SUBJECTIVE SELECTION OF NON-INDIVIDUAL HEAD-RELATED TRANSFER FUNCTIONS

Bernhard U. Seeber and Hugo Fastl

AG Technische Akustik, MMK
TU München, Arcisstr. 21
D-80333 München, Germany
{seeber,fastl}@ei.tum.de

ABSTRACT

The individualization of head-related transfer functions (HRTFs) for virtual acoustics is a key technique for the optimization of the directional reproduction. This paper presents a subjective selection method for a fast, individual selection of one HRTF from a catalogue of non-individual ones. The selection method gives random access to sounds filtered with the HRTFs. In a first selection step a group of HRTFs is chosen out of which a final HRTF is singled out in a second step according to multiple criteria. The results of the two selection-steps were evaluated through a localization experiment. It is found that the selection minimizes the variance of the localization responses and the number of inside-the-head localizations. Localization error as well as the number of front-back confusions is small. As the selection method is fast, easy to implement, and operable for subjects without training, it is suitable for various applications, such as telephone conference systems or computer games.

1. INTRODUCTION

The perception of an auditory event, its direction, externalization, and focus, depends strongly on the individual adaptation of the virtual auditory display, i.e. of the personalization of the utilized set of head-related transfer functions (HRTFs). This can be optimally achieved by an individual measurement of the HRTFs. As the measurement is time-consuming and requires special equipment and know-how it is feasible only for specific applications. Implementing an averaged HRTF or an HRTF of a selected person can be a possible simple solution [1]. This will not, however, always lead to a perfect directional reproduction, as the variation of human physical parameters has a strong influence on the HRTFs [2, 3].

This paper therefore investigates whether an individual subjective selection can be used to optimize the perception and localization performance with non-individual HRTFs. In contrast to a former study, the aim of the current study is to develop a method and questionnaire for a fast, effortless to implement, and easy to handle selection procedure for HRTFs on the basis of common hardware [4]. It is tested on the AUDIS-catalogue of HRTFs [5]. The selection is evaluated in a localization study according to objective *criteria for the directional display:*

- externalization of the auditory image, i.e. a minimization of the number of inside-the-head localizations
- minimization of the number of front-back confusions
- match of the presented and perceived directions
- focused virtual auditory image, i.e. a minimization of the perceived source width.

2. A NEW METHOD FOR A SUBJECTIVE SELECTION OF NON-INDIVIDUAL HRTFS

2.1. Experimental setup

5 pulses of white noise serve as test sounds (20 Hz – 20 kHz, 30 ms duration, 70 ms pause, 5 ms Gaussian slopes, 60 dB SPL). Using the HRTFs, virtual positions are generated at -40, -20, 0, 20, 40° in the frontal horizontal plane and played through an electrostatic headphone[1]. This results in a moving virtual sound source location of the noise pulses.

The AUDIS-catalogue[2] [5] of human HRTFs is used in this study to provide 12 pairs of HRTFs of different subjects. The HRTFs are available as impulse responses[3] which were equalized for the headphone[1]. As the HRTFs were measured at 15°-spacing in the horizontal plane they are interpolated to -50, $-40, \ldots,$ $+50°$. All selection experiments are done in a sound insulated booth at a computer terminal.

2.2. Findings from preliminary tests

In a first approach the test sounds were presented in a specific order. Subjects were asked to evaluate the overall directional impression of the sounds according to the above mentioned criteria with the numbers 0 – 9. The HRTF with the best rating in several repetitions was chosen. The preliminary tests served to develop the right set of questions. 17 subjects without training in acoustics, took part in the preliminary tests. The following results could be obtained:

- An improvement of virtual directional reproduction by subjective selection seems possible.
- Despite of repetitive testing a consistent selection is found.
- An extensive question with all of the above mentioned criteria can not be followed if a great number of test sounds, i.e. HRTFs, is used.
- The term "spaciousness", German "Räumlichkeit", implies the perceived externalization of the sources, i.e. the moving sound source is perceived outside the head as "spacious".

[1] STAX Lambda Pro, STAX Ltd., Japan.
[2] Available from the European Acoustics Association (EAA), www.euracoustics.org.
[3] Length 256, sampling frequency f_s=44.1 kHz.

