Experiments on a System for the Synthesis of Virtual Acoustic Sources*

P. A. NELSON, F. ORDUÑA-BUSTAMANTE, AND D. ENGLER

Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, UK

AND

H. HAMADA

Department of Information and Communication Engineering, Tokyo Denki University, Tokyo 101, Japan

The results are presented of a series of experiments designed to evaluate the subjective effectiveness of a digital signal processing system for the production of virtual acoustic sources. The signal processing system, described in an earlier paper, ensures that the acoustic signals produced in the region of a listener's ears by a pair of loudspeakers are substantially equivalent to the signals that would be produced by a virtual source at a prespecified angular location. In this work, attention is restricted to the horizontal plane containing the listener's ears, the pair of loudspeakers, and the intended location of the virtual source. The signal processing system is designed to be able to produce the desired signals at the listener's ears irrespective of the acoustical environment in which the sound is reproduced. The approach used automatically incorporates the design of inverse filters which compensate for the reverberation in a given acoustical environment. Experiments are therefore undertaken not only in anechoic conditions, but also in a listening room and in the interior of an automobile. The system is shown to be remarkably effective in producing accurately localized images over a wide range of angles to the front of a large population of experimental subjects. The system is not so effective in producing images to the sides of listeners and is not effective in producing images to the rear, these images usually being perceived at "mirror" angular locations to the front of listeners.

0 INTRODUCTION

The creation of the illusion in a listener that a sound source is located in a given spatial position has long been a goal of acoustical engineers. It has been appreciated for many years [1] that relatively simple signal processing schemes can be used to operate on signals fed to a pair of loudspeakers in order to produce the illusion in a listener that the sound originates from a phantom or virtual source placed somewhere between the loudspeakers. Such techniques form the basis of conventional stereophony, the psychoacoustical basis for which has been thoroughly reviewed by Blauert [2] under the category of "summing localization." Simply providing a difference in level (or time delay) between the two signals input to a pair of loudspeakers placed appropriately with respect to the listener enables the image of the virtual source to be shifted in position between the two loudspeakers. A more sophisticated signal processing scheme is that generally attributed to Atal and Schroeder [3] (although a similar procedure had previously been investigated by Bauer [4] within the context of the reproduction of dummy-head recordings). Atal and Schroeder devised a localization network, which processed the signal to be associated with the virtual source prior to being input to the pair of loudspeakers. The principle of the technique was to process the virtual source signal via a pair of filters which were designed in order to ensure that the signals produced at the ears of a listener were substantially equivalent to those produced by a source chosen to be in the desired location of the virtual source. The filter design procedure adopted by Atal and Schroeder assumed that the signals produced at the listener's ears by the virtual source were simply related

* Manuscript received 1996 January 26.

by a frequency-independent gain and time delay. This frequency-independent difference between the signals at the ears of the listener was assumed to be dependent on the spatial position of the virtual source. These assumptions resulted in the analytically tractable design of a localization network whose parameters could be varied in order to provide apparently different locations of the virtual source. Although a comprehensive subjective evaluation of this technique does not appear to have been undertaken by the inventors, the method was reported [3] to be effective in producing the illusion in the listener of virtual sources located in the horizontal plane at angles of azimuth of up to ±60° (that is, outside the range of angular locations of ±30° typically achieved with intensity stereo [2]). However, the inventors also reported that beyond ±60° "localization is less well defined since it is more strongly dependent on frequency."

Schroeder et al. [5] later applied the essence of this method to the loudspeaker reproduction of dummy-head recordings. In this case, the signals recorded at the ears of a dummy head were processed via a filter network, which ensured that substantially the same signals were reproduced at the ears of a listener by a pair of loudspeakers. This network ensured the cancellation of the "crosstalk" between the right loudspeaker and the left ear, and vice versa. Again, no thorough subjective experiments were presented, but it was reported that "virtual sound sources can be created far off to the sides and even behind the listener."

The results of subjective experiments on the same type of system (that is, dummy-head recordings reproduced via a pair of loudspeakers after processing via a crosstalk cancellation network) were, however, reported by Damaske and Mellert [6], who dubbed the technique TRADIS (true reproduction of all directional information by stereophony). The results of localization experiments in both the horizontal and the median planes clearly demonstrate the effectiveness of the technique. More recently the essence of this approach has been used by Hamada et al. [7], who implement the crosstalk cancellation network digitally and refer to it as the orthostereophonic system (OSS). Again, the results of subjective experiments are presented, which show remarkable accuracy in the localization of virtual sources generated by first recording the signals produced at the ears of a dummy head and subsequently processing them via a 2×2 matrix of digital filters prior to transmission via a pair of loudspeakers. Further subjective experiments have also been presented recently by Neu et al. [8] and Urbach et al. [9], who again use a digital implementation of a crosstalk cancellation system to process the signals recorded at the ears of a dummy head. Again, good results are shown to be achievable, especially for virtual source positions in the horizontal plane. This general approach to the production of virtual acoustic sources has also been discussed by Cooper and Bauck [10], who refer to the technique as "transaural stereo" and who also discuss its generalization to reproduction for multiple listeners [11]. Work on transaural stereo has also been presented by Møller [12] and by Kotorynski [13].

The filter design procedures used by all these authors generally involves the deduction of the matrix of filters comprising the crosstalk cancellation network from either measurements or analytical descriptions of the four head-related transfer functions (HRTFs) relating the input signals to the loudspeakers to the signals produced at the listener's ears under *anechoic* conditions. The crosstalk cancellation matrix is the inverse of the matrix of four HRTFs. As recognized by Atal and Schroeder [3], this inversion runs the risk of producing an unrealizable crosstalk cancellation matrix if the components of the HRTF matrix are nonminimum phase. The presence of nonminimum-phase components in the HRTFs (due to reflections from room surfaces, for example [14]) can be dealt with by using the filter design procedure presented by Nelson et al. [15]–[17], which is described in more detail in an earlier paper [18] also appearing in this issue. In that work the sound reproduction problem is formulated in a very general way (accounting for a multiplicity of recorded signals and reproduced signals) and uses a least-squares approach to the design of the inverse filter matrix. A number of possibilities are suggested for making use of such signal processing techniques. In particular the compensation for poorly positioned loudspeakers in conventional stereophonic reproduction is shown to be a possible application, together with further extensions to the case of multiple listeners. The realization of these possibilities depends on the specification of the "desired" signals that are to be reproduced at the ears of a listener, and in [18] it was shown that a convenient way to specify the desired signals is to assume that they emanate from a virtual source in a prespecified position. In the case of a "loudspeaker position compensation" system, therefore, the locations of two virtual sources are specified, and the signal processing system has been shown in principle to be able to reproduce the desired signals accurately. Before evaluating the subjective performance of such systems, however, it makes sense to evalute first the effectiveness of the techniques in reproducing the desired signals from a single virtual source. In this work, therefore, we consider only the production of single virtual source images.

In the work described here we present the results of subjective experiments on a virtual source imaging system that is capable of producing the illusion in a listener of virtual sources located in the horizontal plane, but which has been found to operate effectively in a variety of acoustical environments. In this work we revert to the original intention of Atal and Schroeder, that is, we devise a signal processing scheme that is capable of operating on a single signal to be associated with a virtual source and we do not make explicit use of dummy-head recordings. However, we do make *implicit* use of a dummy head and use a set of measurements of the transfer functions between a loudspeaker input and the outputs of the ears of a dummy head. This database of dummy-head HRTFs is used to filter the virtual source signal in order to produce the signals that would be produced at the ears of the dummy head by a virtual source in a prescribed spatial position. These two signals

are then passed through a matrix of crosstalk cancellation filters, which ensure the reproduction of these two signals at the ears of the same dummy head placed in the environment in which imaging is sought. Full details of the signal processing techniques used to produce the desired effect are given in [18] and also in the Ph.D. dissertation presented by Orduña-Bustamante [19]. Further details of the subjective experiments are also given in the M.Sc. dissertation presented by Engler [20]. These techniques are also described in a patent application [21] and subsequently were first published in [22]. The results of experiments are presented here for listeners in an anechoic room, in a listening room (built to IEC specifications), and inside an automobile. It is concluded that the generality of the signal processing technique used has considerable promise for future applications in the production of virtual acoustic images.

1 PRINCIPLE OF OPERATION OF THE SYSTEM

Fig. 1 illustrates the principle of operation of the system used. The problem that we wish to solve is illustrated in block diagram form in Fig. 1(a), which is a special case of the general block diagram illustrated in [18, Fig. 1]. In this case, however, the desired signals $d_1(n)$ and $d_2(n)$, which are the elements of the vector $\boldsymbol{d}(n)$, are the signals produced at the ears of a listener by the virtual source whose input signal is $x(n)$. We can therefore represent the generation of these desired signals by passing the signal $x(n)$ through the pair of transfer functions $A_1(z, \theta)$ and $A_2(z, \theta)$, which specify the HRTFs of the listener. These transfer functions comprise the vector $\boldsymbol{a}(z, \theta)$ given by $[A_1(z, \theta), A_2(z, \theta)]^T$. In the experiments described here these transfer functions are evaluated by compiling a database of dummy-head HRTFs under anechoic conditions. This involves the measurement of the impulse response functions relating the input sequence $x(n)$ to the sequences $d_1(n)$ and $d_2(n)$. We wish to determine the vector of filters $\boldsymbol{h}(z, \theta)$ given by $[H_1(z, \theta), H_2(z, \theta)]^T$, which ensures the minimization of the sum of squared errors between the two desired signals at the listener's ears and the two reproduced signals at the listener's ears. The error signals $e_1(n)$ and $e_2(n)$ are thus specified by $e_1(n) = d_1(n) - z_1(n)$ and $e_2(n) = d_2(n) - z_2(n)$, where the signals $z_1(n)$ and $z_2(n)$ are the elements of the vector $\boldsymbol{z}(n)$ of reproduced signals. These are in turn related to the signals $y_1(n)$ and $y_2(n)$ input to a pair of loudspeakers by the matrix $\boldsymbol{C}(z)$ of electroacoustic transfer functions. Note that these transfer functions contain the effects of both the HRTF of the listener and the environment in which the listener is placed. Again, in the experiments described, the elements of the matrix $\boldsymbol{C}(z)$ are deduced by measurement of the impulse responses relating the loudspeaker inputs to the outputs of the ears of a dummy head placed in the environment in which imaging is sought (and therefore contain, for example, the effect of room reflections).

