Using a typical human subject for binaural recording

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

Published in:
Proceedings of 100th Audio Engineering Society Convention, Copenhagen, May 11-14, 1996

Publication date:
1996

Using a Typical Human Subject for Binaural Recording

Preprint 4157 (C-10)

Henrik Møller, Clemen Boje Jensen, Dorte Hammershøi & Michael Friis Sørensen
Aalborg University, Aalborg, Denmark

Presented at the 100th Convention 1996 May 11–14 Copenhagen

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

USING A TYPICAL HUMAN SUBJECT FOR BINAURAL RECORDING

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

Acoustics Laboratory, Aalborg University
Fredrik Bajers Vej 7 B4, DK-9220 Aalborg Ø, Denmark.

Abstract

Previous investigations have shown that use of non-individual binaural recordings result in a substantially reduced localization performance when compared to real-life listening and listening to individual binaural recordings. The deteriorations are most pronounced in the median plane, where the two ears receive the same direct sound. The present study comprised a localization experiment, in which 20 subjects listened to binaural recordings from the ears of 30 humans. The "recording head" which resulted in the best performance for the group of subjects was selected as "typical" and used in further experiments. The number of median plane errors was 36.3%, when the non-individual recordings originated from a random subject, and this figure was reduced to 21.2% with recordings from the typical subject. This value is not far from the real life value of 15.5%, although still significantly higher. Front/back confusions in the horizontal plane were almost eliminated with recordings from the typical subject, whereas these were numerous with recordings from a random subject. These results were obtained with individual headphone equalization. A marginal, although statistically significant increase of errors was seen, when a suitable common equalization was used.

0. Introduction

Previous investigations have shown that *individual binaural recordings* (recordings made in the listener's own ears) can offer a reproduction in which the localization performance from real life has been maintained, whereas *non-individual binaural recordings* (recordings made in the ears of another person) result in reduced localization performance (Laws and Platte [1], Møller et al. [2]). The deteriorations are mainly seen for sound sources in the median plane, for which the two ears receive the same direct sound. Reduced performance has also been observed with binaural signals synthesized using non-individual head-related transfer functions, HRTFs (Morimoto and Ando [3], Wenzel et al. [4] (more thoroughly reported by Wenzel [5]), Kawaura et al. [6], [7], Begault and Wenzel [8], Begault [9], Wenzel et al. [10], Hammershøi [11] (review)).

The non-individual recordings, which the subjects listened to in our previous study, originated from randomly selected other subjects. The present investigation was carried out to show whether improved results can be obtained, if the non-individual recordings originate from a selected, typical human subject.

The total transmission in a binaural system is determined not only by the recording situation, but also by the playback system, which normally means the headphone and its equalization. It is possible to use *individual equalization* and thus compensate for the frequency response of the headphone measured on each individual listener, or to use *non-individual equalization* and only compensate for a mean or typical headphone response. Use of individual equalization is impractical for most applications, and it was also the aim of the investigation to examine the effect of using non-individual equalization.

The experimental method in the present study was in most respects identical to that of our previous investigation [2]. The description will therefore be confined to a summary of the procedures supplemented by complete descriptions of issues specific for the present investigation. More details may thus be found in our earlier report.

1. Method

The experiments were carried out in an IEC listening room, where 19 loudspeakers were located around the subject. 14 were positioned in various directions on a sphere with a radius of 1 m, 7 of these in the median plane. The remaining 5 were at more distant positions. The subjects listened to a 5-second recording of a female voice, either directly from the loudspeakers or indirectly as a binaural recording made in the same set-up and reproduced by means of headphones. The subject was sitting in the set-up in either cases and kept his head still during stimuli. The loudspeakers were visible to the subjects, and the experiments were carried out as identification experiments, where the subjects responded from which loudspeaker they perceived the sound.

1.1 Subjects

20 paid students with controlled normal hearing participated as listeners, 10 of each sex, aged 20-30 years. They were all skilled in psychoacoustic experiments, but they were not in any way selected for their hearing or localization proficiency. 8 of the subjects had already participated in the previous experiment. An initial test showed that there was no difference in the real life localization performance between this group of 8 and the group of 12 new subjects. The data and test are reported in Appendix A.

The binaural recordings were made at the blocked entrance to the ear canals of 30 humans, including the 20 subjects participating as listeners (the extra 10 were staff members and others who were unsuitable or unavailable as listeners). The reason for selecting such a large number of "recording heads" was to give the subjects a "range" as wide as possible from which they could select a typical recording head. The reason for choosing a larger group of listeners than previously was to give the listeners' choice of recording head more general validity.

