Binaural technique

Do we need individual recordings?

Møller, Henrik; Sørensen, Michael Friis; Jensen, Clemen Boje; Hammershøi, Dorte

Published in:
Journal of the Audio Engineering Society

Publication date:
1996

Binaural Technique: Do We Need Individual Recordings?*

HENRIK MØLLER, *AES Member,* **MICHAEL FRIIS SØRENSEN,** *AES Member,*
CLEMEN BOJE JENSEN, *AES Member,* **AND DORTE HAMMERSHØI,** *AES Member*

Acoustics Laboratory, Aalborg University, DK-9220 Aalborg Ø, Denmark

The localization performance was studied when subjects listened 1) to a real sound field and 2) to binaural recordings of the same sound field, made a) in their own ears and b) in the ears of other subjects. The binaural recordings were made at the blocked ear canal entrance, and the reproduction was carried out with individually equalized headphones. Eight subjects participated in the experiments, which took place in a standard listening room. Each stimulus (female speech) was emitted from one of 19 loudspeakers, and the subjects were to indicate the perceived sound source. When compared to real life, the localization performance was preserved with individual recordings. Nonindividual recordings resulted in an increased number of errors for the sound sources in the median plane, where movements were seen not only to nearby directions, but also to directions further away, such as confusion between sound sources in front and behind. The number of distance errors increased only slightly with nonindividual recordings. Earlier suggestions that individuals might localize better with recordings from other individuals found no support.

0 INTRODUCTION

The idea behind the binaural technique is the following. The input to the hearing consists of two signals—the sound pressures at each eardrum. If these are recorded in the ears of a listener and reproduced exactly as they were (usually through headphones), then the complete auditory experience is assumed to be reproduced, including spatial aspects such as direction and distance to sound sources.

In most practical applications recordings cannot be made individually for each listener. During recording the listener is replaced by another person or, more often, by an acoustical mannequin, called an artificial head. As the transfer characteristics from a sound field to the eardrums differ between individuals, the use of nonindividual recordings will introduce an error. It is the purpose of the present investigation to explore the effect of this error on localization performance, when recordings are made in a natural sound field with reflections and reverberation.

* Manuscript received 1995 November 4; revised 1996 April 2.

0.1 Limitations of the Investigation

Only recordings from other real persons are considered. Artificial heads are disregarded, since they would ideally be "typical" subjects and thus, in principle, be included in the "other person" concept. And if the artificial heads are not ideal, their "quality"—the extent to which they replicate humans—will introduce an additional and uncontrolled variable. The quality of commercially available artificial heads is the subject of another investigation at our laboratory [1].

Head movements are known to affect localization performance in real life by resolving, for instance, front–back ambiguities (see, for example, Blauert [2, sec. 2.5]). For any practical purpose the orientation of the recording head cannot be controlled subsequently by the listener in the playback situation, and only recordings with a stationary head are considered in the present investigation.

Binaural signals need not originate in a recording with a physical head. It is possible to synthesize binaural signals on a computer by filtering with head-related transfer functions (HRTFs) corresponding to the directions of incidence of the relevant sound waves. Most often the HRTFs will originate in measurements from a person other than the listener or even from an artificial head. Thus the question of nonindividual binaural signals is important in binaural synthesis, too.

If binaural synthesis should simulate anything more complicated than a free field without reflections, it must rely on a room simulation program to specify the incoming sound waves. Relying on a room simulation program would introduce an additional and uncontrolled variable, and synthesized binaural signals are therefore not included in the present investigation. Binaural free-field synthesis with individual HRTFs is treated in a parallel work at our laboratory [3], [4].

0.2 Previous Investigations

It seems to be a general understanding that the use of nonindividual binaural recordings results in localization errors such as front–back confusions, elevations, and in-the-head localization. However, most of these observations seem to originate in investigations with artificial-head recordings or with synthesized binaural signals. Only few investigations have been made with human-head recordings in a real sound field with reflections and reverberation. This summary of previous investigations is not restricted to those dealing with individual versus nonindividual recordings, but it covers in general reproduction of recordings made in the ears of human subjects.

Laws and Platte [5] performed an experiment using recording and playback techniques, as described by Platte et al. [6]. Eight subjects listened to recordings made with probe microphones 4 mm inside the ear canals of a selected "typical" subject. The signals were adequately equalized for the frequency responses of the probe microphones and the headphones (Sennheiser HD 414). The headphone frequency response was taken as the mean of the response measured 4 mm inside the ear canal on many subjects. The exact positioning of the headphone was considered critical, and prior to each measurement the subject adjusted the position of the headphone to obtain the brightest sound from a 6–9-kHz noise applied to the headphone.

Sound sources were located in 12 directions equally spaced in the horizontal plane. The localization performance was slightly poorer than the subjects' localization performance in the real-life situation observed in the same setup. Especially front-to-back reversals occurred, and a higher rate of in-the-head localization was seen for the frontal direction.

A second experiment was carried out with the same recordings and the same equalization, but now the headphones were positioned more accurately before reproduction using the same procedure that was used prior to measurements of headphone transfer functions. Results with four subjects showed almost the same performance as in real life. No front-to-back reversals occurred now, and the rate of in-the-head localization for the frontal direction was reduced.

In a study by Searle et al. [7] five subjects listened to recordings made in their own ears by means of miniature microphones placed approximately 10 mm inside the ear canals. Five directions in the median plane were included (elevations 0°, 45°, 90°, 135°, 180°). No report was given about equalization of the headphone (Koss ESP-9). Four subjects had external localization, but not until after a few minutes of training. One subject could neither externalize nor localize. The study did not include a reference experiment, but it was noted that the performance of the four subjects corresponded to literature reports of real-life localization.

In a second experiment the recording from one of each subject's own ears (left or right) was presented to both of his or her ears. For four of the five subjects this resulted in an increased error score, a finding that is claimed to support a theory that pinna asymmetry plays a role in median-plane localization.

In a study by Butler and Belendiuk [8] eight subjects listened to recordings made in their own ears with microphones "embedded in ear defenders" and inserted into the ear canals. The recordings were made for five directions in the median plane (frontal sound incidence, elevations 0°, ±15°, and ±30°). No report was given about equalization of the headphone (AKG K180). It was found that the localization performance was only slightly degraded compared to real-life performance obtained in the same setup.

A second experiment was carried out where four subjects listened to their own recordings and the recordings of the three others. This experiment showed that two subjects had the best localization performance with their own recordings, and two subjects performed best with recordings from another person (each performing best with one of the former two). This finding was interpreted to mean that the pinnae of some persons provide more adequate cues for median-plane localization than do the pinnae of others.

Weinrich [9] made an experiment where four subjects listened to recordings made in their own ears by means of miniature microphones located approximately 5 mm inside the ear canal. Recordings were made for four directions in the horizontal plane (azimuths ±30° and ±150°). It was understood that in principle the headphone (Beyer DT 440) needed compensation, but referring to undescribed pilot experiments this was claimed not to be necessary, and no equalization was applied. It was reported that no errors were made during listening to binaural recordings. No errors were also reported for an informal test of the subjects' localization in the real-life situation using the same setup.

The investigations cited here differ very much with respect to recording technique and headphone equalization, directions included, the subjects' way of responding, statistics reported, and so on. It is difficult if not impossible to compare the results, and we do not find it adequate to make any conclusions at this point.

0.3 Aim and Strategy of the Investigation

It is the aim of the investigation to answer the following questions:

1) How flawless a reproduction can be obtained with individual recordings, that is, recordings made in the

listener's own ears? (Can we make a true copy of the real-life experience?)

2) What happens if we use nonindividual recordings, that is, recordings made in the ears of a person other than the listener? (Does the quality degrade? Are some ears better than others? Better than our own?)

3) If nonindividual recordings result in degraded quality, do we then adapt to the ears of a specific other person?

The quality of the reproduction will be evaluated in terms of localization performance in a normal room. The experiments will be of the identification type, where the "resolution" in responses is identical to that of the stimuli. Sound source positions will include various distances and various directions in the horizontal, the frontal, and the median planes. The questions will be answered by comparing performance in real life with individual binaural recordings and with nonindividual recordings.

0.4 Calibration Methods

For individual binaural recordings, that is, when the recordings are made in the listener's own ears, the total transmission should include a compensation for the transfer functions of 1) the microphone and 2) the headphone measured at the point in the ear canal where the recording is made. This seems intuitively correct, and it was shown formally by Møller [10] that this leads to the correct sound at the eardrums. The determination of this correction (and its realization) is called transmission calibration or just calibration (or headphone equalization, since if the recording microphone has a flat frequency response, the calibration simply equalizes for the headphone transfer function).

In [10] it was also shown that recordings may be made at the entrance to the blocked ear canal since full spatial information is present at this point (further supported by Hammershøi and Møller [11]). When the recording is made at the blocked ear canal, the transmission calibration should in principle include an extra term to account for the difference in pressure divisions at the entrance to the ear canal in two situations, 1) when the ear is in free air and 2) when the headphone is placed over the ear [10]. However, we have seen that for many headphones this correction is minimal and can be disregarded. Headphones for which this is the case are called FEC headphones (headphones with *f*ree air *e*quivalent *c*oupling to the ear) [12].

Up to this point we have used the terms *individual* and *nonindividual* recordings without defining them precisely. When we try to define them more accurately, it turns out that the concept of individual and nonindividual recordings is to some extent ambiguous. The binaural transmission basically includes three procedures: 1) recording, 2) transmission calibration, and 3) reproduction. It is evident that each of these steps can be carried out in its own way. The only step that, for natural reasons, is always carried out using the final listener, is the reproduction. The recording and the transmission calibration (including determination of the headphone transfer function) may be carried out using the same listener, another human, or an artificial head. Even literature data may be used, such as for HRTFs in computer simulation of a binaural recording (binaural synthesis or binaural auralization).

The Appendix presents an analysis of errors due to nonindividual recording and calibration for various recording points. It is shown that the smallest error is obtained with the following procedure, which is therefore chosen in the present investigation: 1) recordings are made at the blocked ear canal entrance, 2) an FEC headphone is used, and 3) individual headphone equalization (namely, following the listener) is used for individual as well as nonindividual recordings. The equalization should be [from Eq. (29)]

$$G = \frac{1}{M \cdot \mathrm{PTF}} \tag{1}$$

where M is the transfer function of the recording microphone and PTF (head*p*hone *t*ransfer *f*unction) is the electroacoustical transfer function of the headphone from voltage at the terminals to sound pressure at the blocked ear canal entrance.

1 METHOD

By means of psychoacoustic experiments the localization performance was evaluated in real life and when listening to binaural recordings of the same sound field. The experiments included four listening conditions: 1) subjects listening in real life, 2) subjects listening to recordings made in their own ears, 3) subjects listening to a mixture of recordings from four subjects including themselves, and 4) each subject listening to recordings from one other subject. The experimental procedure was the same in all situations, including that the subjects were sitting in the real-life setup also when listening to the recordings.

