant can accidentally be small, leading to ill-conditioned equations and unrealistically large amplitudes in the synthesis signals. Such events are said to be "pathological." Takeuchi and Nelson (2002) provided a simple geometrical illustration for the generation of pathological synthesis amplitudes.

Pathologies were observed by Zhang and Hartmann (2010), who created noise bands using a 2×2 system and 256 frequency components. The $\mathbf{H}$ matrix needed to be inverted for each frequency, and synthesis amplitudes sometimes became large—standing out from the rest of the noise as discrete tones—a clear cue to the listener that the synthesis was not real. Therefore, it was necessary to test each synthesis component amplitude, and when an amplitude became too large, the component was omitted from the signal. Figure 10 of Zhang and Hartmann (2010) shows the spectrum of a noise where four components needed to be

omitted. At that time, it became evident that had there been a third loudspeaker, there would have been enough degrees of freedom in the synthesis to avoid some of these pathological situations. The pseudoinverse technique tested in this article shows how this can be done.

A. The pseudoinverse

The pseudoinverse of a transfer function matrix **H** is denoted as H^+. When **H** has more columns than rows, the pseudoinverse is given in terms of **H** as

$$H^+ = \boldsymbol{H}^*(\mathbf{H}\boldsymbol{H}^*)^{-1}, \tag{4}$$

where $\boldsymbol{H}^*$ is the complex conjugate transpose (Hermitian transpose) of **H**. For ETS with two ear canals and N loudspeakers, **H** has two rows and N columns, whereas $\boldsymbol{H}^*$ has N rows and two columns. As a right-hand multiplier, matrix H^+ is a true inverse in the sense that $\mathbf{H}H^+$ is the identity matrix—a 2×2 identity matrix.[2] Therefore H^+ can be substituted for H^{-1} in Eq. (2), and y' will again be the signals sent to the N loudspeakers with the intention of producing signals x' in the ear canals. Because matrix H^+ is a true inverse, Eq. (3) will continue to hold, and the intended signals will be realized. Appendix A includes a user's guide to the pseudoinverse.

B. Maximum synthesis amplitudes

This section tests the minimum-norm property of the Moore-Penrose pseudoinverse matrix using randomly generated desired ear canal signals (x'). Each such signal is a sine function, which might be one of the Fourier components of an arbitrary broadband noise. We used randomly generated transfer functions (**H**) to simulate the filtering of signals on their path from synthesis loudspeakers to ear canals. Random transfer functions like this approximately simulate responses in a room environment with standing waves. Because of the computational nature of the tests, there was no physical role for frequency in either the model signal or the transfer functions. The computational tests included the usual 2×2 system as well as the 2×3 and 2×4 systems.

Before running the tests, we made some predictions. We predicted that the 2×2 system would produce many instances where the amplitude in one of the synthesis speakers (y') would be uncomfortably large. Such a large amplitude would be heard as an added, isolated tone. We predicted that changing to a 2×3 system with three synthesis loudspeakers would eliminate almost all of those instances. Additionally, we conjectured that the underdetermined system with three synthesis speakers would have enough freedom to completely optimize the solution. Therefore, we further predicted that changing to a 2×4 system would result in negligible additional benefit beyond the 2×3 system.

We developed computational algorithms to determine the synthesis amplitudes (of y') for one million trials for each of the three systems. Comparing the systems in this way immediately posed the problem that the procedure computed two million synthesis amplitudes for the 2×2 system, three million for the 2×3 system, and four million for the 2×4 system, leading to an unfair comparison. We adopted the sensible solution of retaining only the maximum amplitude across the two, three, or four synthesis signals in a given trial. That largest amplitude is the problem we are trying to solve, and this approach led to a fair comparison—one million amplitudes for each system.

The amplitudes and phases for the signals intended for the ear canals x' and all the transfer function matrix elements were randomized. The intended signal amplitudes were Rayleigh distributed with a standard deviation of 1.0, and the intended signal phases were uniformly distributed over 360°. The real and imaginary parts of the transfer function matrix elements were independently normally distributed with unit variance. Therefore, the mean square amplitude of transfer function matrix elements was 2.0. These properties of ear canal signals and matrix elements established the amplitude scale for the tests and ensured a fair comparison of synthesis amplitudes (of y') for the different systems.

Figure 2 shows the distributions of synthesis amplitudes for the three systems. Each histogram has 200 bins for maximum amplitudes between 0 and 20. The first bin gives the number of trials on which the maximum amplitude was

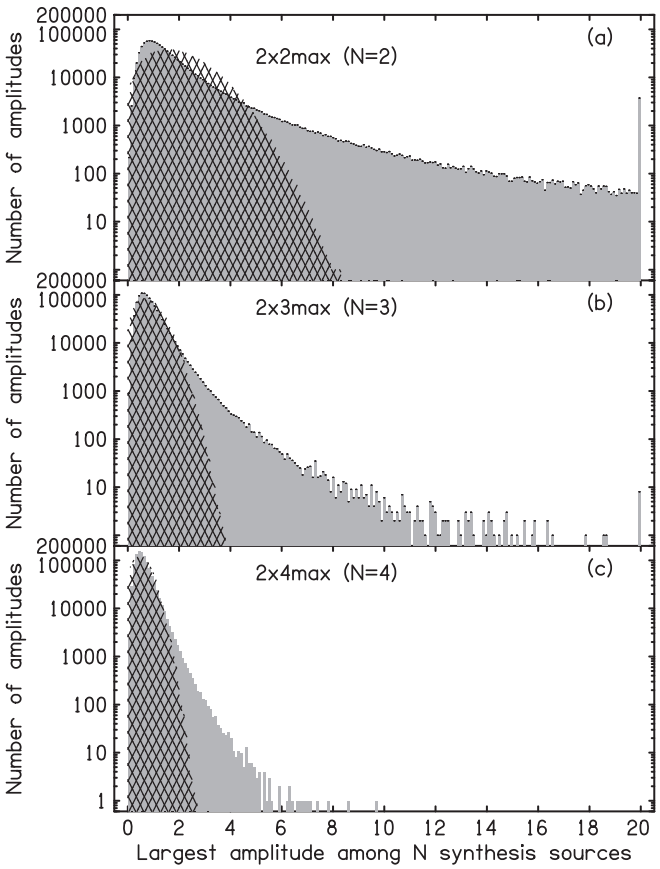


FIG. 2. Distributions of maximum amplitudes (among two or three or four) synthesis signals from the random-matrix models for three systems are shown in gray. The bin on the far right includes all the amplitudes greater than 20. The cross-hatched regions show the Rayleigh distributions with the same means. The mean amplitude for the 2×2 system is 2.0, which sets the scale for all the plots. An amplitude of 20 is 10 times the mean or 20 dB higher.

between 0 and 0.1, the second bin is for the maximum amplitude between 0.1 to 0.2, and so on. The rightmost bin of the histogram enumerates the number of trials where the maximum amplitude was greater than 20, i.e., out of range. For the 2×2 system there were 3741 out of range; for the 2×3 system there were 8 out of range, and for the 2×4 system there were none out of range.

The hatched regions in Fig. 2 show Rayleigh distributions having the same mean values as the (gray) histograms for synthesis (y') amplitudes. Plotted on a log vertical scale, Rayleigh distributions resemble downward parabolas. The Rayleigh distribution for the 2×2 system is exactly the distribution for the synthesis amplitudes in the special case that all the **H** matrices (and their inverses) are identity matrices. The mean amplitude is 2.0. Obviously the actual inverse solution produces a very different distribution with a much longer tail. The long tail arises from multiplication by the inverse random matrix, which is sometimes large. The Rayleigh distributions for the 2×3 and 2×4 systems are less easily interpreted. As they are shown in Fig. 2, they have the same means as the corresponding y' histograms, but those distributions have no connection to **H** matrices reduced to the identity matrices, however defined. The case can be made that a more appropriate comparison for the 2×3 and 2×4 systems would be the Rayleigh distribution for the 2×2 system, obtained by setting all the elements of the 2×3 and 2×4 matrices equal to zero apart from those in the original 2×2 portion.

Table I shows the percentiles for the cumulative distribution of maximum amplitudes. There is no 99.9 percentile point for the 2×2 system because 0.37% of the amplitudes were greater than 20, i.e., out of range, and 0.37 is greater than 0.1.

One of our initial predictions was clearly confirmed by the computational experiments. Application of the pseudoinverse successfully reduced the number of large amplitudes when there were three synthesis loudspeakers instead of two. All percentile values for the 2×3 system were considerably lower than the corresponding values for the 2×2 system. More important, the number of amplitudes greater than 10 decreased from 14 560 to 113.

Our second initial prediction held good in the sense that adding the fourth synthesis speaker led to less dramatic percentile changes than adding the third (Table I). However, adding the fourth synthesis speaker did reduce the extreme amplitudes and quite effectively. With the 2×3 system some amplitudes were greater than 20, but the largest amplitude ever seen with the 2×4 system was less than 10.

TABLE I. Percentiles for maximum amplitudes when the distributions of Fig. 2 are turned into cumulative distributions. For instance, the upper left entry shows that for the 2×2 system, 90%of the maximum amplitudes were less than 3.7. The mean amplitude for the 2×2 system was 2.0, which sets the scale for all three systems. Therefore, the amplitude of 3.7 is 5.3 dB above the mean.

Percentile	2×2	2×3	2×4
90.0	3.7	1.6	1.1
99.0	12.2	3.2	1.8
99.9	—	5.7	2.8

III. LOUDSPEAKER EXPERIMENTS

Experiments using an acoustical manikin were done to compare with the plotted synthesis amplitudes from the computational modeling in Sec. II.

A. Methods

Manikin experiments with two or three synthesis loudspeakers were conducted in the Michigan State P-Lab, a rectangular, variable-acoustics room with dimensions $4.3 \times 5.5 \times 3.0$ m. The ceiling is acoustical tile and the floor is vinyl tile. The walls are plaster but three of them were treated with absorption (Auralex Sunburst, Auralex, Indianapolis, IN)—a total of 13 m^2 of absorption at mid frequencies. The absorbing panels had been removed from the wall nearest the manikin (on the left side) to produce a more challenging room transfer function. This room arrangement was room setup 1. The reverberation time averaged 239 ms in the 250 and 500 Hz octave bands and averaged 144 ms in the four octave bands from 1000 to 8000 Hz.

The manikin was a KEMAR (G.R.A.S., Holte, Denmark) with large pinnae and wearing a cotton tee shirt. The manikin was mounted with its ears 117 cm from the floor, and its internal microphones simulated a listener's eardrums. For this experiment, the internal microphones were used instead of probe microphones.