A very convenient technique for deducing the vector of filters $\boldsymbol{h}(z, \theta)$ is to evaluate first the matrix of crosstalk cancellation filters $\boldsymbol{H}_x(z)$ depicted in Fig. 1(b). This matrix of filters is designed to ensure that, when spectrally broad training signals $\boldsymbol{x}_x(n)$ are applied to the inputs of the loudspeakers, then the sum of the squared error signals comprising the vector $\boldsymbol{e}_x(n)$ is minimized [see Fig. 1(b)]. As described earlier [18], a modeling delay of Δ samples is incorporated in order to ensure that the crosstalk cancellation matrix $\boldsymbol{H}_x(z)$ can be realized. Thus again referring to Fig. 1(b), the crosstalk cancellation matrix for a given environment is designed to ensure that, to a good approximation,

$$\boldsymbol{C}(z)\boldsymbol{H}_x(z) \approx z^{-\Delta}\boldsymbol{I} \tag{1}$$

where $\boldsymbol{I}$ is the identity matrix. Multiplying both sides of this equation by the vector $\boldsymbol{a}(z, \theta)\, x(z)$ then shows that

$$\boldsymbol{C}(z)\boldsymbol{H}_x(z)\boldsymbol{a}(z, \theta)x(z) \approx z^{-\Delta}\,\boldsymbol{a}(z, \theta)x(z) \tag{2}$$

which can be written simply as

$$\boldsymbol{z}(z) \approx z^{-\Delta}\,\boldsymbol{d}(z)\,. \tag{3}$$

Thus the reproduced signals are, to a good approximation, simply delayed versions of the desired signals. This is depicted in block diagram form in Fig. 1(c). It therefore becomes apparent that the vector of filters

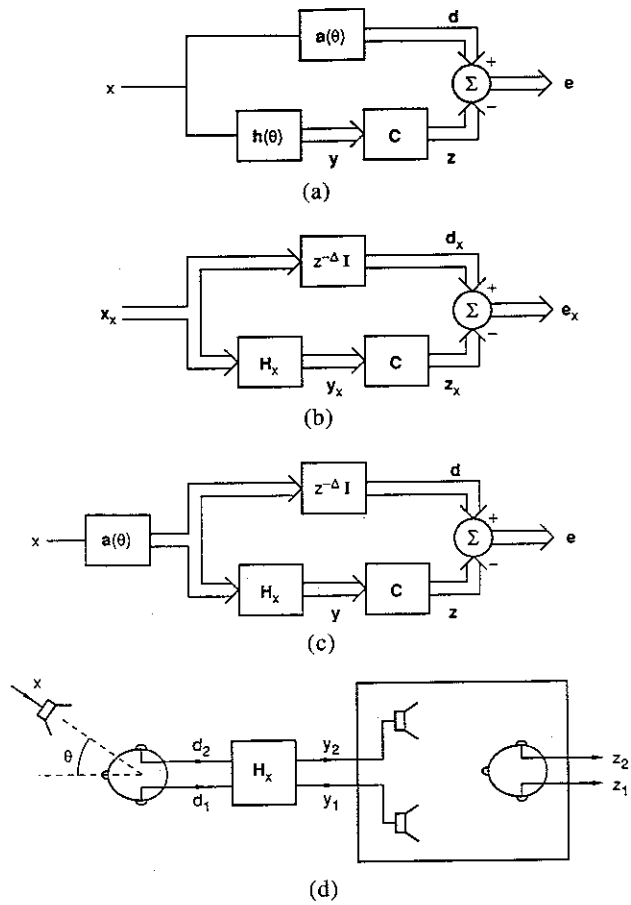


Fig. 1. Signal processing scheme used for synthesis of virtual acoustic sources. (a) Block diagram. (b) Design of crosstalk cancellation matrix. (c) Implementation of system used. (d) Graphic illustration of principles of system operation.

$\boldsymbol{h}(z, \theta)$ that we wish to deduce is given simply by

$$\boldsymbol{h}(z, \theta) = \boldsymbol{H}_x(z)\boldsymbol{a}(z, \theta) . \quad (4)$$

This very conveniently separates the inversion of the electroacoustic system from the generation of position-dependent desired signals associated with the virtual source. Adopting this approach necessitates only one inverse filter design step, that of the crosstalk cancellation matrix $\boldsymbol{H}_x(z)$, as illustrated in Fig. 1(c). Of course one could equally well work with the block diagram of Fig. 1(a) and design $\boldsymbol{h}(z, \theta)$ by minimizing the sum of the squared errors generated for each value of $\boldsymbol{a}(z, \theta)$, but this requires a separate inverse filter design step for each desired location of the virtual source and would involve a very large amount of computation if many virtual source locations were required.

Finally Fig. 1(d) provides a more graphic illustration of the procedure implied by the block diagram of Fig. 1(c). One could imagine the desired signals to be those produced at the ears of a first dummy head by a virtual loudspeaker whose input signal is $x(n)$. One can think of applying these desired signals $d_1(n)$ and $d_2(n)$ to the inputs of the crosstalk cancellation matrix $\boldsymbol{H}_x(z)$. This further matrix ensures that the outputs of the microphones at the ears of the second dummy head are delayed versions of the signals input to the crosstalk cancellation matrix in accordance with Eq. (3), that is, that $z_1(n) \approx d_1(n - \Delta)$ and $z_2(n) \approx d_2(n - \Delta)$, thereby generating the required signals at the listener's ears. In the work described hereafter the measurements of $\boldsymbol{a}(z, \theta)$ (and thus, by implication, the desired signals) were undertaken first under anechoic conditions enabling the compilation of a database of HRTFs. For each environment dealt with the crosstalk cancellation matrix was computed separately, then enabling the generation of the filters $H_1(z, \theta)$ and $H_2(z, \theta)$ by the convolution of the vector $\boldsymbol{a}(z, \theta)$ with the matrix $\boldsymbol{H}_x(z, \theta)$ in accordance with Eq. (4).

All the experiments described here made use of the dedicated signal processing system described earlier [18], which was based on a Texas Instruments TMS320-C30 microprocessor and housed an array of Motorola DSP-56200 FIR convolution circuits. The DSP system was controlled from a PC-386DX/33-MHz personal computer and, in the subjective experiments described, was used to implement the filters $H_1(z, \theta)$ and $H_2(z, \theta)$. The database of dummy-head HRTFs was first compiled using the MLSSA system [23] running on an IBM PC. These measurements were made in the ISVR large anechoic chamber (volume 295 m^3) by placing a KEF C35 SP3093 loudspeaker at a 2-m distance in the horizontal plane of a KEMAR DB 4004 artificial head and torso. The latter made use of Zwislocki DB100 ear-canal simulators with B&K UA 0122 ear-canal adapters and B&K type 4165 microphones. Measurements were made at 36 angles, covering the right half of the horizontal plane of the head in increments of 5°. The HRTFs for the left half of the horizontal plane were derived from these measurements by reflection in the median plane. The sample rate for these measurements was 48.19 kHz and used a bandwidth of 20 kHz. The original impulse response measurements made use of 1024 samples, but these were later reduced to 256 samples by removal of a fixed part of the initial delay (i.e., the same initial delay was removed from each impulse response) and by windowing the tail of the impulse response.

2 EXPERIMENTS UNDER ANECHOIC CONDITIONS

2.1 Experimental Arrangement

Fig. 2 shows the geometrical arrangement of the sources and dummy head used in first designing the

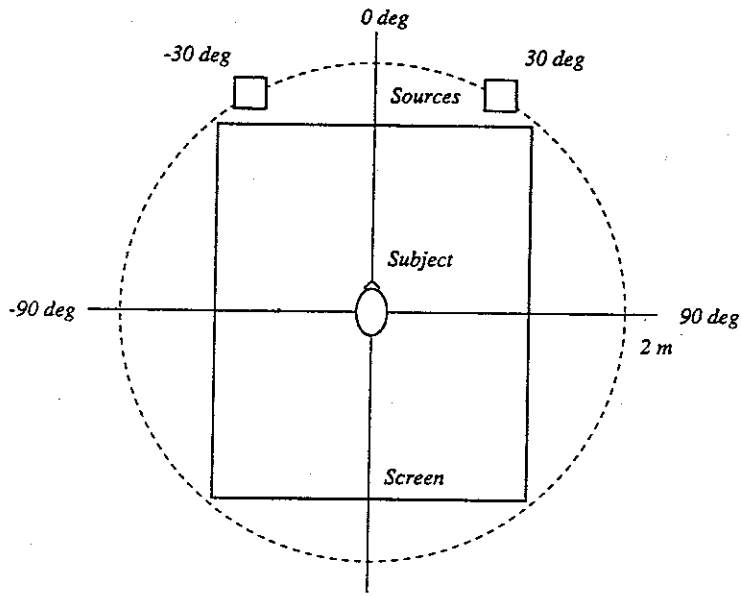


Fig. 2. Layout used during tests for subjective localization of virtual sources. Virtual sources were emulated via the pair of sound sources shown facing the subject. A dark screen was used to keep sound sources out of sight. Circle drawn outside screen marks distance at which virtual and additional real sources were placed for localization at different angles.

crosstalk cancellation matrix $\boldsymbol{H}_x(z)$ for the experiments undertaken in anechoic conditions. The loudspeakers used were KEF type C35 SP3093, and the dummy head used was the KEMAR DB 4004 artificial head and torso, which of course was the same head as that used in the HRTF database compilation described in Sec. 1. The elements of the matrix $\boldsymbol{C}(z)$ relating the loudspeaker input signals to the signals output from the ears of the dummy head were also determined by using the MLSSA system [23]. The HRTF measurements were made at a 72-kHz sample rate, and the resulting impulse responses were then downsampled to 48 kHz. The results are depicted in Fig. 3, which shows the impulse responses corresponding to the elements of the matrix $\boldsymbol{C}(z)$ once the first 280 samples (corresponding to a pure delay due to the acoustic travel time) had been removed and also after truncating the responses to 1024 samples with a half-Hanning window applied to the last 16 points. Fig. 4 shows the impulse responses corresponding to the elements of the crosstalk cancellation matrix $\boldsymbol{H}_x(z)$ that was designed using the procedures described. Again, these impulse responses are those measured at a 48-kHz sample rate. Finally Fig. 5 shows the results of convolving the matrix $\boldsymbol{H}_x(z)$ with the matrix $\boldsymbol{C}(z)$, in both the time domain and the frequency domain. The time-domain results [Fig. 5(a)] show the effectiveness of the crosstalk

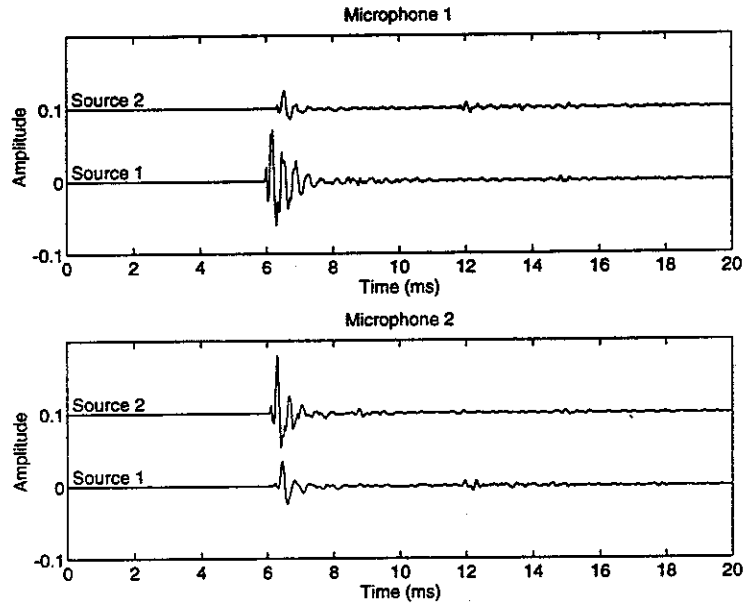


Fig. 3. Impulse responses of electroacoustic system in anechoic chamber. Results are shown for impulse responses relating inputs of left (source 1) and right (source 2) loudspeakers to outputs of left (microphone 1) and right (microphone 2) ears of dummy head.