1.2 Headphone equalization

A headphone (Beyerdynamic DT 990 Professional) with FEC proporties [12] was used and equalized to a flat frequency response when measured at the blocked ear canal entrance. The headphone was equalized individually for each listener using a 32nd order IIR filter as described earlier for the 8 subjects [2]. The equalizations for the remaining 12 subjects were made similarly.

An exception was the experiment with non-individual equalization (experiment D, see Section 1.3.4), where the target function for a common equalization was the reciprocal of the mean (power averaging) of 30 individual headphone transfer functions (for all recording heads). The individual equalization filters and the non-individual filter are shown in Figure 1(a). The error introduced for each individual when using the non-individual filter is seen in Figure 1(b).

1.3 Experimental design

Four different experiments were made. For each subject they were accomplished on 5 days.

1.3.1 Experiment A: Real life

Each subject listened to each loudspeaker 6 times. The experiment was divided into two sessions with each 3 repetitions. The stimulus order was random in each session. The sessions had a duration of approximately 10 minutes, and they were separated by a short coffee break. The number of stimuli for each subject was 114, giving a total of 2280 stimuli for the 20 subjects. The data for the 8 "old" subjects were taken from the previous investigation [2], which was carried out only some weeks before the present investigation.

1.3.2 Experiment B: Recordings from random subjects

Each subject listened to each of the recording heads once for each loudspeaker position. This makes 570 stimuli for each subject (19 loudspeakers $\times$ 30 recording heads). These were presented in random order and divided into 6 sessions of approximately 16 minutes, accomplished on 2 days. The total number of stimuli was 11400 (20 listeners $\times$ 30 recording heads $\times$ 19 loudspeakers).

1.3.3 Experiment C: Recordings from typical subject

From the results of experiment B, the recording head which resulted in the best overall performance for the group of subjects was selected as "typical". In experiment C only recordings from this selected subject were used. Each subject listened to each loudspeaker 6 times. The experiment was divided into two sessions with each 3 repetitions. The sessions had a duration of approximately 10 minutes, and the stimulus order was random in each session. The number of stimuli for each subject was 114, giving a total of 2280 stimuli for the 20 subjects.

1.3.4 Experiment D: Recordings from typical subject, non-individual equalization

Experiment D was identical to experiment C, except that non-individual headphone equalization was used.

2. Results and discussion

The results from each experiment will be presented in a 19 by 19 matrix with the stimulus position as abscissa and the responded position as ordinate. 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. Correct answers are found at the diagonal.

For statistical analysis errors are classified into four groups. 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 called a *within-cone error*, except when stimulus and response are in the median plane, in which case it is designated a *median-plane error.* A response given in the same direction as the stimulus, but at an incorrect distance, is denoted a *distance error.*

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 e. g. [13]). In order to give the most powerful tests, only stimuli that actually can lead to errors in a certain category are included in each test, and in the calculation of error percentages.

2.1 Real life

The results of real life listening (experiment A) are shown in Figure 3. Correct answers are to be found at the diagonal, and 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 judgement of distance is a common error.

These observations are similar to the observations made for real life listening in our earlier investigation, and they are now confirmed for a larger group. Other observations that can be confirmed are: Sound sources in the FRONT direction are almost always perceived in the correct direction. The same applies to the source at BACK.

2.2 Recordings from random subjects

The results for non-individual recordings (experiment B) including all recording heads are shown in Figure 4. It is obvious that considerably more errors are made than in real life. The number of errors have increased for the low median plane sources (FRONT LOW, BACK LOW), and more confusions are seen between the upper median plane directions (FRONT HIGH, ABOVE and BACK HIGH).

Additional directions have also come up with errors, of which the most notable should be mentioned. In real life there were almost no errors for the sound sources in the FRONT direction (except for distance errors), whereas these sources are now frequently perceived in other directions in the median plane, quite often in the BACK direction. Correspondingly, the sound source in the BACK direction is frequently perceived in other median plane directions, quite often in the FRONT direction.

Table I shows the results of real life listening and listening to binaural recordings from a random subject divided into the four error categories. In general, only few out-of-cone and within-cone errors are seen, whereas a lot more median-plane and distance errors are seen. A highly significant increase of errors is seen for the binaural recordings (0.1% level for all error categories). When the actual number of errors are taken into account, the most notable difference is the increase of median-plane errors by more than a factor of 2 from 15.5% to 36.3%. The number of distance errors increased by approximately a factor of 2 from 10.4% to 20.7%.