- The criterion "spaciousness" leads to a lateral overestimation of the sound direction, i.e. "spaciousness" is higher if the moving sound source travels over a wider angle.
- The evaluation of predetermined HRTFs using numbers $0-9$ is difficult for the following reasons:
  - The subject adapts quickly to the given HRTF.
  - A direct comparison of similarly reproducing HRTFs is hampered because of the specific order of HRTFs.
  - Within a test the assessment factors change.

2.3. Selection procedure for non-individual HRTFs

As the preliminary tests showed that a pre-determined, algorithmic presentation of HRTFs is unfavorable for the selection, a random access method is used instead which has proven successful in sound quality research (e.g. [6]). With random access the subject can select and listen to each sound/HRTF via entering the corresponding number on a keyboard. So the subject can omit inappropriate HRTFs quickly. Similar sounding HRTFs can be played repetively and can be compared back-to-back. The subject may take notes about each HRTF.

A two-step procedure can solve the problem of the non-applicability of extensive questions. Through a preselection, a resonable number of HRTFs can be chosen quickly, from which a final optimal HRTF selection can be singled out. The average duration for a complete selection is 10 minutes. The two selection steps are as follows:

Preselection: The question in the first selection step extracts 5 out of 12 HRTFs which evoke the greatest spatial perception in the frontal area.

Final selection: In the second part of the selection procedure one HRTF is singled out of the 5 previously chosen HRTFs which best matches the following criteria:

- The direction of the sound is perceived from $-40°$ left to $+40°$ right, but not further outside.
- The sound moves horizontally in equally-spaced steps.
- The sound has a constant elevation at all times.
- The sound is perceived in the frontal plane,
- at a constant distance,
- and preferably far away.

3. PSYCHOPHYSICAL EVALUATION OF THE SELECTION METHOD

3.1. Overview and method

To evaluate the selection method 3 experiments were carried out:

1st selection: The first verification experiment is identical to the first part of the selection method, the *preselection*, except that the 5 selected HRTFs are ranked best to worst. This ranking order is evaluated later.

2nd selection: The second test is comparable to the *final selection* of the selection method, but again a ranking order is established. This refines the ranking order of the best 5 HRTFs attained by the 1st selection.

Localization experiment: The localization ability is examined with the 5 HRTFs of the 1st selection. Subjects localize virtual sources from $\pm 50°$, $\pm 40°$, $\pm 20°$, and $0°$ in the frontal horizontal plane using a laser-pointer method. With this method subjects point to the sound source position by means of a computer-controlled laser pointer which is adjustable via a trackball [7]. The three buttons of the trackball code the position of the sound as (1) in front, (2) inside the head, and (3) in the rear. Test sounds are wideband noise pulses as described in section 2.1. 8 subjects conducted 10 trials per direction and HRTF (7∗5), and 2 subjects carried out 5 trials per condition. The localization experiment yields objective criteria for the localization ability with the selected HRTFs. On the basis of the criteria in section 1 the influence of the 1st and 2nd selection on the objective quality of the chosen HRTFs can be assessed.

3.2. Results

The 1st selection of 5 out of 12 HRTFs can be done easily because obvious inside-the-head localizations or front-back-confusions usually occur with $5-7$ HRTFs. The rejected HRTFs evidently reproduce directions non-optimally. The 5 pre-selected HRTFs instead separate from the average of the catalogue positively. If the top-ranking HRTF of these 5 HRTFs is superior in localization tests to the other 4 pre-selected HRTFs it will supposably reproduce directions better than all HRTFs from the catalogue. Therefore it is sufficient to test only the 5 HRTFs of the 1st selection.