The desired signal, expected to appear at the KEMAR's "eardrums," was a noise with 211 equal-amplitude, random-phase components regularly spaced from 200 Hz to 15 855 Hz. Signals were generated by TDT System-3 hardware (Alachua, FL) with RP2.1 digital-to-analog converters (DACs) controlled by a laptop computer. Triggering of three channels of the DACs was done via a Zbus trigger executed from RPVdS software (TDT Alachua, FL). The DAC outputs were sent to loudspeakers *A*, *B*, and *G*. Loudspeakers *A* and *B* were Mackie HR824mk2 Studio Monitors, and *G* was a Mackie HR824 Studio Monitor (LOUD Technologies, Woodinville, WA). Loudspeakers were mounted on portable stands and pointed directly at the head with their centers at the height of the ear canals.

Six different geometrical arrangements were used to expand the variety of transfer functions. First was the "120-degree-reference set," as shown in Fig. 1, where loudspeakers *A* and *B* were placed on opposite sides of the manikin, 1 m away, and at approximately $-120°$ and $120°$ from the manikin's forward direction, respectively. Loudspeaker *G* was located at $180°$ and was also 1 m away. Since loudspeaker *G* was not used in the 2×2 experiment, the effect of moving it was merely to change the position of a reflecting object. In a second arrangement, loudspeaker *G* was moved to $-140°$. These two arrangements were crossed with three positions of the manikin—the standard, a displacement of 0.1 m forward, and a displacement of 0.1 m backward. Then, the entire set of six arrangements was repeated except that loudspeakers *A* and *B* were at $\pm 90°$ to make the "90-degree-reference set." In the end there were 12 configurations.

The same set of 211 amplitudes was used for the desired signal at the eardrums and the measurements of the transfer functions—a convenience but not a necessity. However, for

each of the six arrangements of a reference set, a different set of random phases was used.

Transaural synthesis does not require loudspeaker gains or signal levels in the ears to be the same during the calibration phase, but to ensure a good signal-to-noise ratio for transfer function measurements, gains of the loudspeakers were adjusted to produce a level of ∼74 dBA at KEMAR's reference position, as measured by a sound level meter.

KEMAR internal microphone outputs were given an additional stage of amplification and digitized by the RP2.1 analog-to-digital converters. Recordings were manually downloaded from RP2.1 random access memory to the host computer by the click of a button in RPVdS.

B. Results

Similar to the computational modeling, the experiments were designed to find the largest of the spectral amplitudes among the two or three synthesis loudspeakers. With 12 different geometrical configurations and 211 frequencies, there were 2532 amplitude values for each system. In order to compare with the random-matrix computational modeling in Sec. II, the measured amplitudes for both systems were multiplied by a single scale factor so that the mean of the measured distribution for the 2×3 system was the same as the mean of the model distribution for the 2×3 system in Fig. 2(b). Therefore, the measured distributions of maximum amplitudes shown in Fig. 3 can be directly compared with the top two panels of Fig. 2 from the random-matrix modeling.

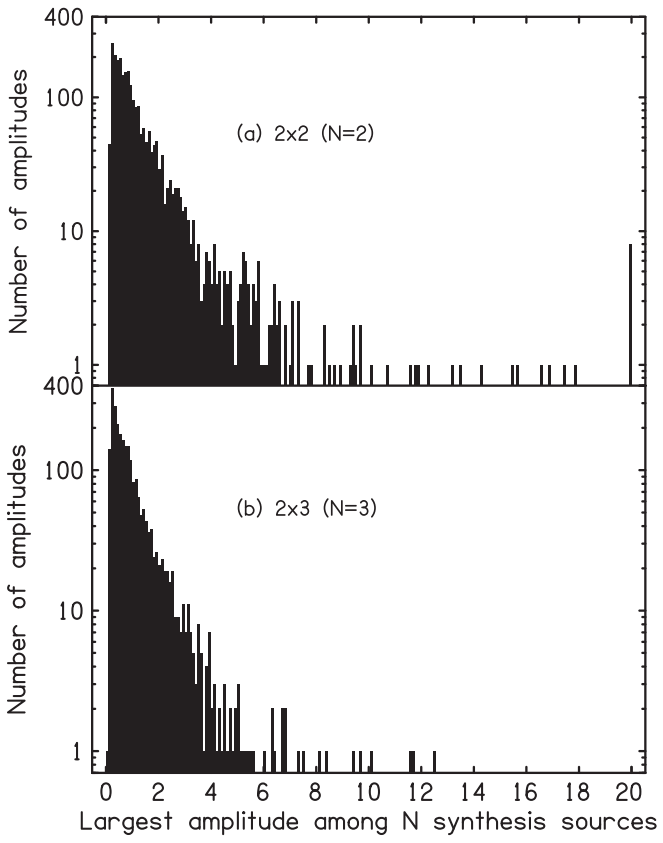


FIG. 3. Histogram of maximum synthesis spectral amplitudes (of 2 or 3)—experimental. Amplitudes were scaled so that the means of the 2×3 distributions in Figs. 2(b) and 3(b) coincide. That enables a fair comparison of the figures. Data were combined over 120° and 90° reference sets for a total of 2532 values.

TABLE II. Percentiles for maximum amplitudes when the distributions of Fig. 3 are turned into cumulative distributions. For instance, the upper left entry shows that for the 2×2 system, 90% of the maximum amplitudes were less than 2.8.

Percentile	2×2	2×3
90.0	2.8	2.0
99.0	9.6	5.1
99.9	—	10.2

Figure 3 for the 2×2 system indicates eight amplitudes off the plot. The largest amplitude occurred for the 90° reference set. That is an expected result. When a source is at 90° there is a bright spot at the ear on the opposite side of the head tending to enlarge the off-diagonal terms of the transfer function matrix (Macaulay *et al.*, 2010). When both sources are at 90° the effect is doubled. According to Appendix A, off-diagonal terms as large as diagonal terms lead to small denominators and large synthesis amplitudes. Adding the third loudspeaker completely eliminated the troublesome effect of the bright spot.

The mean amplitude for the 2×2 system was 1.48. It can be fairly compared with the value 2.02 for the random-matrix calculation. The difference indicates that the manikin measurements revealed physical constraints on the size of the crosstalk and consequent limitation on the size of synthesis amplitudes.

Table II shows percentiles, the experimental analog of Table I with random-matrix calculations. For the 2×2 system, the 90% and 99% points occur at smaller values of amplitude for the experiment than for random matrices. That is another indication that the random-matrix calculation was not realistically constrained.

Table II shows that adding the third synthesis speakers substantially reduced the experimental percentile amplitudes as would be expected from looking at Fig. 3. However, comparison with Table I shows that the experimental amplitudes were larger than the corresponding amplitudes for the random-matrix calculation. Apparently the experimental benefit of the third speaker, though substantial, was less than the theoretical benefit seen in Table I.

C. Discussion

Comparing Figs. 2 and 3 for corresponding systems (e.g., the 2×2 system) shows that the synthesis amplitudes appear to be similarly distributed for the computational model and the experiment, although there are many fewer points in the experiment. We interpret this observation to mean that the random-matrix model is a reasonable model for the stimulus noise components as modified by the inverse transfer functions in a room, though Table II shows that the experiment encountered less extreme cases than the model did.

Both the computational modeling and experiment show that the distributions of maximum amplitudes are

progressively skewed toward smaller values as the number of synthesis speakers increases. Such an effect might be expected because when there are three synthesis loudspeakers instead of two, the signal in each one can be smaller on the average and still achieve the same power at the ears. That argument leads to a reduction in amplitude by a factor of $\sqrt{3/2} = 1.2$. However, the average amplitude reduction was larger—a factor of about 1.5. The skew in the distribution is mainly caused by the advantage of the pseudoinverse. More important from an experimental standpoint is that there are fewer instances of very large amplitudes when the number of synthesis speakers is increased. We attribute that reduction to a reduced number of ill-conditioned inverse matrices. Nevertheless, there remain a few large synthesis amplitudes in the experimental plot. We attribute those to standing wave nulls in the room or anti-resonances in the ear canals. Improvements in the matrix algebra cannot solve those problems. They will be seen in more detail in Sec. IV.

IV. SYNTHESIS ACCURACY

We performed three tests of synthesis accuracy for both systems, 2×2 and 2×3. The first tested a dichotic, invented signal. The second tested a broadband noise from a real-source loudspeaker in the room and the third tested speech in a room.

A. Invented signals

The invented signals had frequency dependent amplitudes, increasing by 20 dB in the left ear and decreasing by 20 dB in the right ear with increasing frequency. Both amplitude dependences were straight-line functions of the frequency. There were 211 components from 200 to 15 855 Hz. The phases were random variables, independently random in each ear. This signal was intended to be challenging for the synthesis method. The *A* and *B* speakers were at $\pm 120°$ and 0.8 m from the head. The *G* speaker was at 180° and 1 m from the head.

The transfer function measurements again used the manikin internal microphones but a different method compared to that in Sec. III. In order to explore greater generality, the transfer functions were measured using a maximum length sequence (MLS) generated by a 17-stage shift register leading to $2^{17} - 1 = 131\,071$ values. At our sample rate (48 828.125 samples per second) the duration was about 2.7 s—adequate for synthesis of a brief sentence. Transfer function matrices were correspondingly large with frequency spacing of about 1/2.7 Hz, but most of the matrix elements were unimportant. Only the elements with frequencies of the 211 components were important.

After the inverse matrix was applied to the desired signal [Eq. (2)], the resulting signals sent to the loudspeakers (y') looked nothing like the desired signals (x') because the left and right desired spectra were so different. However, the spectra recorded by KEMAR internal microphones (**x**) were similar to x', as shown by Figs. 4 and 5.

The 211 measured amplitudes appear as open circles in Figs. 4 and 5. They are plotted on top of the desired (invented) amplitudes shown by filled circles. When a filled circle is not seen it is because the corresponding open circle obscures it. The phase differences shown in these figures were obtained by subtracting the desired phases from the measured phases. The differences were then reduced to the range $-180°$ to 180° by adding or subtracting multiples of 360°.

