Audio Engineering Society
Conference Paper
Presented at the 21st Conference
2002 June 1–3 St. Petersburg, Russia

# Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems

Alois Sontacchi[1], Piotr Majdak, Markus Noisternig and Robert Höldrich

*Institute of Electronic Music and Acoustics, University of Music and Dramatic Arts, A-8010 Graz, AUSTRIA, http://iem.at*
*[1] Correspondence should be addressed to alois.sontacchi@kug.ac.at*

**ABSTRACT**

In general hearing tests are necessary to assess the properties of spatialisation systems. To speed up the procedure of testing different system parameters an objective model of localisation in binaural sound reproduction system is introduced [7]. In the following the localisation properties of an auditory system based on playback via head phones is investigated. Therefore a new design for the experiment setup to investigate the perception of virtual sources is introduced. This paper reports the subjective validation of the objective model by informal listening tests. In addition to the localisation and localisation blur further properties are evaluated. Significant differences between different system setups are studied and depicted.

## 0 INTRODUCTION

Concerning the reproduction of virtual sound sources, several parameters will influence the quality of perception [1]. The most important parameter for convincing sound reproduction, the localisation of sounds, can be described using following definitions:

- Localisation function, that describes the perceived position of a virtual sound source depending on the actual position of the target source.
- Localisation blur, that describes the "width" of the perceived stimulus and also depends on the actual position of the target stimulus.

Constitutive on these definitions, further definitions can be deduced:

- The localisation error function, which refers to the deviation of the localisation function from the actual position of the target stimulus.
- The mean localisation error, as a mean value of the localisation error function over the examined range.
- And the mean localisation blur, as a mean value of the localisation blur over the examined range.

The localisation blur must be normalised to the minimum audible angle (MAA) before calculating the mean value.

These functions and values can be derived by analysing a binaural reproduction system using an objective model, introduced in [6] and [7]. However, there are another important properties of perception of sounds, as well. Which will be discussed in the following.

The externalisation describes the perception of the distance of the virtual sound source and can range from "inside-the-head" to "far-outside-the-head". The goal of virtual acoustic synthesis is usually to produce sound externalised, that is, outside the listeners body. Thus, the externalisation is an important criterion in the development of virtual acoustic systems.
Another phenomenon in reproduction virtual sources is the ambiguity, also termed "front-back-confusion" or reversal error. It refers to the judgement of the sound stimulus as located on the opposite side of the interaural axis than the target source. To

prevent this mistake, humans proceed a slight head movement in natural hearing conditions [5]. To benefit from this phenomenon in reproduction of virtual sound sources, a head tracker can be incorporated [2] [3].
The timbre distortion is a further property of perception of virtual sound sources. Although it is a property of monaural hearing, it has a dynamic character by using time-varying reproduction systems with moved sound sources, and cannot be equalised with a fixed time invariant filter. Additionally, the judgment of timbre is very difficult, due to the varying description possibility of the perceived stimulus. Simplified a comparison of the distorted timbre to the real stimulus can be carried out e.g. by using a „reality rating".

Analysing time-varying reproduction systems with moved sources, the consistency of movement should be evaluated. There is a direct connection between the spatial resolution of the rendering algorithm (e.g. database with HRIRs) and the consistency of movement.
In order to optimise these reproduction systems a judgement of all those properties is necessary. Estimation of localisation properties was shown in [6] and [7] by introducing an objective model of localisation. It is based on a comparison of the behaviour of the evaluated system to a database of reference HRIRs, regarding psychoacoustical effects. Assuming, that a given set of HRIRs offers an optimal localisation, an estimation of localisation error can be achieved. Due to the previous assumption, the results are valid only for the used HRIRs set, and so the estimation is individualised. The localisation of other reproduction systems, not based on HRIR filtering, cannot be estimated by using this objective model. An objective judgment of all other properties of reproduction systems in virtual acoustics seems to very difficult and not satisfactory.
Thus, hearing studies are necessary to achieve some properties of the used reproduction systems. Using evaluated data the reproduction systems can be optimised.

To design a hearing experiment two general approaches can be used:

- Presentation of a stimulus and judgement of the properties of the perception by a subject. Basic difficulty for a subject arises due to the transmission of the perceived properties to the experiment leader. Localisation studies, especially, suffers from the problem, that the subject cannot give an exact position of perceived stimuli. A systematic error occurs in the evaluated data.
- The subject is requested to achieve an aim, the error between the goal and the result is evaluated by the experiment leader. The advantage of this method is, that the systematic error of the previous method can be compensated. However, an erroneous interpretation of the aim can falsify the results.

