Evaluation of artificial heads in listening tests

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

Published in:
Proceedings of 102nd Audio Engineering Society Convention, March 22-27, 1997, Munich, Germany

Publication date:
1997

Citation for published version (APA):
Møller, H., Jensen, C. B., Hammershøi, D., & Sørensen, M. F. (1997). Evaluation of artificial heads in listening tests. In Proceedings of 102nd Audio Engineering Society Convention, March 22-27, 1997, Munich, Germany

Evaluation of Artificial Heads in Listening Tests **Preprint 4404 (A1)**

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

Presented at the 102nd Convention 1997 March 22–25 Munich, Germany

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

EVALUATION OF ARTIFICIAL HEADS IN LISTENING TESTS

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

The localization performance was studied when subjects listened 1) to a real sound field, and 2) to artificial head recordings of the same sound field. The experiments took place in a standard listening room where each stimulus (female speech) was emitted from one of 19 loudspeakers, and the subjects were to indicate the perceived sound source. The artificial head recordings were made a) by the artificial heads' built-in microphones, and b) by blocked ear canal microphones. The reproduction was carried out by carefully equalized headphones. Eight artificial heads were included in the investigation, and twenty subjects participated, except for the experiment with recordings from built-in microphones which was performed for eight subjects.

When compared to real life, the localization performance with the artificial heads resulted in an increased number of errors independent of recording technique. In general, the directions in the median plane were frequently confused, not only with nearby directions, but also with directions further away. For some artificial heads there was also an increase in confusions of directions outside the median plane. The number of errors was significantly higher than what can be obtained with recordings from a carefully selected human head. This encourages the design and production of improved artificial heads.

0. Introduction

During the last decade the binaural technique has been used increasingly for recordings of noise and music. The main reason is the capability of the technique to store and reproduce three-dimensional aspects of a sound field by means of only two channels.

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

In most practical applications recordings cannot be made in the listener's own ears, and during recording the listener is usually replaced by a copy of a human head, called an *artificial head* or a *dummy head*. Several artificial heads are commercially available, and they do produce recordings with a surprisingly natural reproduction. Nevertheless, it is often claimed that the three-dimensional space is distorted, and that localization errors occur. Often mentioned imperfections are: Confusions between front and back, elevation of sound sources, sound sources perceived too close, and in-the-head localization.

These shortcomings have mainly been observed in rather informal listening tests, and their existence has not been proven in large-scale scientific experiments. Nevertheless, it is generally accepted that the imperfections do exist, and this fact seems to limit the success of the artificial head recording technique. It is the purpose of this investigation to study the quality of artificial head recordings. Note that the investigation thus constitutes an evaluation of the binaural technique when the recordings are made with *artificial heads*. We have previously carried out similar evaluations of the technique with recordings made in *human ears* [1], [2].

0.1 Aim of investigation

It is the aim of the present investigation to clarify whether the claimed imperfections of artificial head recordings are real and - if they are confirmed - to clarify to what extent they occur for a range of artificial heads. It is furthermore the aim - if feasable - to explain possible imperfections in the reproduction with specific defects in the design of the artificial heads.

The investigation includes two parts: 1) a study of the localization performance with artificial head recordings (listening tests), and 2) a study of head-related transfer functions (HRTFs) for the artificial heads. Only the first part is reported in the present paper. The second part is to be included in a complete report of the study [4].

0.2 Artificial heads included in the investigation

Seven different artificial heads from five producers were included in the investigation. Some of them could be modified in various details, which made a total of eleven different configurations.

0.2.1 Knowles Electronics Inc.

The *K*nowles *E*lectronics *M*annikin for *A*coustic *R*esearch (KEMAR) consists of a head and a torso originally designed from anthropometrical data and for the purpose of general acoustic research. It includes ear simulators according to IEC 711 [5] (and ANSI S3.25 [6]) for which the sound pressure at the microphone (Brüel & Kjær 4134) is intended to correspond approximately to the sound pressure existing at the human eardrum. Four different human-like pinnae are provided, ordered from smaller to larger (two sides): DB060/061, DB065/066, DB090/091, DB095/096. In the following the KEMAR equipped with these pinnae (in the same order) are denoted: KEMAR 1, KEMAR 2, KEMAR 3, and KEMAR 4.

0.2.2 Georg Neumann GmbH

Both of Georg Neumann GmbH's artificial heads, KU80i and KU81i, have head and pinnae with a human-like shape. As the only ones in the investigation they do not include a torso. The heads have a 4 mm ear canal leading to acoustical networks and microphones (KM83 for KU80i and KK83 for KU81i). The KU80i is designed to simulate the recording of the sound at a certain point in a human ear canal, whereas the acoustical network of KU81i is designed for equalizing the frequency response to a flat response for diffuse sound incidence.

It should be noted that at the time of the investigation Neumann had announced their artificial head KU100. Unfortunately the company was not able to lend us a sample for a few days to make recordings. The head was not yet commercially available, hence we were not able to include it in the investigation.

0.2.3 HEAD acoustics GmbH

The two artificial heads of Head Acoustics GmbH are marketed as parts of complete recording and playback systems. The heads are very different. They both include head, shoulder and pinnae, but the head of the recording system HMS I is a replica of a specific human, whereas the head of the recording system HMS II is a model of very simplified geometry. Even the pinnae have been highly simplified. For both heads, a box with electronic equipment make up the part of the torso below the shoulders. These heads have a 4 mm ear canal terminated by the recording microphones Brüel & Kjær 4166 (HMS I) or 4165 (HMS II). The HMS I head was constructed at Rheinisch-Westfälischen Technischen Hochschule, Aachen, Germany, and it is sometimes denoted the "Aachener Kopf".

The playback systems (HPS II for HMS I, HPS III for HMS II) include playback equalizers to be used with a Stax SR lambda professional headphone. Recording and playback equalizers have settings for free-field equalization as well as a so-called "*i*ndependent of *d*irection (ID)" equalization.

0.2.4 Brüel & Kjær

Brüel & Kjær's two artificial heads have the same outer shape, consisting of a head and a torso, both of simplified geometry, and equipped with human-like pinnae. The 4128 includes ear simulators according to IEC 711 [5] (and ANSI S3.25 [6]) in which the simulated eardrum pressure is recorded by a Brüel & Kjær 4134 microphone. At the 5930 no attempt is made to simulate the ear canal, and studio microphones 4009 are mounted nearly flush with the ear canal entrance.

0.2.5 University of Toronto

This head is designed at the Institute of Biomedical Engineering at University of Toronto. It has a human-like shape, includes human-like pinnae (DB065/066 from a KEMAR), and fits onto a KEMAR torso, with which it was used in the present investigation. It is originally made for measurement of earplug attenuation and hence has an accurate simulation of ear canal tissue, but no microphones are provided with the head. For the purpose of this investigation miniature microphones (Sennheiser KE 4-211-2) were mounted in the ear canals, flush with the ear canal entrances. The head is only produced on request and has no type number, hence in the following it is simply denoted Toronto.

1. Method

By means of psychoacoustic experiments the localization performance was evaluated in real life and when listening to artificial head recordings of the same sound field. The experimental method was in most respects identical to that used in our previous investigation with binaural recordings made in the ears of human subjects [1]. 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.

The experiments included three listening conditions: 1) listening in real life, 2) listening to recordings made with the built-in microphones of the artificial heads, and 3) listening to recordings made with miniature microphones mounted flush with the entrance to the blocked ear canal of the artificial heads.

The latter condition constitutes a situation where the recordings are only influenced by the "outer" geometry of the head. The fact that this results in recordings with full spatial information was shown theoretically by Møller [7] and verified from measurements by Hammershøi and Møller [8], and in listening tests by Møller et al. [1].

1.1 Setup and procedure

The experiments were carried out in a standard listening room [9], 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 an artificial head recording, made in the same setup and reproduced by means of headphones. The subject was sitting in the setup in both situations 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.

Due to practical circumstances a few more curtains were present in the listening room during the recordings than during the experiments. Hence the reverberation time was slightly lower (approximately 0.3 s in contrast to the normal approximately 0.4 s). The possible consequence of this difference is discussed in Section 2.3, and it is verified that did not affect the results. Hence, the difference is ignored elsewhere in this paper.

1.2 Subjects

Twenty paid students with controlled normal hearing participated as listeners, ten 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 same subjects participated in a previous investigation on the selection of a typical human subject for binaural recordings [2]. Prior to that eight of the subjects had participated in our investigation on individual and nonindividual binaural recordings [1].

Some of the experiments (experiments B, Ax, and Ay) were only carried out for eight of the subjects. These subjects were the same eight subjects who had participated in the first experiment [1]. The real-life performance of these eight subjects did not deviate from the real-life performance of the other twelve subjects [2, appendix].