Errors: / Condition:	out of cone	within cone	median plane	distance	total number of stimuli
Real life (experiment A)	**0.4%** 10 (2280)	**0.7%** 6 (840)	**15.5%** 186 (1200)	**10.4%** 87 (840)	2280
Binaural, random subject (experiment B)	**1.4%*** ** 153 (11020)	**2.3%*** ** 94 (4060)	**36.3%*** ** 2103 (5800)	**20.7%*** ** 839 (4060)	11020

Table I

*Comparison of non-individual binaural recordings (experiment B, 380 individual stimuli disregarded) and real life (experiment A). Errors are given in percentage (bold) and numbers. Number of stimuli that can result in errors in a category are given in brackets. *** indicates significance at 0.1% level in a one-sided Fisher-Irwin test.*

The increased number of errors with non-individual binaural recordings confirms our earlier result [2], but it is now shown for a larger group of listeners, and for a much wider range of combinations between listener and recording head. (Each of the 8 subjects in the previous study listened only to recordings from 3 other subjects, and the likelihood of bias due to coincidentally good and bad "matches" between recording head and listener is significantly reduced if not virtually eliminated in the present investigation).

2.3 Selection of typical subject

In the previous section, results for all 30 recording heads were combined. Considerable variation is present between the results from different recording heads. Since median-plane errors constitute the largest group of errors, we have chosen to use these for the selection of a typical head. Figure 2 shows the median-plane errors for the individual recording heads, ranked according to error percentage. The errors range from 25.5% to 47.4% depending on recording head. The recording head "avh" was selected as the "typical" head to be used in experiment C, since the recordings from this head resulted in the lowest number of median-plane errors. The same recording head would have been chosen, if the ranking had been made according to the total number of directional errors (sum of out-of-cone, within-cone and median-plane errors).

2.4 Recordings from typical subject

The results from recordings with the selected subject "avh" (experiment C) are shown in Figure 5. An immediate look at the figure seems to indicate that the number of errors has decreased when compared to the results for recordings from randomly chosen other subjects as seen in Figure 4. A comparison of the four error categories is given in Table II. Significant improvements are seen for all three categories of directional error.

Errors: / Condition:	out of cone	within cone	median plane	distance	total number of stimuli
Random subject (experiment B)	**1.4%** 153 (11020)	**2.3%** 94 (4060)	**36.3%** 2103 (5800)	**20.7%** 839 (4060)	11020
Typical subject (experiment C)	**0.8%**** 19 (2280)	**0.5%***** 4 (840)	**21.2%***** 254 (1200)	**24.0%** 202 (840)	2280

Table II

*Comparison of non-individual recordings from a selected subject (experiment C) and from a random subject (experiment B, 380 individual stimuli disregarded). Errors are given in percentage (bold) and number. Number of stimuli that can result in errors in a category are given in brackets. ** indicates significance at 1% level in a one-sided Fisher-Irwin test, *** at 0.1% level.*

A comparison of Figure 5 and Figure 4 reveals improvements almost everywhere. Most of the circles outside the diagonal have become smaller and in several cases even disappeared. It is especially worth noting that the sources in the FRONT direction are now almost always perceived in the correct direction. A further analysis of these shows 96% responses in the correct direction with the typical head in contrast to 86% with a random head (99% in real life). If only the FRONT (1 M) source is considered (the more distant sources may be identified partly by their distance), corresponding figures are 92% for the typical head in contrast to 71% for a random head (99% for real life). Also the BACK sound source is now more often perceived in the correct direction: 86% with the typical head in contrast to 63% with a random head (96% in real life).

Table II reveals that slightly more distance errors are present in experiment C than in experiment B. This observation can be explained by the method used for classifying the errors. A response is only classified as a distance error, if the direction is correct. Thus a reduction of directional errors will lead to an increase in potential distance errors. A further analysis has estimated that nearly 2% extra directional errors should be expected on this account[1].

The rate of median-plane errors in experiment C is 21.2%, an even lower value than was obtained for the same recording head in experiment B (25.5%). This indicates that the low value for this recording head in experiment B was not just a matter of coincidence. It might also suggest that some adaptation takes place, when a subject listens to recordings from only one other subject for some time. It should be noted, however, that no feedback was given about right or wrong responses that could facilitate a possible adaptation to the cues of the recording head. Moreover, a statistical analysis (not shown) has revealed that the difference between experiment C and the "avh" part of experiment B was not significant (two-sided Fisher-Irwin test at 5% level).

2.5 Effect of equalization

The results with non-individual equalization (experiment D) are shown in Figure 6. These results are not obviously different from those with individual equalization in Figure 5, but the statistical tests shown in Table III reveal that a minor increase in median-plane errors was seen (significant at 5% level).

---

[1] When the direction is correct, 23-25% distance errors occur (estimated from experiment C and B, respectively). For the relevant sources there were approximately 7% more correct directions, and an increase of approximately $7\% \times 0.24 \approx 1.7\%$ should be expected.