The introduced experimental design uses both methods to evaluate several properties of virtual reality reproduction systems via headphones.
This setup was primary used to validate the objective model of localisation [6]. However, it depicts a general approach to the estimation of perceived properties in virtual acoustics. In this experiment only the 2D case was investigated, concerning the variable azimuth angle.

## 1 EXPERIMENTAL DESIGN

### 1.1 Used Methods

The aim of the experimental setup is to achieve the estimation of some perceived properties of given binaural 2D reproduction systems (in the following called settings). The estimated properties are localisation (including ambiguity), externalisation, tone distortion and consistency of movement. These properties should be given for each setting. Through evaluating the localisation function, equalisation function can be obtained to achieve better localisation for the listener.
At the beginning of the experiment the subject can hear a virtual source via headphones. The subject can move the source over the whole azimuth range (360 degrees). After a short training phase the subject has to answer some questions about the timbre, externalisation, perceived width of the source and consistency of the movement. Two or three minutes are required to complete this task in common.
The evaluation of the localisation can be divided into two sections: the estimation of localisation function and the estimation of localisation blur. The estimation of localisation blur can be reduced to an estimation of the localisation function using narrow-band stimuli with different frequencies. Thus a single experiment setup can be used to estimate both parameters.
To obtain the estimation of localisation function the subject has to achieve a certain target. The difference between the instructed target and the obtained result is evaluated. A stimulus is played back via a single loudspeaker from a known position. The stimulus is presented to the subject via an open dynamic headphones thus the real sound source can be localised too (ignoring some details, compare chapter "spatial sampling" in [4]). Simultaneous a virtual sound source is reproduced via the mentioned headphones. The subject can move the virtual stimulus over the whole azimuth range and has to match both perceived positions. Also the head position of the subject is measured with a head tracking device and is used to equalise the position of the virtual sound source. Thus the subject is able to make some small head movements during the experiment, the limit is set to $\pm 5°$ . To keep the head movement in that range, graphical symbols are displayed on the screen. So the subject is requested to move to the head back to the initial position. After achieving the target, the subject has to push a button and the chosen position of the virtual source is recorded. This procedure is repeated several times for each position. In this way, a complete localisation function for a setting could be estimated.
Stimuli are presented from the front to keep the reversal errors low. By improving the symmetry of evaluated reproduction systems to the median plane, its possible to halve the number of real source positions to evaluate. The desired resolution of the localisation function is set to 10°, in the range of $\pm 70°$ . Therefore, eight speakers are arranged with the same radius around the subject. The eight positions, folded to one quadrant yield the desired resolution of 10° (Fig 1). Each position is evaluated several times to reduce the spread errors.

A pre-experiment was carried out, to evaluate the influence of carrying open dynamic headphone when localising the real sound source. It could be shown, that no distortion could be registered working with the spatial resolution of 10°. Just one subject had a

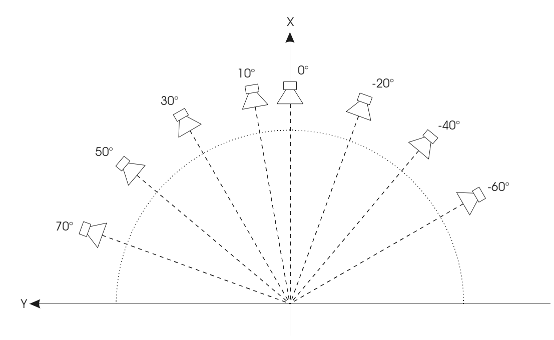


Fig. 1: Speaker arrangment

front-back-confusion of given stimuli. That behaviour appeared as grown reversal error in localisation data.
Evaluating the localisation blur can be proceeded the same way – the kind of stimuli must be changed to narrow-band noise. Although it is theoretically possible to carry out the localisation blur, the number of measurement would increase dramatically (multiplied by the number of frequency groups).

### 1.2 Stimuli

Brief segments of speech stimuli are used in the experiment. The duration is adjusted to the measurement time (approx. 20 seconds). In order to distinguish between the virtual and the real source, different stimuli are used (female and male speech signals). The reason of using speech signals is given by the application field itself, e.g. teleconferencing or information systems. Another reason is, the long-term average speech spectrum does not contain a significant level of power at those frequencies, where HRIRs yields elevation cues that can be used by a subject. Therefore its possible to use speech signals for localisation tests.