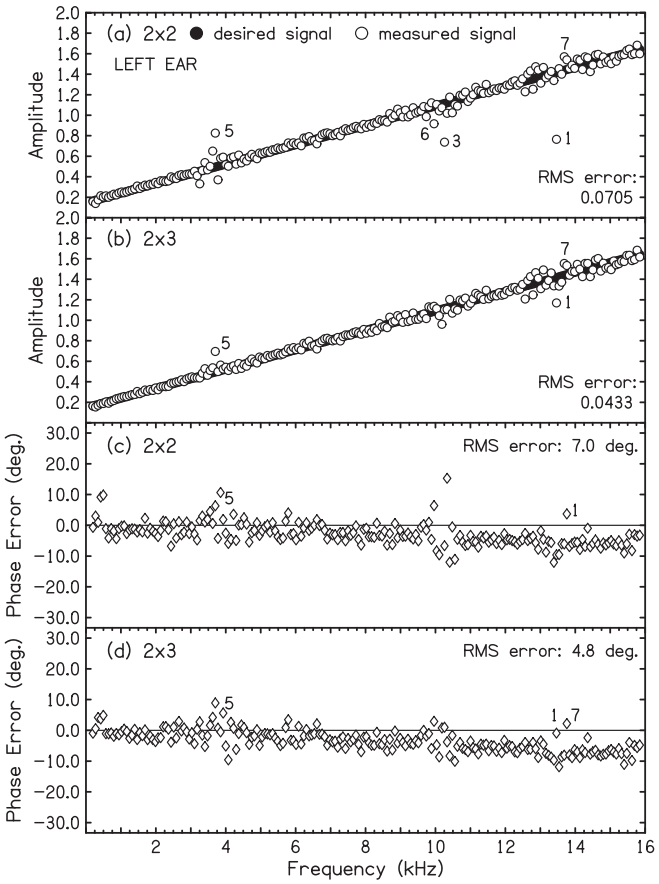


FIG. 4. Left ear: Desired amplitudes (x') are shown by filled circles. They are straight-line functions of frequency. Measured amplitudes (**x**) are shown by open circles. Small numbers, 1–7, track particular components of interest. Desired phases were random variables. Desired phases were subtracted from measured phases to find phase errors shown by diamonds.

Measured spectral amplitudes for both ears show anomalous values between 3 and 4 kHz and near 10 kHz. These were likely caused by the first and second ear canal resonances. Also, discrepancies tend to be larger when the amplitudes were smaller.

It is interesting to try to track the discrepant amplitudes through Figs. 4 and 5. For each of the four amplitude plots there, the largest ten discrepancies were found. Seven of them were given numbered labels. The largest discrepancy ever found was given the label "1," and it appears as one of the ten largest discrepancies in all the amplitude plots except for Fig. 5(a).

For a given synthesis, a component with a discrepant amplitude in one ear can be expected to be discrepant in the other because of the interaction on synthesis through the matrix equations. This effect occurred three times in Figs. 4 and 5 (points 1, 3, and 5).

For a given ear, a component with a discrepant amplitude for the 2×2 system might be expected to be discrepant for the 2×3 system if head/pinnae diffraction leads to a small amplitude on calibration. Figures 4 and 5 show four such instances (points 1, 2, 5, and 7).

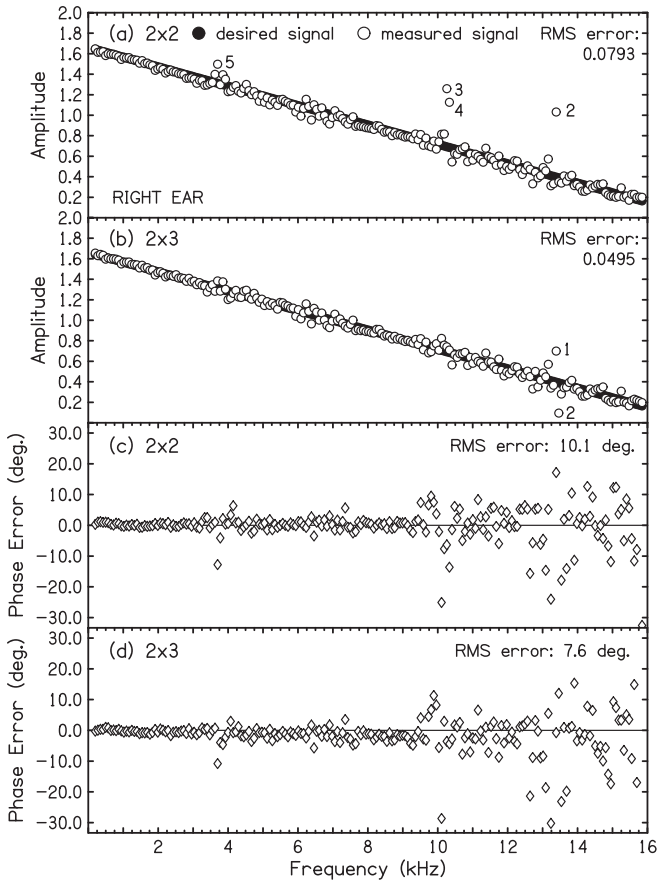


FIG. 5. Same as Fig. 4 but for the right ear. Larger phase errors at high frequencies arise from smaller amplitudes.

The figures make it clear that adding a third loudspeaker to make the 2×3 system led to decreased amplitude discrepancies. Especially important, the largest amplitudes, which become problems for a 2×2 synthesis, were significantly reduced with the 2×3 synthesis. These outsized amplitudes (signals $\mathbf{x}$) often corresponded to large amplitudes in the synthesis signals (y', not shown here) and may be attributed to pathological inverse transfer functions.

The phase plots in Figs. 4 and 5 agree with the amplitude plots in the sense that discrepancies occur at, or near, the same frequencies for both kinds of plots. Phase discrepancies tend to increase with increasing frequency as expected because phase is the product of delay and frequency. Phase errors in Fig. 4 have a decreasing linear component indicating a simple delay. Similar to our experience with the amplitudes, discrepancies for the phases were smaller and fewer for the 2×3 system.

B. Signal from a real source

The real-source experiments were practical tests of transaural synthesis. In practice, an experimenter may want to synthesize signals at a listener's eardrums based on signals from a remote source as measured in the ear canals. In these experiments, the synthesis was based on probe microphone measurements and tested by KEMAR internal microphone recordings. The experiments tested the idea that if a synthesis got it right in the probe microphones, then it would also get it right at the eardrums (internal microphones). The relevant mathematics appears in Appendix B.

1. Method

The real-source experiment used a variation on the synthesis system described in Sec. IV A. Loudspeakers A and B were at $\pm 120°$ and at a distance of 1 m from the center of the head. Loudspeaker G was at 180° and at 1 m (Fig. 1). The real-source loudspeaker was 28° to the right of the forward direction at 3.8 m to enhance relative room effects. The room was arranged in room setup 2, in which the acoustical foam was removed from all walls. In addition, porcelain tile panels (2.7 m^2) were placed along the wall behind the synthesis loudspeakers. Setup 2 provided a longer reverberation time and a more challenging test environment for synthesis. The reverberation time averaged 463 ms in the six octave bands 250–8000 Hz. The probe microphones were Etymotic ER-7 s (Etymotic, Elk Grove Village, IL) inserted with their tips close to the eardrums of the KEMAR ears.

2. Target and standard

In the real-source experiments, the *target* was the measurement at the *probe* microphones of the signal from the real-source loudspeaker. The first signal was again a 211-component noise. The target was used to create the synthesized signals. The *standard* was the measurement at the *internal* KEMAR microphones of that same signal. The standard was used to evaluate the quality of the ultimate synthesis.

A straightforward approach to the target and standard would be to turn on the real source and make the recordings at the two sets of microphones. However, because the microphones (especially the probe microphones) and the environment were somewhat noisy, we chose a different approach. We used the MLS to measure the impulse response between the real source and the probe microphones, and determined the target by convolving the original signal with the impulse response. We used the MLS to measure the impulse response between the real source and the internal microphones, and determined the standard by convolving the original signal with the impulse response.[3] Measurements showed that this latter method, with the impulse response averaged over eight repetitions of the sequence improved the signal-to-noise ratio by 33 dB over direct recording—an enormous advantage. Comparison between desired and measured signals used only a 211-component subset of frequency components so that results could be conveniently displayed.

3. Results

The results of the experiment are shown in Fig. 6 for the right ear as measured in the KEMAR internal microphone. Figure 6 compares measured and standard amplitudes for the 2×2 system and the 2×3 system (top two panels). Differences between measured and standard phases are shown in the bottom two panels. Root-mean-square (RMS) amplitude and phase errors data are listed in Table III. We came to the following conclusions:

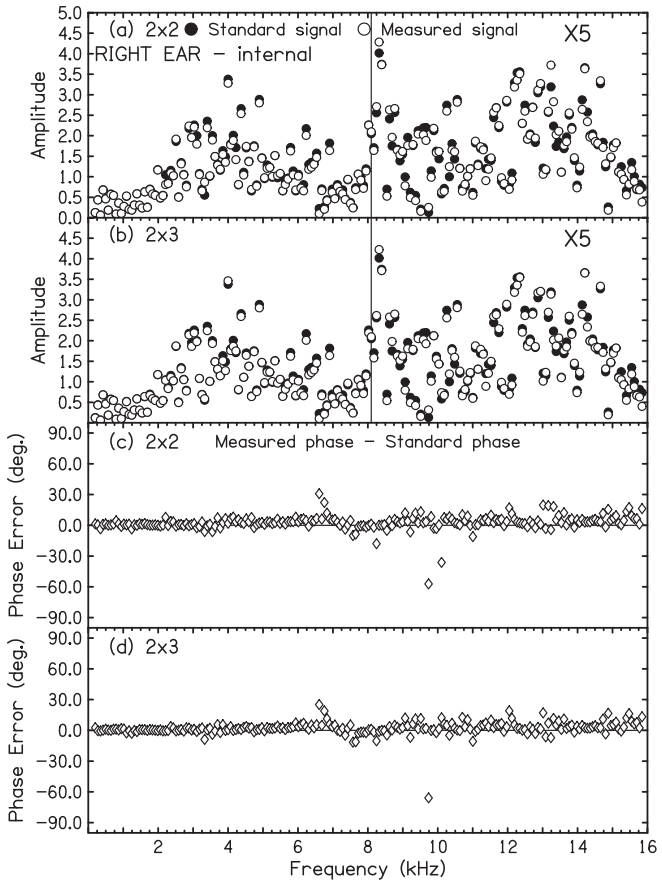


FIG. 6. Amplitudes and phase errors measured by an internal microphone in the right ear for the 2×2 and 2×3 systems. The real source was a 211-component white noise. Top two panels: The standard amplitudes are shown by filled circles. They are the same for 2×2 and 2×3 systems. The measured amplitudes are shown by open circles. Amplitudes above 8097 Hz are multiplied by five for better viewing. Bottom two panels: Differences (in degrees) between measured and standard phases.