1.1 Subjects

Eight paid students with controlled normal hearing participated, four of each sex, aged 20–30 years. They were all skilled in psychoacoustic esperiments, but they were not in any way selected for their hearing or localization proficiency.

1.2 Listening Setup

The listening setup was an approximate copy of the setup used by Nielsen in his investigation of human distance perception [13], [14]. In the present investigation, though, the loudspeakers were uncovered and thus visible to the subjects.

The setup consisted of 19 small loudspeakers (70-mm-diaphragm Vifa M10MD-39, mounted in 155-mm-diameter hard plastic ball) in a standard listening room with a reverberation time of approximately 0.4 s (IEC 268-13 [15] (used without a carpet); for the specific room, see Langvad et al. [16]). Fig. 1 shows the setup.

Fourteen loudspeakers were placed at a distance of 1 m from the center of the subject's head (that is, on a sphere). In FRONT, LEFT, BACK, and RIGHT (azimuths 0°, 90°, 180°, and −90°) loudspeakers were placed 45° below the horizontal plane (LOW), in the horizontal plane, and 45° above the horizontal plane (HIGH) (elevations −45°, 0°, and 45°). (The designations in small caps will be used in the following for identification). One loudspeaker was placed right ABOVE the subject and one at −45° azimuth in the horizontal plane. The remaining five loudspeakers were placed at other distances, all in the horizontal plane—in FRONT at distances of 1.7, 2.9, and 5.0 m, and at −45° at distances of 1.7 and 2.9 m. For the two directions where loudspeakers were behind each other (FRONT and −45°), the loudspeakers were slightly displaced vertically by less than the minimum audible angle in order to reduce disturbance of the direct sound from the more distant loudspeakers [13], [14].

The loudspeakers were equalized by means of a graphic equalizer (Klark-Teknik DN 30/30) in order to maintain an approximately flat free-field response (Fig. 2). The equalization was the same for all loudspeakers, and control measurements did not reveal any considerable variation between the responses. The subject was seated in a chair, which was adjusted vertically and horizontally so that the center of his or her head was at the correct position at the center of the loudspeaker sphere.

A digital recording and playback system using a sampling frequency of 48 kHz formed the core of the electrical setup. The system used a PC with a large hard disk and a floating-point signal processor system (TMS320C30 processor board and Pro-Audio interface board, Loughborough Sound Images Limited). The digital output was sent to a digital preamplifier (Roland E 660, equalizer in flat mode) and a power amplifier (Pioneer A-616, modified to calibrated gain setting). A relay switch box controlled by the computer directed the signal to one of the loudspeakers. Switch noise from the relays was inaudible for the subjects.

The source signal was a 5-s recording of a female voice. The recording was made in an anechoic chamber [13], and the total transmission was calibrated to give the same free-field on-axis output from the loudspeakers as was present from the talker during recording. When the differences in the directional patterns of a human talker and the loudspeakers are disregarded, the setup replicates the situation where the human talker is placed at each loudspeaker position, facing the listener.

1.3 Binaural Recordings

Binaural recordings were made at the blocked ear canals of each subject, while the subject was seated in the listening setup. The microphones (Sennheiser KE 4-211-2) were mounted in an ear plug, with the end of the earplug and the microphone being flush with the ear canal entrance. Details of the microphone technique are described in [17] (including the frequency responses of

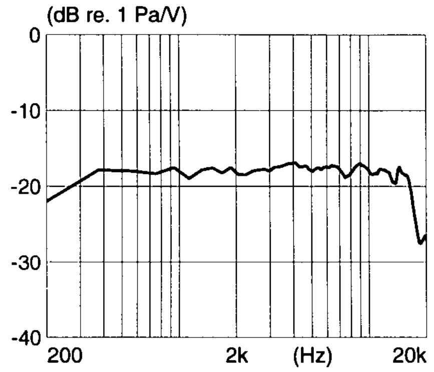


Fig. 2. Free-field frequency response for sample loudspeaker with equalization used.

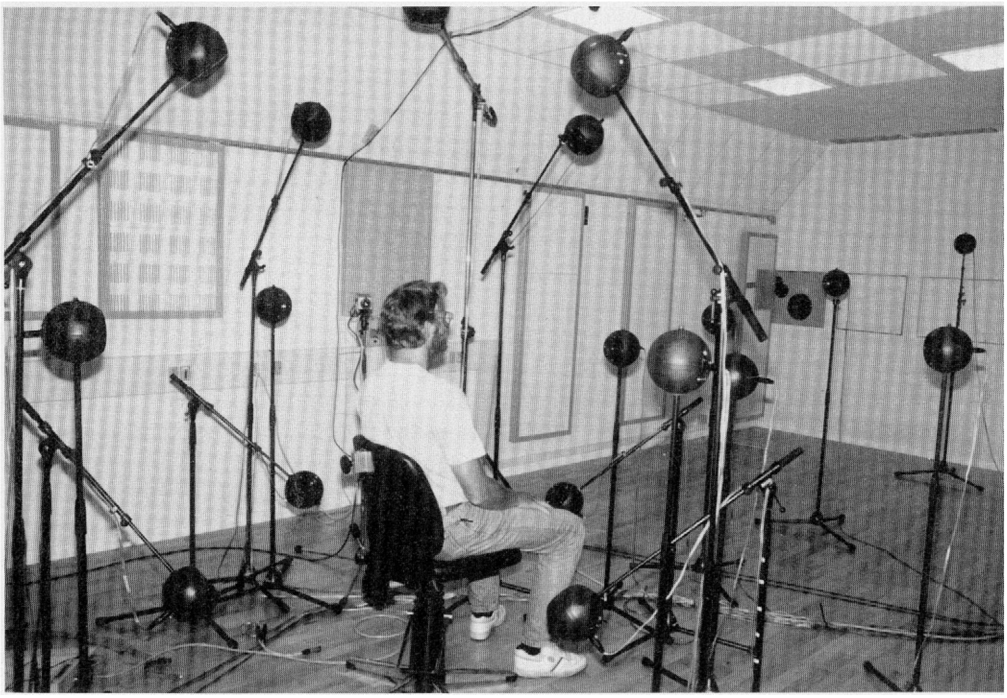


Fig. 1. Setup in listening room with 19 loudspeakers in different directions and at various distances.

the very two microphones used in this study for recordings [17, fig. 4]). Recordings were made directly to disk for each of the loudspeakers using a Rostec LMA 4 preamplifier, the analog-to-digital converter of a Panasonic SV-3700 DAT recorder, and AES/EBU digital transmission to the Pro-Audio board. In order to improve the signal-to-noise ratio on the recordings, the playback level from the loudspeakers was increased by 10 dB relative to the true level used in the real-life listening experiment. The total gain was corrected accordingly. In this way the noise at playback was reduced to an unnoticeable level.

1.4 Binaural Reproduction

A Beyerdynamic DT 990 professional headphone was chosen for reproduction of the binaural signals. We showed earlier that this headphone has approximate FEC properties [12]. Individual equalization filters were implemented on the signal processor board during reproduction of the individual as well as the nonindividual recordings.

The filter transfer function should ideally be the reciprocal of the PTF. The filter design was based on five repeated measurements of PTFs on each subject and each ear. (Measurements of PTFs are described in detail in [12].) The headphone was positioned by the subject to give best comfort, and it was replaced between measurements. Fig. 3 shows an example of the five measured PTFs for both ears of a subject. The repeatability is seen to be excellent, and this was the case for most of the subjects. The very same microphones were used for recording and for measuring PTFs. Nothing was therefore done to correct for their frequency responses, as they cancel out by the calibration procedure [10, sec. 4.5].

The target frequency response for each filter was obtained in the following way. The results of the five repeated PTF measurements were averaged on a decibel basis, frequency by frequency, and the result was inverted. A minimum-phase IIR-filter approximation of the target amplitude response was made with a Yule–Walker design procedure (maximum 32nd order, carried out in MATLAB, The MathWorks Inc.). The measured

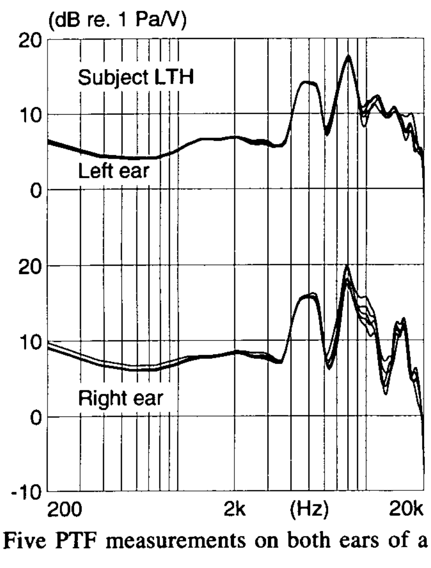


Fig. 3. Five PTF measurements on both ears of a subject.

PTFs were in general not minimum-phase transfer functions, and the resulting total transmission was thus only correct with regard to the amplitude response, since the inverted all-pass sections from the PTFs were not accounted for.

A few modifications of the target responses were necessary to stabilize and optimize the digital filters. Lack of dc transmission in the headphone transmission would result in an infinite dc gain of the target response. Dc values were inserted manually in the target response to obtain a fairly flat frequency response at low frequencies. A corresponding problem exists close to the half sampling frequency, where the target responses were manually flattened. Manual modification was also needed where the amplitude response of the headphone had a narrow dip, resulting in a very high and sharp peak in the target response. The gain of such peaks was limited manually in a few cases. To give better noise performance the filters were split into second-order sections by pairing poles and zeros. Fig. 4 shows the target responses and the designed filter characteristics for both ears of all subjects.

1.5 Experimental Procedure

The experiments were controlled by the computer. For each stimulus the subject responded from which loudspeaker he or she perceived the sound. The answers were collected using a digitizer (Océ G6421) held by the subject on his or her lap. Fig. 5 illustrates this arrangement. The subject selected a loudspeaker by pressing the digitizer pen in the appropriate zone (Fig. 6). Answers were not accepted during stimulus, and the experiment was not continued until a response had occurred. The subjects were not given any feedback in terms of correct or incorrect responses. Before the experiments, all subjects received a short training session, including control of the correct use of the digitizer.

The subjects were instructed to keep their heads still in a natural upright position and to look straight ahead during stimuli. A warning was displayed on the answering board just before each stimulus, but the subjects found very soon a rhythm where they held their heads correctly immediately after each answer. Cameras were placed to the left of and above the subject to monitor position. The experiment was halted if the experimenter observed head movements during a stimulus. Then the subject was reminded about the instruction before the experiment was continued. This control was the only task for the experimenter during sessions, and the situation occurred very rarely.

1.6 Experimental Design

Four different experiments were carried out with each subject on five separate days.