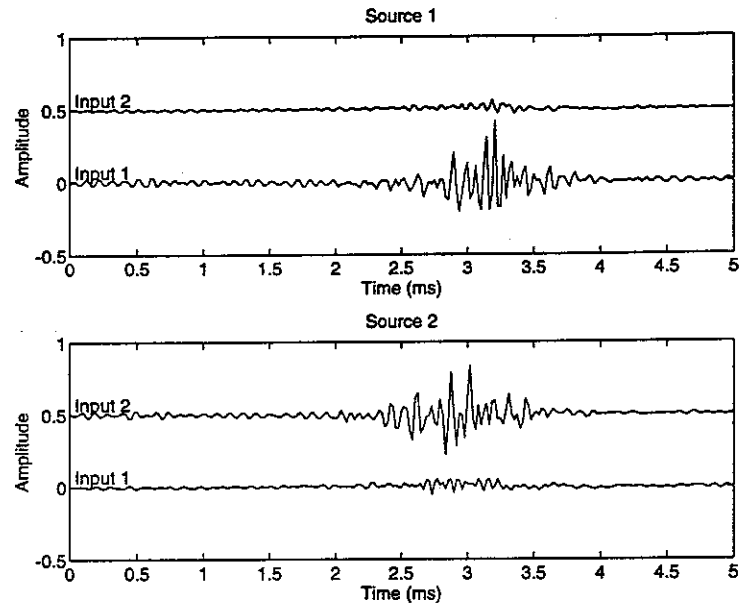


Fig. 4. Impulse responses of matrix of crosstalk cancellation filters used in anechoic chamber. Results are shown for impulse responses relating inputs $d_1(n)$ (input 1) and $d_2(n)$ (input 2) to loudspeaker inputs $y_1(n)$ (source 1) and $y_2(n)$ (source 2). [See Fig. 1(d) for definition of signals involved.]

cancellation and clearly illustrate that only the diagonal elements of the product $\boldsymbol{H}_x(z)\boldsymbol{C}(z)$ are significant and that Eq. (1) is, to a good approximation, satisfied. Note that the modeling delay Δ chosen was the order of 150 samples. The frequency-domain results [Fig. 5(b)] show excellent crosstalk cancellation up to about 8 kHz, but at frequencies above this, the effectiveness of the inversion is a little erratic. Nevertheless, below 8 kHz the crosstalk is canceled by a margin of about 20 dB.

2.2 Subjective Experiments

Having designed the matrix of crosstalk cancellation filters as described, the HRTF database that was previously captured was then used to operate on various virtual source signals $x(n)$ in order to generate the desired signals $d_1(n)$ and $d_2(n)$ corresponding to a chosen virtual source location. These were then passed through the crosstalk cancellation filter matrix to generate the loudspeaker input signals. Listeners were then seated such that their heads were, as far as possible, in the same position relative to the loudspeakers as that occupied by the dummy head when the crosstalk cancellation matrix was designed. Listeners were surrounded by an acoustically transparent screen (Fig. 2), and a series of marks were made inside the screen at 10° intervals along a line in the horizontal plane (that is, the plane containing the center of the loudspeakers and the listener's ears). Listeners were asked to look straight ahead at the mark

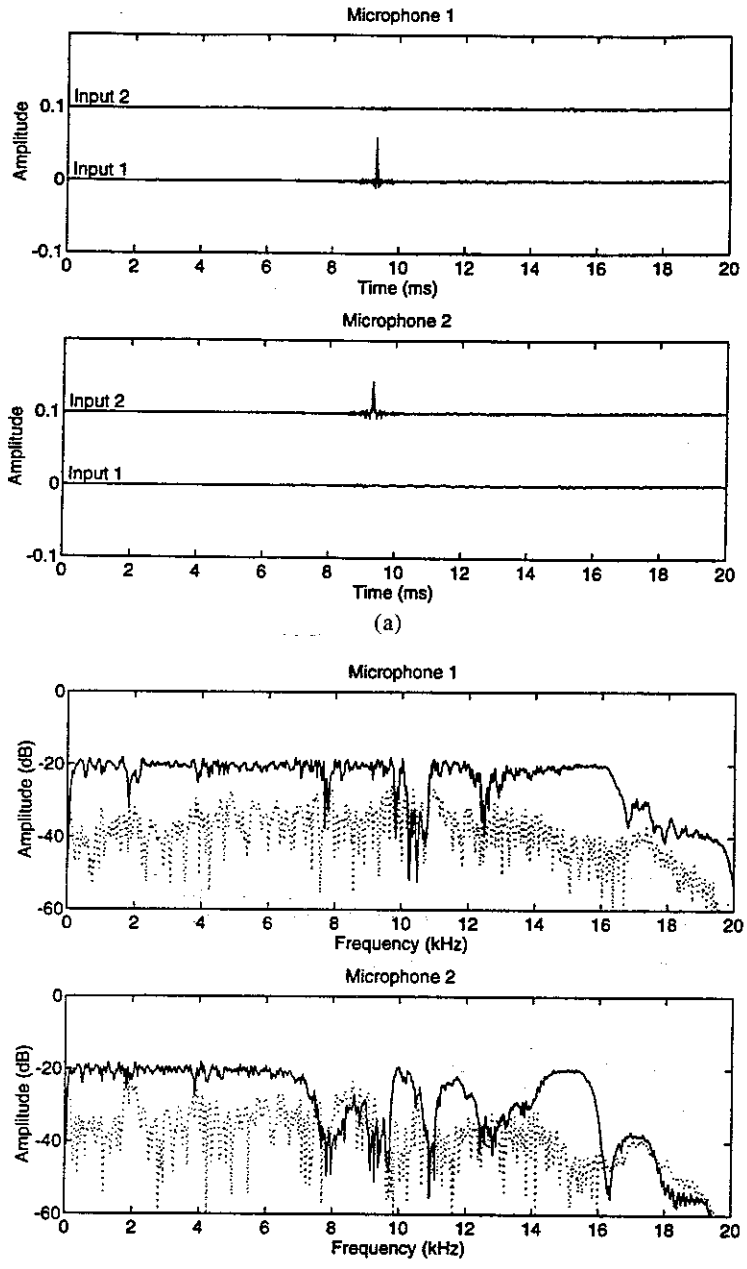


Fig. 5. Matrix of filters resulting from convolution of impulse responses of electroacoustic system in anechoic chamber with matrix of crosstalk cancellation filters (a) in time domain and (b) in frequency domain. (a) Results for impulse responses relating inputs $d_1(n)$ and $d_2(n)$ to outputs $z_1(n)$ (microphone 1) and $z_2(n)$ (microphone 2). [See Fig. 1(d) for definition of signals involved.] (b) Responses at microphone 1 due to input $d_1(n)$ (solid line) and due to input $d_2(n)$ (dashed line). Similarly for responses at microphone 2.

corresponding to 0°, the loudspeakers being positioned symmetrically relative to the listener behind the screen at azimuthal locations of ±30° (Fig. 2). After presentation of a given virtual source stimulus [that is, some combination of input signal $x(n)$ and choice of filters $A_1(z, \theta)$ and $A_2(z, \theta)$ corresponding to a given virtual source location] the listeners were asked to decide upon the angular location of the virtual source. Listeners were asked to make this decision while still looking straight ahead and then (if necessary) turn their heads to nominate the mark on the screen that most closely corresponded to their choice of virtual source location. No attempt was made to restrain the motion of the listener's head otherwise.

In order to provide a direct evaluation of the effectiveness of the system in producing the illusion of virtual sources in a given location, a series of experiments were also undertaken using real loudspeaker sources. These were placed at various locations on a circle of 2-m radius surrounding the listener. For each set of experiments undertaken with virtual sources, an equivalent set of experiments were undertaken with real sources. Each subject was presented with both sets of stimuli. The real sources were presented first to the subjects, with the duration of a typical experimental session being on the order of 50 min. The subjects were asked to return two days later for the experiments with virtual sources.

The types of signal $x(n)$ used as inputs to both real and virtual sources consisted of speech, one-third-octave bands of random noise centered at 250 Hz, 1 kHz, and 4 kHz, and also pure tones at 250 Hz, 1 kHz, and 4 kHz. A summary of all the experiments undertaken is shown in Table 1. The presentation of different angular locations of both real and virtual sources was divided into three "sets" of angles. These are also defined in Table 1. Set 0 consisted of angles to both the front and the rear of the listener, whereas set 1 and set 2 contained angles only in the forward half of the horizontal plane. In each of the experiments defined in Table 1, the angles from a given set were presented in a particular sequence. Thus, for example, sequence 0A refers to a specific order of presentation of angles from set 0, whereas sequence 1A refers to another sequence of presentations of angles from set 1. The particular sequences used are specified in Table 2. Note that the order of presentation of the angles in a given sequence was chosen randomly in order that subjects could not learn from the order of presentation. In addition, an attempt was made to minimize any bias produced in the subjective judgments caused by order of presentation by ensuring that each sequence

Table 1. Summary of all experiments undertaken.*

		One-Third-Octave			Pure Tone		
Experiment	Speech	250 Hz	1 kHz	4 kHz	250 Hz	1 kHz	4 kHz
1	0A	0B*	0C*	0D*	0E*	0F*	0G*
	1A	1B	2A	2B	2C	1C	1D
2	0A	0B*	0C*	0D*	0E*	0F*	0G*
	1Ar†	1Br	2Ar	2Br	2Cr	1Cr	1Dr
3	0Ar	0Br*	0Cr*	0Dr*	0Er*	0Fr*	0Gr*
	2A	2B	1A	1B	1C	2C	2D
4	0Ar	0Br*	0Cr*	0Dr*	0Er*	0Fr*	0Gr*
	2Ar	2Br	1Ar	1Br	1Cr	2Cr	2Dr
Sets of angles							
0	60	100	150	180	−60	−100	−150
1	30	60	90	−30	−50	−90	
2	20	50	80	−20	−60	−80	

* All experiments used both real and virtual sources except those marked *, which were undertaken with real sources only. The sequences of angles used are defined in Table 2. Each experiment was undertaken by three subjects.

† r denotes presentation of a sequence in reverse order.

Table 2. Specification of all sequences of angles used in subjective experiments.

0A:	0	180	−150	−100	60	−60	100	0												
	150	180	60	−150	−100	100	−60	150	(Virtual sources)											
0A:	0	100	150	60	180	−60	−150	−100	(Real sources)											
0B:	100	−60	180	150	60	−100	0	−150												
0C:	150	180	0	100	60	−60	−150	−100												
0D:	60	150	−60	0	100	−150	150	−100												
0E:	−150	−60	0	100	−100	60	180	150												
0F:	100	−60	60	180	0	−150	150	−100												
0G:	−60	0	−150	180	150	60	−100	100												
1A:	−30	0	60	90	−50	−90	30	0	−30	60	30	−90	90	−30	−50	60	90	−90	0	30	−50
1B:	60	30	−50	30	0	−90	90	60	−50	−90	−30	30	0	−30	−90	−50	90	60	0	90	−30
1C:	90	−90	30	0	60	90	−90	−50	−30	30	0	−50	60	90	30	−50	0	−90	−30	60	−30
1D:	30	60	−90	−50	−30	−50	0	−90	90	60	0	30	60	−30	−90	0	−50	90	30	−30	90
2A:	−80	20	50	0	−20	80	−60	−80	20	50	0	−20	80	−20	−80	−60	80	20	0	50	−60
2B:	0	20	−60	0	−20	80	50	−60	50	−80	−20	0	80	50	−60	20	−20	−80	20	80	−80
2C:	−20	80	20	50	−60	−20	0	20	0	−80	80	−60	0	50	80	−80	−20	50	−60	−80	20
2D:	20	50	−80	0	−20	−60	0	−80	80	50	0	20	50	−20	−80	0	−60	80	20	−20	80

was also presented in reverse order. Thus sequence 1Ar denotes the presentation of sequence 1A in reverse order.