- Synthesis was somewhat more successful for the right ear than for the left. The difference is particularly noticeable for the phases.
- Amplitudes and phases in the probe microphones agreed better with the desired values compared to the internal microphones. This might have been expected because the synthesis was based on the probe microphones.
- For the left ear (not shown in Fig. 6) adding the third loudspeaker (G) to the synthesis hardly mattered for the amplitudes, either for the probe microphone or the internal microphone. Phase errors were modestly reduced. In contrast, for the right ear, adding the third loudspeaker reduced amplitude errors considerably. The difference between ears may be attributable to a worse signal-to-noise ratio in the left ear because it was farther from the source. This implies there is a signal-to-noise threshold below which a third loudspeaker may confer only minimal benefit.

C. Application to speech signal

The experiment described in Sec. IV B was repeated, but the target and standard were female speech instead of white noise. The goal was to demonstrate the utility of ETS in perceptual experiments. The utterance was the brief sentence, "Cats hate dogs." Its duration was 2.68 s, which corresponds to a frequency spacing of 0.37 Hz. Again, there were 131 071 frequency components. All calculations were done in the frequency domain, which means that the order of the three words was determined by the phases in the Fourier transform.

TABLE III. RMS errors for synthesis of the 211-component noise from the real source. RMS amplitude errors are in dB re the RMS amplitudes of the target (probes) or the standard (internal). Phase errors are in degrees.

	Left ear		Right ear	
	Probes	Internal	Probes	Internal
2×2 (dB)	−27.4	−23.8	−24.4	−23.8
2×3 (dB)	−27.0	−24.0	−30.0	−26.1
2×2 (°)	7.19	16.24	4.13	7.88
2×3 (°)	5.32	11.48	3.51	7.22

Results of synthesis in the right ear are shown in Fig. 7 for the internal microphones. Comparison between measured and standard signals again used only the 211-component subset of frequency components so that results could be

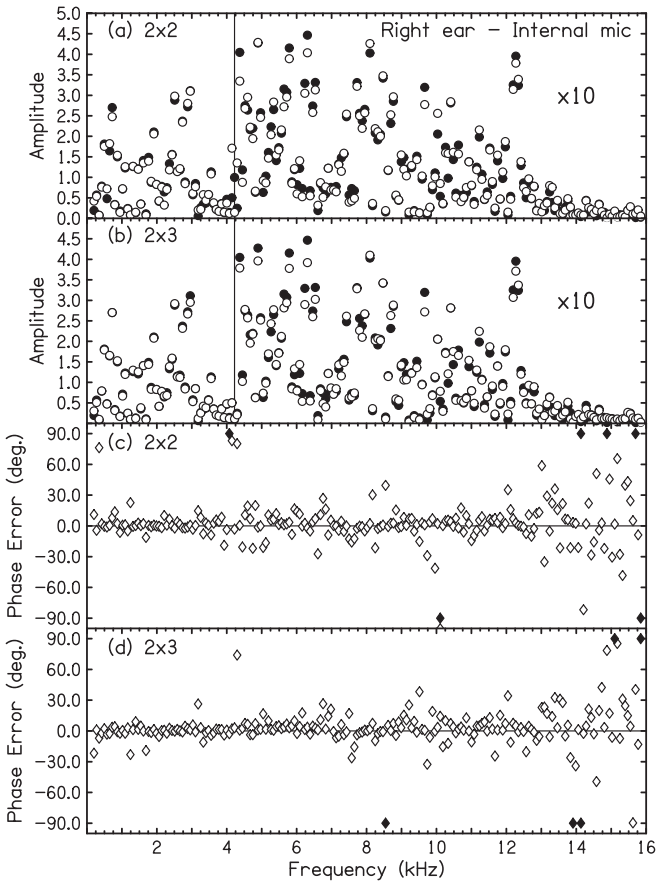


FIG. 7. Same as Fig. 6 but the real source was female speech ("Cats hate dogs."). Comparison between standard amplitudes (filled circles) and measured amplitudes (open circles) show only a 211-component subset of frequency components for a convenient display. The amplitude scale for frequencies above 4 kHz is expanded by a factor of 10. Phase errors (diamonds) for the same set of frequencies are the difference: measured-standard. Phase errors outside the ±90° range are shown by solid diamonds at ±90°.

TABLE IV. RMS error values for synthesis of "Cats hate dogs." RMS amplitude errors are in dB re the RMS amplitudes of the target (probes) or the standard (internal). Phase errors are in degrees. Errors were calculated for the 10 202 frequency components between 200 and 4000 Hz—the range of the speech energy.

	Left ear		Right ear	
	Probes	Internal	Probes	Internal
2×2 (dB)	−19.8	−23.4	−20.6	−24.5
2×3 (dB)	−21.2	−27.1	−22.3	−30.0
2×2 (°)	29.9	17.7	29.9	17.0
2×3 (°)	29.1	14.0	28.6	11.8

conveniently displayed. Further, the amplitude scale for frequencies above 4 kHz was expanded to facilitate visual comparison. Phase errors increased considerably above 4 kHz because the measured signals were so small at those frequencies. RMS errors in Table IV were calculated using only frequency components between 200 and 4000 Hz, because almost all the speech energy lies in that range.

The RMS average results for the speech experiment are shown in Table IV. It is evident that adding the third loudspeaker improved synthesis accuracy. Improvement was most dramatic in the internal microphones. Amplitude errors were as much as 5.5 dB smaller in the 2×3 system compared to the 2×2 system. Phase errors were reduced by 31% (right) and 27% (left). Compared to the white noise experiment (cf. Sec. IV B) for which improvement was only observed in the right ear, synthesis accuracy was ameliorated in both ears for the speech source. However, the frequency ranges were different for these two tables. Enhanced low- and middle-frequency spectral content in the speech target is a possible explanation.

Table IV shows that for both ears, both systems, and both amplitude and phase, the RMS errors were smaller for the internal microphones than for the probe microphones. This result is opposite to the corresponding result for the noise source in Table III, and it is initially surprising. How can the internal microphone results be better than the probe microphone results when the stimuli for the internal microphone recordings were made from the probe microphone signals? The answer lies in the final measurement process. The probe microphones, with very thin probe tubes, were much noisier than the internal microphones. Although the effective noise from the probe microphones could be reduced by our repeated MLS technique in producing the target and standard signals, the final measurements were simple recordings of the synthesized signals. The probe microphone measurements were thus contaminated by noise. The frequency range used for speech signal measurements was different from that for the noise source, and the speech signal had intervals of smaller signal level.

V. NOISE IMMUNITY

Transaural synthesis is vulnerable to changes in the transfer functions that occur after the calibration of the synthesis speakers but before, or during, the presentation of a test sound for an experiment. Such changes may be caused by small, inadvertent motions of the listener's head, and they appear as noise in the synthesis.

We conjectured that synthesis with an increasing number of synthesis loudspeakers would reduce the effects of this kind of noise. The idea was that the pseudoinverse technique leads to a minimum in the multidimensional space of matrix elements. Therefore, the sensitivity to changes, as represented by the slopes in this space near the minimum, should be smaller with a larger number of synthesis speakers. This section describes computational experiments to test that conjecture. Distributions for the sensitivity of both the signal amplitude (measured in dB) and signal phase (measured in degrees) were computed.

A. Method

The procedure began with desired signals for the ear canals (x', randomly chosen) and initial transfer functions from the synthesis speakers (**H**, also randomly chosen). The program computed the synthesis signals necessary (y') to produce the desired signals in the ear canals. Then, random changes were made in the synthesis transfer functions (**H**; simulating listener motion or other noise) while the synthesis signals (y') remained *unchanged*. The computational experiment monitored the changes in the signals in the ear canals (**x**) as the procedure was done 500 000 times. In the first experiment, the percentage changes in transfer functions were uniformly distributed from −5% to +5%. In

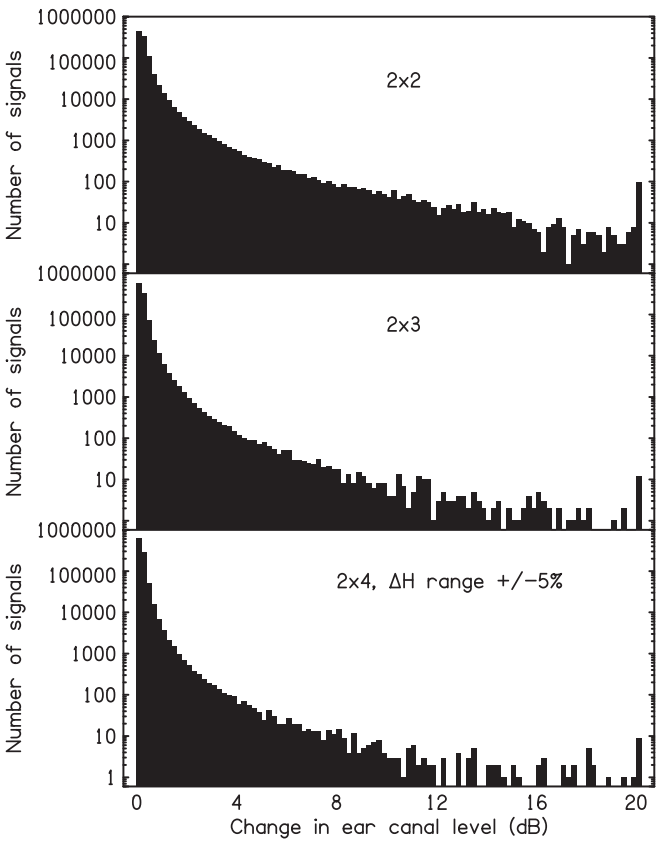


FIG. 8. Distributions of the magnitudes of changes in component level (dB) caused by a 5% randomization of transfer functions for three different model systems. There are 100 bins with widths of 0.2 dB. The last bin on the right includes all the signals where the change was more than 20 dB.

the second experiment the range was −20% to +20%. We predicted that the changes in the signals in the ear canals (**x** vs x') would be smaller when there were more synthesis speakers.

B. Results

With two ears × 500 000 trials, the experiments led to distributions of one million ear canal amplitudes (levels in dB) and phases (degrees). Figures 8 and 9 show the distributions for the 5% experiment.

The amplitude change plot (Fig. 8) refers to both positive and negative level changes, e.g., a change of 2 dB and a change of −2 dB were both added to the 2-dB histogram bin. The plot gives visual evidence that the conjecture holds good—amplitude levels in the ear canals are less sensitive to noise in the transfer functions when there are more synthesis speakers. The cumulative distribution, calculated from Fig. 8, leads to a similar conclusion: For the 2 × 2 system there were 10 405 signals (out of one million) for which the level change was greater than 3 dB. The corresponding numbers for the 2 × 3 and 2 × 4 systems were, respectively, 2361 and 1339. For the 2 × 2 system there were 965 signals for which the level change was greater than 10 dB. The corresponding numbers for the 2 × 3 and 2 × 4 systems were, respectively, 154 and 84.