1.6.1 Experiment A: Real Life

Each subject listened to each loudspeaker six times. The experiment was divided into two sessions with three repetitions each. The stimulus order was random in each session. The sessions had a duration of approximately

10 min, and they were separated by short coffee breaks. The number of stimuli for each subject was 114, giving a total of 912 for the eight subjects.

1.6.2 Experiment B: Individual Binaural

The design was the same as for experiment A, except that the stimuli were the subject's own recordings presented via headphones.

1.6.3 Experiment C: Mixed Individual and Three Nonindividual Binaural

Each subject listened to recordings of his or her own and of three other subjects. The other subjects were in principle selected at random but balanced in a systematic rotation, so that the recordings of each subject were used for three other subjects. The experiment was divided into six sessions, each holding a full randomized set of 19 loudspeakers and four "recording heads." The sessions had a duration of 12–13 min, and only three sessions, separated by short breaks, were carried out on one day. The number of stimuli for each subject was 456, giving a total of 3648 for the eight subjects.

1.6.4 Experiment D: Nonindividual Binaural

Each subject listened to recordings of one other subject. The other subject was selected from the "three other" in experiment C in a systematic rotation, so that the recordings of each subject were used for one other subject. In other respects, the design was as in experiments A and B.

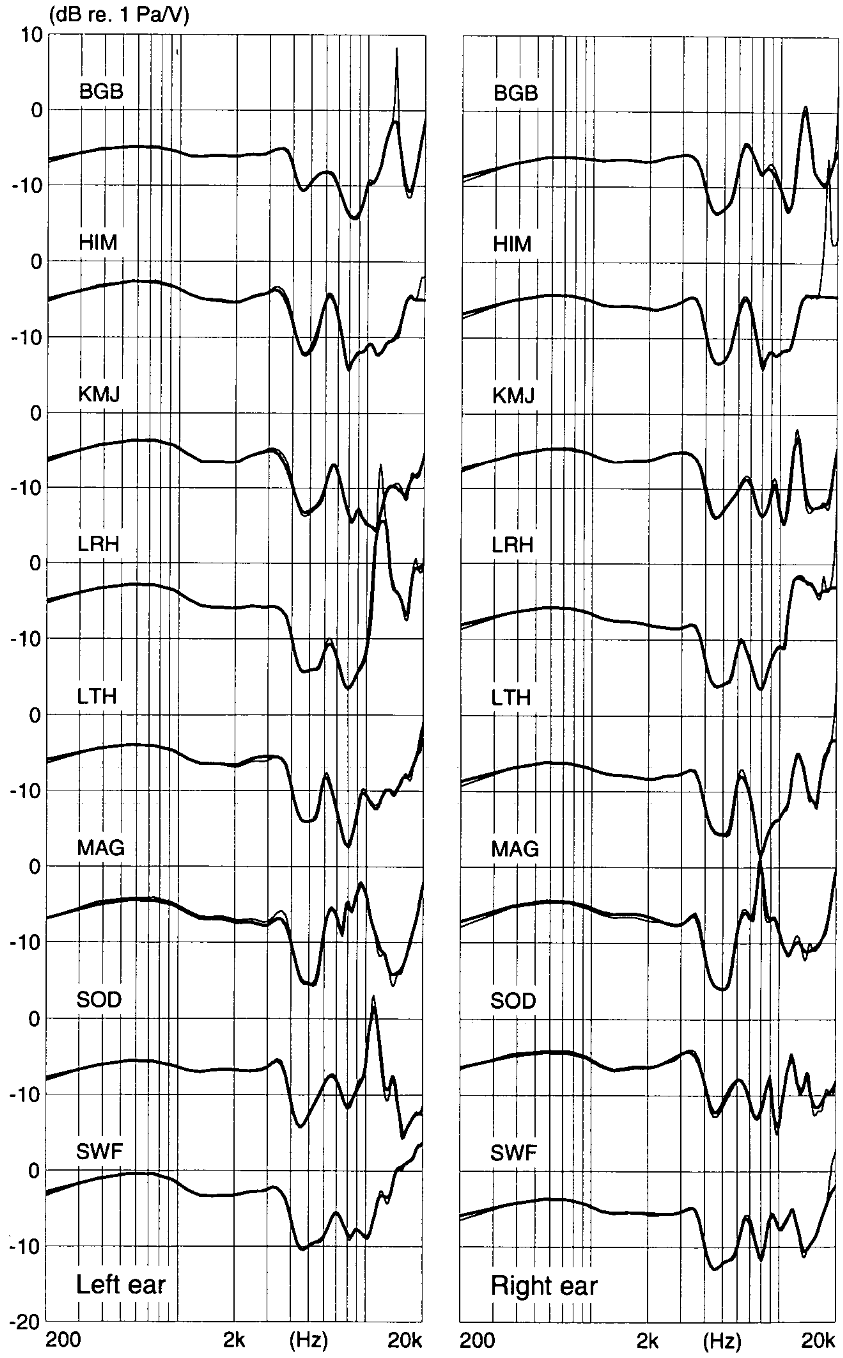


Fig. 4. Design of headphone equalizing filters for both ears of eight subjects. Target responses (inverse of measured PTFs) are shown in thin line; designed filter characteristics are shown in heavy line.

2 RESULTS AND DISCUSSION

The method of presenting the results is most easily shown by way of an example, and Section 2.1 presents the results of the real-life listening experiment. In subsequent sections the performances under different experimental conditions will be compared, and for that reason Section 2.2 introduces error categories and gives appropriate reference to the statistical method. Results for individual recordings are presented and discussed in Section 2.3, whereas Section 2.4 covers nonindividual recordings. Some general comments are given in Section 2.5.

2.1 Real-Life Listening

The results from the real-life listening experiment are shown in Fig. 7(a) in a 19 × 19 matrix with the stimulus position as abscissa and the responded position as ordinate. The black circles represent answers, and the area of each circle is proportional to the number of answers for the particular combination of stimulus and response. A full-size circle (as seen, for example, at the (LEFT, LEFT) position) corresponds to the number of stimuli given for each direction (in this case 48). The smallest circles correspond in this case to one answer (as seen, for example, at the (LEFT HIGH, LEFT) position).

Correct answers are found at the diagonal, and it is encouraging to see that most of the responses are indeed seen here. However, it is also obvious that the subjects do not localize sound sources perfectly. The major part of the errors are seen for sources in the median plane. Directions in the upper median plane (FRONT HIGH, ABOVE, and BACK HIGH) are often confused, and sound coming from FRONT LOW and BACK LOW are frequently perceived at various other directions in the median plane. Also wrong judgment of distance is a common error.

Difficulties in median-plane localization are well known from the literature, and the lack of interaural

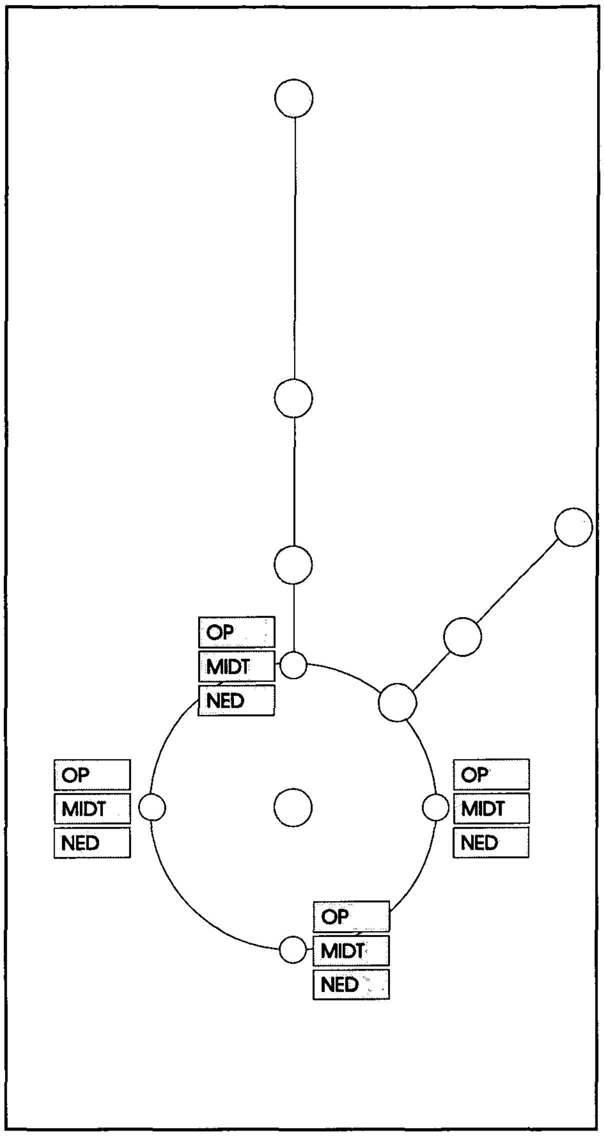


Fig. 6. Sketch of loudspeaker setup, which was attached to digitizer plate. Responses were given by pressing one of the grey zones with digitizer pen. Grey zone at position of subject indicates loudspeaker ABOVE subject.

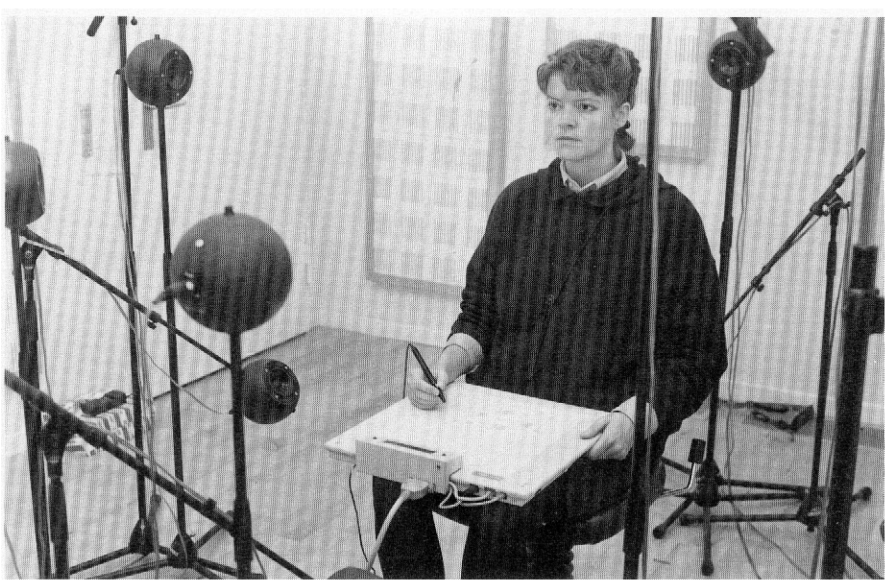


Fig. 5. Subject in setup, ready to answer with digitizer.

differences is assumed responsible (see, for example, Blauert [2, secs. 2.1 and 2.3]). It is worth noting that the upper median plane is a region where the HRTFs vary only little between directions (see, for example, Møller et al. [17, middle column of fig. 13]).