### 1.3 Subjects

Ten volunteers (four female and six male, age range 25 to 40) participated. They had partly miscellaneous experiences with virtual acoustic systems, although they had never attended any acoustical experiment. Prior to the experiment, they were introduced to the system and test procedure. They were also instructed to the used terms and the head position was referenced to 0°. The experiment series were merged to get units of half an hour duration, subjects were encouraged to make a break between the units.

### 1.4. Evaluated Systems

The reproduction systems are based on an approach termed ambisonic [6], [8], [9]. The main goal of the experiments was the validation of the objective model of localisation using reproduction systems based on the ambisonic approach. Accordingly the used systems correspond to the evaluated systems in [7] varying the following parameters:

- order of ambisonic
- weights of individual ambisonic channels
- two different sets of HRIRs: KEMAR [10] and proprietary
- length of HRIR filters
- arrangement of virtual speakers used in ambisonic

Detailed description of used parameters and their theoretical effects are given in [6] and [9]. Furthermore direct implementations of HRIR filtering were used to investigate the influence of using the ambisonic model as well. In order to compare the results the same sets of HRIR were used for these systems: KEMAR and proprietary. The sets were given in different spatial resolutions (KEMAR: 5°, proprietary: 15°). The single HRIRs were selected depending on the chosen azimuth angle. No interpolation algorithm between different directions was used.
The head movements of subjects were determined by a head-tracking device (‘Flock of Birds’, Ascension Tech.). The binaural sound reproduction was rendered by an evaluation module containing a DSP (Texas Instruments TMS320C6711), run with a self-made software. A personal computer controlled the sound cues and the synchronisation between the head-tracker and the DSP. Therefore a maximum latency of 30ms was guaranteed. In addition all evaluated data were collected by that computer. The virtual sound stimuli were presented to subjects over open electro-static headphones (‘HE 60’, Sennheiser). The real sound stimuli were presented over a set of speakers (‘system 800’, Tannoy) controlled by a digital mixing console (‘d8b’, Mackie).

## 2. RESULTS

### 2.1 Localisation

The data obtained by the listening test results in a four-dimensional matrix given in Eq. 1.

$$L(sub, seti, spk, ex) \qquad (1)$$

where *sub* denotes the subject, *seti* selects used setting, *spk* is the position of real stimuli and *ex* the number of measurement. In front of further calculation, all data were folded to one quadrant and ambiguity problems were corrected. The amount of reversal errors was collected and evaluated especially. It can be shown, that the reversal error increased with the growing azimuth angle, independent of the subject. Furthermore the evaluation of the reversal error on settings showed, that it behaves very subject-dependent and there is possibility to obtain a generalised statement over all subjects (Fig. 2).

Further evaluation on the data was done by calculating the mean values and standard deviations over the measurement number *ex*

$$\begin{matrix} std(sub, seti, spk) \\ \overline{L}(sub, seti, spk) \end{matrix} \qquad (2)$$

where *std* denotes the standard deviation and $\overline{L}$ the mean values of $L$. The standard deviation *std* can be interpreted as the uncertainty to localise the same position. If the ability to localise a source increases the standard deviation will decrease. Additionally, the standard deviation describes the localisation blur. Reproduction systems with high localisation blur produce high standard deviation and high uncertainty, as well. Regarding the dependence of the localisation blur to the azimuth angle, which is defined by the minimum audible angle MAA [1], the individual values of uncertainty cannot be compared to each other. Thus, in front of further evaluation each value of the uncertainty was normalised to the MAA of the corresponding localisation $\overline{L}$

$$s = \frac{std_{sub,seti,spk}}{MAA(\overline{L}_{sub,seti,spk})} \qquad (3)$$

where *s* is the spread parameter. The spread parameter can be further evaluated by averaging over the source positions. In Fig. 3

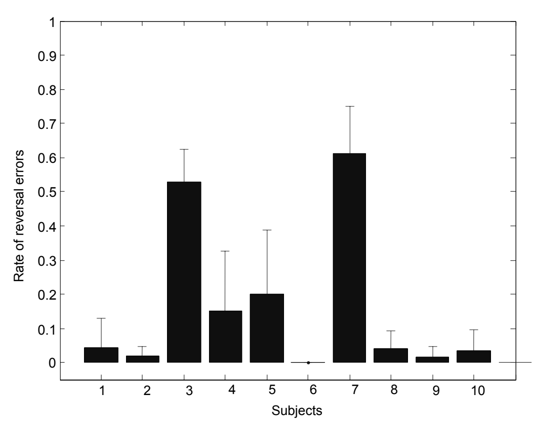



Fig. 2 : Reversal errors of subjects

the mean values of scatter parameters, calculated over each subject are depicted.