The phase change plot (Fig. 9) refers to both positive and negative phase change, e.g., a change of 15° and a change of −15° were both added to the 15-deg histogram bin. The plot indicates that decreased sensitivity to transfer function noise with increasing number of synthesis speakers occurred for phases as well as for amplitudes. The cumulative distribution, calculated from Fig. 9, leads to a similar conclusion: For the 2 × 2 system there were 5725 signals (out of one million) for which the phase change was greater than 30°. The corresponding numbers for the 2 × 3 and 2 × 4 systems were, respectively, 1172 and 670. For the 2 × 2 system there were 938 signals for which the phase change was greater than 90°. The corresponding numbers for the 2 × 3 and 2 × 4 systems were, respectively, 169 and 82.

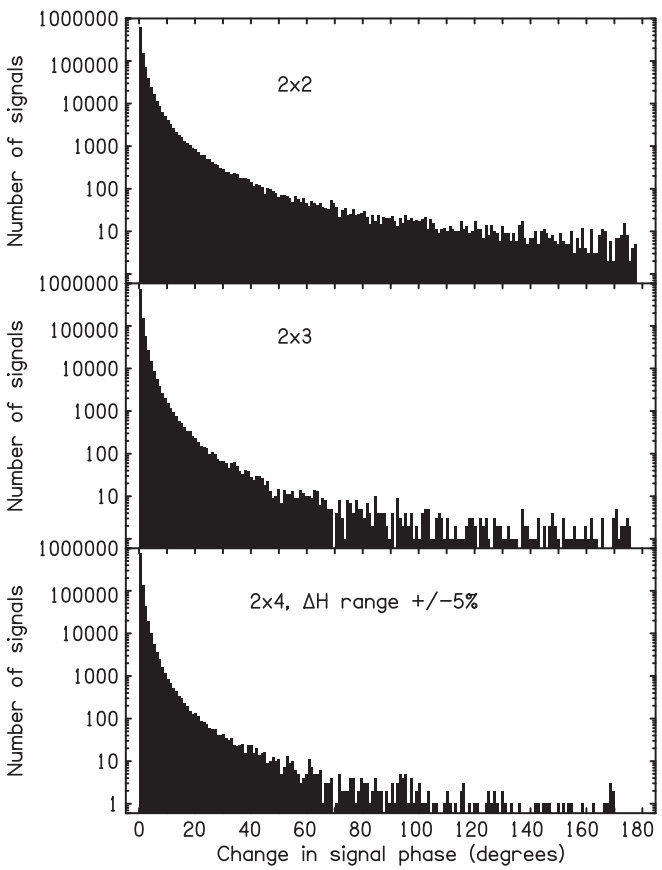


FIG. 9. Distributions of the magnitudes of changes in component phases (degrees) caused by a 5% randomization of transfer functions for three different model systems.

The calculations with a ±20% randomization in matrix elements also indicated that increased numbers of synthesis speakers led to increased noise immunity, but the effect was smaller. The cumulative numbers above show that in the ±5% experiment there was a 87% reduction in the number of amplitude changes greater than 3 dB when the 2 × 2 system was replaced by the 2 × 4 system. In the ±20% experiment the corresponding reduction was 75%. This result is consistent with the notion that reduced sensitivity to changes in the vicinity of a minimum in a multidimensional space is better seen when the changes are small (range ±5%) than when they are larger (range ±20%).

VI. SENSITIVITY TO LISTENER ROTATIONS

Whereas, Sec. V studied the sensitivity of synthesis systems to random variations in the transfer functions (noise), this section studies systematic variations caused by a small rotation of the head.

A. Computational experiments

The computational experiments began with a model head and two, three, or four synthesis speakers together with a model real source in a free-field environment. The head was a sphere with antipodal, point-like ears. Signals and transfer functions at the ears were calculated using the Legendre polynomial expansion for finite source distances (Rabinowitz *et al.*, 1993; Duda and Martens, 1998; Brungart and Rabinowitz, 1999). The synthesis sources were given different azimuths with respect to the forward direction of the model head, and they were always 1 m away from the center of the head. Computations proceeded as follows: (1) Transfer functions (**H**) were computed from the synthesis sources to each ear ("synthesis transfer functions") and the pseudoinverse matrix (H^+) was computed. (2) The signals at the two ears caused by a model source ("target signals") were computed. (3) The target signals were taken as the desired signals x', and the pseudoinverse matrix [from step (1)] was used to compute synthesis signals y'. (4) Synthesis signals were processed by the synthesis transfer functions (from step 1) to verify correct synthesis at the ears ($\mathbf{x} = x'$). (5) The model head was rotated by 5° and the *same* synthesis signals (y' from step 3) were processed by the rotated synthesis transfer functions to find altered signals at the ears (**x**). (6) The discrepancies (**x** vs x') in amplitudes and phases of spectral components at the ears, caused by the rotation, were calculated and averaged. Amplitude

discrepancies were converted to decibel magnitudes before averaging. Phase discrepancies were computed as RMS values in degrees.

Averages were computed over 91 different azimuths for the model real source from 0° to 90° with respect to the forward direction. The averaging was also over 24 spectral components in an octave band (quarter-tone spacing). The computation was repeated for six octave bands with bottom frequencies of 125, 250, 500, 1000, 2000, and 4000 Hz.

The broad-brush results of the computational head-rotation experiment are not hard to describe because of the major role of symmetry for the spherical head:

- Symmetry applied in the special case when the synthesis speaker azimuths were ±90°. In this case, amplitudes and phases at the ears were at minima or maxima and the derivatives near these points were small leading to especially small sensitivity. With two synthesis speakers at ±90°, adding a third or fourth synthesis speaker elsewhere *increased* sensitivity to rotation. An increased sensitivity by a factor of 10 for the low-frequency band and by a factor of 2 for the high-frequency band was observed.
- With two synthesis speakers at ±120°, there was less sensitivity to rotation than with two synthesis speakers at ±135°. There was no advantage to adding a third speaker, but adding third and fourth speakers at ±90° reduced sensitivity. The simple explanation for all these results is that symmetrical synthesis azimuths closer to ±90° have an immunity advantage.
- With two synthesis speakers at ±175°, there was less sensitivity than with two synthesis speakers at ±120°. There was no advantage to adding third and fourth speakers at ±90°.
- There were occasional exceptions in one or two frequency bands for most of the statements above. Most dramatic, with two synthesis speakers at ±175°, there was less sensitivity than with two synthesis speakers at ±90°, but only in the highest frequency band.

The two conclusions that can be drawn from the computational rotation experiments are that there is an immunity advantage to putting two synthesis speakers as close as practical to ±90° with respect to the forward direction. This was the synthesis configuration used by Moore *et al.* (2010) in a two-loudspeaker synthesis system, though Koring and Schmitz (1993) recommend 45°. Our experiments find that with two sources at ±90° there is little additional rotation insensitivity upon adding a third or fourth synthesis speaker.

The advantage that we observed for widely spaced synthesis speakers is contrary to the concept of the stereo dipole, which attempts to minimize the discrepancies caused by listener motion by putting the speakers close together. Takeuchi *et al.* (2001) studied the change in interaural differences as a function of head displacement in six degrees of freedom. Their synthesis sources were at ±30° or ±5°. Their calculations generally revealed an important advantage for the smaller angular separation. Their localization experiments with seven successful listeners showed a modest advantage for the smaller separation. It remained for us to test rotation with the manikin—a less symmetrical receiver and more like human listeners.

B. Manikin rotation experiments

The rotation experiments used the manikin and the setup-1 environment described in Sec. III. The side synthesis loudspeakers, *A* and *B*, were initially placed at ±120° at a distance of 1 m and then moved to ±90°. Loudspeaker *G* was at −140°, 180°, or 140°, also at a distance of 1 m. Again, the invented desired signal was constant amplitude, random phases noise, and this signal was also used to measure the transfer functions.

Transfer functions were measured and synthesis waveforms were computed, played, and recorded in internal microphones with the head facing the forward direction (0° reference condition). Then the KEMAR's head was rotated 5° to the left and the (unchanged) synthesis was replayed and recorded again. It was expected that the 2 × 3 system would be less sensitive to the rotation than the 2 × 2 system because of the pseudoinverse minimum property. A second prediction was that the ±90° reference set would be more robust to rotation because of flattening of the transfer function at the point of approximate symmetry.

The changes caused by rotation for one of the reference sets are shown by the amplitudes and phases in Fig. 10. Reference 0° amplitudes are plotted as filled circles and rotated amplitudes as open circles. An overall deterioration

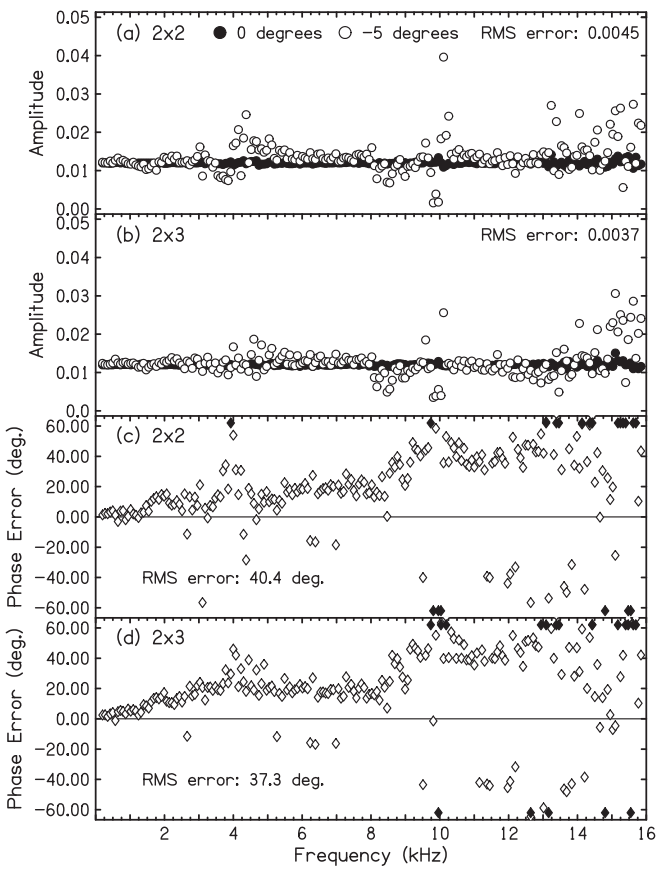


FIG. 10. Comparison of amplitudes measured at the left eardrum for 211 components before manikin rotation (filled symbols) and after rotation by 5° (open symbols) for the (a) 2 × 2 system and (b) 2 × 3 system. Corresponding phase differences (after–before) appear in plots (c) and (d). Differences outside the ±60° range are plotted by diamonds at ±62°. Synthesis loudspeakers *A* and *B* were at ±120° and *G* was at 180°.