Each of the experiments defined in Fig. 3 was undertaken by three subjects, a total of 12 subjects being tested in all. The subjects were all aged in their 20s and had normal hearing. A roughly equal division between male and female subjects was used, with at least one female being included in each group of three subjects. More details of these subjective experiments are presented by Engler [20].

2.3 Experimental Results

The first point to be made regarding the performance of the system was that it was generally unable to produce a convincing illusion of virtual sources located to the rear of the listener. This is clearly shown by the results depicted in Fig. 6, which compares the localizations of real and virtual sources. Here the squares have a side length that is directly proportional to the number of times a given "response" angle was recorded for a particular "presented" angle, that is, the number of times that the subjects responded to a given stimulus by answering that the source was located in a given angular location. The results in Fig. 6 (which are for speech signals) show that while the localizations of the real sources to the rear of the listener are remarkably accurate, presentations of virtual sources to the rear of the listener were very often "mirrored" to their equivalent angular locations to the front of the listener. Thus, for example, a presented angle of 150° would result in a response angle of 30°. It is worth pointing out, however, that although there were very few such "front–back confusions" in the case of real sources with a speech signal, these were very much in evidence when other types of stimulus signals were used with real sources, particularly so in the case of pure tones. (The reader is referred to [20] for the data on these test cases.)

Fig. 7 shows more clearly the ability of the system to generate convincing illusions of virtual sources to the front of the listener. This is particularly so for angles within the range of ±60°, although occasionally subjects again exhibited front–back confusions within this angular range. For angles outside ±60° there was a tendency for the subjects to localize the image slightly forward of the angle presented (that is, presented angles of 90° would be localized at 80, 70, or 60°). This is more clearly shown by the data in Figs. 8–10, which present the results for source signals consisting of one-third-

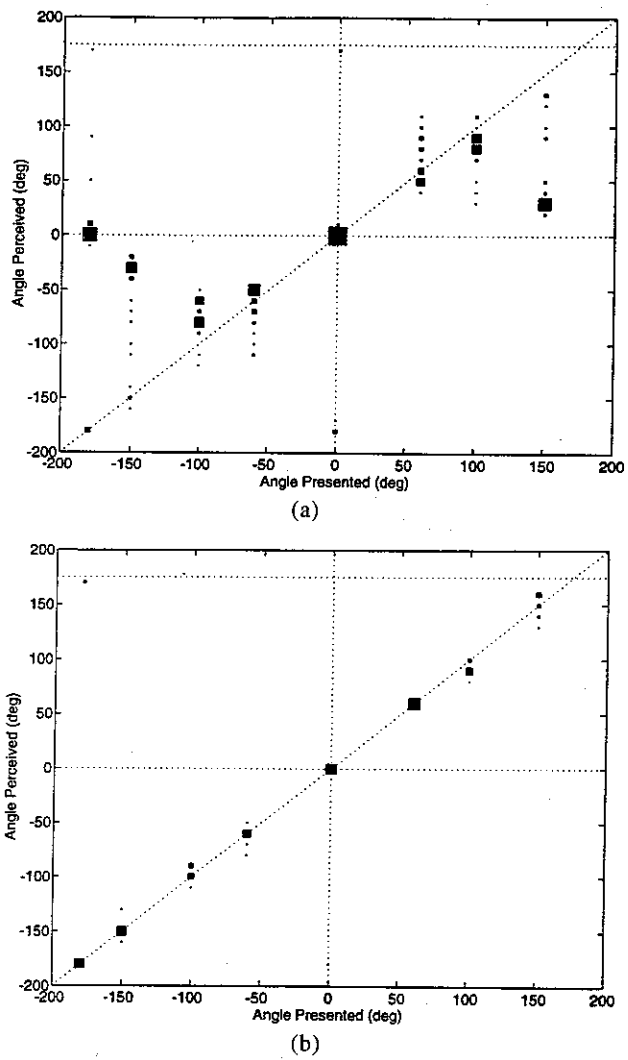


Fig. 6. Results of localization experiments in anechoic chamber using speech signal. (a) Virtual sources. (b) Real sources.

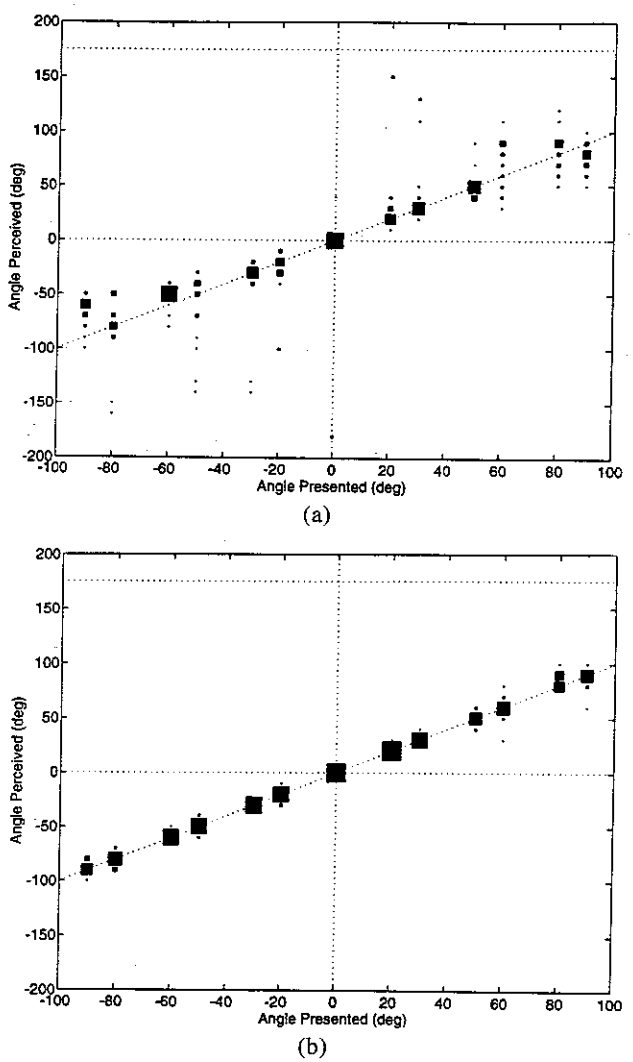


Fig. 7. Results of localization experiments in anechoic chamber using speech signal. (a) Virtual sources. (b) Real sources.

octave bands of white noise centered at 250 Hz, 1 kHz, and 4 kHz, respectively. Again occasional front–back confusion occurs, but these data show principally that there is some frequency dependence of the effectiveness of the system. Thus the data at 4 kHz (Fig. 10) show a larger degree of "forward imaging" of virtual sources when sources are localized to the front of their intended locations at the sides of the listener. The results for pure tones (see [20]) showed similar trends, although the scatter in the data was considerably greater than in the case of one-third-octave bands of noise.

3 EXPERIMENTS IN A LISTENING ROOM

3.1 Experimental Arrangement

An experiment arrangement identical to that used under anechoic conditions was also used under reverberant conditions, except that the experiments were undertaken inside a listening room built to IEC specifications. The geometrical arrangement of loudspeakers, listeners, and screen was identical to that illustrated in Fig. 2. The response of the electroacoustic system to be inverted was, however, markedly different and is shown in Fig. 11. Comparison with Fig. 3 shows that the signals input to the loudspeakers produced a significantly stronger series of reflections at the ears of the dummy head as a result of the surfaces of the listening room. Fig. 12 shows the impulse responses of the matrix of crosstalk cancellation filters, and Fig. 13 presents the results of convolving these with the measured impulse responses shown in Fig. 11. Again, the filter design procedure was very effective in deconvolving the system and producing a significant net response only in the diagonal terms of the matrix product $\boldsymbol{C}(z)\boldsymbol{H}_x(z)$, although the effectiveness of inversion was less at frequencies above about 8 kHz. It is also notable that the frequency-domain results depicted in Fig. 13(b) show a less effective inversion than the results presented in Fig. 5 measured under anechoic conditions.

3.2 Subjective Experiments

A series of experiments identical to those performed under anechoic conditions were undertaken. All the tests listed in Table 1 (using the sequences specified in Table 2) were repeated in the listening room. However, a different set of 12 subjects was used for the listening room tests, but the same procedures of testing with real and virtual sources were adhered to. Again, the listeners

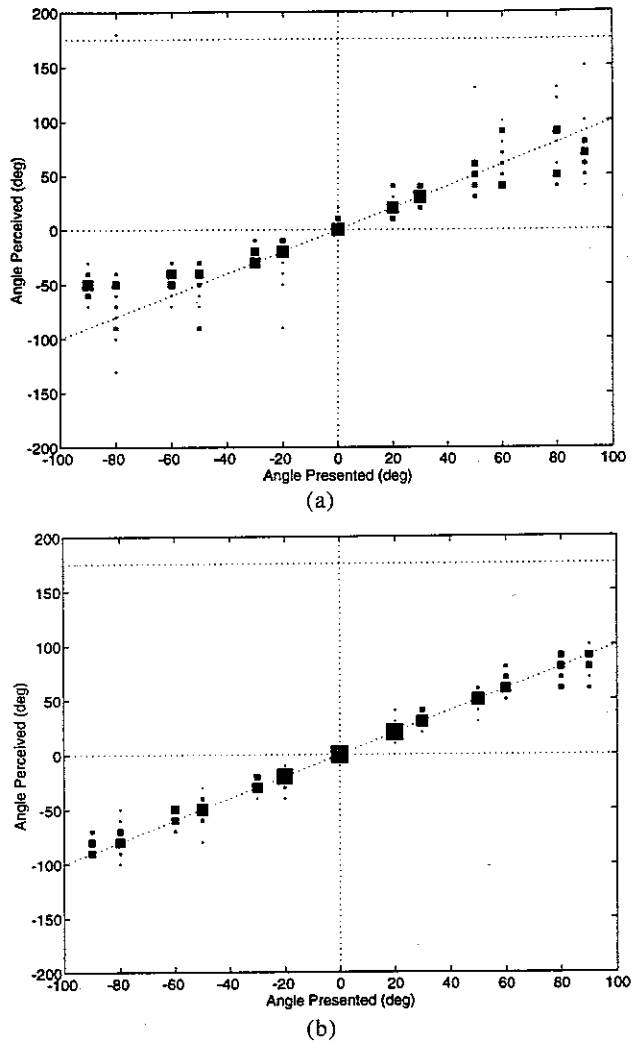


Fig. 8. Results of localization experiments in anechoic chamber using 250-Hz one-third-octave filtered white noise signal. (a) Virtual sources. (b) Real sources.