Condition: / Errors:	out of cone	within cone	median plane	distance	total number of stimuli
Typical subject, individual equalization (experiment C)	**0.8%** 19 (2280)	**0.5%** 4 (840)	**21.2%** 254 (1200)	**24.0%** 202 (840)	2280
Typical subject, non-individual equal. (experiment D)	**0.7%** 16 (2280)	**1.0%** 8 (840)	**24.0%*** 288 (1200)	**23.6%** 198 (840)	2280

Table III

*Comparison of recordings from a typical subject, non-individual equalization. Errors are given in percentage (bold) and number. Number of stimuli that can result in errors in a category are given in brackets. * indicates significance at 5% level in a one-sided Fisher-Irwin test.*

2.6 How far are we from real life?

Now it would be interesting to compare the results obtained with recordings from a selected subject with the results from real life listening. This is done in Table IV where each of experiments B and C are compared to experiment A. The numbers of median-plane errors and distance errors are still significantly higher than in real life, even for recordings made with a selected subject.

Condition: / Errors:	out of cone	within cone	median plane	distance	total number of stimuli
Real life (experiment A)	**0.4%** 10 (2280)	**0.7%** 6 (840)	**15.5%** 186 (1200)	**10.4%** 87 (840)	2280
Typical subject, individual equalization (experiment C)	**0.8%*** 19 (2280)	**0.5%** 4 (840)	**21.2%***** 254 (1200)	**24.0%***** 202 (840)	2280
Typical subject, non-individual equal. (experiment D)	**0.7%** 16 (2280)	**1.0%** 8 (840)	**24.0%***** 288 (1200)	**23.6%***** 198 (840)	2280

Table IV

*Comparison of recordings from a typical subject (individual and non-individual equalization, experiments C and D, respectively) and real life (experiment A). Errors are given in percentage (bold) and number. Number of stimuli that can result in errors in a category are given in brackets. Each of the rows 2 and 3 have been tested against row 1. * indicates significance at 5% level in a one-sided Fisher-Irwin test, *** at 0.1% level.*

A comparison of Figure 5 and Figure 6 with Figure 4 shows that the additional errors occur very scattered. The most obvious differences are that the FRONT LOW sound source is now quite often perceived at the BACK position, and that the assessment of distance has become more uncertain for the sound sources in FRONT. In addition, we have also more confusions between the upper median plane sources (FRONT HIGH, ABOVE, and BACK HIGH), and more confusions between the FRONT and the BACK directions. However, the occurrence of these errors has obviously been reduced as compared to the results with a random recording head.

2.7 Additional comments

In our previous investigation [2] we found support for a general understanding that use of non-individual recordings tend to cause frontal sound sources to be perceived in the back, and that movements the other way round are seen more rarely. The present investigation confirms this for a larger group and a much wider range of combinations between listener and recording head.

With recordings from a random subject (experiment B), stimuli at the three frontal sound sources at 1 m distance were perceived behind the frontal plane in 23% of the cases, whereas stimuli at the corresponding three sound sources in the back were perceived in front of the frontal plane in only 11% of the cases. These figures were reduced to 13% and 4%, respectively, when the typical subject was used (experiment C), and to 6% and 2%, respectively, in real life. The BACK direction was responded much more often than stimuli were given there (an increase by a factor of 1.73 for a random recording head, 1.48 for the typical recording head, and 1.14 in real life).

Another general understanding that was examined in our previous investigation is that non-individual binaural recordings should give rise to elevations, i. e. sources in the horizontal plane would be perceived above that. For a random recording head stimuli at FRONT 1 M and BACK were moved 45° up in 8% of the cases and 45° down in 6% of the cases. Corresponding figures were 2% and 1% for the typical head, 2% and 1% for real life. As concluded also in the earlier report, the trend is in the claimed direction, but the relative occurrence of these errors is low.

3. Conclusion

It has been confirmed that use of non-individual binaural recordings results in reduced localization performance. The reduced performance is observed most clearly for sound sources in the median plane, where movements are seen to nearby directions as well as to directions further away. The results also support the general understanding that non-individual recordings tend to cause frontal sources to be perceived in the back more often than the other way round. A hypothesis of non-individual recordings as responsible for elevations is not supported.

If the non-individual recordings originate from a carefully selected typical subject, it is possible to reduce the number of errors substantially as compared to non-individual recordings from a random subject. The investigation has demonstrated a reduction of median-plane errors down to a level not far from that of real life listening (although still significantly higher). Front/back confusions can be almost eliminated for sound sources in the horizontal plane, whereas they are numerous with recordings from a random subject.

Only a minor increase of errors is seen, when a common headphone equalization is used instead of individual equalization for each listener. This observation should not be misinterpreted to indicate that headphone equalization is unimportant, since the common equalization in the experiment was carefully constructed from information of the headphone's performance on real ears.