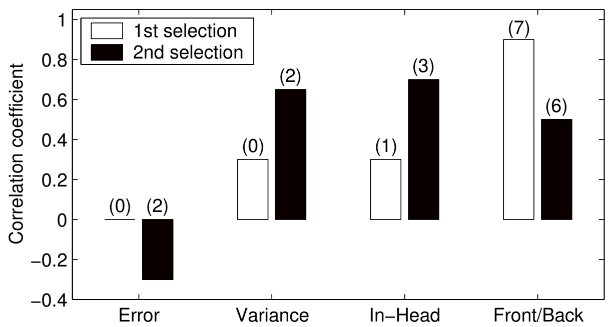

Figure 1: *Correlation of the ranking order of the localization results and the ranking order of the 1st (white) or 2nd (black) selection for the 5 HRTFs obtained in the 1st selection. The rank correlation coefficients (Spearman) averaged over subjects are given for the criteria: deviation of the localized from the presented direction (error), interquartile range of the localization responses (variance), relative number of inside-the-head localizations (in-head), and front-back confusions (front/back). The numbers in brackets represent the number of subjects with significant correlation at 5 %, i.e. subjects with one single reversal of neighboring ranks. Data of 10 subjects, age 24 – 32 years, 4 female, 6 male, are shown.*

Which objective criteria of section 1 can be obeyed by the subjective selection can be assessed through a correlation of the ranking order of the HRTFs by the selection and its ranking in the objective localization test. The ability of establishing a rank order of the 5 pre-selected HRTFs according to different criteria can be seen in figure 1. For the criterion "error" no correlation is given between the ranking order of the 1st (white) or 2nd (black) selection and the ranking order of the results of the localization test. The correlation is also small for the ranking of the 1st selection for the criteria ”variance” and ”inside-the-head localizations”. The 2nd selection improves the average correlation to $r_s > 0.6$ for these two criteria, i.e. on average HRTFs causing less localization variance are judged as better within the pre-selected group. The average correlation for the relative number of ”front-back reversals” is $r_s \approx 0.9$ for the 1st selection but decreases to $r_s \approx 0.5$ through the 2nd selection. It still is significant for 6 out of 10 subjects.

Considering the ranking order of the HRTFs in the selections against the ranking order in the objective localization test reveals that it is possible to subjectively categorize 5 HRTFs according to the objective criteria ”variance”, ”inside-the-head localizations” and ”front-back confusions”. It further on shows that these criteria are, on average, fulfilled by the same HRTFs. They can also be mostly optimized together through the selection.

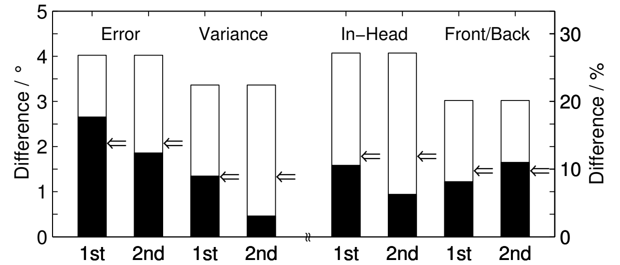

Figure 2: *Localization results of the 5 pre-selected HRTFs with respect to their ranking in the 1st and 2nd selection. The results are given as a difference of the individually top-ranking HRTF in the selection and the individually best HRTF in the localization test averaged over subjects. The average results of the individually worst HRTF out of the 5 pre-selected ones are given for comparison (white). The arrow marks the average results of all 5 HRTFs. The results are broken down for: deviation of the localized from the presented direction (error), interquartile range of the localization responses (variance), relative number of inside-the-head localizations (in-head), and front-back-confusions (front/back). Data of 10 subjects, age 24 – 32 years, 4 female, 6 male, are shown.*

As the selection method is aimed at finding only the most appropriate HRTF, the #1 ranking HRTF will be considered next. Figure 2 shows the localization results of the #1 ranking HRTF (black) of the 1st and 2nd selection relative to the results of the individually best HRTF in the objective localization test. Although the relative localization error of the #1 HRTF improves through the 2nd selection by about 0.8° it is still in the range of the average error of all pre-selected HRTFs (arrow). I.e., for this test no improvement can be seen against the pre-selected HRTFs. However, a comparison of localization results from the free-field and with selected HRTFs shows an average increase of relative error with virtual presentation of only 1° [8]. Therefore an optimization seems to take place through the selection process.

The average variance in the localization responses is minimized with the #1 HRTF through the 2nd selection process to a value of 0.5° (interquartile range) worse than the individually best HRTF in the test (figure 2). With the 1st selection using the spaceciousness-criterion instead no optimization of the variance can be seen.