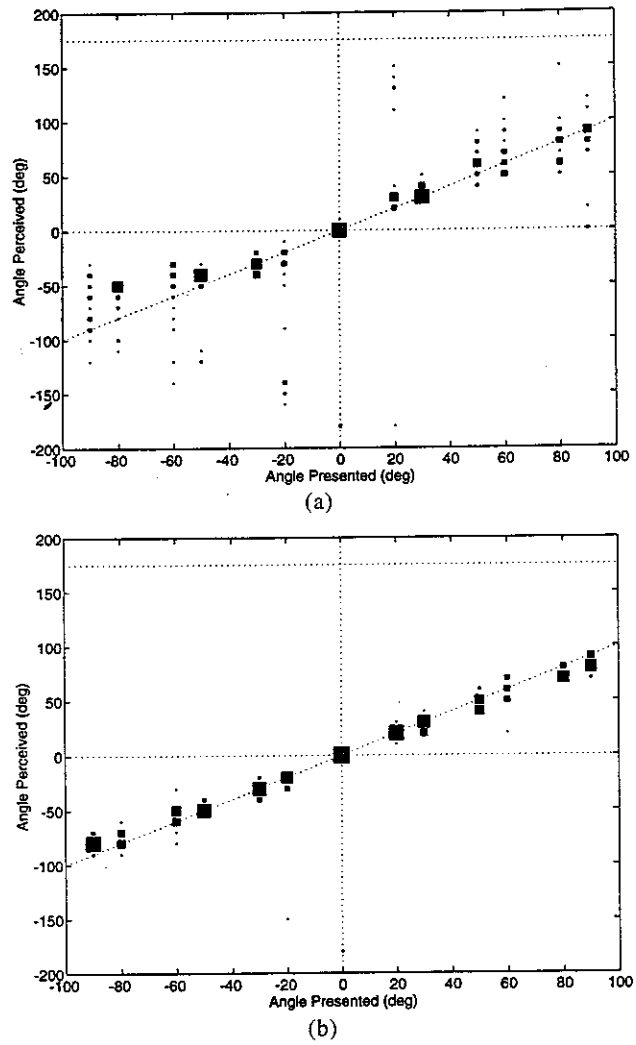


Fig. 9. Results of localization experiments in anechoic chamber using 1-kHz one-third-octave filtered white noise signal. (a) Virtual sources. (b) Real sources.

were generally in their 20s with normal hearing and distributed evenly in numbers between male and female.

3.3 Experimental Results

Fig. 14 compares the effectiveness of the virtual source imaging system and the ability of the listeners to localize real sources. Again, the system was found to be incapable of producing convincing images to the rear of the listener, with almost all virtual source presentations in the rear of the horizontal plane being perceived in their mirror image positions in the front. The results shown in Fig. 14 were again undertaken for speech signals, and it should be noted that although the results are not presented here, the localization of real sources with other signal types (pure tones and one-third-octave bands of noise) was far less accurate than with the speech signals and showed significant numbers of front–back confusions.

Again, however, the system was highly effective in producing accurately located images to the front of the listener, especially in the range of ±60°. This is illustrated in Fig. 15, which also shows fewer front–back confusions than observed in the equivalent experiments performed under anechoic conditions (Fig. 7). The results in Fig. 15 also show the tendency of the system to produce "forward images" of the virtual sources to either side of the listener. This tendency is again shown by the results produced by one-third-octave bands of noise (Figs. 16–18), being especially marked at 4 kHz (Fig. 18). It is also interesting to note that at 250 Hz (Fig. 16) the data show significantly greater scatter than at the same frequency under anechoic conditions. In the additional data presented by Engler [20] it is also shown that the localization of pure-tone virtual sources in a reverberant environment was generally poor, with results at 1 and 4 kHz being scattered similarly to those measured under anechoic conditions and those at 250 Hz showing a degree of scatter that was markedly greater than that measured under anechoic conditions.

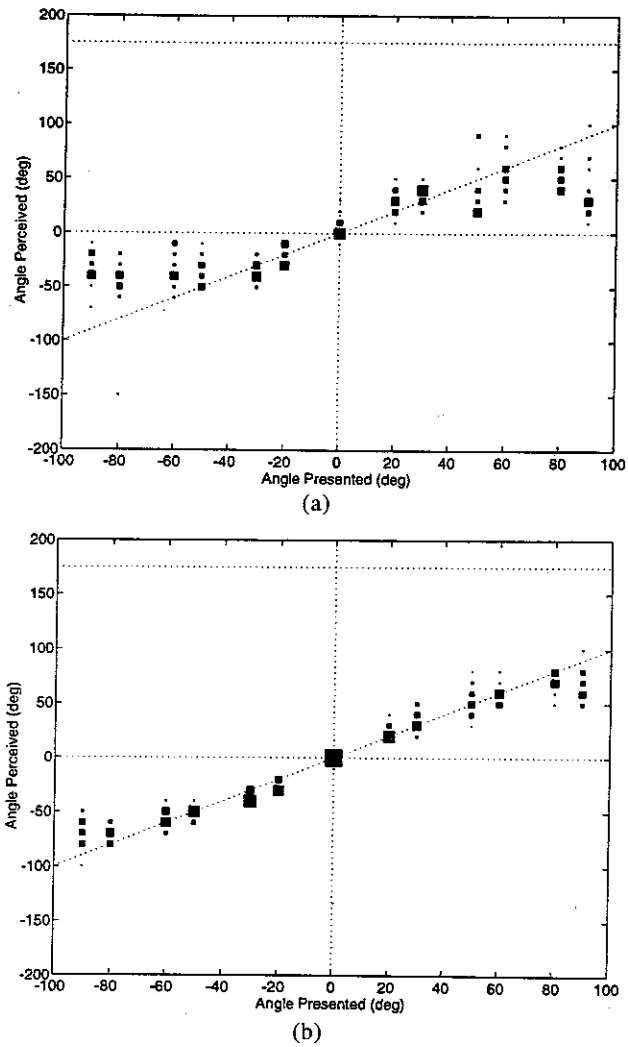


Fig. 10. Results of localization experiments in anechoic chamber using 4-kHz one-third-octave filtered white noise signal. (a) Virtual sources. (b) Real sources.

4 EXPERIMENTS INSIDE AN AUTOMOBILE

4.1 Experimental Arrangement

As a final, and more challenging, test of the ability of the system to produce convincing virtual acoustic sources, some brief experiments were undertaken in the interior of an automobile. The car used was an ISUZU I-Mark XS left-hand drive vehicle. The existing audio system loudspeakers were used to generate the signals presented to the listeners, these loudspeakers being fitted into the underside of the vehicle dashboard facing downward at an angle of approximately 45° to the horizontal. An approximate dimensional drawing of the arrangement is shown in Fig. 19. The loudspeakers were placed in a position well below the horizontal plane containing the listener's ears. Both the dummy head used to design the matrix of crosstalk cancellation filters and the listener were placed in equivalent positions in the driver seat on the left-hand side of the vehicle.

The impulse response of the loudspeaker–vehicle interior combination proved quite difficult to invert satisfactorily, the design of the matrix of crosstalk cancellation filters being made difficult by the limited number of filter coefficients available. Some attempt was made to ease this situation through damping the car interior by adding anechoic wedges to the boot space at the rear of the vehicle. The impulse responses comprising the matrix of electroacoustic transfer functions once this treatment was installed are shown in Fig. 20. The form and duration of these impulse responses are clearly very different from those measured in the anechoic room and the listening room, with substantial energy in the impulse response arriving well after the direct sound. This, of course, is a natural consequence of the highly reflective nature of the vehicle interior surfaces, which are placed very close to the listener. The crosstalk cancellation filters were consequently also of very long duration, and these impulse responses are shown in Fig. 21. The truncation of these impulse responses produced a less effective inversion than in the cases described, which is evident in the detailed frequency analysis of the deconvolved system transfer functions. The corresponding frequency responses of the deconvolved system are presented in Fig. 22, showing that the crosstalk cancellation was basically effective despite these difficulties.

4.2 Subjective Experiments

The environment being dealt with precluded a direct comparison between real and virtual sources, and therefore experiments were conducted only with virtual sources. The experiments described above showed that the system was at its most effective when using speech signals for the virtual source, and therefore only speech was used in these experiments. Essentially the same approach was taken in these experiments as in those described earlier, with subjects being asked to look directly in front, decide upon an angular location of the virtual source, and then nominate a marker placed in the horizontal plane outside the car.

In addition to the judgment of angular location, the subjects were also asked to give a judgment of elevation of the virtual source, either "above," "below," or "level" with the horizontal plane. This simple test was included since, unlike in the previous experiments, the loudspeakers used to generate the signals were well below the horizontal plane. The desired signals at the listeners' ears were of course due to virtual sources in the horizontal plane.

A total of 12 subjects was again used, all having normal hearing. These subjects were again different from those participating in the experiments undertaken in both the anechoic and the listening rooms. A total of 38 randomly chosen angular locations of virtual source were presented to each listener.

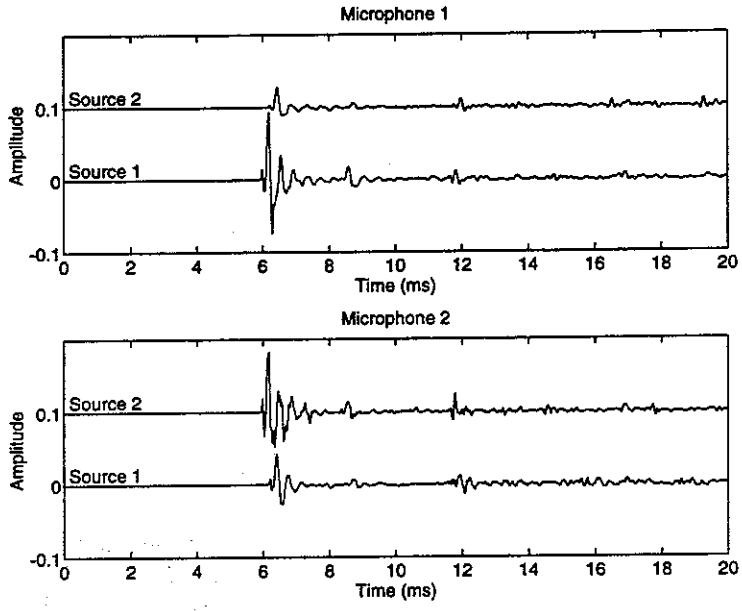


Fig. 11. Impulse responses of electroacoustic system in listening room. Results are shown for impulse responses relating inputs of left (source 1) and right (source 2) loudspeakers to outputs of left (microphone 1) and right (microphone 2) ears of dummy head.

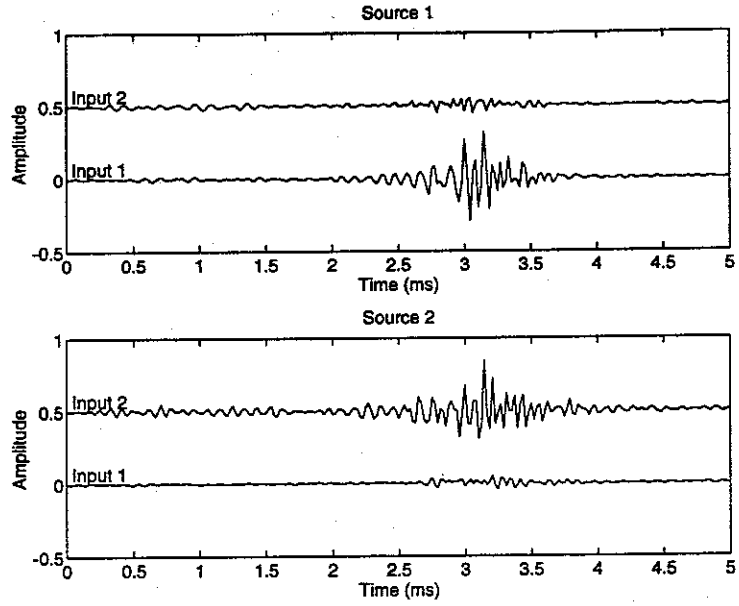


Fig. 12. Impulse responses of matrix of crosstalk cancellation filters used in listening room. Results are shown for impulse responses relating inputs $d_1(n)$ (input 1) and $d_2(n)$ (input 2) to loudspeaker inputs $y_1(n)$ (source 1) and $y_2(n)$ (source 2). [See Fig. 1(d) for definition of signals involved.]