Artificial heads should ideally be constructed to resemble the acoustics of a typical subject. The performance of existing artificial head recording systems is the subject of a subsequent investigation in our laboratory [14].

Acknowledgements

Economic support from Brüel & Kjær A/S, Perceptive Acoustics A/S, the National Agency of Industry and Trade, Denmark, Aalborg University, and the Danish Technical Research Council is greatly acknowledged. 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. We would also like to thank Anne Kirstine Andersen, our coordinator, for her handling of all appointments with the test subjects. Also Claus Vestergaard Skipper, the technical employee at our laboratory, deserves appreciation for his valuable help in the practical work. At last we would like to thank our subjects for listening patiently for a total of 15960 sound sources.

References

[1] Laws, P., Platte, H.-J. (1975) "Spezielle Experimente zur Kopfbezogenen Stereophonie", Fortschritte der Akustik, DAGA '75, 365-368.

[2] Møller, H., Sørensen, M. F., Jensen, C. B., Hammershøi, D. "Binaural technique: Do we need individual binaural recordings ? To be published in J. Audio Eng. Soc.

[3] Morimoto, M., Ando, Y. (1980) "On the Simulation of Sound Localization", J. Acoust. Soc. Jpn. (E) **1**(3) 167-174.

[4] Wenzel, E. M., Wightman, F., Kistler, D., Foster, S. (1988) "Acoustic origins of individual differences in sound localization behavior", J. Acoust. Soc. Am. Suppl. **84** S79.

[5] Wenzel, E. M. (1992) "Localization in virtual acoustic displays", Presence **1**(1) 80-107.

[6] Kawaura, J., Suzuki, Y., Asano, F., Sone, T. (1989) "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear", J. Acoust. Soc. Jpn. (J) **45** 756-766.

[7] Kawaura, J., Suzuki, Y., Asano, F., Sone, T. (1991) "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear", J. Acoust. Soc. Jpn. (J) **12**(5) 203-216.

[8] Begault, D., Wenzel, E. M. (1991) "Headphone localization of speech stimuli", Proc. 35th meeting Human Factors Soc. 82-86.

[9] Begault, D. R. (1992) "Perceptual effects of synthetic reverberation on three-dimensional audio systems", J. Audio Eng. Soc. **40**(11) 895-904.

[10] Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L. (1993) "Localization using nonindividualized head-related transfer functions", J. Acoust. Soc. Am. **94**(1) 111-123.

[11] Hammershøi, D. (1995) "Binaural technique - a method of true 3D sound reproduction", Ph. D. thesis, Aalborg University, Denmark, Aalborg University Press, ISBN 87-7307-516-7.

[12] Møller, H., Sørensen, M. F., Hammershøi, D., Jensen, C. B. (1995) "Transfer Characteristics of Headphones Measured on Human Ears", J. Audio Eng. Soc., **43**(4) 203-217.

[13] Ross, S. M. (1987) "Introduction to probability and statistics for engineers and scientists", John Wiley and Sons, pp. 230-234.

[14] Møller, H., Jensen, C. B., Sørensen, M. F., Hammershøi, D. "How well do existing binaural systems work ?", in preparation.

A. Appendix

This appendix presents a comparison of the real life performance of the two groups of subjects: 1) the 8 subjects who participated in the previous investigation, and 2) the 12 new subjects. From Table I it is seen that there was no difference between the two groups.

Errors: / Condition:	out of cone	within cone	median plane	distance	total number of stimuli
Group of 8 subjects	**0.2%** 2 (912)	**0.3%** 1 (336)	**16.0%** 77 (480)	**11.9%** 40 (336)	912
Group of 12 subjects	**0.6%** 8 (1358)	**1.0%** 5 (504)	**15.1%** 109 (720)	**9.3%** 47 (504)	1368

Table I

Comparison of the real life performance (experiment A) for the group of 8 subjects used earlier [2], and the group of 12 new subjects. Errors are given in percentage (bold) and numbers. Number of stimuli that can result in errors in a category are given in brackets. Statistical tests did not show any significant difference between the two groups (two-sided Fisher-Irwin test at 5% level).