2.2 Error Categories and Statistical Tests

The number of stimuli for each loudspeaker position is assumed to be too low to allow statistical testing of individual directions with adequate power. The errors are therefore classified into four groups that have been chosen with an eye to the localization cues that are available for the hearing.

Interaural time differences are considered of great importance in human sound localization. Sound sources that produce the same interaural time difference are positioned on an approximate cone, called a *cone of confusion* (see, for example, [18, p. 352]). As far as the direct sound is concerned, our experiments include stimuli at five such cones. Five sound sources at the right side [RIGHT LOW, −45° (three distances), and RIGHT HIGH] are positioned at one cone of confusion, and two sources at the left side (LEFT LOW and LEFT HIGH) lie on another. The median plane, where 10 of our sound sources are located, constitutes a special "cone" of confusion (with no interaural time difference). Two other special "cones" with only one possible direction each (in our experiment and in general) are the directions LEFT and RIGHT.

If a response is given at another cone of confusion than where the stimulus was given, it is denoted an *out-of-cone* error. A response at the correct cone but at an incorrect direction is denoted a *within-cone* error, except when stimulus and response are in the median plane, in which case it is a *median-plane* error. The reason for making a special group for median-plane errors is that for these stimuli no interaural differences exist. Thus directions within the group are discriminated exclusively by means of monaural cues (as far as the direct sound is concerned). A response given in the same direction as the stimulus, but at an incorrect distance, is denoted a *distance* error. Fig. 7(b) shows the stimulus–response matrix divided into error categories.

With the present experimental design the number of errors in a certain category will follow a binomial distribution. The null hypothesis assumes that the error probability is the same for the two conditions under test. The required test function follows a hypergeometrical distribution, and the test is called a Fisher–Irwin test (see, for example, [19]). One-sided tests are used whenever the sign of a possible difference can be anticipated. In order to give the most powerful test only stimuli that actually can lead to errors in a certain category are included in each test and in the calculation of error percentages.

2.3 Individual Recordings

Results for individual recordings are available from two experiments—experiment B, where only individual recordings were reproduced in a session, and experiment C, where individual recordings were mixed into sessions that also included nonindividual recordings.

2.3.1 Separate Sessions

The results from individual recordings, presented in separate sessions, are shown in Fig. 8(a). The overall impression is that the type and number of errors are comparable to those of the real-life experiment [Fig. 7(a)]. Errors at some stimulus–response combinations have decreased, whereas errors at other combinations have increased. In any case, differences are small.

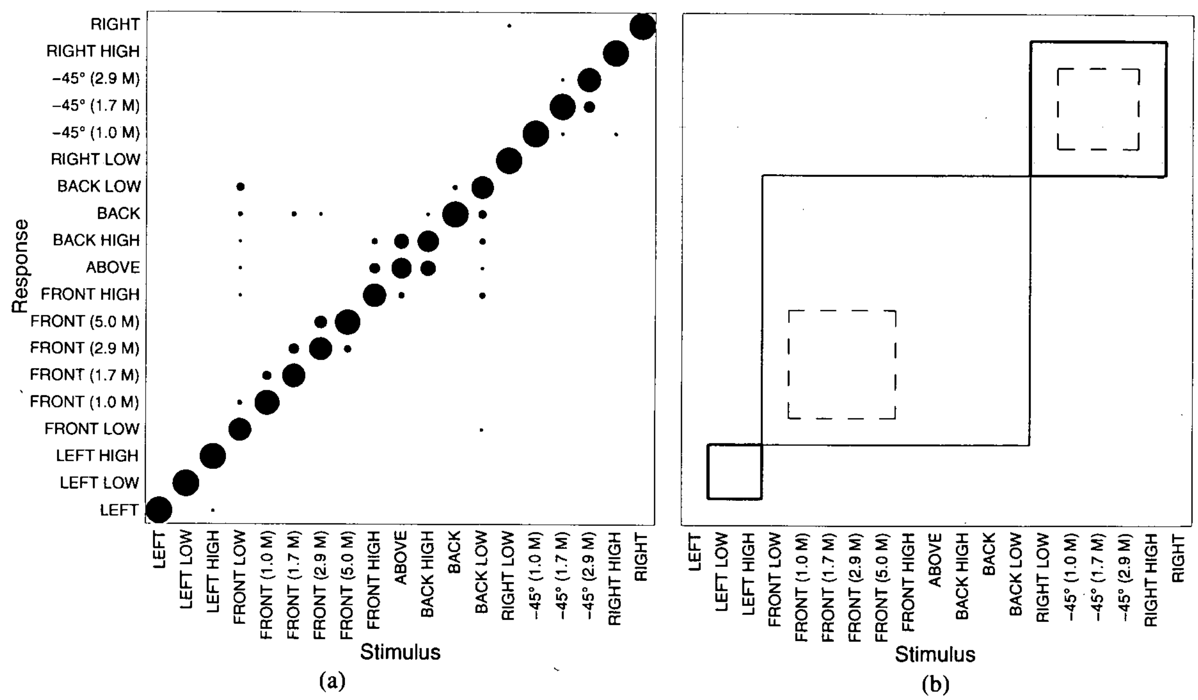


Fig. 7. (a) Result of real-life listening (experiment A, 912 responses). (b) Classification of errors. Correct answers are at diagonal. Distance errors are framed by dashed thin lines, median-plane errors by thin lines (exclusive of distance errors), and within-cone errors by thick lines (exclusive of distance errors). Remaining positions correspond to out-of-cone errors.

Table 1 shows a comparison of the four error categories for real-life and for individual recordings. Considerable numbers of median-plane errors and distance errors are seen, whereas almost no out-of-cone and within-cone errors are observed. The figures seem very comparable in the two situations. The statistical tests show no difference for three error categories but a difference that is significant at the 5% level for within-cone errors. Only a few errors are seen in this category, and in view of the data shown subsequently for individual recordings presented in mixed sessions, we believe, despite the significance, that the difference is a coincidence. It is concluded that the localization performance with the individual recordings is the same as in the real-life situation. As a consequence, in subsequent tests the results of experiments A and B will be pooled in order to improve the power of the test.

2.3.2 Mixed Sessions

The results from individual recordings, presented in mixed sessions, are given in Fig. 8(b). The immediate impression is that the type and the number of errors are comparable to those obtained in real life [Fig. 7(a)] and with individual recordings presented in separate sessions [Fig. 8(a)].

Table 2 shows a comparison of the results using individual recordings presented in mixed sessions and the pooled results of real-life and individual recordings presented in separate sessions. The figures seem very comparable, and the statistical tests showed no significant differences. As a consequence, in subsequent tests the results of the individual part of experiment C will be pooled with the results of experiments A and B. The pooled result matrix is shown in Fig. 9(a).

2.3.3 Discussion

Our experiments have shown that individual binaural recordings are capable of giving an authentic reproduction for which the localization performance is equal to that of real life. Our findings are in contrast to the results of Searle et al. [7]. One of their five subjects had in-the-head perception and could not localize at all with

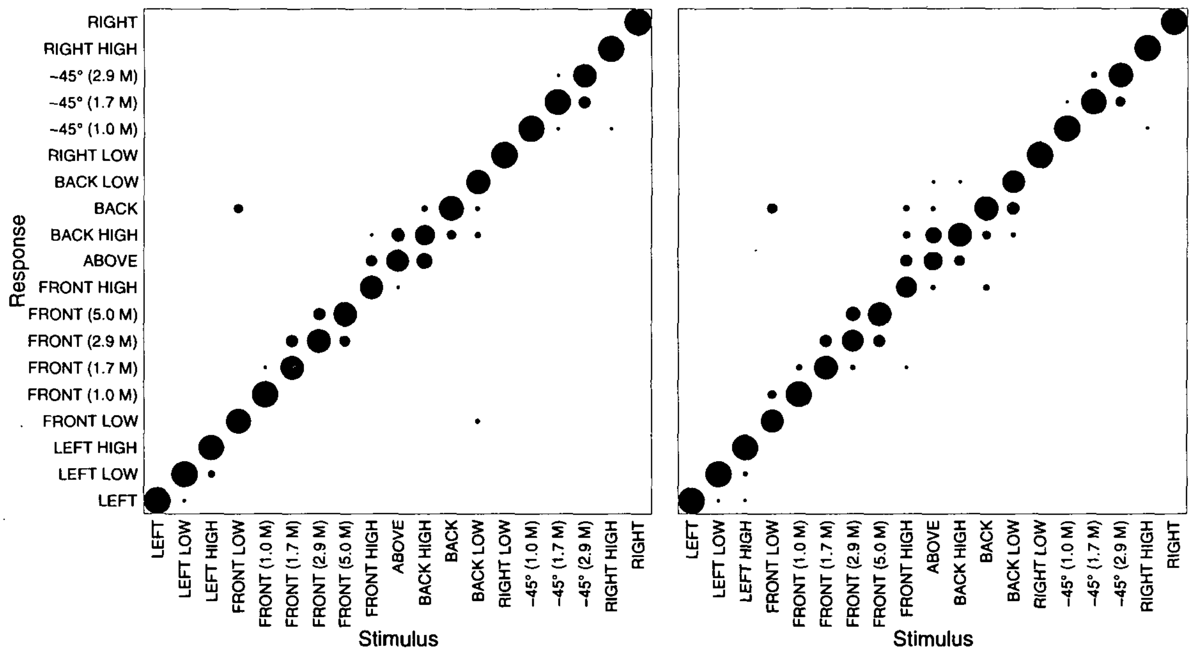


Fig. 8. Individual binaural recordings. (a) In separate sessions (experiment B, 912 responses). (b) In mixed sessions (part of experiment C, 912 responses).

Table 1. Comparison of individual binaural recordings (experiment B) and real life (experiment A). Errors are given as numbers and in percent of the number of stimuli (in parentheses) that can result in errors in the category.

	Errors				
Condition	Out of Cone	Within Cone	Median Plane	Distance	Total Number of Stimuli
	0.2%	0.3%	16.0%	11.9%	
Real life	2	1	77	40	912
(experiment A)	(912)	(336)	(480)	(336)	
	0.1%	1.5%*	13.1%	12.2%	
Individual binaural	1	5	63	41	912
(experiment B)	(912)	(336)	(480)	(336)	

* Significance at 5% level in one-sided Fisher–Irwin test.

individual binaural recordings. The remaining four were able to have out-of-head perception and to localize, but not until after a few minutes of training. Butler and Belendiuk [8] found only slightly degraded localization performance compared to real life; thus their results are closer to ours.

Weinrich [9] reported no errors in real life nor with binaural recordings. Thus he shares the similarity between the two conditions with us. However, he only included directions outside the median plane (and only four), and it is questionable whether his experimental design was sensitive to errors at all.