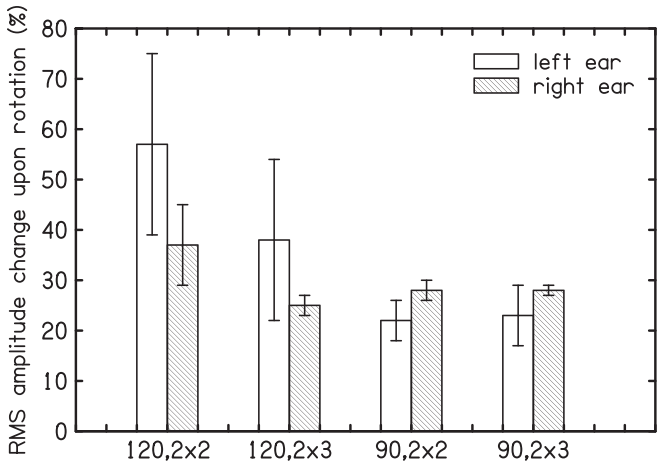


FIG. 11. RMS change in amplitude caused by an uncompensated rotation of 5°, averaged over 211 frequencies and over the 3 azimuths of synthesis speaker *G*. Speakers *A* and *B* were either at ±90° or ±120°. The error bars are two standard deviations in overall length. The data for these histograms came from data sets of which Fig. 10 is an example.

in synthesis occurred for both ears (right ear not shown) and angular reference sets, most notably at high frequencies (≥9.5 kHz). Problematical amplitudes occurred in the usual places—namely, the first and second ear canal resonances (3.5 kHz and 10 kHz). The 2 × 3 system was less sensitive to rotation, at least for the first ear canal resonance, as seen by reduction of large amplitudes in Fig. 10(b).

Figure 11 is a histogram plot of the RMS changes in amplitude caused by an uncompensated 5-deg rotation. These changes are experimental indicators of sensitivity. There are three kinds of comparison: (1) *A* and *B* speakers at ±120° or ±90°, (2) 2 × 2 system or 2 × 3 system, (3) left ear or right ear. The most dramatic effect was the large reduction in sensitivity when the *A* and *B* speakers were at ±90°—an anticipated effect. When the *A* and *B* speakers were at ±90°, adding the third synthesis speaker led to no further reduction in sensitivity. When the *A* and *B* speakers were at ±120°, adding the third speaker led to a considerable reduction in sensitivity for both left and right ears. This effect was larger than could have been anticipated from the computational experiment of Sec. VI A, and it is not known whether it would occur for other azimuths of the *A* and *B* speakers. Although Fig. 11 shows that the left ear was more sensitive than the right when the *A* and *B* speakers were at ±120°, there is little reason to expect the left ear to be more sensitive than the right. The left ear was closer to the nearby wall, and the rotation was to the left. Otherwise, the experiment was left-right symmetrical.

VII. ETS

This section is an interlude prior to the summary and conclusions. It describes our opinions and experience with the potential benefits of ETS.

The ETS technique can be contrasted with the headphone techniques that aim to simulate HRTF (Wightman and Kistler, 1989a,b; Martin *et al.*, 2001). The HRTF-headphone techniques measure head-related impulse responses (HRIR), or they measure the HRTFs and then calculate the HRIRs. In a second phase, the experiments present desired signals—tones, noise, speech, or music—by convolving with the known HRIRs and then presenting the filtered signals by headphones. Inverse filtering to remove the transfer function of the headphones may or may not be applied. The headphone techniques have the enormous appeal that once the HRIRs are measured, any signal can be presented to the listener, and in any environment. The HRTFs are typically measured only once.

In the ETS technique, the HRTFs (or HRIRs) are not saved; they are remeasured at the start of every experiment run, and they may be updated during the course of a run. The ETS technique has *matched*, *contemporaneous* calibration and synthesis. There is no need to remove headphone artifacts because there are no headphones. Typically, ETS generates signals in the frequency domain using signals with discrete frequency components. Consequently, there is no need to compute HRIRs and no need for interpolation. Modern computers running fast Fourier transform routines are fast enough to generate signals lasting several seconds using the frequency domain. Although the probe microphone requirement adds complexity for both the experimenter and the listener, ETS has a number of distinct advantages:

- *Reproducibility*: In experiments using ETS at medium frequencies (e.g., 500 Hz), Hartmann *et al.* (2016) showed that interaural time differences and interaural level differences had small variance—smaller than reported previously for headphones. This result was surprising because headphone presentation is the standard method to gain stimulus control in binaural experiments. However, signals from headphones can vary significantly with different headphone placement (Domnitz, 1975; Pralong and Carlile, 1996; Kulkarni and Colburn, 2000). It is not yet clear whether ETS retains its advantage at high frequencies in the absence of a head restraint such as a bite bar, nor has ETS been critically compared with ear insert phones.
- *Transducer linear distortion immunity*: HRTFs are affected by linear distortions (frequency dependent gain and phase shifts) in the signal processing chain beginning with the measurement microphone. The ETS technique is automatically immune to such linear distortion because the same electronic chain is used during calibration and test phases of an experiment. For example, it is not necessary that the synthesis loudspeakers have a flat frequency response. Irregularities in the loudspeaker response are unimportant for the synthesis of accurate signals because the irregularities are automatically compensated by the calibration procedure determining **H** and hence H^{-1}.
- *Probe insensitivity*: A similar immunity to linear distortions applies to probe microphones because of the method's automatic compensation. Zhang and Hartmann (2010) studied the effect of relocating the probe microphones in the ear canals of a manikin. This manikin had its own internal microphones in Zwislocki couplers, effectively representing the signals at the eardrums, and these

served as a standard. It was found that the signals in the probe microphones changed considerably when the probe microphones were moved in steps of 1 mm. However, the signals at the eardrums, based on the probe microphone measurements, changed very little. The signals at the eardrums were insensitive to motion of the probe microphones because the technique requires that the same probe locations are used for measuring the target and calibrating the synthesis loudspeakers. The insensitivity persisted out to 16 kHz. Self-correction of this kind does not occur with HRTF-headphone techniques.

The assumption implicit in the HRTF-headphone techniques is that the HRTF can be measured from a point in space to a point close to the eardrum (Wightman and Kistler, 1989a,b) or with blocked meatus (Hammershøi and Møller, 1996), and then applied to a signal delivered by headphones to simulate a source at the chosen point in space. The assumption implicit in the ETS loudspeaker technique is only that if two presentation conditions lead to identical signals at the probe microphones in the ear canal, then those two conditions will also lead to identical signals at the eardrums (though different from those in the probe microphones). This assumption is reviewed mathematically in Appendix B. Experiments by Zhang and Hartmann (2010) showed that this assumption was usually valid with occasional failures at special frequencies caused by ear canal resonances/anti-resonances.

- *Reality check*: In a typical ETS experiment, the listener hears real sources, simulated real sources (baseline stimuli), and simulated altered sources. Because every signal is measured in the ear canals, continuous online monitoring is available throughout the experiment.
- *Pre-test*: Given a challenging stimulus task, Zhang and Hartmann (2010) employed a pretest. Every block of runs began with an alternating sequence of sources *real*, *virtual*, *real*, *virtual*. The listener was aware of that sequence and could repeat it as often as desired to try to learn the difference between real and virtual presentation. Then the listener took a real-virtual classification test of 20 forced-choice trials. The experimental block continued on to the main experiment only if the listener failed the real-virtual test. With this pretest, and securing the listener's head with a bite bar, it was possible to solve the front-back confusion problem in the difficult mid-sagittal plane.

VIII. SUMMARY

This article has considered ETS where the listener always has probe microphones in the ear canals. It has primarily explored the consequences of expanding the crosstalk cancellation system that uses two synthesis loudspeakers to systems that use three or four. The first concern was with the amplitudes of the signals sent to the synthesis loudspeakers. It is the nature of inverse problems, such as transaural synthesis, to produce anomalously large amplitudes, and these can wreck an otherwise perceptually persuasive virtual reality.

Both computer modeling using random matrices (Sec. II) and experiments using a manikin in a room (Sec. III) showed that increasing the number of synthesis loudspeakers resulted in smaller amplitudes for the synthesis signals on the average. More important, additional loudspeakers reduced the number of especially large synthesis amplitudes. Experimentally, the 2×3 system eliminated all but one of the 17 largest amplitudes seen with the 2×2 system.

A second set of manikin experiments compared the 2×2 and 2×3 systems for accuracy with three tests (Sec. IV). The first test employed an invented, broadband, and challenging dichotic signal presented in a room environment. Adding the third synthesis speaker improved the synthesis overall, but pathologies were identified at several frequencies that were not improved.

The second test required the transaural systems to reproduce broadband noise from a distant real source elsewhere in the room. Table III showed that for one of the two ears the amplitude error seemed to be at a floor level, and adding a third synthesis speaker did not reduce it. For the other ear, adding the third speaker reduced the RMS amplitude error in the internal microphone by 22%.

The third test required the transaural systems to reproduce a brief spoken sentence from the distant real source. Table IV showed a benefit for amplitudes and phases upon adding a third synthesis speaker. This experiment led to the surprising result that errors were smaller for the internal recordings than for the probe recordings used to generate the internal recordings. The surprising result was attributed to the probe-microphone noise in the verification stage.

A second set of computational experiments (Sec. V) studied the immunity of different systems to noise—inadvertent, random variations in the HRTFs. The minimum-norm feature of the pseudoinverse method suggested that systems with more loudspeakers would be less sensitive to such noise. Calculations with transfer function variations in the range of -5% to $+5\%$ showed that changing from a 2×2 system to 2×3 or 2×4 systems reduced the number of amplitude changes greater than 3 dB by 77% or 87%, respectively, thereby supporting the suggestion.

Finally, both computational experiments and manikin experiments explored the sensitivity to systematic head rotations (Sec. VI). These experiments showed that the advantages of an additional loudspeaker in the synthesis system were modest and somewhat haphazard. Although the manikin experiments showed some advantage for the 2×3 system, the dominant effect was the location of the synthesis loudspeakers, and the number of synthesis loudspeakers was of secondary importance.