1.3 Recordings

Two sets of recordings were made for each artificial head: Recordings with the built-in microphones, and recordings with miniature microphones mounted flush with the entrance to the blocked ear canal. The type of miniature microphones and the mounting technique were the same as used earlier for measurements of head-related transfer functions on human subjects [10] and for making binaural recordings in the ears of human subjects [1]. Since the outer geometry is identical for the 4128 and 5930 heads, only one set of miniature microphone recordings was made for these. Recordings with the built-in microphones of the KEMAR were only made with one size of pinnae - the KEMAR 2.

The procedure used during recording was identical to that used in the earlier experiment with recordings in human ears [1], except that a Panasonic SV-3700 DAT recorder was used for intermediate storage before the recordings were put onto the disk. The recordings with built-in microphones of the HMS I head were made including the recording equalizer of the system in position "free-field", whereas the recordings with the HMS II head were made without the recording equalizer of the system.

1.4 Reproduction and equalization

Most of the manufacturers of artificial heads give only very unspecific instructions about the reproduction, which was then left to the experimenters' decision.

Headphones were chosen because of their almost complete channel separation. The total transmission should then include equalization for 1) the transfer function of the recording microphone, and 2) the transfer function of the headphone measured at the point in the ear canal where the recording is made [7]. The equalization was accomplished by digital IIR filters implemented at the signal processor board during reproduction.

1.4.1 Recordings with built-in microphones

There is in general some uncertainty related to which point in the human ear canal, the sound recorded by the built-in microphone represents. It was therefore considered necessary to make the measurement of the headphone transfer function at the artificial head. It was further decided to use an FEC headphone (a headphone with *f*ree-air *e*quivalent *c*oupling to the ambient, see Møller et al. [11]). It was shown earlier that this procedure leads to a cancellation of effects from a possibly incorrect simulation of the ear canal and eardrum [1, appendix].

A Beyerdynamics DT990 Professional headphone was chosen. We showed earlier that this headphone has approximate FEC properties [11]. The headphone was fitted onto each artificial head and the combined transfer function of the recording microphone and the headphone was found by measuring the transfer function from headphone input to microphone output. The inverse of the amplitude response was used as target in a Yule-Walker IIR filter design (maximum 32nd order, carried out in MATLAB, The MathWorks Inc.). The measured transfer functions were in general not minimum-phase transfer functions, and the equalization was thus only correct with regard to the amplitude response, since the inverted all-pass sections 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 transfer function would result in an infinite DC-gain of the target response. DC-values were therefore manually inserted in the target responses to obtain fairly flat responses at low frequencies. A corresponding problem exists close to the half sampling frequency, where the target responses were manually flattened. Manual modifications were 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 2nd order sections by pairing poles and zeros.

The KU80i had a severe lack of high frequency output and in order to avoid an extremely high filter gain at high frequencies, the target response was limited by a third order Butterworth low pass filter at 6.5 kHz. This resulted in an effective restriction of the frequency range for this head.

As mentioned earlier, the manufacturer specified complete playback methods for the HMS I and HMS II artificial heads. A Stax SR lambda professional headphone was to be used in combination with playback filters of the reproduction platforms HPS II and HPS III, respectively. These directions on headphone and equalizing filters were also used for the experiments, but for practical reasons the filters were implemented at the signal processor board in the same way as the filters for the other recording heads. Since the recordings for the HMS II head were made without the recording equalizer, the implemented filters include this equalizer as well. The magnitude of the HPS II playback equalizer in position "free-field" was obtained from the manufacturer in graphic form, whereas the data for the HMS II recording and HPS III playback equalizers (also position "free-field") were obtained from the manufacturer in terms of magnitude at 1/24-octave frequencies for the frequency range 20 Hz to 20 kHz.

For completeness, it should be mentioned that the manufacturer specifies recordings with KU81i to be reproduced without extra equalization using a diffuse-field equalized headphone. We showed earlier that the frequency responses of commercial headphones with a claimed specific equalization vary a lot [11], and none of the headphones included in that investigation fulfilled the design goal to just a reasonable extent. The general equalization procedure described above was therefore used also for KU81i.

Figure 1 shows the target responses and the implemented IIR filters for all artificial heads for the left and right ears. The figure also shows the target responses which would have been used for the HMS I and HMS II heads and the Stax headphone, had the general equalization procedure been used for these heads as well.

1.4.2 Recordings at blocked ear canal

Here the recording point was well defined and the equalization could be based on headphone transfer functions measured at the individual subjects. Also here it was decided to use an FEC headphone. It had been shown earlier [1, appendix] that this procedure minimizes the error which is inevitably introduced by the use of nonindividual recordings, such as artificial head recordings.

The Beyerdynamics DT990 Professional headphone was used also for these recordings. The design of the equalization filters was carried out as described in the preceding section, except that the headphone transfer functions were measured on the human subjects with the miniature microphones mounted at the blocked ear canal, and filters were designed individually for each subject. The filters were reported earlier ([1, Figure 4] eight subjects, [2] twelve subjects).

1.5 Experimental design

As mentioned already there were three experiments, one of real-life listening and two with listening to artificial head recordings accomplished with two different microphone techniques. In addition to the microphone technique, the latter two experiments differed in their statistical focus: one was carried out for a small group of subjects and with several repetitions in order to favor the power of statistical tests, whereas the other was carried out without repetitions but for a larger group of subjects to improve the general validity of the results.

1.5.1 Experiment A: Real life

Each subject listened to each loudspeaker 6 times. The experiment was divided into two sessions with 3 repetitions each. The stimulus order was random in each session. The sessions had a duration of approximately 10 minutes, and they were separated by a short break.

The number of stimuli for each subject was 114, giving a total of 2280 for the 20 subjects.

(The real-life experiment was the very same experiment as used and reported earlier for the whole group of subjects [2], and for the group of eight used in the present experiments B, Ax, and Ay [1]).

1.5.2 Experiment B: Recordings with built-in microphones

Each subject listened to each loudspeaker six times for each of the eight artificial heads included in this experiments. For a particular subject the experiment was carried out on four separate days. The experiments on a particular day were divided into two parts by a long break. Each part covered recordings from one head.

For each head the experiment was divided into two sessions with three repetitions in each session. The stimulus order was random in each session. The sessions had a duration of approximately 10 minutes, and they were separated by a short break. Only eight subjects participated in this experiment. The order of heads was randomized among the subjects.

The number of stimuli per subject was 912, giving a total of 7296 for all subjects.

1.5.3 Experiment C: Recordings at blocked ear canal

Each subject listened to each loudspeaker only once for each of the ten artificial heads included in this experiment. For a particular subject the experiment was divided into three sessions, separated by short breaks. Each session had a duration of approximately 10 minutes. For each subject the stimulus order was random for the whole experiment.

The number of stimuli per subject was 190, giving a total of 3800 stimuli for all subjects.

1.5.4 Experiments Ax and Ay: Real life control experiments

The experiments included a considerable number of stimuli for each subject and even when no feedback was given, the possibility of a learning effect should be considered. To facilitate an assessment of this, an extra real-life experiment, denoted experiment Ax, was carried out for eight of the subjects on a separate day at a late stage in the course of the study. At this time each of these subjects had participated in experiment A of the present study, experiments B, C, and D of our earlier study [1], experiment B of our earlier study [2], and experiments B and Ay of the present study (in the order mentioned).

Another possible source of error is the slightly different reverberation time of the listening room during recording and during the experiments. To facilitate an assessment of possible effects of this, an extra real-life experiment, denoted experiment Ay, was carried out with the same reverberation time as during the recordings. This experiment was carried out for eight of the subjects on a separate day just prior to experiment Ax.

Both of the experiments Ax and Ay were carried out like experiment A. The number of stimuli per subject was 114, giving a total of 912 for the eight subjects (for each of Ax and Ay).

2. Results and discussion

The way of presenting the results, performing statistical analysis etc. will to a wide extent follow the procedures used in our previous studies. Some details are therefore described only briefly and more information is found in [1].

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 test function follows a hypergeometrical distribution, and the test is called a Fisher-Irwin test (see e. g. [12]). One-sided tests are used, whenever the sign of a possible difference can be anticipated. 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.

Results from real-life listening are reported in Section 2.1. Section 2.2 gives the results from listening to artificial head recordings made with built-in microphones, including a comparison with real-life listening. Section 2.3 presents an evaluation of two possible sources of error: learning and deviation of reverberation time during recording. The results with blocked ear canal microphones are given in Section 2.4, including a comparison with real life. The two microphone techniques are compared in Section 2.5, whereas detailed results for each single head are given in Section 2.6. Section 2.7 presents a comparison between the present results with artificial head recordings and earlier results with nonindividual recordings made in human ears. Some further comments are given in Section 2.8.