The relative number of inside-the-head localizations and front-back confusions is not minimized through the 1st selection for the #1 HRTF compared to the other pre-selected HRTFs (figure 2). The 2nd selection reduces the number of inside-the-head localizations for the #1 HRTF, but has on average no further influence on front-back reversals. Inside-the-head localizations occur only 6.2 % more often than with the best HRTF. 50 % of the subjects choose an HRTF which is the test-optimum for inside-the-head localizations and front-back confusions. Further 3 subjects select an HRTF which leads to an increase in variance and inside-the-head localizations of each less than 10 %.

In summary the localization test reveals that the two-step selection method places an HRTF at the first rank which minimizes the variance of the localization responses and the relative number of inside-the-head localizations. For the #1 HRTF the localization error and the average number of front-back-reversals seems not to be optimized if compared to the other 4 pre-selected HRTFs. Nevertheless, 50 % of the subjects choose the test-optimum for front-back confusions. Overall 8 of 10 subjects select a HRTF which is the optimal one in the localization test or which leads to a hardly noticeable deterioration from the optimum.

4. APPLICATION OF THE SELECTION METHOD

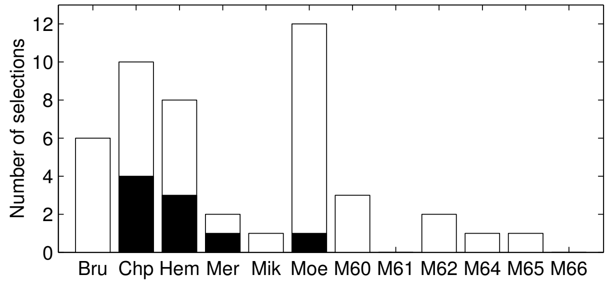

Figure 3: *Statistics on the selection of HRTFs from the AUDIS-catalogue [5] according to the selection method. Black: selections of a group of 9 subjects, age 24 – 28, white: selections of 46 subjects, mostly students.*

The selection method as described in sections 2.1 and 2.3 was conducted so far with 46 subjects. Figure 3 shows the statistics of their HRTF-selections. The selection of a small group of HRTFs by most of the subjects is pronounced. The HRTFs ”moe”, ”chp”, ”hem”, and ”bru” seem to show specific features which are relevant for a preferred selection according to the criteria. The HRTF ”moe” displays the importance of an individual selection: although only one subject of the small group selected this HRTF it is on average favoured most by all subjects.

5. DISCUSSION

5.1. Two-step selection using direct access

The preliminary tests for the development of the selection method showed that for selection a direct access to the different sounds of HRTFs is superior to an algorithmic sequential presentation. The main reason for this advantage is the subject’s ability to compare two sounds back-to-back. The differences between pre-selected HRTFs can be very small and thus only rateable by direct comparison. Multiple criteria can be easily judged and weighted by repeated listening to one or two HRTFs. Using direct access the selection of one HRTF is extremely efficient, as the subject is able to focus quickly on a small sub-set of suitable HRTFs. This also distinguishes the current method against a selection by successive 2-AFC ratings [4].

5.2. Selection criteria

The two-step selection procedure uses different criteria in its first and second part. This stems from the realization that multiple criteria can be followed only on a small set of sounds, i.e. HRTFs. Therefore the first selection is based on a rather broad criterion ("spaceciousness"). The results might be comparable to the single criterion "sound like free-field" [4]. The evaluation has shown that a further selection according to a second set of criteria can improve the virtual directional reproduction of the chosen HRTF clearly. Through the second selection the criteria "localization variance" and "externalization" are optimized.

Although further study [8] shows only a small increase in localization error with selected HRTFs compared to the free-field or individual HRTFs, the criterion "error" is not perfectly optimized by the selection. The increase in error with selected HRTFs is based mainly on an overestimation of lateral sound direction, i.e. the source is localized further towards the side. From this follows that subjects tend to select a HRTF from a subject with a slightly larger head, i.e. larger interaural time difference (cf. [2, 3]). Yet, the localization error between different pre-selected HRTFs is often below 1° which is in the range of the minimum audible angle [9]. Therefore further optimization might be difficult to reach purely on subjective listening without localization tests. A further improvement could be achieved by the presentation of directional "anchors" for comparison with the perceived auditory direction. The presentation of those well-defined acoustical directions though demands complicated hardware and is therefore not feasible for many applications.