4.3 Experimental Results

The results of the angular localization experiment are shown in Fig. 23. Although the general scatter of the data is somewhat larger than with the previous two test conditions using a speech source, very similar trends are evident in the data. Thus, for example, centrally placed images are reliably located and there is a tendency for forward imaging of virtual source locations to the side of the listener. There is also a tendency evident in the data that conflicts somewhat with the forward imaging trend. That is, for a relatively large number of tests, virtual sources presented to the side of the listener (from 60 to 90° and −60 to −90°) were all located at exactly 90 or −90°. It is possible that these results were actually derived from front–back confusions and were located by the listeners at the extremes (±90°) of the angular locations, which could be chosen from on the array of markers outside the car.

The results of the elevation test are presented in Fig. 24. These demonstrate that, on average, the subjects judged the virtual sources to be in the horizontal plane, although there was considerably indeterminacy in this judgment. Significant numbers of subjects judged the virtual sources to be below the horizontal plane for virtual source locations to the left of the listener, which is perhaps not surprising in view of the relatively large angle of elevation of the left-hand loudspeaker situated

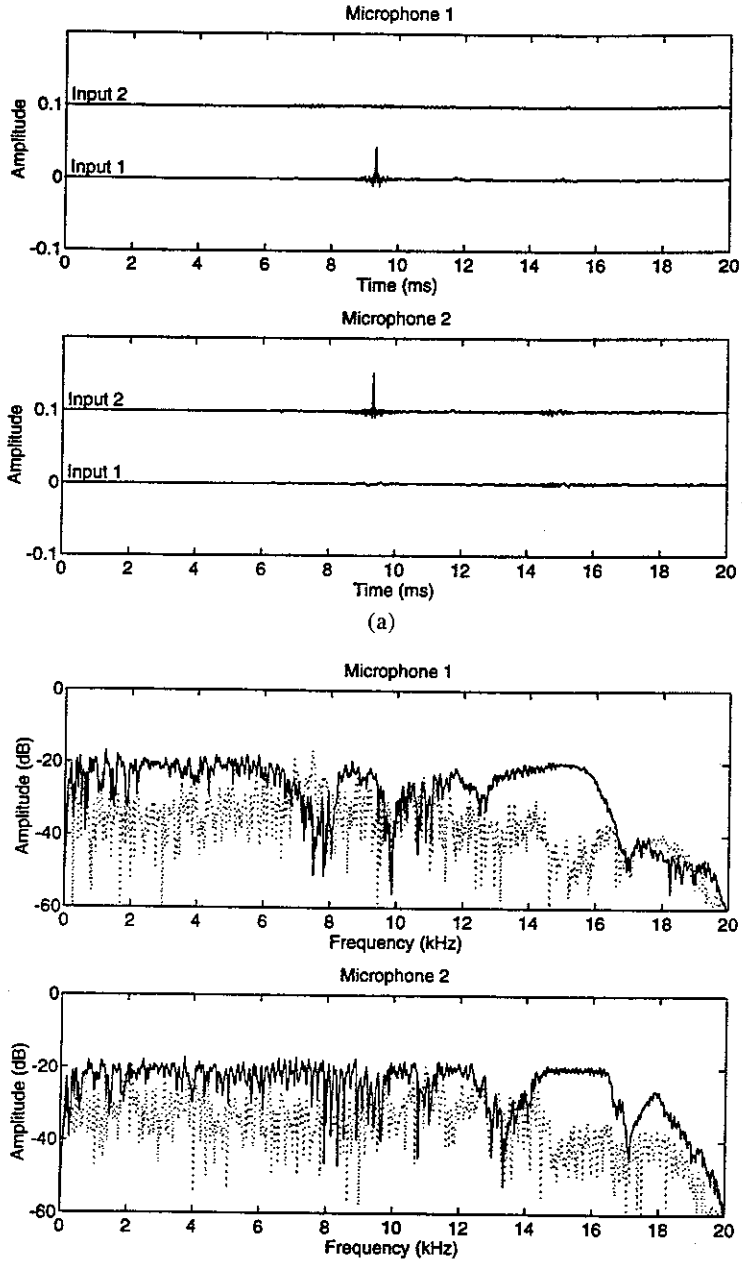


Fig. 13. Matrix of filters resulting from convolution of impulse responses for electroacoustic system in listening room with matrix of crosstalk cancellation filters (a) in time domain and (b) in frequency domain. (a) Results for impulse responses relating inputs $d_1(n)$ and $d_2(n)$ to outputs $z_1(n)$ (microphone 1) and $z_2(n)$ (microphone 2). [See Fig. 1(d) for definition of signals involved.] (b) Responses at microphone 1 due to input $d_1(n)$ (solid line) and due to input $d_2(n)$ (dashed line). Similarly for responses at microphone 2.

below the listener. In retrospect, this elevation test could have been posed better, with subjects being asked to locate the elevation of the virtual source with a range of vertical locations. What is clear from these data, however, is that the subjects did *not* consistently judge the location of the virtual sources to be below the horizontal plane.

5 DISCUSSION

The results of these experiments demonstrate that the signal processing scheme described earlier [18] is a very effective means of reliably producing virtual source images to the front of listeners in the horizontal plane over a range of angles of ±60°. Furthermore, this can be accomplished in a variety of environments, almost irrespective of the complexity of the acoustic response of those environments. It should also be emphasized that this technique has proved consistently effective with a population of subjects with normal hearing, and although the crosstalk cancellation filters used have been environmentally dependent, they have not been designed for individual listeners.

Certain trends have been repeatedly evident in the data. For example, the system failed to produce virtual images to the rear of the subjects tested, with those presentations generally being perceived in their mirror image locations to the front of the listener. Virtual sources presented to the side of the listener suffered from forward imaging and were generally perceived to be to the front of the intended angular location in the horizontal plane. Thus although virtual images to the side of the listener were more difficult to produce consistently, it was found possible to produce them at angular locations outside the ±60° range.

Clearly the explanation of these trends is a complex issue, and much remains to be done to understand fully the workings of acoustical virtual imaging systems of this type. The obvious explanation for the ability of the system to produce good results for sources to the front of the listener is that the real sources generate a sound field in the region of the listener's head that is predominantly propagating from front to back. The natural head movement that subjects used in the listening tests may well be capable of clearly distinguishing this direction of propagation, which is inherent in the sound field. An analysis of the exact form of the sound field produced by such systems in the region of a scattering body (the

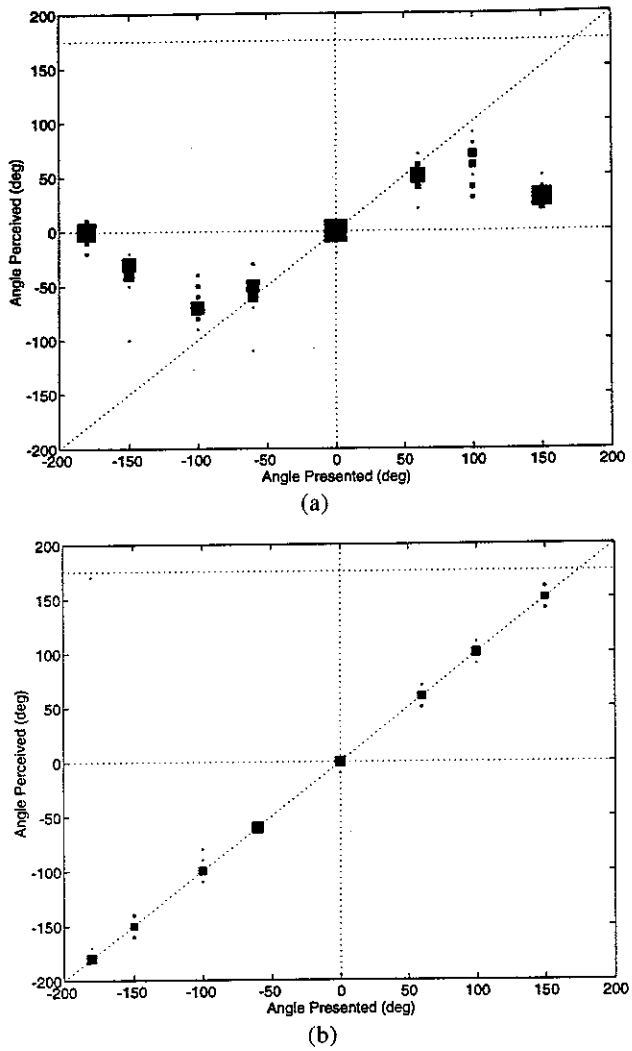


Fig. 14. Results of localization experiments in listening room using speech signal. (a) Virtual sources. (b) Real sources.

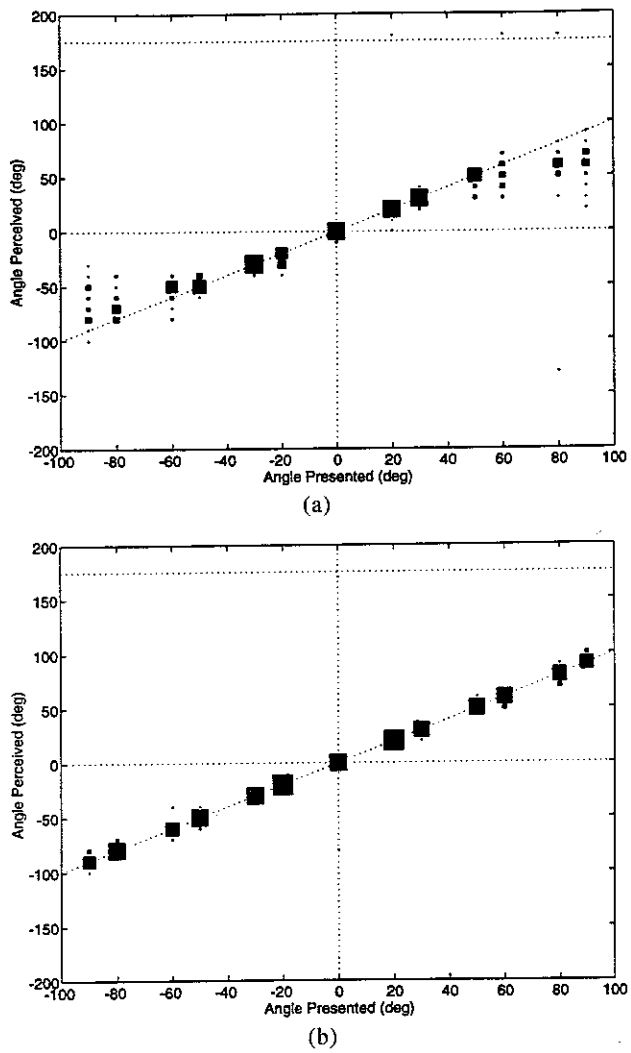


Fig. 15. Results of localization experiments in listening room using speech signal. (a) Virtual sources. (b) Real sources.

listener's head) may well yield further insights into this matter, and this work is currently under way.