With respect to the perceived quality of our recordings, it is worth mentioning some spontaneous comments from the subjects. Most of them expressed their surprise at the realism of the headphone reproduction, and a few subjects more or less accused us of cheating. Why should they wear headphones when the sound came from the loudspeakers?

We used recordings made at the entrance to the blocked ear canal and FEC headphones calibrated accordingly. The results confirm that this technique is suitable for binaural recording. It is advantageous from a practical point of view. Larger microphones can be used, thus allowing a better signal-to-noise ratio, and insertion of microphones deeply into the ear canal is avoided.

Localization in real life (and with individual binaural recordings) is not perfect. Median-plane errors occur, and especially the upper directions are confused. Also distance errors occur, whereas within-cone and out-of-cone errors are rare.

Confusions between front and back are among the median-plane errors, but they are not frequent. In experiments A and B and the individual part of experiment C, the three frontal sound sources at 1-m distance were perceived behind the frontal plane 7% of the times they were presented, whereas the corresponding three sound sources in the back were perceived in front of the frontal plane in only 2% of the cases.

Sound sources in the FRONT direction are almost al-

Table 2. Comparison of individual binaural recordings presented in mixed sessions (part of experiment C) and pooled results of real-life and individual binaural recordings presented in separate sessions (experiments A and B). Errors are given as numbers and in percent of the number of stimuli (in parentheses) that can result in errors in the category.*

Condition	Errors: Out of Cone	Within Cone	Median Plane	Distance	Total Number of Stimuli
Real life and individual (experiments A and B)	0.2% 3 (1824)	0.9% 6 (672)	14.6% 140 (960)	12.1% 81 (672)	1824
Individual, mixed (part of experiment C)	0.2% 2 (912)	0.9% 3 (336)	17.3% 83 (480)	14.9% 50 (336)	912

* Statistical tests did not show any significant difference between the two conditions (one-sided Fisher–Irwin test at 5% level).

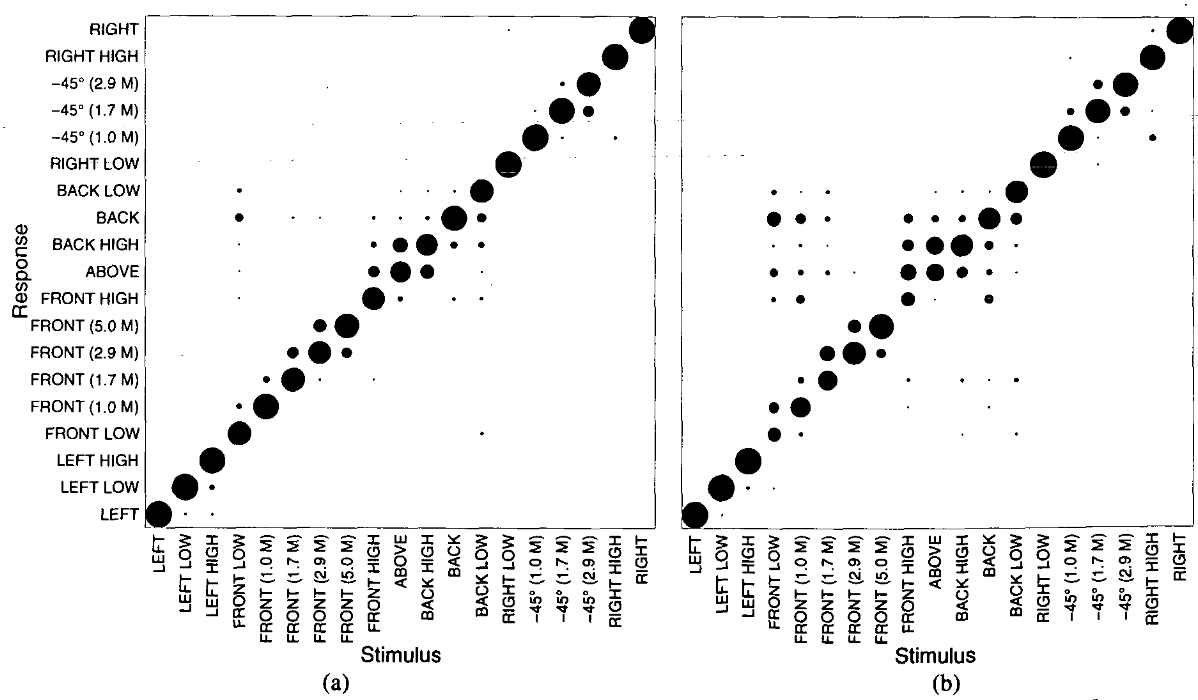


Fig. 9. (a) Pooled results of real-life listening and individual recordings (experiments A, B, and part of C, 2736 responses). (b) Nonindividual binaural recordings, each subject listening to recordings from three randomly selected other subjects in mixed sessions (remaining part of experiment C, 2736 responses).

ways perceived in the correct direction [more than 99% of the times they were presented, 100% if only the FRONT (1 M) source is considered]. The BACK direction is perceived correctly 89% of the times it is presented.

Displacements up or down from the horizontal plane are also rare. For the two 1-m sources in the median plane [FRONT (1 M) and BACK] we observed 4% movements 45° up and 1% movements 45° down. [None of these movements were actually for the FRONT (1 M) sound source.]

2.4 Nonindividual Recordings

Results for nonindividual recordings are available from two experiments—experiment C, where each subject listened to recordings from three other subjects in mixed sessions, and experiment D, where each subject listened to recordings from one other subject in a separate session. The "other" subjects were in principle selected randomly, although balanced among subjects.

2.4.1 Three Other Subjects

Fig. 9(b) shows the results of the experiment in which each subject listened to recordings from three other subjects. The order of stimuli was random and thus did not allow adaptation to a single other subject, and the stimuli included individual recordings as well. [The matrix in Fig. 9(b), however, only includes responses to the nonindividual recordings.] Responses are more scattered than in real life and with individual recordings [compare with Fig. 9(a)]. More errors are seen, especially in the median plane.

Table 3 shows a comparison of the results from nonindividual recordings and the pooled results from real-life and individual recordings. There is an increase in median-plane errors (significant at the 0.1% level). Also the number of distance errors is higher for the nonindividual recordings (significant at the 5% level).

2.4.2 One Other Subject

Each subject listened to recordings from the same other subject under two conditions—in a separate session (experiment D) and mixed into sessions that also included individual recordings as well as recordings from two additional subjects (part of experiment C). Fig. 10(a) shows the results of experiment D, and Fig. 10(b) those of the pertinent part of experiment C. The response patterns of the two frames in Fig. 10 seem very similar.

Table 4 compares the errors in the two situations. As it might be argued that a possible adaptation would only be seen after some time, the table also includes a row with results from only the second half of experiment D. Only minor and statistically nonsignificant improvements are seen for median-plane errors, when the subjects listened to recordings from only one other subject.

2.4.3 Comparison of Individual Subjects

In the preceding sections it was shown that nonindividual recordings result in poorer localization performance than in real life and with individual recordings. This result was obtained for the group of subjects, but it may conceal large individual differences. Table 5 shows the results obtained in experiment C for each individual subject, when he or she listened to each of the three other subjects and to himself or herself. Because of the low number of observations for each combination, we have chosen to report only the total number of errors. The number of errors varies clearly between subjects (whether the "subject" is considered listener or origin of the recording). It can be seen, though, that nonindividual recordings never resulted in fewer errors than individual recordings, although in several cases comparable numbers are seen.

2.4.4 Discussion

Our experiments have shown that more localization errors occur with nonindividual binaural recordings than in real life and with individual recordings. The increase in errors is pronounced and significant for median-plane errors, whereas the increase of distance errors is less distinct, although still significant. Like in real life, within-cone and out-of-cone errors are rare.

Listening to recordings from only one other subject in a separate session does not result in significantly improved results—at least not within the duration of our sessions and without feedback.

The additional errors are concentrated in the median plane, where no interaural differences exist. Many additional errors are seen, not only as movements to neighboring directions but also to directions further away. The increase is especially pronounced for the directions

Table 3. Comparison of mixed binaural recordings from three other subjects (part of experiment C) and pooled results from real-life and individual binaural recordings (experiments A, B, and part of C). Errors are given as numbers and in percent of the number of stimuli (in parentheses) that can result in errors in the category.

Condition	Errors: Out of Cone	Errors: Within Cone	Errors: Median Plane	Errors: Distance	Total Number of Stimuli
Real life and individual (experiments A, B, and part of C)	0.2% 5 (2736)	0.9% 9 (1008)	15.5% 223 (1440)	13.0% 131 (1008)	2736
Nonindividual, three others (part of experiment C)	0.1% 4 (2736)	1.6% 16 (1008)	33.8%* 486 (1440)	16.1%† 162 (1008)	2736

* Significance at 0.1% level in one-sided Fisher–Irwin test.
† Significance at 5% level.

FRONT LOW and FRONT HIGH, slightly less for FRONT. The rate of correct indications of direction for sound sources in the FRONT direction is reduced to 88% [to only 63% for the FRONT (1 M) source]. A correct response to the BACK direction occurs now in only 68% of the cases.

Laws and Platte [5] reported slightly poorer performance with nonindividual recordings than in real life, but similar performance after improvements of their headphone positioning technique. Their results seem in contrast to our findings, but most of their stimuli were outside the median plane, and their procedure differed from ours in several aspects. Thus a direct comparison is doubtful. Furthermore they used recordings from one selected typical subject, whereas we used recordings from randomly chosen subjects. This latter aspect might explain why they did not see confusions between the only two sources they had in the median plane—front and back.

Laws and Platte mentioned that in-the-head localization occurred in some cases, and we have often met colleagues who believe that this phenomenon is a result of nonindividual recordings. However, none of our subjects reported a perception such as this, neither with individual nor with nonindividual recordings. The widespread opinion that nonindividual recordings bring sound sources closer is not supported by our data either. Fig. 9 reveals that sources do not tend to come closer with nonindividual recordings.