IX. CONCLUSION

Although most of this article has been devoted to comparing different synthesis systems, a more important message may be in Sec. VII, which promotes transaural synthesis as an alternative to headphone presentation in critical psychoacoustical experiments. In our experience, and in the experience of others (Mills, 1972; McAnally and Martin, 2002), localization experiments with realistic interaural parameters, and mainly in the

azimuthal plane, find similar listener performance whether the stimuli are presented by loudspeakers in a free field or by headphones. Localization is not a sensitive attribute (Brinkmann *et al.*, 2017). When the percepts under study become more subtle—externalization, assessment of room effect, conflicting interaural cues, distance perception, elevation determination, front-back discrimination, effects of listener or source motion, and virtual reality in general—it becomes increasingly important for the listener to have realistic signals at the eardrums and for the experimenter to know what these signals are.

The requirements and opportunities for PTS and ETS are often different. PTS envisions that the listener may move and therefore is concerned with maximizing the sweet spot. Such concerns led to the *stereo dipole* configurations where the loudspeakers are close together (Kirkeby *et al.*, 1998a,b). PTS may need to be speedy, perhaps even to keep up with real time. By contrast, ETS begins with the advantage that the listener's position is fixed (possibly even rigid), and the computation time requirements on the experiment may be less demanding.

Having attempted to make the case for ETS, we must also acknowledge that it is not easy to do. Although the use of multiple loudspeakers can reduce some of the worst problems of transaural synthesis, there remain pathologies caused by anomalous head diffraction, ear canal resonances, standing waves in rooms, or listener motion. Listener motion is a particular problem in a room where standing waves lead to sharp peaks and valleys in the room transfer function. This sharp structure causes the synthesis to be sensitive to small head displacements. Transaural synthesis in a room normally requires inverting a transfer function that is not minimum phase, which can lead to artifacts as observed by Neely and Allen (1979). These problems are alleviated if the experiment is done in free field, but the problem cannot be avoided if the experiment itself is about listening in rooms. As is evident in the figures of this article, such problems tend to become more serious as the frequency increases. Dealing with these problems through regularization or the selective elimination of malignant spectral components may become necessary. Alternatively, a modification to the pseudoinverse solution, as suggested for an acoustically transparent head by Yang *et al.* (2003), may enhance the robustness of the inverse solution. Although the selection of an optimum inverse solution may depend on individual experimental circumstances, the improved control that an experimenter has with probe-microphone based stimulus delivery can be well worth the effort.

ACKNOWLEDGMENTS

This work was supported by the AFOSR Grant No. 11NL002. Dr. Eric Macaulay and Professor Brad Rakerd assisted with the measurements.

APPENDIX A: THE PSEUDOINVERSE

This appendix expands the formal introduction to the pseudoinverse from Sec. II with the intent of making its operation more transparent. It continues the convention that bold symbols represent physically occurring quantities.

Signals in the ear canals are called $\mathbf{x}$, where

$$\mathbf{x} = \mathbf{H}y, \tag{A1}$$

and y represents signals sent to the synthesis speakers, and $\mathbf{H}$ is the matrix of transfer functions. Then, given desired signals x' in the ear canals, the required synthesis signals can be computed as

$$y' = H^{+}x', \tag{A2}$$

where H^{+} is the inverse or the pseudoinverse of $\mathbf{H}$, and all signals and matrix elements are in the frequency domain.

1. 2 × 2 system

For the 2 × 2 system, Eq. (A1) describes signals x in the left (L) and right (R) ears in terms of the synthesis signals in speakers A and B,

$$\begin{bmatrix} \mathbf{x_L} \\ \mathbf{x_R} \end{bmatrix} = \begin{bmatrix} \mathbf{H_{AL}} & \mathbf{H_{BL}} \\ \mathbf{H_{AR}} & \mathbf{H_{BR}} \end{bmatrix} \times \begin{bmatrix} y_A \\ y_B \end{bmatrix}. \tag{A3}$$

Because matrix $\mathbf{H}$ is square, the inverse is straightforward and Eq. (A2) becomes

$$\begin{bmatrix} y'_A \\ y'_B \end{bmatrix} = \begin{bmatrix} \mathbf{H_{BR}} & -\mathbf{H_{BL}} \\ -\mathbf{H_{AR}} & \mathbf{H_{AL}} \end{bmatrix} \times \begin{bmatrix} x'_L \\ x'_R \end{bmatrix} / (\mathbf{H_{AL}H_{BR}} - \mathbf{H_{BL}H_{AR}}). \tag{A4}$$

The determinant in the denominator is the origin of the occasional pathologies in crosstalk cancellation using only two synthesis speakers.

2. 2 × 3 system

For the 2 × 3 system, there are three synthesis speakers A, B, and G, so that Eq. (A1) becomes

$$\begin{bmatrix} \mathbf{x_L} \\ \mathbf{x_R} \end{bmatrix} = \begin{bmatrix} \mathbf{H_{AL}} & \mathbf{H_{BL}} & \mathbf{H_{GL}} \\ \mathbf{H_{AR}} & \mathbf{H_{BR}} & \mathbf{H_{GR}} \end{bmatrix} \times \begin{bmatrix} y_A \\ y_B \\ y_G \end{bmatrix}. \tag{A5}$$

Then the solution for y' requires the pseudoinverse, H^{+},

$$\begin{bmatrix} y'_A \\ y'_B \\ y'_G \end{bmatrix} = \begin{bmatrix} H^{+}_{LA} & H^{+}_{RA} \\ H^{+}_{LB} & H^{+}_{RB} \\ H^{+}_{LG} & H^{+}_{RG} \end{bmatrix} \times \begin{bmatrix} x'_L \\ x'_R \end{bmatrix}, \tag{A6}$$

where $H^{+} = \boldsymbol{H}^{*}(\boldsymbol{H}\boldsymbol{H}^{*})^{-1}$.

To calculate H^{+} we begin by calculating the 2 × 2 matrix $(\boldsymbol{H}\boldsymbol{H}^{*})^{-1}$.

$$(\boldsymbol{H}\boldsymbol{H}^{*})^{-1} = \left\{ \begin{bmatrix} \mathbf{H_{AL}} & \mathbf{H_{BL}} & \mathbf{H_{GL}} \\ \mathbf{H_{AR}} & \mathbf{H_{BR}} & \mathbf{H_{GR}} \end{bmatrix} \begin{bmatrix} \mathbf{H^{*}_{AL}} & \mathbf{H^{*}_{AR}} \\ \mathbf{H^{*}_{BL}} & \mathbf{H^{*}_{BR}} \\ \mathbf{H^{*}_{GL}} & \mathbf{H^{*}_{GR}} \end{bmatrix} \right\}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}, \tag{A7}$$

where the symbol "*" applied to a matrix element means *complex conjugate*, and a, b, c, and d stand for the matrix products inside the curly brackets. For instance,

$$a = |\mathbf{H_{AL}}|^2 + |\mathbf{H_{BL}}|^2 + |\mathbf{H_{GL}}|^2. \tag{A8}$$

Finally, the pseudoinverse matrix is the 3×2 matrix given by

$$H^+ = \begin{bmatrix} \mathbf{H^*_{AL}} & \mathbf{H^*_{AR}} \\ \mathbf{H^*_{BL}} & \mathbf{H^*_{BR}} \\ \mathbf{H^*_{GL}} & \mathbf{H^*_{GR}} \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \Big/ (ad - bc). \tag{A9}$$

3. 2 × 4 system

The expanded mathematics for the 2×4 system is a straightforward extension of the 2×3 system. The actual mechanics for calculating the pseudoinverse never requires the inversion of a matrix greater than a 2×2 because with ETS there are only two receiving locations—the two ear canals.

APPENDIX B: SYNTHESIS

This appendix describes how a signal sent from a real-source loudspeaker and received at the eardrums can be simulated by synthesis loudspeakers using transfer functions measured with probe microphones in the ear canals. The description takes the form of a test wherein the signals at the eardrums can be a known standard because they are measured using an anatomical manikin with internal microphones for eardrums. Recordings made with those internal microphones are given the subscript k. Recordings made with the probe microphones have subscript p. As before, quantities that occur physically are indicated with bold symbols.

A target stimulus called s_0 is played through a real-source loudspeaker and recordings are made using the manikin's internal microphones to represent eardrum recordings $\mathbf{x_{k0}}$

$$\mathbf{x_{k0}} = \mathbf{H_{k0}} s_0, \tag{B1}$$

where $\mathbf{H_{k0}}$ is the transfer function matrix of s_0 to the eardrums. The subscript *zero* is used to indicate that the signal originated from the real source. Signals originating from the synthesis loudspeakers do not have this subscript. Recorded signals $\mathbf{x_{k0}}$ are the standard for evaluating the subsequent transaural synthesis. In addition, with signal s_0 played through the real source, recordings $\mathbf{x_{p0}}$ are made using probe microphones in the ear canals

$$\mathbf{x_{p0}} = \mathbf{H_{p0}} s_0. \tag{B2}$$

The next step is to determine transfer functions $\mathbf{H_p}$ to the two ear canals for each of the synthesis speakers using signal y from each speaker in turn while recording $\mathbf{x_p}$ in the probe microphones. Signal y is a long MLS. The cross-correlation of $\mathbf{x_p}$ and y is calculated to obtain $\mathbf{H_p}$. This matrix has dimensions $2 \times N$ (two rows and N columns), where N is the number of synthesis loudspeakers. It is used to compute the pseudoinverse matrix, H_p^+, required for synthesis.

The synthesis proceeds by arranging for signal $\mathbf{x_{p0}}$ from Eq. (B2) to appear at the probe microphones during synthesis, i.e., the desired signal at the probe microphones x'_{p0} is set equal to the recorded signal $\mathbf{x_{p0}}$. In order to achieve that $\mathbf{x_{p0}}$ is filtered by the pseudoinverse to obtain y'

$$y' = H_p^+ \mathbf{x_{p0}}, \tag{B3}$$

where y' are the N signals to be sent to the synthesis loudspeakers. Recordings of the synthesis as made in the probe microphones are

$$\mathbf{x_p} = \mathbf{H_p} y' = \mathbf{H_p} H_p^+ \mathbf{x_{p0}}, \tag{B4}$$

and $\mathbf{x_p}$ ought to equal $\mathbf{x_{p0}}$ because $\mathbf{H_p} H_p^+$ equals the identity matrix.