2.1 Real life

The results of real-life listening (experiment A) are shown in Figure 2. The abscissa gives the stimulus position and the responded position is given as the 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 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 frequently confused, and sound coming from FRONT LOW and BACK LOW are sometimes perceived at various other directions in the median plane. Also wrong judgement of distance is a common error.

An important observation is that 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 with built-in microphones

The results from artificial head recordings with built-in microphones are seen in Figure 3. The figure is given for all heads together, thus the observations need not be valid for every single head.

It is obvious that considerably more errors are made than in real life. Much more confusions are seen between the upper median-plane directions (FRONT HIGH, ABOVE and BACK HIGH), and a remarkable increase is seen in the number of errors for the low median-plane sources (FRONT LOW, BACK LOW).

Additional directions have also come up with errors, of which some should be mentioned. In real life there were almost no errors for the sound sources in the FRONT direction (except for distance errors), whereas - for artificial head recordings - these sources are 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 BACK HIGH, ABOVE or FRONT HIGH directions.

More errors are also seen for all directions outside the median plane, except for LEFT and RIGHT, where no errors at all are seen for real life and for the artificial heads. It is believed that these directions are identified merely by the interaural time difference cue.

Figures similar to Figure 3 can be given for each artificial head. However, in order to save space we have chosen to present detailed information for each head only from experiment C, since experiment C included more subjects and thus a wider range of "fits" between humans and artificial heads. Detailed results for single heads are presented in Section 2.6.

At this place the difference between each artificial head and real life is analyzed statistically for each error category, and the results are given in numerical form in Table I, and in graphical form in Figure 4.

Errors: Condition:	out-of-cone	within-cone	median-plane	distance	total number of stimuli
Real life	**0.2%** 2 (912)	**0.3%** 1 (336)	**16.0%** 77 (480)	**11.9%** 40 (336)	912
KEMAR 2	**1.3%**** 12 (912)	**3.6%**** 12 (336)	**38.5%***** 185 (480)	**19.6%**(**) 66 (336)	912
KU80i	**1.3%**** 12 (912)	**21.7%***** 73 (336)	**50.2%***** 241 (480)	**20.8%**(**) 70 (336)	912
KU81i	**1.5%**** 14 (912)	**13.7%***** 46 (336)	**44.0%***** 211 (480)	**22.9%**(***) 77 (336)	912
HMS I	**2.7%***** 25 (912)	**0.9%** 3 (336)	**37.3%***** 179 (480)	**16.1%** 54 (336)	912
HMS II	**1.9%***** 17 (912)	**3.6%**** 12 (336)	**46.7%***** 224 (480)	**12.5%** 42 (336)	912
4128	**1.2%*** 11 (912)	**8.9%***** 30 (336)	**42.1%***** 202 (480)	**17.0%**(*) 57 (336)	912
5930	**1.1%*** 10 (912)	**3.9%***** 13 (336)	**48.5%***** 233 (480)	**16.1%** 54 (336)	912
Toronto	**0.8%** 7 (912)	**3.0%**** 10 (336)	**41.2%***** 198 (480)	**17.6%**(*) 59 (336)	912

Table I

*Comparison of real-life performance (experiment A less 12 subjects who did not participate in experiment B) and performance with recordings from artificial heads (recordings with built-in microphones, experiment B). 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 compared each of the artificial heads with real life (one-sided Fisher-Irwin assuming best performance in real life). *** indicates significance at 0.1% level, ** at 1% level, and * at 5% level. (See text - especially Section 2.3 - for significance levels given in brackets).*

It is seen that there is a higher percentage of errors for all heads and for all error categories. For median-plane errors the increase is significant at 0.1% level for all heads. The increase in out-of-cone errors and within-cone errors is significant at various levels for most of the heads. The same applies to distance errors. However, for distance errors an important observation which affects the statistical tests is reported in Section 2.3, so the reader is requested to ignore this for the moment. (For this reason the significance indications are put in brackets).

2.3 Effect of learning and reverberation time.

A possible effect of learning was studied in a repeated real-life experiment carried out at a late stage in the course of the study. Table II shows a comparison of real-life performance in the initial experiment (experiment A) and in the late experiment (experiment Ax).

Errors: / Condition:	out-of-cone	within-cone	median-plane	distance	total number of stimuli
Real life (experiment A)	**0.2%** 2 (912)	**0.3%** 1 (336)	**16.0%** 77 (480)	**11.9%** 40 (336)	912
Real life (experiment Ax)	**0.0%** 0 (912)	**0.0%** 0 (336)	**8.3%***** 40 (480)	**22.0%***** 74 (336)	912

Table II

*Comparison of real-life performance in the original experiment (experiment A) and at a later stage in the course of the study (experiment Ax). 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 two-sided Fisher-Irwin test.*

A striking decrease is observed for median-plane errors (significant at 0.1% level). This indicates a learning effect for these errors. An equally prominent increase of distance errors is seen (significant at 0.1% level). The latter is quite surprising. One might suspect that the explanation is that the decrease of median-plane errors (and also out-of-cone and within-cone errors) results in more potential distance errors. A detailed analysis has shown that this is not the reason, since the reduction of median-plane errors (plus out-of-cone and within-cone errors) hardly affects the directions where distance errors are possible (FRONT and -45°).

We have no other explanation for the changes of errors with time than learning of the directional cues combined with a reduced attention to the impression of distance.

A possible effect of the reduced reverberation time during recordings was studied in experiment Ay. Table III shows a comparison between results from this experiment and results from the late real-life experiment with "normal" reverberation. It is seen that there is no significant difference at all. This is not unexpected, since the difference in reverberation time was very small.

Errors: Condition:	out-of-cone	within-cone	median-plane	distance	total number of stimuli
Real life (experiment Ax)	**0.0%** 0 (912)	**0.0%** 0 (336)	**8.4%** 40 (480)	**22.0%** 74 (336)	912
Real life (experiment Ay)	**0.3%** 3 (912)	**0.0%** 0 (336)	**9.4%** 45 (480)	**16.4%** 55 (336)	912

Table III

Comparison of real-life performance with different reverberation times (experiment Ax and experiment Ay), both observed at a late stage in the course of the study. 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 conditions (two-sided Fisher-Irwin test at 5% level).

In conclusion of the analysis of these possible sources of errors it is accepted that the reverberation time did not influence the results. It is further recognized that - for real life - the number of median-plane errors have decreased and the number of distance errors have increased during the course of the study. In order to make the most conservative conclusions about artificial head recordings, further statistical comparisons with real life will use the initial value from experiment A for median-plane errors (and out-of-cone and within-cone errors), and a "long term" value for distance errors obtained as the pooled value of experiments Ax and Ay.

The result of a such modified comparison of distance errors is shown in Figure 5, and no significant deterioration of performance is now seen for any of the artificial heads. The indications of significance for distance errors in Table I and Figure 4 were put in brackets in anticipation of this result.

2.4 Recordings at blocked ear canal

The results from recordings at the blocked ear canal are given for the four error categories in Figure 6, which also includes a comparison with real life. The most noticeable observations are 1) an unmistakable increase in number of median-plane errors for all heads (significant at 0.1% level), 2) some increase of out-of-cone errors for most heads (significant at various levels), 3) some increase of within-cone-errors for most heads (significant at various levels), KU80i and KU81i being remarkably different, though, in showing very much increased values, and 4) no difference for distance errors. Detailed results for each head are given in Section 2.6.

2.5 Comparison of recording types

The significant deviations from real life were in general very similar for recordings with built-in microphones and with blocked ear canal microphones. It would therefore be natural to make a comparison of results for the two recording types. The result of this analysis is shown in Figure 7. It is seen that results are very similar for the two situations. Only two comparisons turned out to indicate a significant difference, a significance which we - taking into account the low significance level and the large number of comparisons - consider a coincidence. Hence we conclude that the recording point does not influence the localization performance when proper equalization is accomplished.

2.6 Detailed results for each artificial head

For each artificial head, information on responses to single source positions may be given for each of the recording types. As we have seen that there is no significant difference between the two types, we have - in order to save space - chosen to present results from only one recording type. Since the experiment with recordings at the blocked ear canal, experiment C, included more subjects and thus a wider range of "fits" between human and artificial heads, we have selected the results from this experiment for presentation. The diagrams for individual heads are given in Figure 8, Figure 9, Figure 10, Figure 11, and Figure 12.