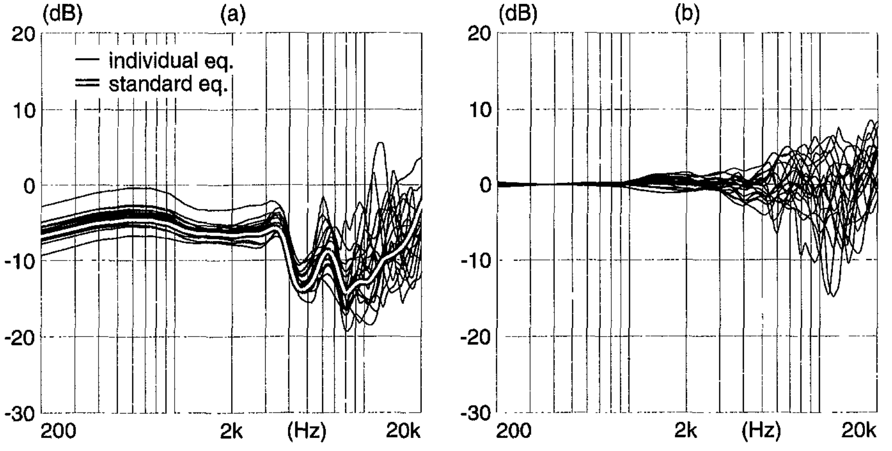

Figure 1

(a) Individual equalization filters for 20 subjects (thin black lines), non-individual equalization filter (heavy white line). (b) Error made for each of the 20 subjects when using non-individual equalization, i. e. difference between non-individual and individual equalization (level differences ignored).

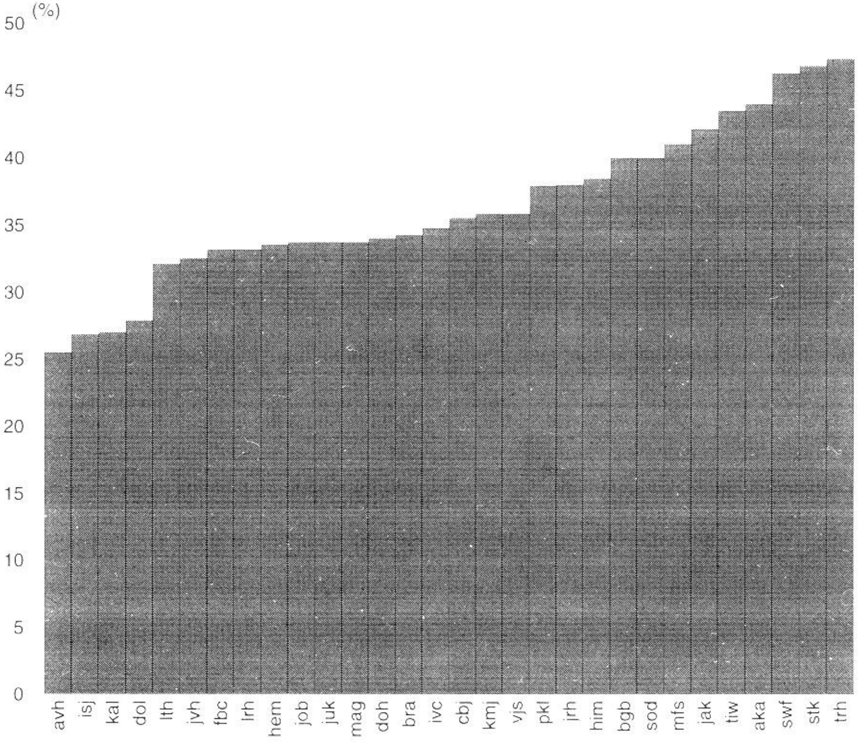

Figure 2

Median-plane errors in experiment B shown for individual recording heads. The recording heads are ranked according to error percentage.