There are many other issues raised by these experiments. For example, it would be interesting to see whether similar results would be obtained if instead of a dummy head, whose HRTF yields a very well defined and characteristic ear canal resonance and has "average" pinnae, another simpler body (a sphere, for example) were used. Furthermore, although in these experiments the effect of the acoustical environment has been corrected for in the design of the inverse filters, it would be interesting to evaluate the deterioration in system performance when instead of correcting for the environment, the crosstalk cancellation filters designed under anechoic conditions were used in a system operating in reverberant conditions. These issues remain to be addressed.

6 CONCLUSIONS

The results have been presented of a series of experiments designed to test the effectiveness of a system for the synthesis of virtual acoustic sources. The signal processing system used proved capable of inverting the responses of a variety of acoustical environments, including a listening room and the interior of an automobile in addition to that of an anechoic chamber. The results of the experiments showed that, irrespective of the acoustical environment, images of virtual acoustic sources could reliably be produced within an angular range of $\pm 60°$ in the horizontal plane. Virtual sources presented in the ranges 60 to 90° and -60 to $-90°$ were generally perceived to be slightly forward of their intended locations, whereas the virtual sources presented to the rear of the listener (in the ranges 90 to 180° and -90 to $-180°$) were generally perceived to be in their mirror angular locations to the front of the listener.

7 ACKNOWLEDGMENT

The authors are grateful to a number of Japanese companies for their support of this work, including Yamaha, MTT Instrumentation, Alpine Electronics, and Nittobo Acoustic Engineering. The contributions of Nissan and Bridgestone for their support of earlier phases of this work is also gratefully acknowledge. The authors would also like to thank I. H. Flindell of ISVR for a number of very useful discussions regarding this work.

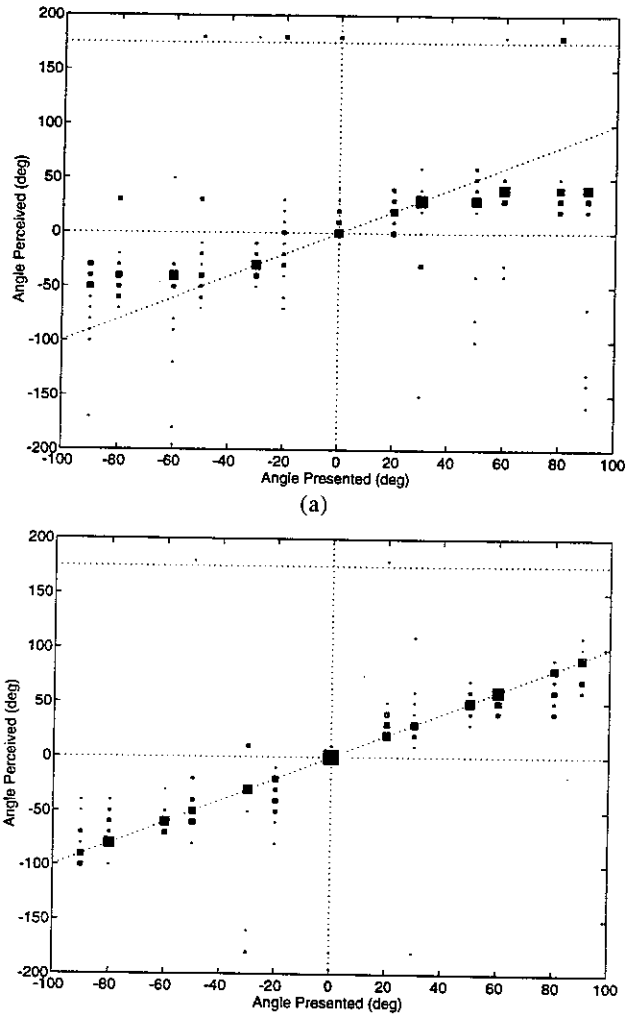


Fig. 16. Results of localization experiments in listening room using 250-Hz one-third-octave filtered white noise signal. (a) Virtual sources. (b) Real sources.

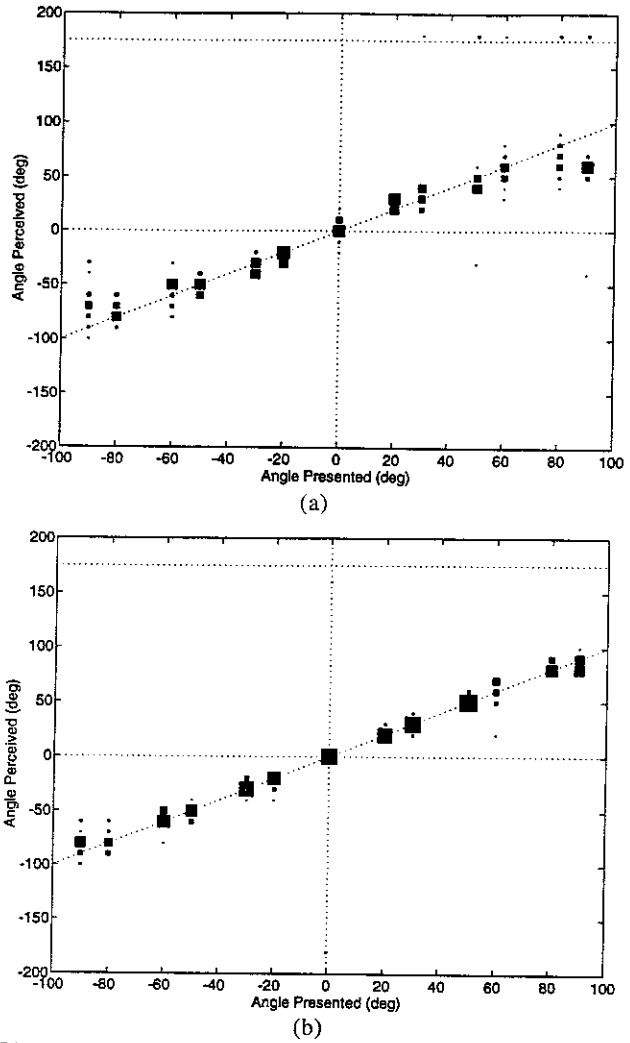
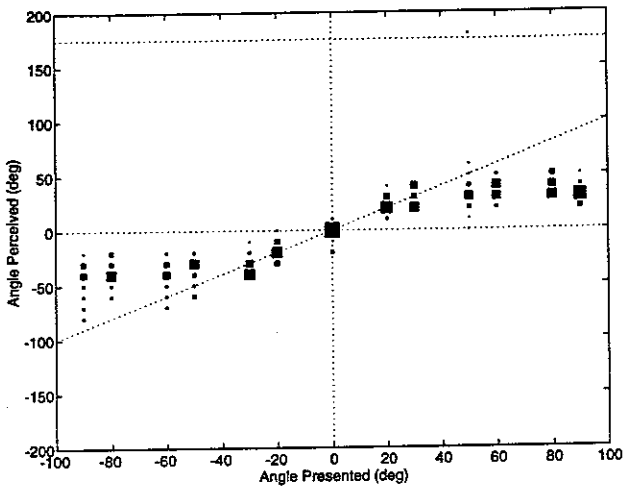
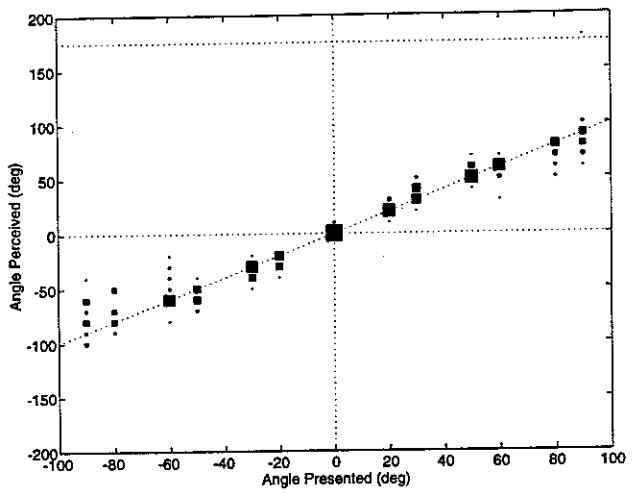


Fig. 17. Results of localization experiments in listening room using 1-kHz one-third-octave filtered white noise signal. (a) Virtual sources. (b) Real sources.



(a)



(b)

Fig. 18. Results of localization experiments in listening room using 4-kHz one-third-octave filtered white noise signal. (a) Virtual sources. (b) Real sources.

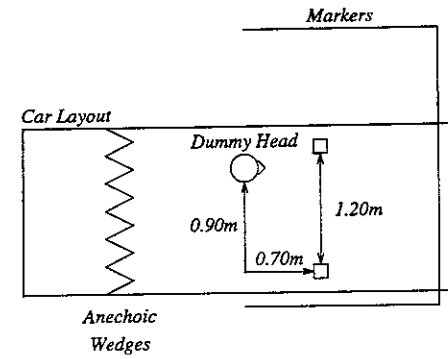
8 REFERENCES

[1] A. D. Blumlein, "Improvement in and Relating to Sound-Transmission, Sound-Recording and Sound-Reproducing Systems," British patent 394325 (1931).

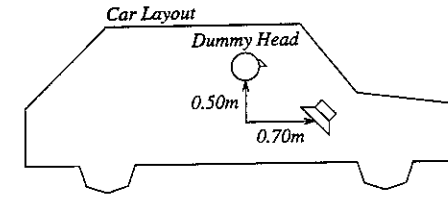
[2] J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, MA, 1983).

[3] B. S. Atal and M. R. Schroeder, "Apparent Sound Source Translator," U.S. patent 3,236,949 (1962).

[4] B. B. Bauer, "Sterephonic Earphones and Binaural Loudspeakers," *J. Audio Eng. Soc.*, vol. 9, pp. 148–151 (1961).



(a)



(b)

Fig. 19. Layout of loudspeakers and position of dummy head in car used for subjective experiments. (a) Top view. (b) Side view.

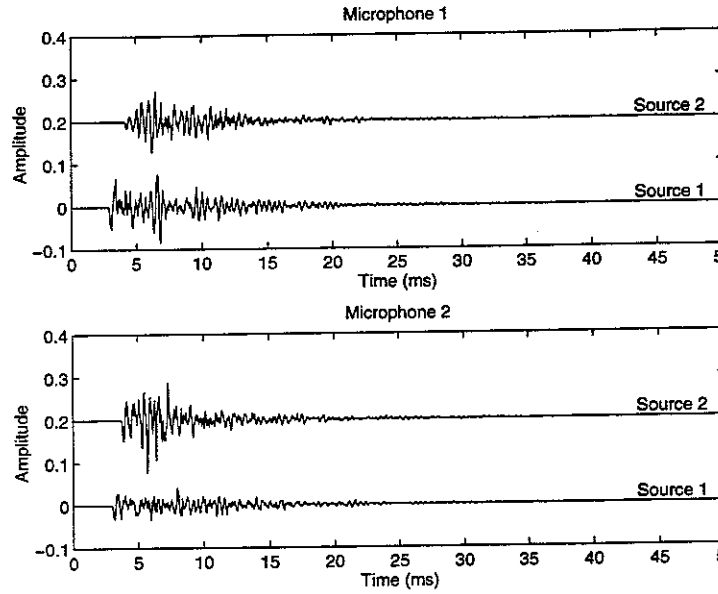


Fig. 20. Impulse responses measured from front pair of loudspeakers in car to microphones at ears of dummy head sitting in driver seat (in left-hand-drive car).