In general the errors seen are similar to those mentioned in Section 2.2. Only a couple of observations specific for certain heads should be mentioned at this point. For KU80i and KU81i sound sources in the directions LEFT LOW and RIGHT LOW are often perceived in the directions LEFT HIGH and RIGHT HIGH, respectively, whereas these directions are almost always perceived correctly for the other heads. The HMS I, however, has also problems with these directions, since responses are often given at LEFT and RIGHT.

2.7 Comparison with recordings from human subjects

It is characteristic for artificial head recordings that they are not made in the ears of the individual listener. We showed earlier that this leads to reduced localization performance [1]. It is therefore interesting to compare the results with information on performance with nonindividual recordings from humans. In Figure 13 the results for the artificial head recordings are compared with the results obtained in a previous investigation for listening to recordings made in the ears of other human subjects chosen at random [2].

The localization performance for the artificial heads is seen to be in the same order of magnitude as for recordings with random human subjects. Some artificial heads, though, show more median plane errors. The many within-cone errors for the KU80i and KU81i still represent exceptions, as does the relatively high number of out-of-cone errors for the HMS I.

We also showed earlier that it is possible to obtain a significantly better localization performance if the nonindividual recordings originate from a carefully selected 'typical' human recording head [2]. Figure 14 compares the results obtained with the 'typical' human head to the artificial head results.

The surprising and encouraging observation is the low number of directional errors for the 'typical' human subject. Especially the number of median-plane errors is much lower than for any of the artificial heads. This means that careful selection of the head geometry should enable a greatly improved artificial head design.

(The number of distance errors for the KU80i turned out significantly lower than the number for the 'typical' human, which - taking into account the significance level and the number of comparisons carried out - is considered a coincidence).

The fact that recordings from a 'typical' human offers such an improvement is due to the relatively large difference between the 'best' and the 'worst' recording heads. In [2] the results for human subjects were displayed, and for comparison the artificial heads are ranked together with the human recording heads in Figure 15. As it can be seen, the artificial heads are found in the poorer part of the human heads with 60% (18) of the human heads being better than any artificial head.

2.8 Further comments

All the listening experiments conducted, required - for obvious reasons - that the listeners kept their heads still during playback. Although the subjects were asked to keep their heads still, and we further monitored the obeyance of this, it can be argued that the subjects might have induced minor head movements in the real-life situation, which could have assisted the localization and improved the performance. It is therefore important to notice that the same experimental procedure was used in our earlier experiments for reproduction of individual binaural recordings [1], in which case the localization performance was the same as in the real-life situation. There is thus no indictment that the localization performance in the real life situation was improved by the use of head movements.

3. Conclusion

We have shown that artificial head recordings do not result in the same localization as observed in real life. This is the conclusion for all artificial heads that were included in the investigation, and it proved valid both for recordings that were made by the artificial heads' built-in microphones, and for recordings made with a common microphone technique (blocked ear canal recordings).

The deteriorated localization performance was reflected in the increase of directional errors. Especially the number of confusions between the sources in the median plane increased much (from 15.5% in real life to 36.0 - 55.0% for the artificial heads).

It is known that nonindividual recordings in general result in more errors in judgement of directions, and the corresponding figure for nonindividual recordings from random human subjects is 36.3% (from a previous investigation [2]). The performance with artificial head recordings is thus either comparable to the performance obtained from recordings with random humans, or poorer.

The artificial heads could, however, be expected to better represent the 'typical' human subject, since the design goal for artificial heads ought to be the geometry that matches the average listener best. In [2] it was found that for the same experimental setup and procedure, a head could be selected which resulted in the better localization performance. The occurrence of median plane errors for this head was as low as 21.2%.

Although it is somewhat disappointing to conclude that the artificial heads at best can be compared to performance obtained with recordings from random human subjects, it is worthwhile to stress the superior reproduction that is obtained from binaural recordings, compared to other reproduction principles in general. It is further encouraging that it was possible to find a human for which recordings gave a better localization performance than for random human subjects or artificial heads. This ought to encourage the design and production of better artificial heads.

Acknowledgements

This work was supported by Brüel & Kjær A/S, Perceptive Acoustics A/S, the National Agency of Industry and Trade, Denmark, the Danish Technical Research Council, and Aalborg University.

The authors are further in dept to a number of persons who have made this investigation possible. First of all, we wish to thank manufacturers and institutions who have made the artificial heads available for the experiments. These are Knowles Electronics (for lending us the KEMAR with 4 different sets of ears), Danish Broadcasting (for lending us the KU80i), Institut für Rundfunktechnik in Munich (for lending us the KU81i), Volvo and Prof. Mendel Kleiner from Chalmers University of Technology in Gothenburg (for lending us the HMS I and bringing it to Aalborg), Head Acoustics GmbH (for lending us the HMS II, and providing equalizer data for HMS II, HPS II and HPS III), Brüel & Kjær (for lending us the 5930), and finally Prof. Hans Kunov from University of Toronto (for lending us the head denoted Toronto).

We further wish to thank our former colleagues Kim Alan Larsen and Jørn Vagn Hundebøll for their participation when the experiments were planned and carried out. We also wish to thank Anne Kirstine Andersen for her handling of all appointments with the subjects who took part in the experiments. Also Claus Vestergaard Skipper, the technical employee at this laboratory deserves appreciation for his valuable help with the practical work.

Finally, we wish to thank the twenty persons who came to our laboratory many times, and listened patiently to the playback of the speech segment.

References

[1] Henrik Møller, Michael Friis Sørensen, Clemen Boje Jensen, Dorte Hammershøi: Binaural technique: Do we need individual recordings ? Journal of the Audio Engineering Society, Vol. 44, No. 6, June 1996, pp. 451-469.

[2] Henrik Møller, Clemen Boje Jensen, Dorte Hammershøi, Michael Friis Sørensen: Using a typical human subject for binaural recording. In preparation. Preliminary report given in [3].

[3] Henrik Møller, Clemen Boje Jensen, Dorte Hammershøi, Michael Friis Sørensen: Using a typical human subject for binaural recording[3] . Proceedings of the 100th Audio Engineering Society Convention, Copenhagen, May 11-14, 1996, preprint 4157, pp. 1-18.

[4] Henrik Møller, Clemen Boje Jensen, Dorte Hammershøi, Michael Friis Sørensen: Evaluation of artificial heads in listening tests and by measurements of HRTFs. In preparation.

[5] IEC 711. Occluded-ear simulator for the measurement of earphones coupled to the ear by ear inserts. International Electrotechnical Commission, Geneva, Switzerland, first edition 1981.

[6] ANSI S3.25-1979 (R1986). American national standard for an occluded ear simulator. American National Standards Institute, New York, 1980.

[7] Henrik Møller: Fundamentals of binaural technology. Applied Acoustics, Vol. 36, No. 3/4, 1992, pp. 171-218.

[8] Dorte Hammershøi, Henrik Møller. Sound transmission to and within the human ear canal: Journal of the Acoustical Society of America, Vol. 100, No. 1, July 1996, pp. 408-427.

[9] IEC 268-13. Sound system equipment, pt. 13, Listening tests and loudspeakers. International Electrotechnical Commission, Geneva, Switzerland (1985).

[10] Henrik Møller, Michael Friis Sørensen, Dorte Hammershøi, Clemen Boje Jensen: Head-related transfer functions of human subjects. Journal of the Audio Engineering Society, Vol. 43, No. 5, May 1995, pp. 300-321.

[11] Henrik Møller, Dorte Hammershøi, Clemen Boje Jensen, Michael Friis Sørensen: Transfer characteristics of headphones measured on human ears. Journal of the Audio Engineering Society, Vol. 43, No. 4, April 1995, pp. 203-217.

[12] Ross, S. M.: Introduction to probability and statistics for engineers and scientists, John Wiley and Sons, pp. 230-234, 1987.