To test the system, the same signals, y', are played through the synthesis loudspeakers and recordings, $\mathbf{x_k}$, are made using internal microphones

$$\mathbf{x_k} = \mathbf{H_k} y', \tag{B5}$$

where $\mathbf{H_k}$ is the transfer function that occurs between the synthesis loudspeakers and the eardrums. Neither $\mathbf{H_k}$ nor $\mathbf{x_k}$ plays any role in the synthesis, but $\mathbf{x_k}$ is the final result for comparison with $\mathbf{x_{k0}}$. Equation (B3) can be substituted for y', resulting in

$$\mathbf{x_k} = \mathbf{H_k} H_p^+ \mathbf{x_{p0}}. \tag{B6}$$

A further substitution from Eq. (B2) can be made for $\mathbf{x_{p0}}$, yielding

$$\mathbf{x_k} = \mathbf{H_k} H_p^+ \mathbf{H_{p0}} s_0. \tag{B7}$$

If it could be shown that $\mathbf{H_k} H_p^+ \mathbf{H_{p0}} = \mathbf{H_{k0}}$, then, according to Eq. (B1), the signals at the eardrums from the synthesis would be the same as the signals at the eardrums from the original real source, namely, $\mathbf{x_k} = \mathbf{x_{k0}}$.

We begin by writing an expression to relate the probe-microphone transfer function to the internal-microphone transfer function. Both transfer functions originate at the synthesis loudspeakers

$$\mathbf{H_p} = Q_p \mathbf{H_k}, \tag{B8}$$

where Q_p is necessarily a 2×2 matrix whatever the number of synthesis speakers. Further, Q_p is diagonal because the relationship between probe microphone and manikin internal microphone that occurs in one ear is unaffected by the relationship in the other ear. An analogous expression can be written that relates the probe-microphone transfer function to the internal-microphone transfer function when both transfer functions originate at the real source.

$$\mathbf{H_{p0}} = Q_{p0} \mathbf{H_{k0}}. \tag{B9}$$

Because the inverse of Eq. (B8) is

$$H_p^+ = H_k^+ Q_p^+, \tag{B10}$$

it follows that Eq. (B7) can be rewritten using Eqs. (B9) and (B10):

$$\mathbf{x_k} = \mathbf{H_k} H_k^+ Q_p^+ Q_{p0} \mathbf{H_{k0}} s_0 \tag{B11}$$

or

$$\mathbf{x_k} = Q_p^+ Q_{p0} \mathbf{H_{k0}} s_0 \tag{B12}$$

because $\mathbf{H_k} H_k^+ = I$. It is a common and reasonable assumption that the relationship between signals as measured at two different points within an ear canal depends only on the signal spectrum and is independent of the direction from which the original signal originates (Mehrgardt and Mellert, 1977; Hammershøi and Møller, 1996; Middlebrooks *et al.*, 1989). Therefore, $Q_p = Q_{p0}$ and $Q_p^+ Q_{p0} = I$. Then

$$\mathbf{x_k} = \mathbf{H_{k0}} s_0, \tag{B13}$$

or, by substituting Eq. (B1) for $\mathbf{H_{k0}} s_0$,

$$\mathbf{x_k} = \mathbf{x_{k0}}. \tag{B14}$$

In the end, the signals at the eardrums resulting from the synthesis are found to be equal to the signals at the eardrums from the original real source. Although this appendix refers to manikin recordings as a standard for comparison, information from those recordings played no role in the synthesis process. Therefore, the results of this appendix apply to human listeners.

[1]The probe microphones should not interfere with normal acoustics of the ear canal, and they should be safe for the listener. There is an advantage at high frequency if the probe tips are close to the eardrums. There are no special requirements for the frequency response of the microphones.

[2]A feature of the pseudoinverse for an N-loudspeaker system is that although $\mathbf{H}H^+$ is a 2×2 identity matrix, $H^+\mathbf{H}$ is an $N \times N$ matrix, and it is not an identity matrix ($N > 2$). The best that can be said about $H^+\mathbf{H}$ is that it is Hermitian.

[3]The response of the room was considered to be part of the transfer function from the source to the listener's eardrums. Consequently, impulse responses were about 3 s in duration. Longer may be needed for more reverberant spaces.

Akeroyd, M. A., Chambers, J., Bullock, D., Palmer, A. R., Summerfield, A. Q., Nelson, P. A., and Gatehouse, S. (**2007**). "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," J. Acoust. Soc. Am. **121**, 1056–1069.

Bai, M. R., and Lee, C. C. (**2006**). "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," J. Acoust. Soc. Am. **120**, 1976–1989.

Bai, M. R., Tung, C. W., and Lee, C. C. (**2005**). "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," J. Acoust. Soc. Am. **117**, 2802–2813.

Bauck, J., and Cooper, D. H. (**1996**). "Generalized transaural stereo and applications," J. Audio Eng. Soc. **44**, 683–705.

Bauer, B. B. (**1961**). "Stereophonic earphones and binaural loudspeakers," J. Audio Eng. Soc. **9**, 148–151.

Brinkmann, F., Lindau, A., and Weinzierl, S. (**2017**). "On the authenticity of individual dynamic binaural synthesis," J. Acoust. Soc. Am. **142**, 1784–1795.

Brungart, D. S., and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. HRTFs," J. Acoust. Soc. Am. **106**, 1465–1479.

Cooper, D. H., and Bauck, J. L. (**1989**). "Prospects for transaural recording," J. Audio. Eng. Soc. **37**, 3–19.

Damaske, P. (**1971**). "Head-related two-channel stereophony with loudspeaker reproduction," J. Acoust. Soc. Am. **50**, 1109–1115.

Domnitz, R. H. (**1975**). "Headphone monitoring system for binaural experiments below 1 kHz," J. Acoust. Soc. Am. **58**, 510–511.

Duda, R. O., and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

Hartmann, W. M., Rakerd, B., Crawford, Z. D., and Zhang, P. X. (**2016**). "Transaural experiments and a revised duplex theory for the localization of low-frequency tones," J. Acoust. Soc. Am. **139**, 968–985.

Hartmann, W. M., and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678–3688.

Kirkeby, O., and Nelson, P. A. (**1999**). "Digital filter design for inversion problems in sound reproduction," J. Audio Eng. Soc. **47**(7/8), 583–595.

Kirkeby, O., Nelson, P. A., and Hamada, H. (**1998a**). "Local sound field reproduction using two closely spaced loudspeakers," J. Acoust. Soc. Am. **104**, 1973–1981.

Kirkeby, O., Nelson, P. A., and Hamada, H. (**1998b**). "The 'stereo dipole'—A virtual source imaging system using two closely spaced loudspeakers," J. Audio Eng. Soc. **46**(5), 387–395.

Kirkeby, O., Nelson, P. A., Hamada, H., and Orduna-Bustamante, F. (**1998c**). "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**(2), 189–195.

Koring, J., and Schmitz, A. (**1993**). "Simplifying cancellation of crosstalk for playback of head-related recordings in a two-speaker system," Acustica **79**, 221–232.

Kulkarni, A., and Colburn, H. S. (**2000**). "Variability in the characterization of the headphone transfer-function," J. Acoust. Soc. Am. **107**, 1071–1074.

Macaulay, E. J., Hartmann, W. M., and Rakerd, B. (**2010**). "The acoustical bright spot and mislocalization of tones by human listeners," J. Acoust. Soc. Am. **127**, 1440–1449.

Majdak, P., Masiero, B., and Fels, J. (**2013**). "Sound localization in individualized and non-individualized crosstalk cancellation systems," J. Acoust. Soc. Am. **133**, 2055–2068.

Martin, R. L., McAnally, K. I., and Senova, M. A. (**2001**). "Free-field equivalent localization of virtual audio," J. Audio Eng. Soc. **49**(1/2), 14–22.

McAnally, K. I., and Martin, R. L. (**2002**). "Variability in the headphone-to- ear-canal transfer function," J. Audio Eng. Soc. **50**(4), 263–266.

Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.

Mills, A. W. (**1972**). "Auditory localization," in *Foundations of Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 301–345.

Moore, A. H., Tew, A. I., and Nicol, R. (**2010**). "An initial validation of individualized crosstalk cancellation filters for binaural perceptual experiments," J. Audio Eng. Soc. **58**(1/2), 36–45.

Moore, E. H. (**1920**). "On the reciprocal of the general algebraic matrices," Bull. Amer. Math. Soc. **26**, 394–395.

Morimoto, M., and Ando, Y. (**1980**). "On the simulation of sound localization," J. Acoust. Soc. Jpn. (E) **1**(3), 167–174.

Neely, S. T., and Allen, J. B. (**1979**). "Invertibility of a room impulse response," J. Acoust. Soc. Am. **66**, 165–169.

Nelson, P. A., and Rose, J. F. W. (**2005**). "Errors in two-point sound reproduction," J. Acoust. Soc. Am. **118**, 193–204.

Parodi, Y. L., and Rubak, P. (**2010**). "Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers," J. Acoust. Soc. Am. **128**, 1045–1055.

Penrose, R. (**1955a**). "A generalized inverse for matrices," Proc. Cambridge Philos. Soc. **51**, 406–413.

Penrose, R. (**1955b**). "On the approximate solution of linear matrix equations," Proc. Philos. Soc. **52**, 17–19.

Pralong, D., and Carlile, S. (**1996**). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," J. Acoust. Soc. Am. **100**, 3785–3793.

Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (**1993**). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," Presence **2**, 125–129.
Rose, J., Nelson, P. A., Rafaely, B., and Takeuchi, T. (**2002**). "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," J. Acoust. Soc. Am. **112**, 1992–2002.
Schroeder, M. (**1975**). "Models of hearing," Proc. IEEE **63**, 1332–1354.
Schroeder, M. R., and Atal, B. S. (**1963**). "Computer simulation of sound transmission in rooms," IEEE Intl. Conv. Rec. **11**, 150–155.
Takeuchi, T., and Nelson, P. A. (**2001**). "Optimal source distribution for binaural synthesis over loudspeakers," Acoust. Res. Let. Online **2**, 7–12.
Takeuchi, T., and Nelson, P. A. (**2002**). "Optimal source distribution for binaural synthesis over loudspeakers," J. Acoust. Soc. Am. **112**, 2786–2797.
Takeuchi, T., Nelson, P. A., and Hamada, H. (**2001**). "Robustness to head misalignment of sound imaging systems," J. Acoust. Soc. Am. **109**, 958–970.
Ward, D. B., and Elko, G. W. (**1999**). "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," IEEE Signal Process. Lett. **6**(5), 106–108.
Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.
Yang, J., Gan, W. S., and Tan, S. E. (**2003**). "Improved sound separation using three loudspeakers," Acoust. Res. Lett. Online **4**(2), 47–52.
Yin, T. C. T., Kuwada, S., and Sujaku, Y. (**1984**). "Interaural time sensitivity of high-frequency neurons in the inferior colliculus," J. Acoust. Soc. Am. **76**, 1401–1410.
Zhang, P. X., and Hartmann, W. M. (**2010**). "On the ability of human listeners to distinguish between front and back," Hear. Res. **260**, 30–46.