Binaural recording using a typical human subject

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

Published in:
Proceedings of Nordic Acoustical Meeting, NAM'96, Helsinki, Finland, June 12-14, 1996

Publication date:
1996

NAM 96 Nordic Acoustical Meeting
Helsinki, 12 – 14 June 1996

E3

BINAURAL RECORDING USING A TYPICAL HUMAN SUBJECT

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

Aalborg University, Acoustics Laboratory
Fredrik Bajers Vej 7 B4, DK-9220 Aalborg Ø, Denmark
acoustics@kom.auc.dk

1 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.

The present paper only reviews the study, which is presented in more detail in Møller et al. [12]. In the original publication also results with individual and non-individual equalization can be seen. The reader is also referred to Møller et al. [2] and [12] for details in methods and results.

2 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.

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.

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).

A headphone (Beyerdynamic DT 990 Professional) with FEC proporties [13] 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.

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

2.1 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.

2.2 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 × 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 × 30 recording heads × 19 loudspeakers).

2.3 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.

3 RESULTS AND DISCUSSION

For statistical analysis errors have been 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. [14]). In order to give the most powerful tests, only stimuli that could actually lead to errors in a certain category were included in each test, and in the calculation of error percentages.

In general for the present experimental design, *out-of-cone errors* and *within-cone errors* are quite rare, whereas *distance errors* and especially *median-plane errors* are frequent. The number of *median-plane errors* for a particular experiment has been found to be indicative for the outcome, and for simplicity only comparisons of median-plane errors are presented in this paper (the complete statistical analysis is presented in the original publication [12]).

The results for the median plane sources are presented in 10 by 10 matrices in Figure 1.

3.1 Real life

The results of real life listening (experiment A) can be found in Figure 1(a). Correct answers are to be found in 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. 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.

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.

3.2 Recordings from random subjects

The results for non-individual recordings (experiment B) including all recording heads are shown in Figure 1(b). 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).

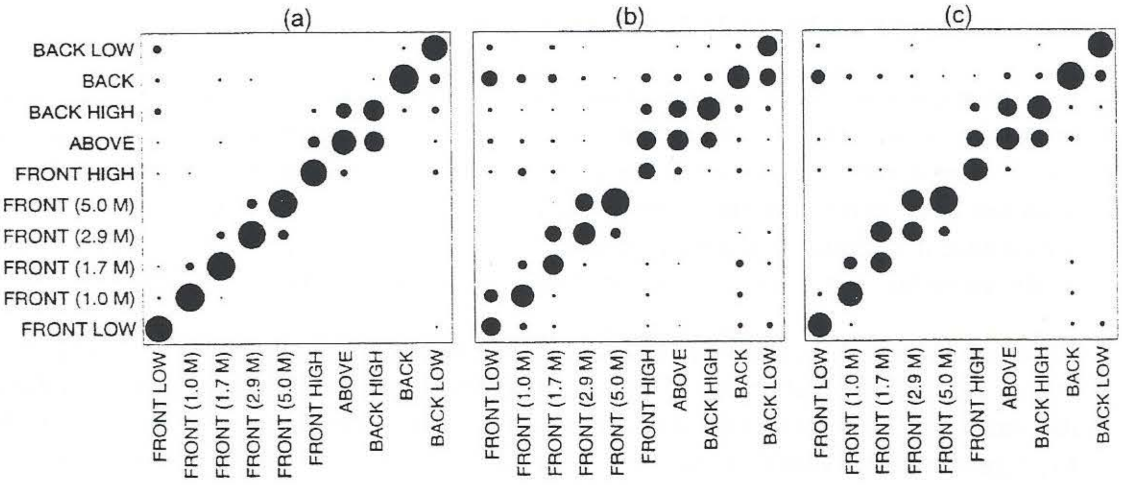

Figure 1. Response (ordinate) versus stimulus (abscissa) for the median plane sources for (a) real life, (b) recordings for random subjects, and (c) recording from typical subject. The area of each circle is proportional to the number of answers for the particular combination of stimulus and response. Correct answers are found in the diagonal.

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, 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.

The actual number of median-plane errors is 2103 out of 5800, which is equivalent to 36.3%. This is significantly higher (at 0.1% level in a one-sided test) than the corresponding number of errors in the real life situation (186 out of 1200, which is equivalent to 15.5%).

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.

Considerable variation is present between the results from different recording heads (ranging from 25.5% to 47.4% median-plane errors). Since median-plane errors constitute the largest group of errors, these were used for the selection of a typical head. 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).

3.3 Recordings from typical subject

The results from recordings with the selected 'typical' subject (experiment C) are shown in Figure 1(c). 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 1(b). Further analysis 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).

The rate of median-plane errors in experiment C is 21.2% (254 out of 1200), 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 part of experiment B belonging to the typical head was not significant (two-sided Fisher-Irwin test at 5% level).

3.4 How far are we from real life ?

It is also interesting to compare the results with the typical head with the results from the real life situation to see how far we are from real life. In the real life situation there were 15.5% median-plane errors, and statistical test reveal that this is significantly less (one-sided test at 0.1% level) than the 21.2% for the experiment with the typical head.

A comparison of Figure 1(c) and Figure 1(a) 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. 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.

3.5 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.

4 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 typically 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.

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 [15].

5 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.

6 REFERENCES

1. Laws, P., Platte, H.-J. 1975. Spezielle Experimente zur Kopfbezogenen Stereophonie. *Fortschritte der Akustik, DAGA '75*, pp. 365-368.

2. Møller, H., Sørensen, M. F., Jensen, C. B., Hammershøi, D. Binaural technique: Do we need individual binaural recordings ? *J. Audio Eng. Soc* (to be published).

3. Morimoto, M., Ando, Y. 1980. On the Simulation of Sound Localization. *J. Acoust. Soc. Jpn. (E)*, vol. 1, no. 3, pp. 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.*, vol. 84, S79.

5. Wenzel, E. M. 1992. Localization in virtual acoustic displays. *Presence*, vol. 1, no. 1, pp. 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)*, vol. 45, pp. 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)*. vol. 12, no. 5, pp. 203-216.

8. Begault, D., Wenzel, E. M. 1991. Headphone localization of speech stimuli. *Proc. 35th meeting Human Factors Soc.*, pp. 82-86.

9. Begault, D. R. 1992. Perceptual effects of synthetic reverberation on three-dimensional audio systems. *J. Audio Eng. Soc.*, vol. 40, no. 11, pp. 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.*, vol. 94, no. 1, pp. 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., Jensen, C. B., Hammershøi, D., Sørensen, M. F. 1996. Using a typical human subject for binaural recording. *Proc. 100th Audio Eng. Soc. Conv.*, Copenhagen, May 11-14 1996.

13. Møller, H., Sørensen, M. F., Hammershøi, D., Jensen, C. B. 1995. Transfer Characteristics of Headphones Measured on Human Ears. *J. Audio Eng. Soc.*, vol. 43, no. 4, pp. 203-217.