5.3. Catalogue of HRTFs

The selection statistics in figure 3 show a clear preference for a small number of HRTFs. One reason might be that the AUDIS-catalogue of HRTFs is not compiled to cover a wide range of different subjects [5]. Another cause might be the population of participating subjects. To optimize the catalogue, the HRTFs "m61" and "m66", which were never chosen, could be substituted by different HRTFs. The build-up of a catalogue based on the often selected HRTFs "moe" and "chp" would instead cover only a small sub-set of subjects. For example "moe" was chosen by only 25 % of the subjects – 75 % prefer a different HRTF. The omission of rarely chosen HRTFs comprises the risk of a specialization of the catalogue for average subjects – uncommon features might in this case no longer be reflected in the catalogue although they are preferable for some subjects.

5.4. Application fields of the selection method

The selection procedure can be easily implemented without the need for special hardware as it requires only the playback of sounds on demand. The questions posed for selection were developed for and with subjects without training in acoustics. The complete selection is fast and takes about 10 minutes. In this time a HRTF is found which supports an externalized, focused image, and reproduces directions in a predictable manner. Therefore, the selection method is suitable for a variety of applications. It could be included into the setup of teleconferencing systems or computer games to optionally replace a default HRTF. The measurement of direction-dependent speech understanding in audiology could be done with virtual presentation after a selection of a suitable HRTF. Cockpit-applications could profit from the clear externalization of the sound image. As users are able to adapt to different HRTFs (e.g. [10]), the selection supports and speeds up the adaptation process by finding a HRTF which already allows a near-normal perception. In localization tests with selected HRTFs subjects often reported that the externalization of the virtual sound source improves greatly over about 5 minutes of use.

6. ACKNOWLEDGEMENTS

We like to thank Dr. C. v. Rücker and two anonymous reviewers for comments on the manuscript. B.S. is indebted to Prof. E. Hafter for his support for a participation at this conference. This work was supported by the Deutsche Forschungsgemeinschaft.

7. REFERENCES

[1] H. Møller, C.B. Jensen, D. Hammershøi, and M.F. Sørensen, "Selection of a typial human subject for binaural recording," *Acustica*, vol. 82, pg. 215, Supplement, 1996.

[2] W.M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3678–3688, 1996.

[3] J.C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493–1510, 1999.

[4] J.C. Middlebrooks, E.A. Macpherson, and Z.A. Onsan, "Psychophysical customization of directional transfer functions for virtual localization," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 3088–3091, 2000.

[5] J. Blauert, M. Brüggen, K. Hartung, A.W. Bronkhorst, R. Drullman, G. Reynaud, L. Pellieux, W. Krebber, and R. Sottek, "Der AUDIS-Katalog menschlicher Auenohrübertragungsfunktionen (The AUDIS-catalogue of human head-related transfer functions)," in *Fortschritte der Akustik – DAGA '98*, Oldenburg, 1998, pp. 174–175, DEGA.

[6] H. Fastl, "Psychoacoustics and sound quality," in *Fortschritte der Akustik – DAGA '02*, Oldenburg, 2002, pp. 765–766, DEGA.

[7] B. Seeber, "A new method for localization studies," *Acta Acustica – Acustica*, vol. 88, no. 3, pp. 446–450, 2002.

[8] B. Seeber, "Localization in the anechoic environment and with virtual acoustical directional presentation investigated with a laser-pointing method," in *Abstracts of the 25th annual midwinter research meeting*, Mt. Royal, NJ, 2002, pg. 14, Abs.: 52, Assoc. Res. Otolaryngol.

[9] A.W. Mills, "On the minimum audible angle," *J. Acoust. Soc. Am.*, vol. 30, no. 4, pp. 237–246, 1958.

[10] P.M. Hofman, J.G. Van-Riswick, and A.J. Van Opstal, "Relearning sound localization with new ears," *Nature Neuroscience*, vol. 1, no. 5, pp. 417–421, 1998.