[5] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative Study of European Concert Halls: Correlation of Subjective Preference with Geometric and Acoustic Parameters," *J. Acoust. Soc. Am.*, vol. 56, pp. 1195–1201 (1974).

[6] P. Damaske and V. Mellert, "Sound Reproduction of the Upper Semispace with Directional Fidelity Using Two Loudspeakers" (in German), *Acustica*, vol. 22, pp. 153–162 (1969).

[7] H. Hamada, N. Ikeshoji, Y. Ogura, and T. Miura, "Relation between Physical Characteristics of Orthostereophonic System and Horizontal Plane Localization," *J. Acoust. Soc. Jpn.*, vol. 6, pp. 143–154 (1985).

[8] G. Neu, E. Mommertz, and A. Schmitz, "Investigations on True Directional Sound Reproduction by Playing Head-Referred Recordings over Two Loudspeakers: Part I" (in German), *Acustica*, vol. 76, pp. 183–192 (1992).

[9] G. Urbach, E. Mommertz, and A. Schmitz, "Investigations on the Directional Scattering of Sound Reflections from the Playback of Head-Referred Recordings over Two Loudspeakers: Part II" (in German), *Acustica*, vol. 77, pp. 153–161 (1992).

[10] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.*, vol. 37, pp. 3–19 (1989 Jan./Feb.).

[11] J. Bauck and D. H. Cooper, "Generalized Transaural Stereo," presented at the 93rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 1050 (1992 Dec.), preprint 3401.

[12] H. Møller, "Reproduction of Artificial-Head Recordings through Loudspeakers," *J. Audio Eng. Soc.*, vol. 37, pp. 30–33 (1989 Jan./Feb.).

[13] K. Kotorynski, "Digital Binaural/Stereo Conversion and Crosstalk Cancelling," presented at the 89th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 38, p. 869 (1990 Nov.), preprint 2949.

[14] S. T. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *J. Acoust. Soc. Am.*, vol. 66, pp. 165–169 (1979).

[15] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive Inverse Filters for Sterephonic Sound Reproduction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 40, pp. 1621–1632 (1992).

[16] P. A. Nelson, "Active Control of Acoustic Fields and the Reproduction of Sound," *J. Sound Vibration*, vol. 177, pp. 447–477 (1994).

[17] P. A. Nelson, F. Orduña-Bustamante, and H. Hamada, "Inverse Filter Design and Equalization Zones in Multi-Channel Sound Reproduction," *IEEE Trans. Speech Audio Process.*, vol. 3, pp. 1–8 (1995).

[18] P. A. Nelson, F. Orduña-Bustamante, and H. Hamada, "Multichannel Signal Processing Techniques in the Reproduction of Sound," *J. Audio Eng. Soc.*, this issue, pp. 973–989 (1996 Nov.).

[19] F. Orduña-Bustamante, "Digital Signal Processing for Multi-Channel Sound Reproduction," Ph.D. dissertation, University of Southampton, Southampton, UK (1995).

[20] D. Engler, "Subjective Testing of a Localization System," M.Sc. dissertation, University of Southampton, Southampton, UK (1995).

[21] P. A. Nelson and F. Orduña-Bustamante, "A New Technique for Recording and Reproducing Stereophonic Sound," U.K. patent application (1994).

[22] F. Orduña-Bustamante and P. A. Nelson, "Use of a Crosstalk Cancellation Network as a Generalized Inverse Filter Matrix in Multi-Channel Sound Reproduction Systems," in *Proc. 2nd Int. Conf. on Motion and Vibration Control* (Yokohama, Japan, 1994), vol. 1, pp. 424–427.

[23] D. D. Rife and J. Vanderkooy, "Transfer-Function Measurement with Maximum-Length Sequences," *J. Audio Eng. Soc.*, vol. 37, pp. 419–444 (1989 June).

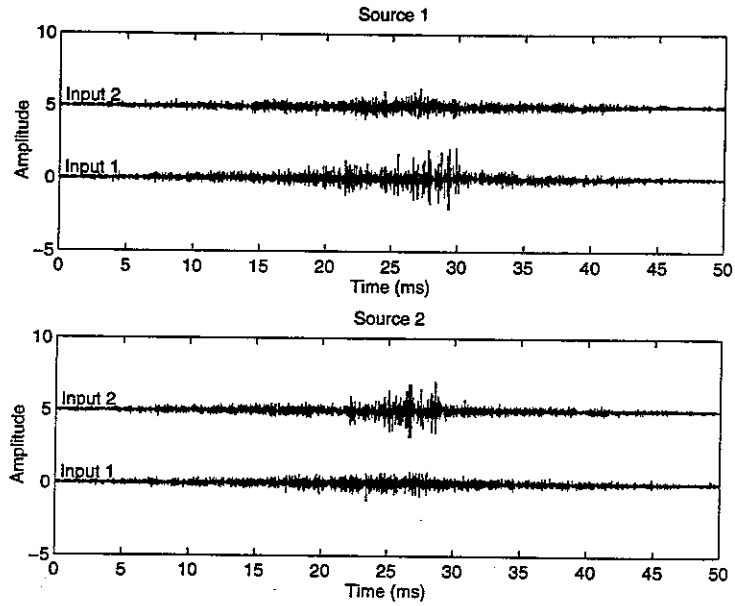


Fig. 21. Impulse response of crosstalk cancellation filters used for in-car experiments.

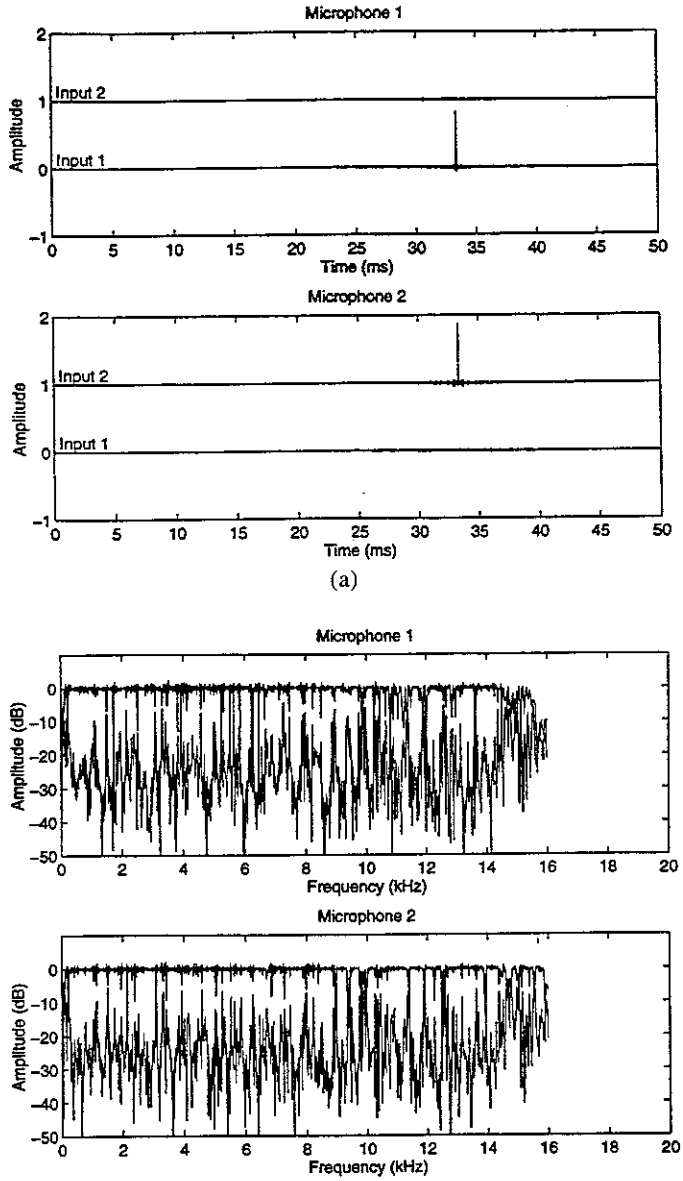


Fig. 22. Matrix of filters resulting from convolution of impulse responses of electroacoustic system in anechoic chamber with matrix of crosstalk cancellation filters (a) in time domain and (b) in frequency domain. (a) Results for impulse responses relating inputs $d_1(n)$ and $d_2(n)$ to outputs $z_1(n)$ (microphone 1) and $z_2(n)$ (microphone 2). (b) Responses at microphone 1 due to inputs $d_1(n)$ and $d_2(n)$. Similarly for the responses at microphone 2.

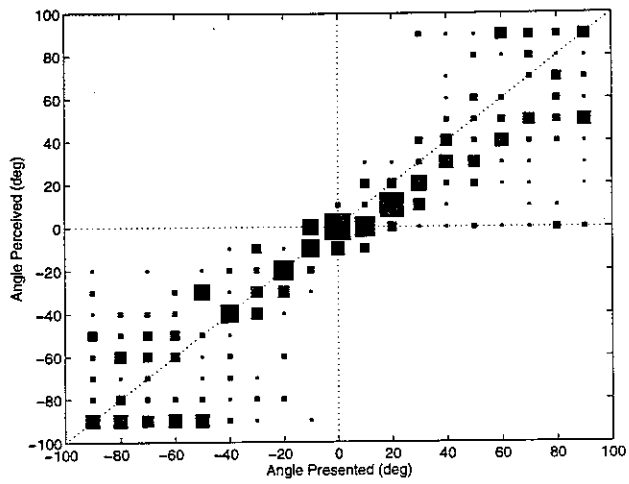


Fig. 23. Subjective evaluation of virtual-source angular location for in-car experiments.

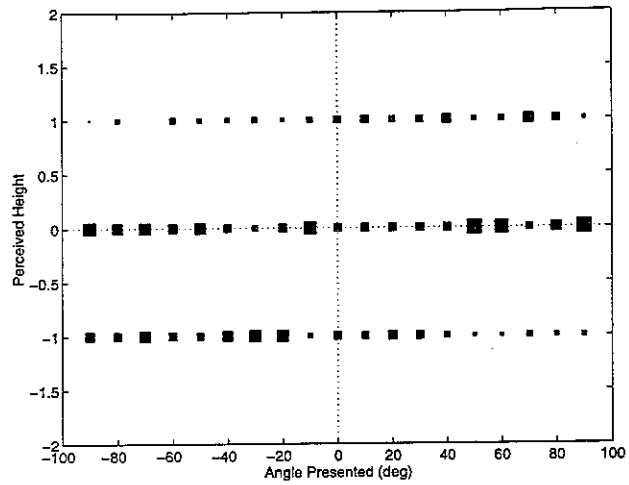


Fig. 24. Subjective evaluation of virtual-source elevation for in-car experiments. Elevation is represented on vertical scale as follows: below = −1, level = 0, above = 1.

THE AUTHORS

David Engler born in Paris, France, in 1970. He graduated in mechanical engineering in 1993 from the Ecole Nationale Supèrieure des Arts et Mètiers in Paris. He worked for one year with Renault on the identification of structure transmission paths of road–tire induced vehicle interior noise. He received the M.Sc. degree in sound and vibration studies in 1995 from the Institute of Sound and Vibration Research.

Biographies for P. A. Nelson, F. Orduña-Bustamante, and H. Hamada are published on p. 989 of this issue.