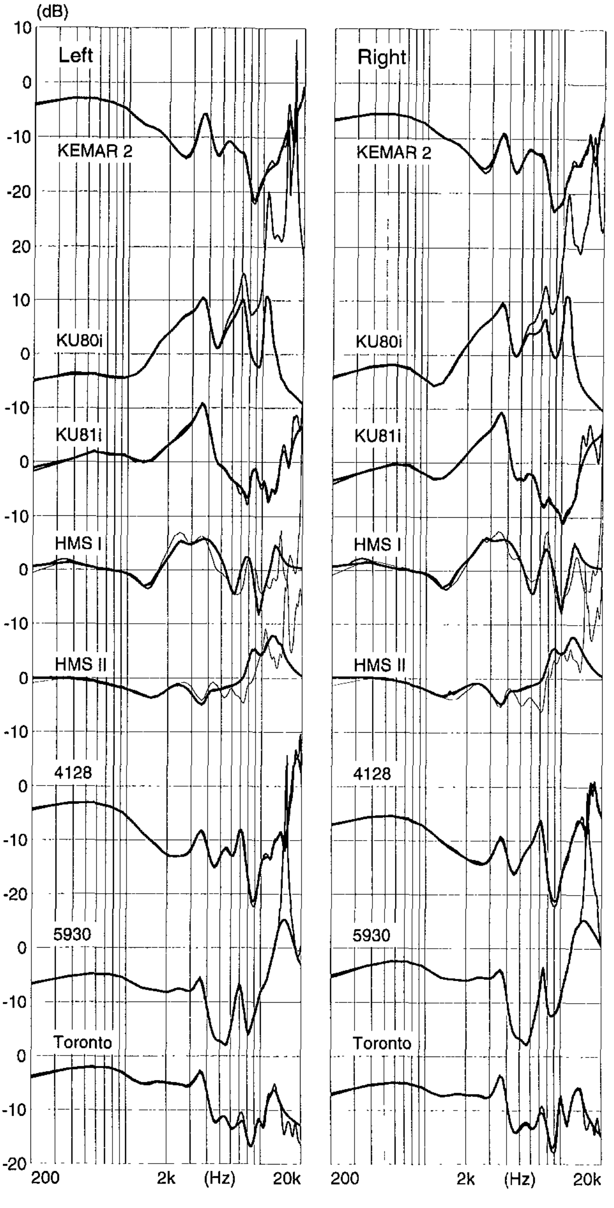

Figure 1

Equalization filters for reproduction of recordings with built-in microphones. Medium line indicates target, heavy line the implemented IIR filter. Headphones were Beyerdynamics DT990 Professional (except for HMS I and HMS II), or Stax SR lambda professional (for HMS I and HMS II). Thin lines indicate targets for HMS I and HMS II with the Stax headphone, had the general equalization procedure been used (see text) (for HMS I taking into account that the recording was made including the HMS I recording equalizer).

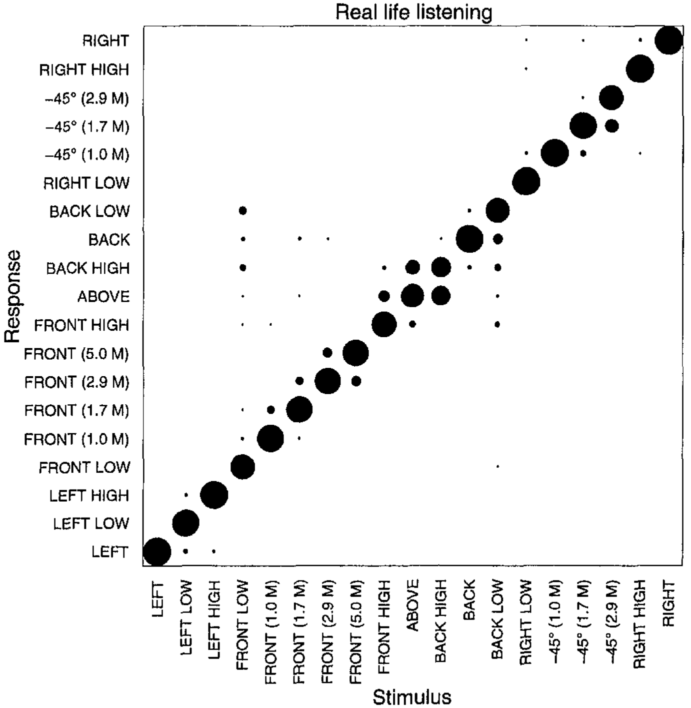

Figure 2

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 direction (which is 120 in this case).

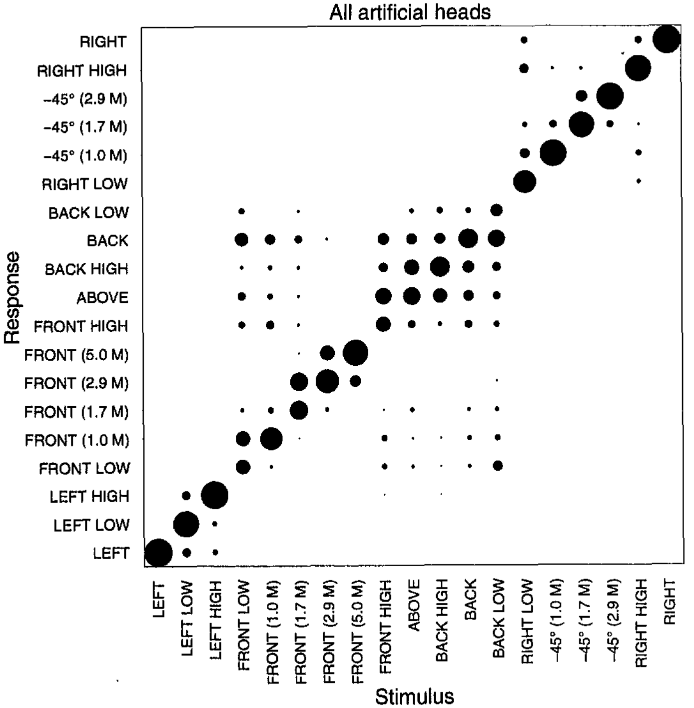

Figure 3

Artificial head recordings (experiment B, all heads, 7296 stimuli).

4404

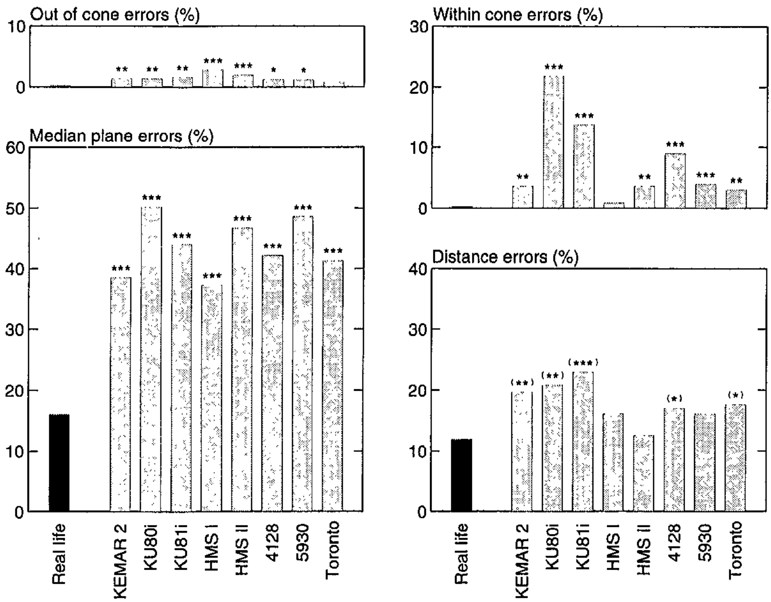

Figure 4

*Comparison of real-life performance (experiment A less 12 subjects who did not participate in experiment B, 912 stimuli) and performance with recordings from artificial heads (recordings with built-in microphones, experiment B, 912 stimuli for each head). Statistical tests compared each of the artificial heads with real life (one-sided Fisher-Irwin assuming best performance in real life). *** indicates significance at 0.1% level, ** at 1% level, and * at 5% level. (See text - especially Section 2.3 - for significance levels given in brackets).*

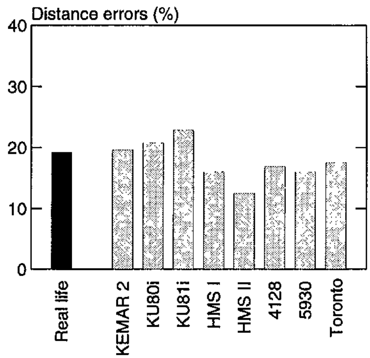

Figure 5

Comparison of real-life distance performance (experiments Ax and Ay, 1824 stimuli) and performance with recordings from artificial heads (recordings with built-in microphones, experiment B, 912 stimuli for each head). Statistical tests did not show significant difference between any of the artificial heads and real life (one-sided Fisher-Irwin test at 5% significance level assuming best performance in real life).