There seems to be a general understanding that use of nonindividual recordings tends to cause frontal sound sources to be perceived in the back, and that movements the other way around are seen more rarely. This is supported by our results. A closer analysis of the nonindividual part of experiment C reveals that the three frontal sources at 1-m distance were perceived behind the frontal plane 30% of the times they were presented, whereas

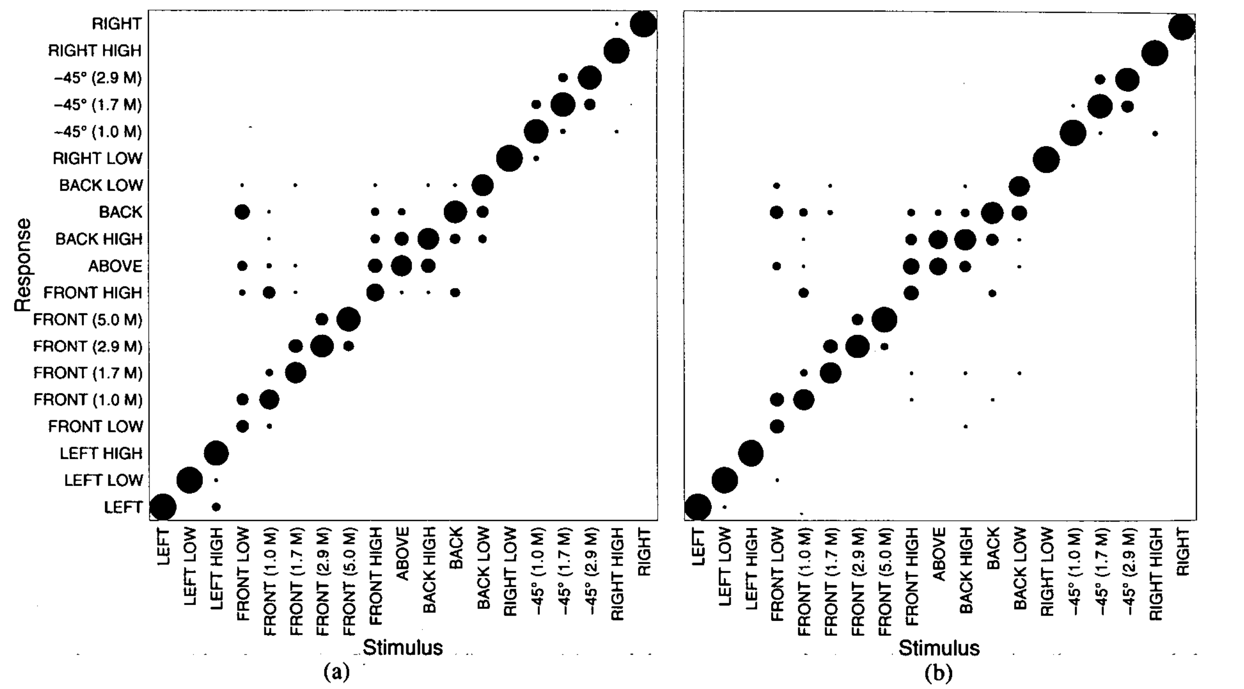


Fig. 10. Nonindividual binaural recordings, each subject listening to recordings from one randomly selected other subject. (a) In separate sessions (experiment D, 912 responses). (b) In mixed sessions (part of experiment C, 912 responses).

Table 4. Comparison of binaural recordings from one other subject presented in separate sessions (experiment D as a whole and second half) and in mixed sessions (part of experiment C). Errors are given as numbers and in percent of the number of stimuli (in parentheses) that can result in errors in the category.*

	Errors				
Condition	Out of Cone	Within Cone	Median Plane	Distance	Total Number of Stimuli
One other subject, mixed session (part of experiment C)	0.2% 2 (912)	0.6% 2 (336)	33.5% 161 (480)	14.9% 50 (336)	912
Same other subject, separate session (experiment D)	0.7% 6 (912)	1.2% 4 (336)	30.4% 146 (480)	17.9% 60 (336)	912
Same other subject separate session (experiment D, 2nd half)	0.2% 1 (456)	1.8% 3 (168)	28.8% 69 (240)	14.9% 25 (168)	456

* Statistical tests comparing each of rows 2 and 3 to row 1 did not show any significant difference between the two conditions (one-sided Fisher–Irwin test at 5% level).

the three sources in the back were perceived in front of the frontal plane in only 7% of the cases. Especially the FRONT LOW source was often perceived at the BACK position. And the BACK position was responded much more often than it was presented (an increase by a factor of 1.69).

It also seems to be a general understanding that nonindividual binaural recordings give rise to elevations, that is, sources in the horizontal plane would be perceived above that. It was already reported for the median plane that movements up or down from the horizontal plane were rare for real life (and individual binaural recordings). In the nonindividual part of experiment C we observed for the sources at FRONT (1 M) and BACK 11% movements 45° up and 2% movements 45° down. The trend is in the claimed direction, but the relative occurrence of these errors is not high.

Butler and Belendiuk [8] observed that their four subjects did not always perform best with recordings of their own. Similar observations are not seen in our results. All our eight subjects performed best with their own recordings, although in several cases comparable performance was obtained with recordings from a specific other subject. A possible explanation for the observations of Butler and Belendiuk could be that their headphone was not equalized. All their source positions were in the median plane, where only vulnerable monaural cues are available. Inadequate equalization—or complete lack of equalization—might have caused structures of the headphone transfer function to interfere with and modify the directional cues.

2.5 Additional Comments

A few further comments are in order. The question of the stability of the hearing's interpretation of physical cues has been dealt with in two ways—by mixing individual and nonindividual recordings, and by having subjects listen to the same nonindividual recordings in a separate session. For the median-plane errors a minor deterioration was seen in the former case (from 14.6 to 17.3%) and in the latter case a minor improvement was seen (from 33.5 to 28.8% in the second half of the session). Both results may be interpreted as adaptation, but the differences were not statistically significant.

We have chosen an identification experiment where the "resolution" in responses is identical to that of the stimuli. This means that we would not be able to detect errors, unless they are in the same order of magnitude as, or larger than, the stimulus resolution. Thus the significance of this limitation depends on the resolution of the hearing. For the median plane we have observed many confusions even in real life. The similarity of errors in real life and with individual binaural recordings is thus strong evidence with regard to the similarity of the auditory percept in these cases. However, it is still an open question whether the very fine structure of the directional impression has been preserved for directions where few errors occurred at all, such as the frontal direction.

Within-cone and out-of-cone errors are rare in real life and with binaural recordings, individual as well as nonindividual. All that we can conclude is that errors of these kinds are below the resolution of our setup in any case.

3 CONCLUSION

We have shown that individual binaural recordings are capable of giving an authentic reproduction for which localization performance is preserved when compared to that of real life. The recordings were made at the entrance to the blocked ear canal, and reproduction was carried out with FEC headphones calibrated accordingly. The results confirm that this technique is suitable for binaural recording. It is advantageous from a practical point of view. Larger microphones can be used, thus allowing a better signal-to-noise ratio, and insertion of microphones deeply into the ear canal is avoided.

Use of nonindividual recordings results in an increased number of median-plane errors, occurring not only in terms of movements to nearby directions but also to directions further away. Our results support a general understanding that nonindividual recordings tend to cause frontal sound sources to be perceived in the back. Nonindividual recordings do not result in in-the-head localization, nor in sound sources being perceived closer

Table 5. Total number of errors in experiment C, split up by origin of recording and by listener. The number of stimuli was 114 for each combination. First row is the number of errors made by subject BGB when listening to her own recordings and to those of subjects LRH, LTH, and SOD, and so on.

Listener \ Recording	BGB	HIM	KMJ	LRH	LTH	MAG	SOD	SWF
BGB	**8**			22	13		28	
HIM	29	**26**					37	45
KMJ			**17**	19	17			28
LRH		37		**25**		33		37
LTH			26	33	**16**		34	
MAG	14	20	15			**8**		
SOD	37		30			23	**11**	
SWF		32			27	32		**27**

than they are in real life. Movements from the horizontal plane up or down to neighboring positions are not very frequent, so even when the majority of these did go upward, our results do not support a supposition about elevations in general as a result from nonindividual recordings.

Nonindividual recordings also result in an increased number of distance errors, although this effect is less pronounced than the increase of median-plane errors. Within-cone and out-of-cone errors are rare and unaffected by the use of nonindividual recordings.

Signs of a minor adaptation to the cues of other individuals were seen, but the trend was not statistically significant. Earlier suggestions that individuals might localize better with recordings from other individuals than with their own recordings was not supported by our data.

In the present investigation all nonindividual recordings were taken from another randomly selected subject. In order to improve localization it is a logical and often suggested approach to use recordings from a "typical" subject. The selection of a typical subject and the performance that can be obtained in this way are the subjects of a subsequent investigation at our laboratory [20], [21].

4 ACKNOWLEDGMENT

This work was supported by Brüel & Kjær A/S, Perceptive Acoustics A/S, the National Agency of Industry and Trade, Denmark, Aalborg University, and the Danish Technical Research Council. The authors would like to thank their former colleagues Kim Alan Larsen and Jørn Vagn Hundebøll for their participation when the experiments were planned and carried out. They would also like to thank Anne Kirstine Andersen, their coordinator, for her handling of all appointments with the test subjects. Claus Vestergaard Skipper, the technical employee at this laboratory, deserves appreciation for his valuable help in the practical work. At last the authors would like to thank their subjects for listening patiently to a total of 6384 sound sources—of which only 912 were really there!

5 REFERENCES

[1] H. Møller, C. B. Jensen, M. F. Sørensen, and D. Hammershøi, "How Well Do Existing Binaural Systems Work?" (in preparation).

[2] J. Blauert, *Spatial Hearing—The Psychophysics of Human Sound Localization* (M.I.T. Press, Cambridge, MA, 1983).

[3] D. Hammershøi and J. Sandvad, "Binaural Auralization. Simulating Free-Field Conditions by Headphones," presented at the 96th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 42, p. 395 (1994 May), preprint 3863. (To be more thoroughly reported in [4].)

[4] D. Hammershøi and J. Sandvad, "Binaural Auralization. Simulating Free-Field Conditions by Headphones" (in preparation).

[5] P. Laws and H.-J. Platte, "Spezielle Experimente zur kopfbezogenen Sterephonie," in *Forschritte der Akustik* (DAGA '75), pp. 365–368 (1975).

[6] H.-J. Platte, P. Laws, and H. vom Hövel, "Anordnung zur genauen Reproduktion von Ohrsignalen," in *Fortschritte der Akustik* (DAGA '75), pp. 361–363 (1975).

[7] C. L. Searle, L. D. Braida, D. R. Cuddy, and M. F. Davis, "Binaural Pinna Disparity: Another Auditory Localization Cue," *J. Acoust. Soc. Am.*, vol. 57, pp. 448–455 (1975).

[8] R. A. Butler and K. Belendiuk, "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane," *J. Acoust. Soc. Am.*, vol. 61, pp. 1264–1269 (1977).

[9] S. Weinrich, "The Problem of Front-Back Localization in Binaural Hearing," *Scand. Aud.*, suppl. 15, pp. 135–145 (1982).

[10] H. Møller, "Fundamentals of Binaural Technology," *Appl. Acoust.*, vol. 36, pp. 171–218 (1992).

[11] D. Hammershøi and H. Møller, "Sound Transmission to and within the Human Ear Canal," *J. Acoust. Soc. Am.*, vol. 100, no. 2, pp. 408–427 (1996 July).

[12] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, vol. 43, pp. 203–217 (1995 Apr.).