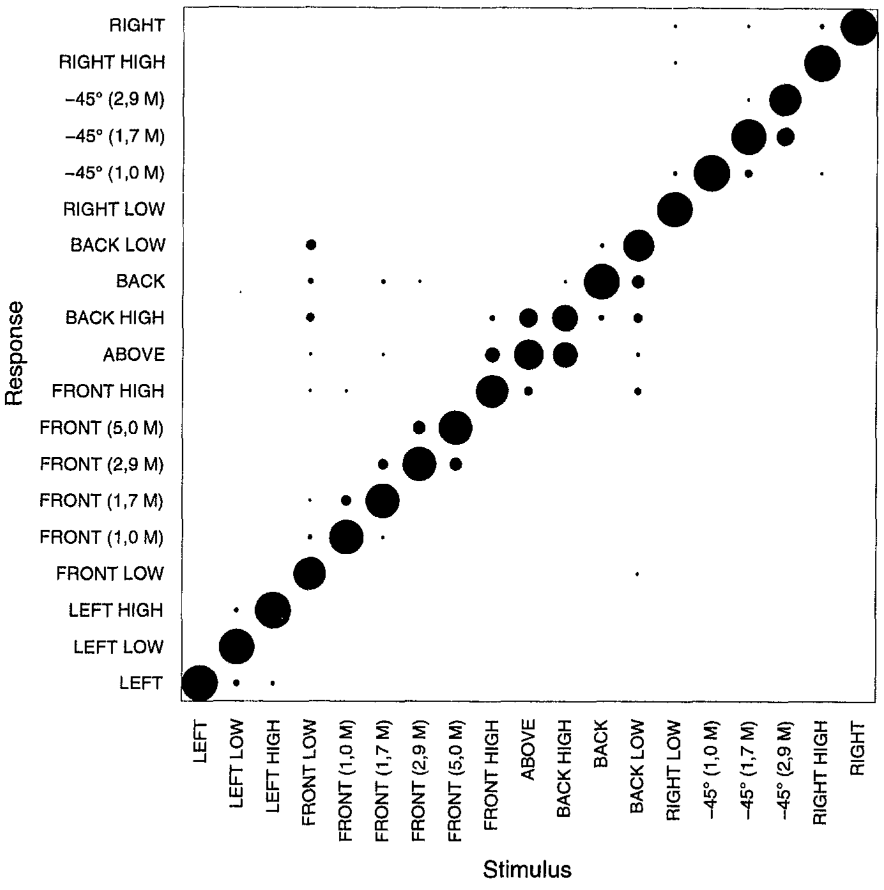

Figure 3

Real life listening (experiment A, 2280 stimuli). The area of each circle is proportional to the number of answers for the particular combination of stimulus and response. A "full" circle, e. g. at (LEFT, LEFT), corresponds to the total number of stimuli for each stimulus (which is 120 in this case).

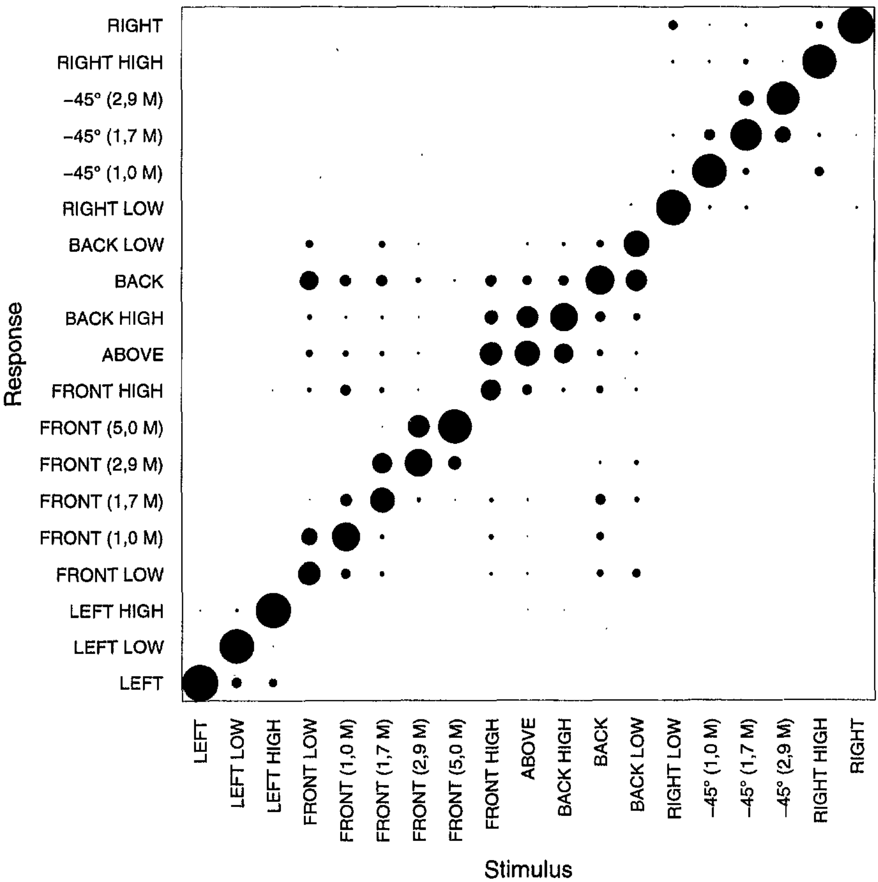

Figure 4

Non-individual recordings using a random subject (11020 stimuli from experiment B, excluding 380 individual stimuli).