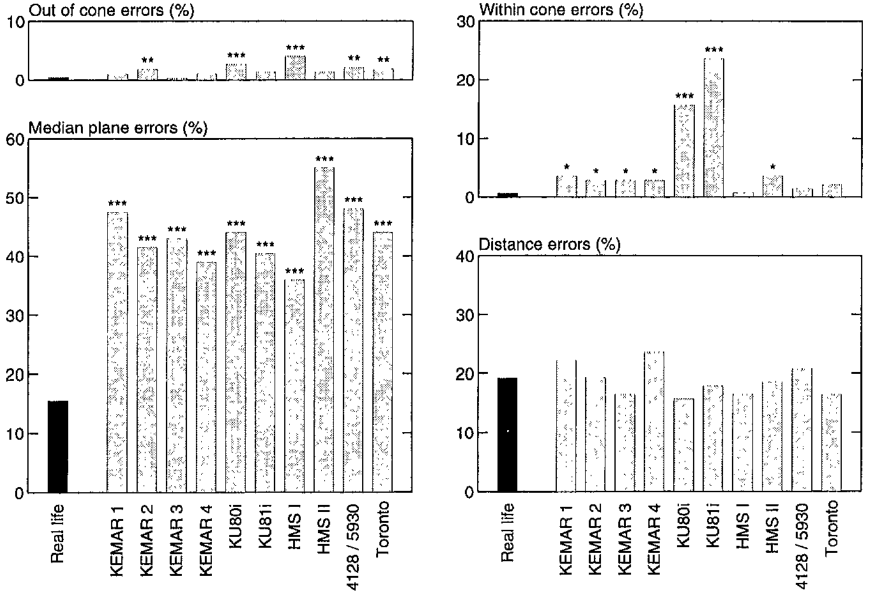

Figure 6

*Comparison of real-life performance (experiment A, 2280 stimuli) and performance with recordings from artificial heads (recordings with blocked ear canal microphones, experiment C, 380 stimuli for each head). Statistical tests compared each of the artificial heads with real life (one-sided Fisher-Irwin assuming best performance in real life). *** indicates significance at 0.1% level, ** at 1% level, and * at 5% level. (The comparison of distance errors was made with the pooled experiments Ax and Ay, 1824 stimuli, see Section 2.3).*

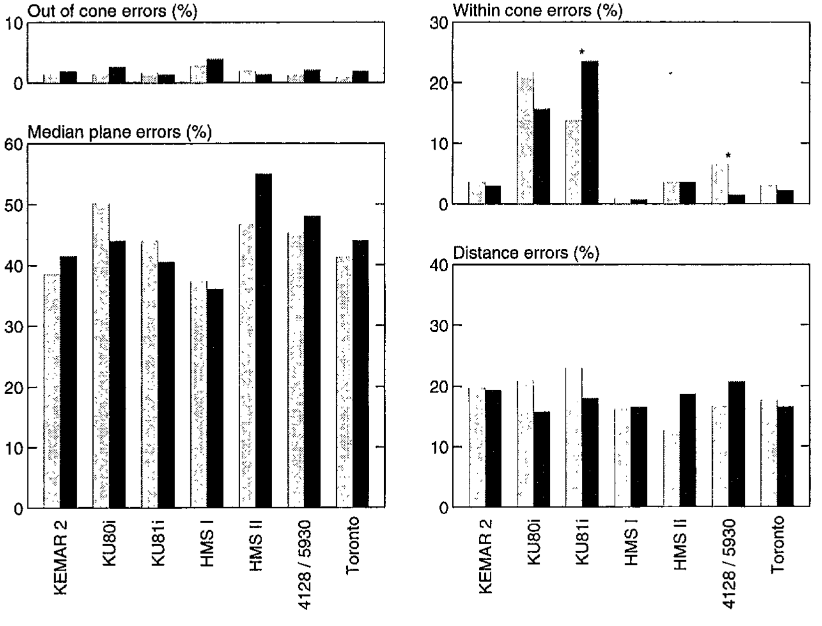

Figure 7

*Comparison of recording types. Light grey columns: Built-in microphones (experiment B, 912 stimuli for each head). Dark grey columns: Blocked ear canal microphones (experiment C, 380 stimuli for each head). Statistical tests compared the two recording types for each of the artificial heads (two-sided Fisher-Irwin tests). * indicates significance at 5% level.*

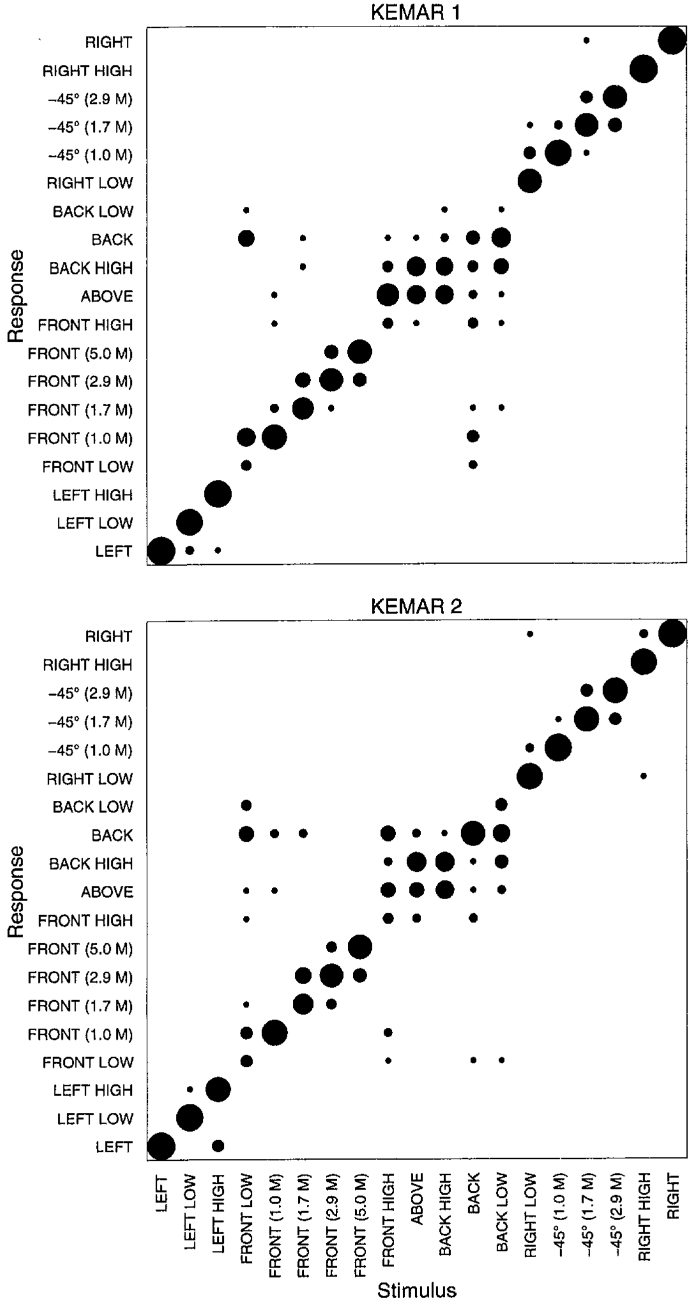

Figure 8

KEMAR 1 and KEMAR 2 with blocked ear canal microphones (380 stimuli from experiment C in each frame).

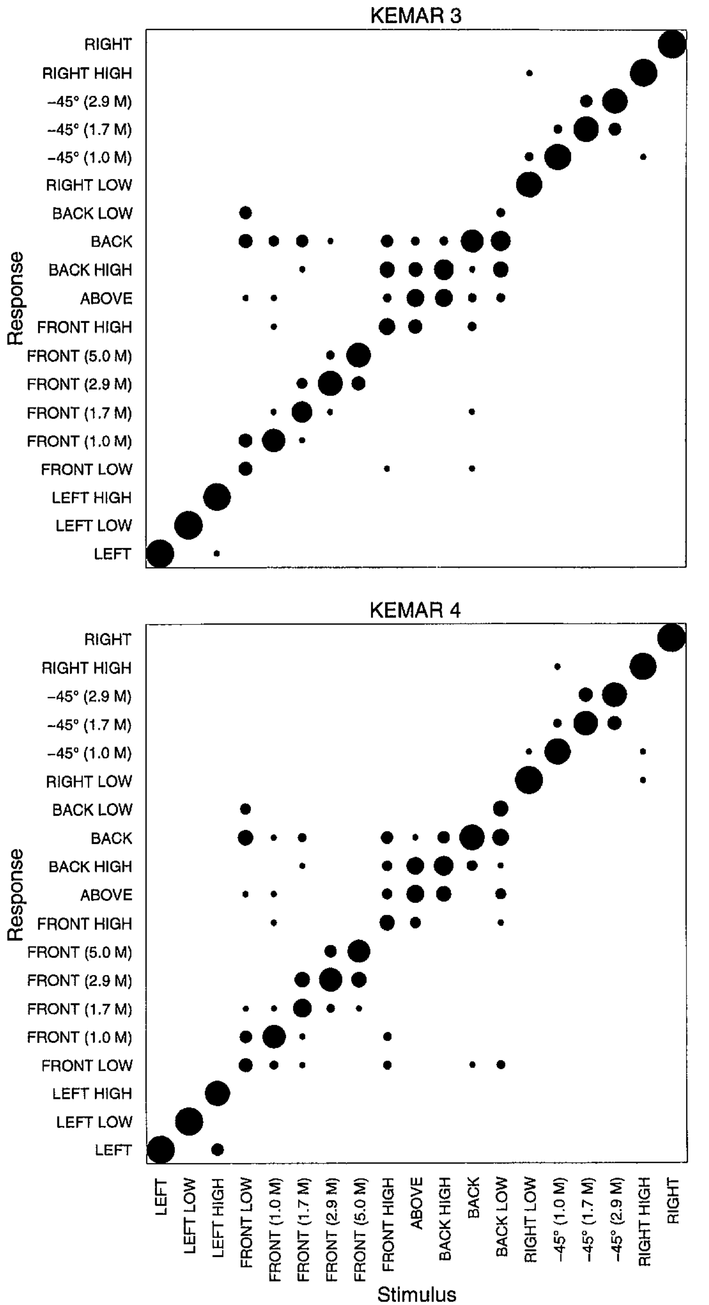

Figure 9

KEMAR 3 and KEMAR 4 with blocked ear canal microphones (380 stimuli from experiment C in each frame).