[13] S. H. Nielsen, "Distance Perception in Hearing," Ph.D. dissertation, ISBN 87-7307-447-0, Aalborg University Press, Aalborg University, Aalborg, Denmark (1991).

[14] S. H. Nielsen, "Auditory Distance Perception in Different Rooms," *J. Audio Eng. Soc.*, vol. 41, pp. 755–770 (1993 Oct.).

[15] IEC 268-13, "Sound System Equipment," pt. 13, "Listening Tests and Loudspeakers," International Electrotechnical Commission, Geneva, Switzerland (1985).

[16] B. Langvad, H. Møller, and G. Budzynski, "Testing a New Listening-Room," *Arch. Acoust.*, vol. 14, no. 1–2, pp. 45–60 (1989).

[17] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321 (1995 May).

[18] R. S. Woodworth and H. Schlosberg, *Experimental Psychology*, 3rd ed. (Methuen, London, UK, 1955).

[19] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists* (Wiley, New York, 1987), pp. 230–234.

[20] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Selection of a Typical Human Subject for Binaural Recording," presented at the Forum Acusticum, Antwerpen, Belgium, 1996 Apr. 1–4.

[21] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Using a Typical Human Subject for Binaural Recording," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44 (1995 July/Aug.), preprint 4157.

APPENDIX NONINDIVIDUAL RECORDING AND CALIBRATION ERROR ANALYSIS

In most situations where the binaural technique is used it is not possible for each listener to have recordings made in his or her own ears, and it may also cause practical difficulties to obtain a headphone calibrated at the listener's own ears. Therefore each of the two steps, recording and transmission calibration, may be made by means of another person or by means of information obtained from another person. Recording should be understood in a broad sense. It may literally mean recording, but it may as well refer to binaural synthesis. This appendix presents a calculation of the error introduced by replacing the listener with other persons during one or both of these steps.

The terminology and the designation of the variables in this appendix are as introduced by Møller [10], and reference to equations and expressions in that article are given here in italics.

An important term in some of the calculations is the ratio between the pressure divisions at the entrance to the ear canal in two situations: 1) when the ear is in free air and 2) when the headphone is placed over the ear. In Møller et al. [12] this term was denoted by PDR (*p*ressure *d*ivision *r*atio), and from *eq. (24)* we have

$$\mathrm{PDR} = \frac{[P_3/P_2]}{[P_6/P_5]} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}} . \tag{2}$$

In Møller et al. [12] it was found that for many headphones the PDR is approximately unity. A headphone for which this is the case was defined as an FEC headphone (a headphone with *f*ree air *e*quivalent *c*oupling to the ear). (The term is equivalent to the term "open" used in [10]. The reason for changing the term is that "open" is used commerically to describe a headphone that does not exclude sound from the outside.)

A.1 General Expressions of Error

In the following, general expressions will be derived for the error when one person is used for recording (terms relating to this person are indexed "record"), a second person is used for calibration (terms relating to this person are indexed "calibrate"), and a third person is listening to the result (terms relating to this person are indexed "listener"). When the reference point is the blocked ear canal entrance, the PDR is used in the transmission calibration. In principle, information on the PDR may originate from a fourth person (terms relating to this person are indexed "pdr").

The error terms derived in this section are very general, and it might seem superfluous to assume that four different persons could be involved, including the listener. However, the term "person" should not be taken too literally, but rather be regarded as indicative of a specific source of information. Thus it is not pure speculation that three "persons" other than the listener could be involved. For instance, in a virtual reality system using binaural synthesis, head-related transfer functions may come from a database (corresponding to "record"), the headphone transfer function may have been measured with an artificial ear ("calibrate"), and information on the PDR may be from the manufacturer's data sheet ("pdr").

A.1.1 Method A: Recording at the Eardrum

The actual transmission to the listener's eardrum is obtained by inserting *eq. (30)* into *exp. (28)* and adding appropriate indexes,

$$\left[\frac{P_4}{P_1}\right]_{\text{record}} \cdot \frac{[P_7/E_{\text{headphone}}]_{\text{listener}}}{[P_7/E_{\text{headphone}}]_{\text{calibrate}}} . \tag{3}$$

The transmission should have been

$$\left[\frac{P_4}{P_1}\right]_{\text{listener}} \tag{4}$$

The error given as a factor can be obtained by dividing the actual transmission [Exp. (3)] by the wanted transmission [Exp. (4)],

$$\varepsilon_{\mathrm{A}} = \frac{[P_4/P_1]_{\text{record}}}{[P_4/P_1]_{\text{listener}}} \cdot \frac{[P_7/E_{\text{headphone}}]_{\text{listener}}}{[P_7/E_{\text{headphone}}]_{\text{calibrate}}} . \tag{5}$$

A.1.2 Method B: Recording at the Entrance to the Open Ear Canal

The actual transmission to the listener's eardrum is obtained by inserting *eq. (34)* into *exp. (31)* and adding appropriate indexes,

$$\left[\frac{P_3}{P_1}\right]_{\text{record}} \cdot \frac{[P_7/E_{\text{headphone}}]_{\text{listener}}}{[P_6/E_{\text{headphone}}]_{\text{calibrate}}} . \tag{6}$$

The error can be obtained by dividing the actual transmission [Exp. (6)] by the wanted transmission, which is again given by Exp. (4), splitting up the "listener" transmission terms, and utilizing that $[P_7/P_6] = [P_4/P_3]$,

$$\varepsilon_{\mathrm{B}} = \frac{[P_3/P_1]_{\text{record}}}{[P_3/P_1]_{\text{listener}}} \cdot \frac{[P_6/E_{\text{headphone}}]_{\text{listener}}}{[P_6/E_{\text{headphone}}]_{\text{calibrate}}} . \tag{7}$$

A.1.3 Method C: Recording at the Entrance to the Blocked Ear Canal

The actual transmission to the listener's eardrum is obtained by inserting *eq. (38)* into *exp. (35)*, adding appropriate indexes, and using Eq. (2),

$$\left[\frac{P_2}{P_1}\right]_{\text{record}} \cdot \frac{[P_7/E_{\text{headphone}}]_{\text{listener}}}{[P_5/E_{\text{headphone}}]_{\text{calibrate}}} \cdot [\mathrm{PDR}]_{\text{pdr}} . \tag{8}$$

The error can be obtained by dividing the actual transmission [Exp. (8)] by the wanted transmission still given by Exp. (4), splitting up the "listener" transmission terms, and utilizing Eq. (2) and that $[P_7/P_6] = [P_4/P_3]$,

$$\varepsilon_C = \frac{[P_2/P_1]_{\text{record}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_5/E_{\text{headphone}}]_{\text{listener}}}{[P_5/E_{\text{headphone}}]_{\text{calibrate}}} \cdot \frac{[\text{PDR}]_{\text{pdr}}}{[\text{PDR}]_{\text{listener}}} . \tag{9}$$

If, in the calibration, unity is inserted for the PDR (because it is known that the headphone has FEC properties—whether or not that is true—due to a lack of data, due to ignorance, or whatever the cause may be), then the error becomes

$$\varepsilon_C^* = \frac{[P_2/P_1]_{\text{record}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_5/E_{\text{headphone}}]_{\text{listener}}}{[P_5/E_{\text{headphone}}]_{\text{calibrate}}} \cdot \frac{1}{[\text{PDR}]_{\text{listener}}} . \tag{10}$$

A.1.4 Comments

For each of the methods the error terms are easily recognizable as the error made during recording and calibration. In Eq. (10) the last term is the error introduced by neglecting the PDR (regardless of how close to unity it may be).

A.2 Error Terms in Selected Cases

In the following, error terms are derived for some selected specific combinations of "persons." The cases chosen are those considered especially relevant for the present investigation, but they may also have some relevance for practical applications of the binaural technique. Only two "persons" are considered, the listener and one other person. Each of the two steps recording and calibration (including determination of the headphone transfer function as well as determination of the PDR) are carried out by means of one of these. Parts of the recording–calibration–reproduction process for which the listener is used are called *individual*, whereas parts for which the other person is used are called *nonindividual*.

A.2.1 Individual Recording, Individual Calibration

Here the recordings are made in the listener's own ears, and the calibration is carried out using the listener himself or herself. By inserting "listener" for "record," "calibrate," and "pdr" in Eqs. (5), (7), and (9) we get

$$\varepsilon_A = \varepsilon_B = \varepsilon_C = 1 . \tag{11}$$

As expected—perfect transmission. By similar insertion, Eq. (10) becomes

$$\varepsilon_C^* = \frac{1}{[\text{PDR}]_{\text{listener}}} . \tag{12}$$

A precondition for using Eq. (10) was that an FEC headphone be used. Thus strictly speaking, the error term in Eq. (12) includes the error introduced if the precondition is violated. If an FEC headphone *is* actually used, Eq. (12) naturally reduces to

$$\varepsilon_C^* = 1 . \tag{13}$$

A.2.2 Nonindividual Recording, Individual Calibration

In this case another person is used for recording, whereas the listener himself or herself is used for calibration. By inserting "other" for "record," and "listener" for "calibrate" and "pdr" in Eqs. (5), (7), (9), and (10) we get

$$\varepsilon_A = \frac{[P_4/P_1]_{\text{other}}}{[P_4/P_1]_{\text{listener}}} = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_4/P_2]_{\text{other}}}{[P_4/P_2]_{\text{listener}}} \tag{14}$$

$$\varepsilon_B = \frac{[P_3/P_1]_{\text{other}}}{[P_3/P_1]_{\text{listener}}} = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_3/P_2]_{\text{other}}}{[P_3/P_2]_{\text{listener}}} \tag{15}$$

$$\varepsilon_C = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \tag{16}$$

$$\varepsilon_C^* = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{1}{[\text{PDR}]_{\text{listener}}} . \tag{17}$$

If an FEC headphone is actually used, Eq. (17) reduces to

$$\varepsilon_C^* = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} . \tag{18}$$

A.2.3 Nonindividual Recording, Nonindividual Calibration

Here the listener is only used for listening. By inserting "other" for "record," "calibrate," and "pdr" in Eqs. (5), (7), and (9), and by using Eq. (2) and that $[P_7/P_6] = [P_4/P_3]$, we obtain the same expression regardless of the reference point,

$$\varepsilon_A = \varepsilon_B = \varepsilon_C = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_5/E_{\text{headphone}}]_{\text{listener}}}{[P_5/E_{\text{headphone}}]_{\text{other}}} \cdot \frac{[\text{PDR}]_{\text{other}}}{[\text{PDR}]_{\text{listener}}} . \tag{19}$$

A slightly different expression is found by similar insertion in Eq. (10),

$$\varepsilon_C^* = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_5/E_{\text{headphone}}]_{\text{listener}}}{[P_5/E_{\text{headphone}}]_{\text{other}}} \cdot \frac{1}{[\text{PDR}]_{\text{listener}}} . \tag{20}$$