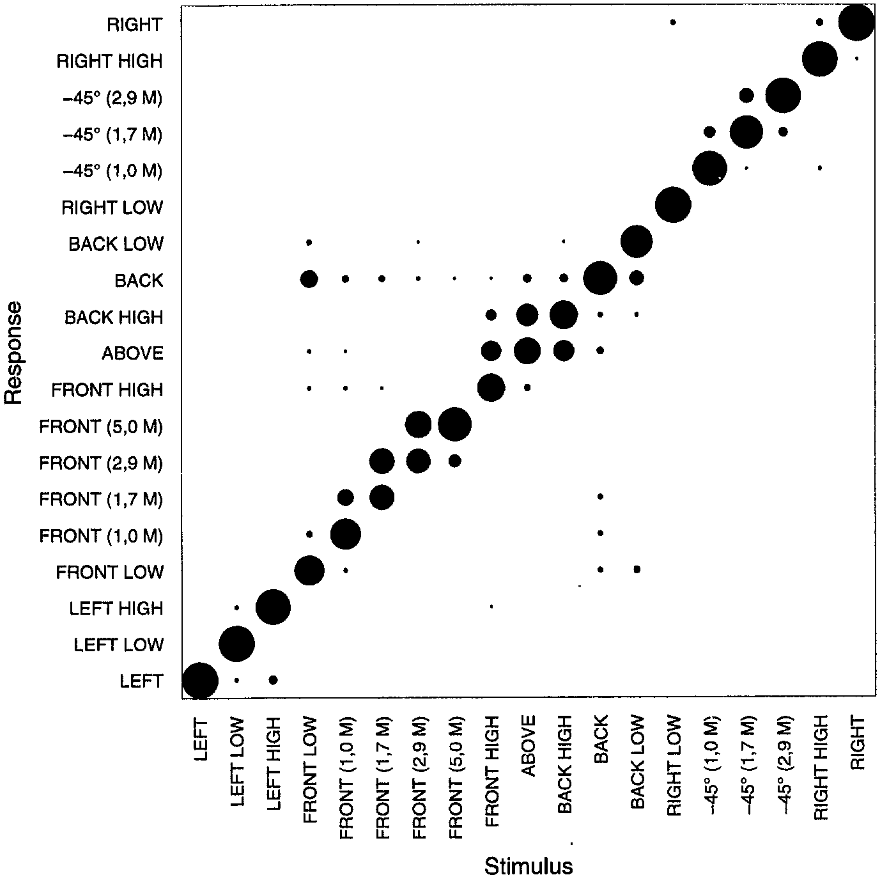

Figure 5

Non-individual recordings using a typical subject and individual equalization (experiment C, 2280 stimuli).

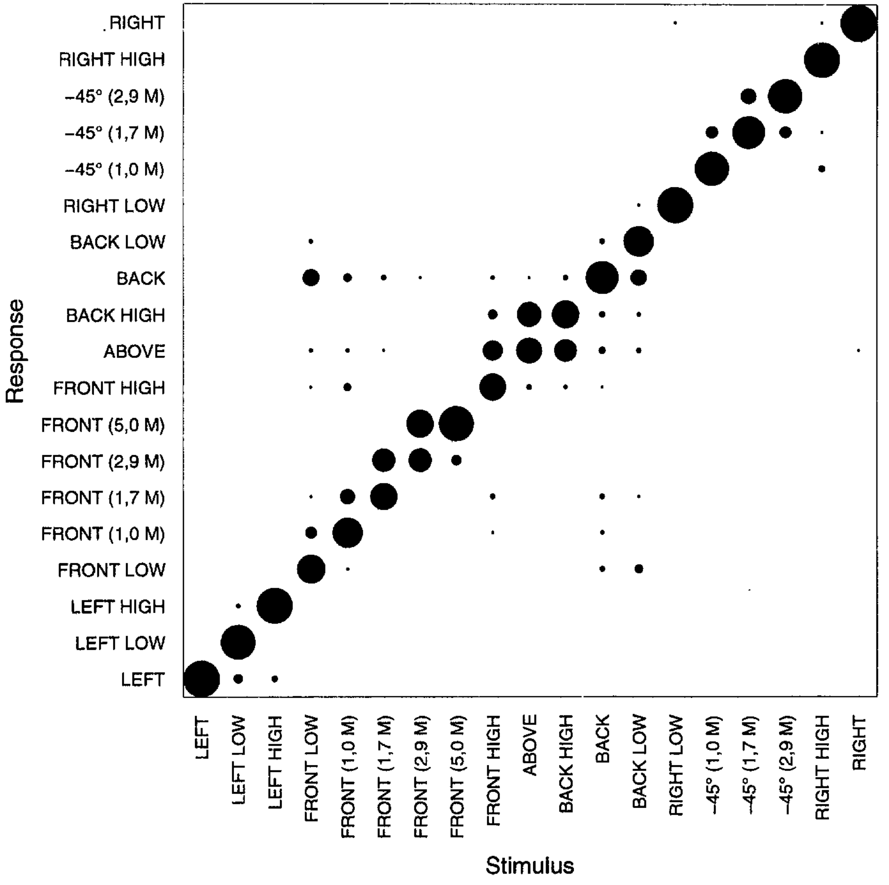

Figure 6

Non-individual recordings using a typical subject and non-individual equalization (experiment D, 2280 stimuli).