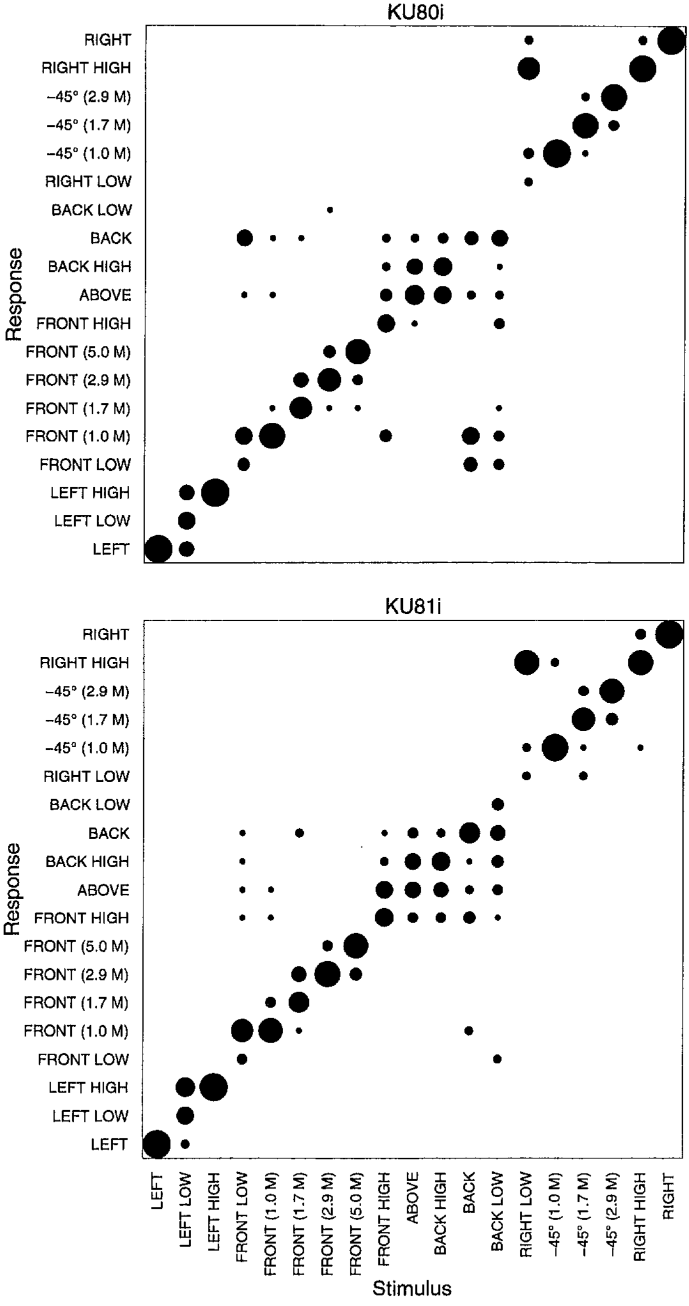

Figure 10

KU80i and KU81i with blocked ear canal microphones (380 stimuli from experiment C in each frame).

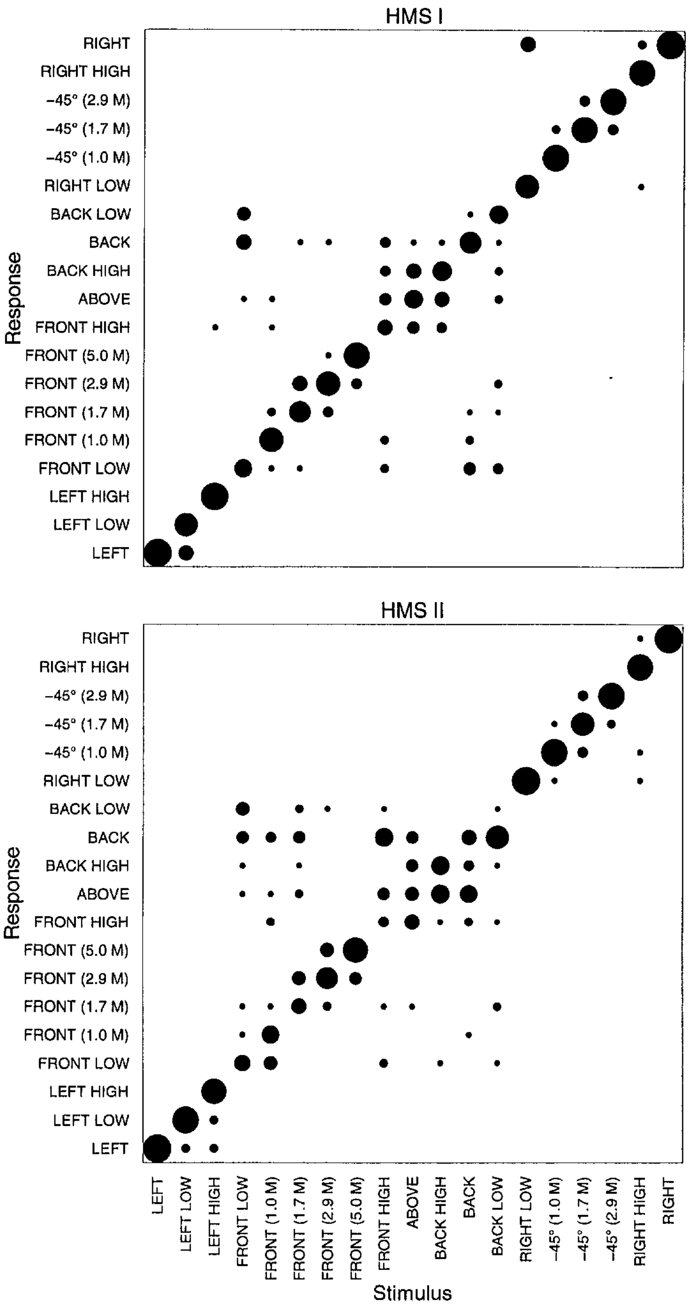

Figure 11

HMS I and HMS II with blocked ear canal microphones (380 stimuli from experiment C in each frame).