If an FEC headphone is actually used, all error terms of Eqs. (19) and (20) reduce to the same expression,

$$\varepsilon_A = \varepsilon_B = \varepsilon_C = \varepsilon_C^* = \frac{[P_2/P_1]_{\text{other}}}{[P_2/P_1]_{\text{listener}}} \cdot \frac{[P_5/E_{\text{headphone}}]_{\text{listener}}}{[P_5/E_{\text{headphone}}]_{\text{other}}} . \tag{21}$$

A.2.4 Individual Recording, Nonindividual Calibration

This seems to be a rather awkward case. If the listener is available for recording, then why not use him or her for calibration? Nevertheless it is included for completeness—and it might have some practicality by avoiding a tiresome process when fitting an individual equalization. By inserting "listener" for "record," and "other" for "calibrate" and "pdr" in Eqs. (5), (7), (9), and (10), we get

$$\varepsilon_A = \frac{[P_7/E_{headphone}]_{listener}}{[P_7/E_{headphone}]_{other}} = \frac{[P_5/E_{headphone}]_{listener}}{[P_5/E_{headphone}]_{other}} \cdot \frac{[P_7/P_5]_{listener}}{[P_7/P_5]_{other}} \tag{22}$$

$$\varepsilon_B = \frac{[P_6/E_{headphone}]_{listener}}{[P_6/E_{headphone}]_{other}} = \frac{[P_5/E_{headphone}]_{listener}}{[P_5/E_{headphone}]_{other}} \cdot \frac{[P_6/P_5]_{listener}}{[P_6/P_5]_{other}} \tag{23}$$

$$\varepsilon_C = \frac{[P_5/E_{headphone}]_{listener}}{[P_5/E_{headphone}]_{other}} \cdot \frac{[\mathrm{PDR}]_{other}}{[\mathrm{PDR}]_{listener}} \tag{24}$$

$$\varepsilon_C^* = \frac{[P_5/E_{headphone}]_{listener}}{[P_5/E_{headphone}]_{other}} \cdot \frac{1}{[\mathrm{PDR}]_{listener}} \tag{25}$$

If an FEC headphone is actually used, Eqs. (24) and (25) reduce to

$$\varepsilon_C = \varepsilon_C^* = \frac{[P_5/E_{headphone}]_{listener}}{[P_5/E_{headphone}]_{other}} . \tag{26}$$

A.3 Selection of Procedure for Investigation

In Section A.2.1 it was shown that when individual recordings are used, impeccable transmission can be obtained by using individual calibration. This applies irrespective of the reference point. Nonindividual calibration results in an error that depends on the reference point (Section A.2.4).

In Sections A.2.2 and A.2.3 it was shown that when nonindividual recordings are used, correct transmission is not possible. Various expressions apply for the error [Eqs. (14)–(21)]. One term is present in all of them, namely,

$$\frac{[P_2/P_1]_{other}}{[P_2/P_1]_{listener}} . \tag{27}$$

This is quite natural since nonindividual recordings rely on the "other person" for the transmission of a three-dimensional sound field into two signals. The term reflects exactly the imperfection in this transformation. Additional terms in the expression for the error cannot counterbalance this error since they are independent of direction.

In the experiments the smallest possible error is aimed at, and the procedures leading to Eqs. (16) and (18) are attractive since only the unavoidable error term is present. These equations represent individual calibration in combination with the use of the blocked ear canal as reference point (either accounting for the PDR or using an FEC headphone).

The following choice stands to reason:

1) Recordings for individual and nonindividual use are made at the blocked ear canal.

2) An FEC headphone is used.

3) Individual equalization (that is, following the listener) is used for individual as well as nonindividual recordings.

Under these conditions the equalization should be determined as given in *eq. (39)*,

$$G_C^* = \frac{1}{M_1 \cdot [P_5/E_{headphone}]} . \tag{28}$$

For convenience we introduce for the equalization: G ($= G_C^*$), for the headphone transfer function: PTF ($= [P_5/E_{headphone}]$), and for the transfer function of the recording microphone: M ($= M_1$),

$$G = \frac{1}{M \cdot \mathrm{PTF}} . \tag{29}$$

A.4 General Comments

Some more general guidelines about implementations of the binaural technique can be derived from the preceding calculations. For almost any practical purpose recordings cannot be made individually, and the following assumes nonindividual recordings (or nonindividual binaural synthesis).

If individual calibration is possible, then the method with recordings at the blocked ear canal entrance (as chosen for our experiments) is the best choice, since it only gives the unavoidable error from the "other" subject [Eq. (16)]. FEC headphones are practical since they make the calibration simpler, but the use of such headphones is not a prerequisite.

Individual calibration may be of questionable value if a reference point in the open ear canal is chosen. The last term in each of Eqs. (14) and (15) should then be compared to what can be obtained by nonindividual calibration, for example, with an FEC headphone [Eq. (21)]. It is our experience that for the frequency range of 2–7 kHz individual variations, especially in the pressure division [last term of Eq. (15)] may be larger than individual variations in headphone transfer functions [last term of Eq. (21)]. (See [11, fig. 13] and some of the better headphones in [12, fig. 7].)

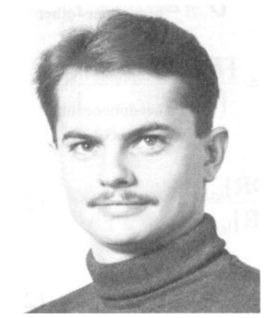
If individual calibration is not possible—and this is the case for most commercial applications—then the use of FEC headphones is especially recommended since it reduces the error to the minimum [Eq. (21)], regardless of the reference point.

The blocked ear canal entrance has practical advantages for recording with human subjects. Larger microphones can be used, allowing a better signal-to-noise ratio, and insertion of microphones deeply into the ear canal is avoided. As far as artificial heads are concerned, they can be constructed without ear canal simulations, and for use in binaural synthesis human HRTFs can be measured utilizing the same advantages with respect to microphones and their positioning as for recordings. Furthermore, individual variations in HRTFs are lower at the blocked ear canal [17], [11], and "typical" HRTFs are more easily yielded. Lower variations are also seen in headphone transfer functions measured at the blocked ear canal [12], and a typical headphone equalization for general use is thus more easily obtained.

THE AUTHORS

H. Møller

M. F. Sørensen



C. B. Jensen

D. Hammershøi

Henrik Møller was born in Århus in 1951. He studied electrical engineering (Danish Engineering Academy) and received a B.Sc. degree in 1974. He worked as a development engineer for Brüel & Kjær from 1974 to 1976. Since then he has been at Aalborg University. He became associate professor in 1980, received a Ph.D. degree in 1984, and was appointed professor in 1988. During the period 1991–94 he was partly on leave from the university to work as a director of Perceptive Acoustics A/S, a research subsidiary company of Brüel & Kjær.

Dr. Møller's previous and current research reflect his long-time experience with sound, its influence on humans, acoustical measurement techniques, signal processing, hearing, and psychometric methods. His research areas include effects of infrasound and low-frequency noise on humans, investigations of hearing thresholds and loudness assessment, and exploitation of binaural techniques. He is the author of numerous scientific publications and invited as well as contributed conference papers.

When new high-quality acoustical laboratories were built at Aalborg University in 1987, Dr. Møller was responsible for the design as well as control of the work. As head of the Acoustics Laboratory, he is now the manager of research and education in a wide range of areas such as human sound perception, audiology, psychometry, electroacoustics, recording and playback techniques, auralization in acoustic room modeling and virtual reality, acoustical measurement techniques, electronics, and signal processing.

Dr. Møller has organized conferences on Low Frequency Noise and Hearing (Aalborg 1980) and general acoustics (Nordic Acoustical Meeting, Aalborg 1986). He is convener of ISO Technical Committee 43: "Acoustics," Working Group 1: "Thresholds of Hearing," a member of the Nordic Research Group on Noise Effects, and of the editorial board of Journal of Low Frequency Noise & Vibration. He holds membership in the Danish Engineering Society, Audio Engineering Society, Acoustical Society of America, IEEE, Danish Acoustical Society, Danish Technical-Audiological Society, Danish Biomedical Society, and Danish Standardisation Organisation (board of Acoustics and Working Group of Audiometry and Hearing).

Dr. Møller spends hours off work (too few) by playing big band music on his baritone saxophone or by keeping his classic British cars in good shape. Now and then, he also drives them.

•

Michael Friis Sørensen was born in 1966. He studied electrical engineering at Aalborg University, where he received his M.Sc.E.E., specializing in acoustics, in 1991. From 1991 to 1995 he worked as a research engineer at Aalborg University.

His major fields of interest are psychoacoustics, noise measuring techniques, psychometry, electroacoustics,audiology, and experimental design. From 1991 he has worked on research projects within various aspects of binaural technique and audiometry. Since 1995 August he has been working as a consultant in noise and vibration control at Ødegaard & Danneskiold-Samsøe, Jylland Aps.

Mr. Sørensen is a member of the Danish Acoustical Society, Danish Engineering Society, Audio Engineering Society, and Danish Technical Audiological Society.

•

Clemen Boje Jensen was born in Aalborg, Denmark, in 1965. He studied electrical engineering at Aalborg University. In 1991 he received a Master of Science in electrical engineering, specializing in acoustics.

From 1991 to 1994 he was employed as a research engineer at the Acoustics Laboratory of Aalborg University, where he worked on various aspects of binaural techniques, including sound measurements in the human ear canal, design of artificial heads, calibrated binaural sound reproduction, design of measuring and experimental setups, and artificial generation of binaural signals by means of digital signal processing. Since 1995 he has been working in the R&D department of Dancall

Telecom A/S, developing DECT telephones.

Mr. Jensen holds membership in the Danish Acoustical Society, Danish Engineering Society, and the Audio Engineering Society.

●

Dorte Hammershøi was born in 1965. She studied electrical engineering in Aalborg (Aalborg University). In 1989 she received a Master of Science in electrical engineering, with specialization in biomedical engineering. During her study, she worked part time as an engineer at Synaps Electronic Aps., developing communication aids for motionally disabled people.

From 1990 to 1995 she worked as a research engineer at the Acoustics Laboratory at Aalborg University, and in 1995 she received the Ph.D. degree in acoustics. Since 1995 she has worked as an assistant professor at Aalborg University.

Dr. Hammershøi is experienced in electronics, sound measuring techniques, digital signal processing, hearing, and psychometry, and is familiar with neuro and sensory physiology. From 1990 to 1992 she worked on research projects on the improvement of sound reproduction techniques by means of binaural techniques. Since 1992, she has been working with computer generation of binaural signals for auralization of room models and for generation of auditory virtual environments in virtual reality systems.

Dr. Hammershøi holds membership of the Danish Acoustical Society, Danish Engineering Society, and Audio Engineering Society, and is the author of many scientific papers.