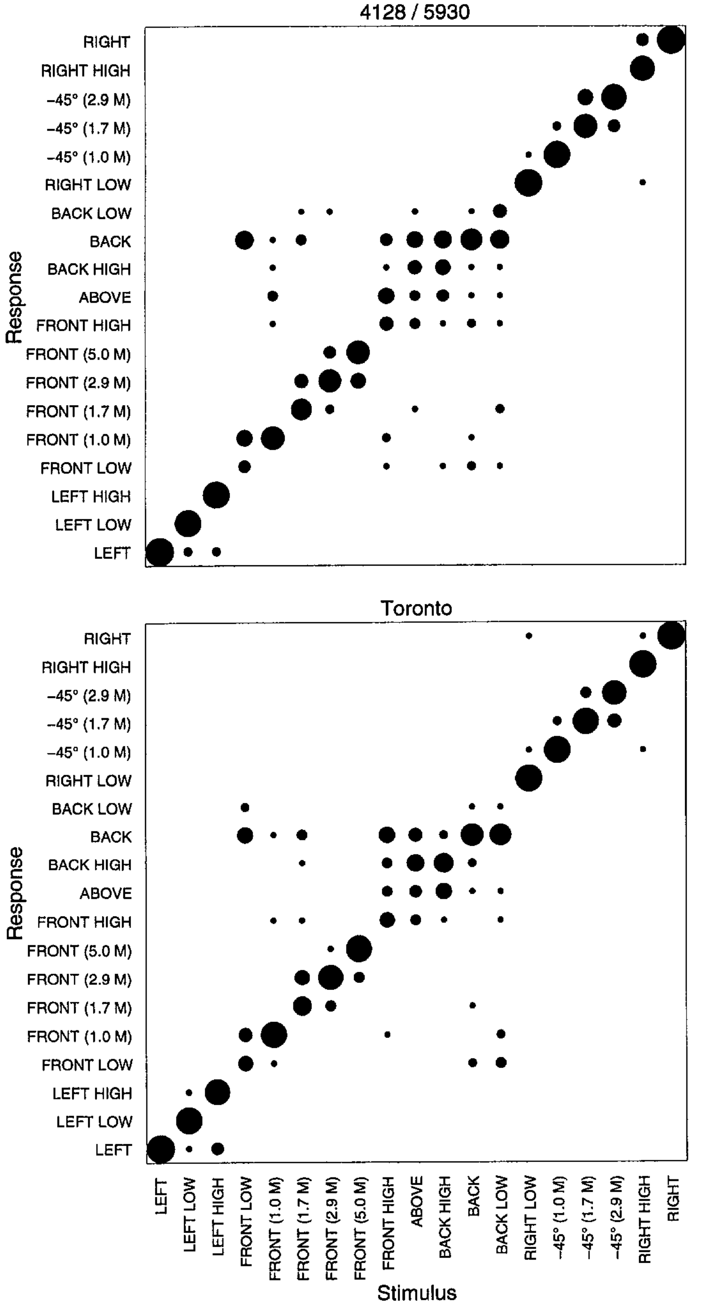

Figure 12

4128 / 5930 and Toronto with blocked ear canal microphones (380 stimuli from experiment C in each frame).

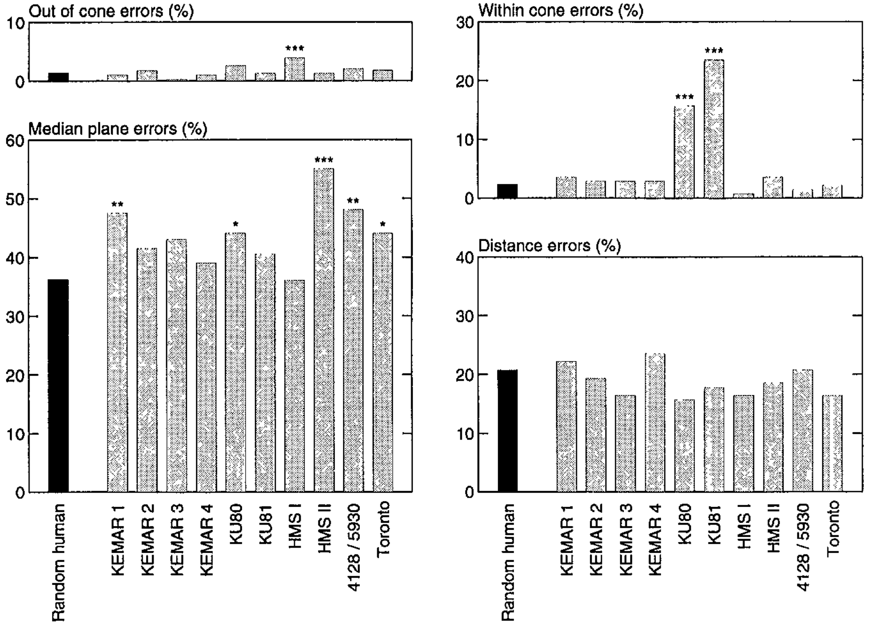

Figure 13

*Comparison of performance with recordings from artificial heads (recordings with blocked ear canal microphones, experiment C, 380 stimuli for each head) and earlier observed performance with recordings from random human subjects (experiment B of [2], 11020 stimuli). Statistical tests compared each of the artificial heads with random human subjects (two-sided Fisher-Irwin tests). *** indicates significance at 0.1% level, ** at 1% level, and * at 5% level.*

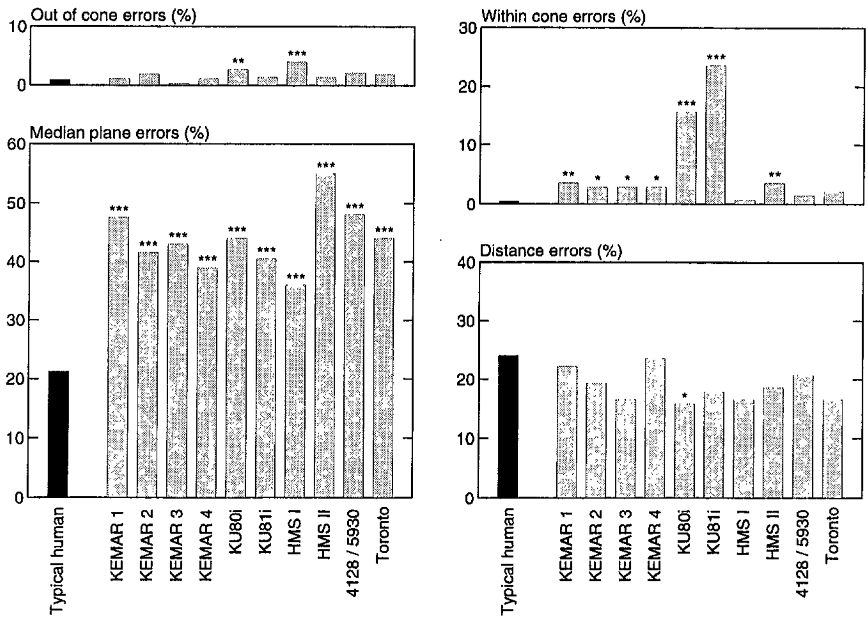

Figure 14

*Comparison of performance with recordings from artificial heads (recordings with blocked ear canal microphones, experiment C, 380 stimuli for each head) and earlier observed performance with recordings from a typical human subject (experiment C of [2], 2280 stimuli). Statistical tests compared each of the artificial heads with the typical human subject (two-sided Fisher-Irwin tests). *** indicates significance at 0.1% level, ** at 1% level, and * at 5% level.*

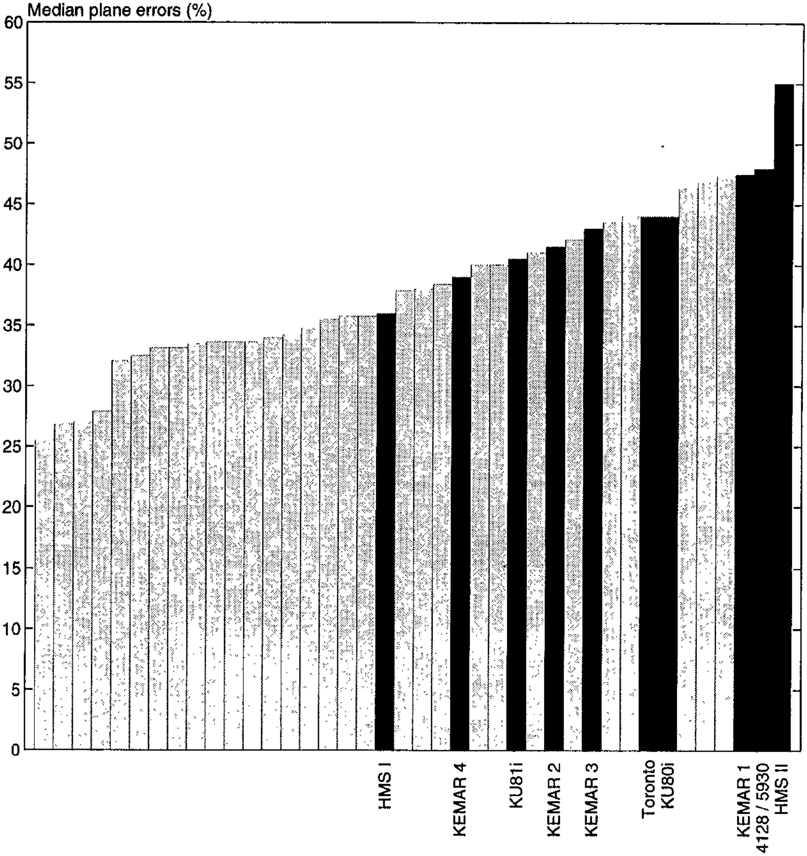

Figure 15

Median-plane errors of human and artificial recording heads (experiment B of [2] and experiment C of present investigation, respectively). The recording heads are ordered according to error percentage.