It can be shown, that the subjects have different localisation uncertainty. Subject no.2 supplies a relative small error. The localisation uncertainty of subject no.7 is 2.5 times larger compared to subject no.2. Furthermore, the standard deviation shows that the localisation uncertainty of subject no.2 and 6 differs negligible due to different settings. The enormous localisation uncertainty of subject no.7 can be properly explained by problems during the experiment (e.g. loss of concentration, troubles handling the user interface etc.). Therefore, for further investigations the data set produced by subject no.7 is omitted.

In the following significant differences between different settings were investigated by conducting the univariate repeated-measures analysis of variance (ANOVA).

Following significant differences can be found:

- Dependence of HRIRs (see Fig. 4)
- Dependence of the used ambisonic order (see Fig. 5)
- Dependence of different weighted ambisonic orders (see Fig. 6)

### 2.2 Equalisation function

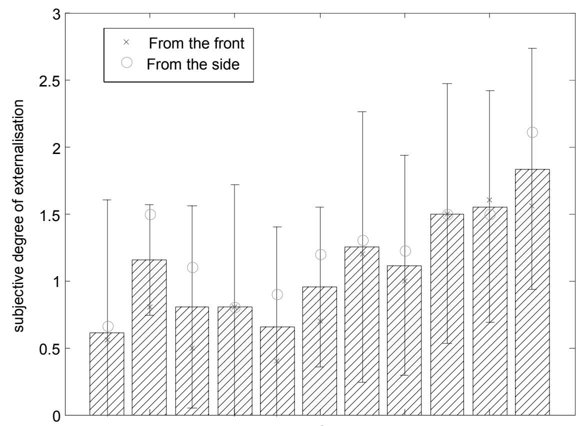
To obtain an equalisation function for a specific system, first the corresponding localisation function $\overline{L}$ has to be evaluated. This function refers to each evaluated setting and subject. The average of this function over all subjects should result in a localisation function for each setting. To get the equalisation function we have to invert the localisation function, therefore the localisation function must be strictly monotonous. In addition, data should be statistically independent. The evaluation results showed, that this criteria was missed. Therefore the equalisation function can be calculated just for each subject and setting separately.

### 2.2 Verbal statements

First the verbal data was categorised to obtain a list of numerical values for further statistical calculations. The following example shows the classification of the different terms in the case of externalisation.

- 0: inside-the-head
- 1: edge of the head
- 2: 10cm outside the edge of the head
- 3: 30cm outside the edge of the head
- 4: far outside-the-head

The results for the averaged classified externalisation data is shown in Fig. 7. Analogue to this example it is possible to evaluate other properties.




| Seti | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Len | 256 | 128 | 128 | 96 | 64 | 128 | 128 | 256 | 128 | 128 | 256 |
| HRIR | P | P | P | P | P | P | K | K | K | K | K |
| Sparr | - | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - |
| Ord | - | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | - |
| Wght | - | 1 | 1 | 1 | 1 | 1 | 1 | K-25 | K-25 | 0.4 | - |

Fig 7: Subjective degree of externalisation for each setting. Depicted values are given for directions from the front, the side and a mean of both of them, where *Seti* is the setting, *Len* is the filter length, *HRIR* is the kind of HRIR-set used (*P*: proprietary, *K*: KEMAR), *Sparr* is the number of speakers near the median axis, *Ord* is the number of ambisonic order and *Wght* is the weighting of ambisonic orders.

### 3. CONCLUSIO

Studying literature shows that previous listening tests focus on human localisation properties, whereby in this study the properties of reproduction systems is emphasised. The informal subjective listening test validates the proposed objective model of localisation in binaural sound reproduction systems [6]. The subjective perception correlates well with the results obtained by the mathematical model. Furthermore it is possible to calculate an equalisation function of localisation for a single person. To obtain a general equalisation function the experimental setup has to be further optimised to reduce the subject dependency. Measurements concerning the localisation blur require extended listening tests.
It must be emphasised that these data apply only to speech stimuli and that experimental results may differ when broadband stimuli such as noise or clicks are used.

### 4. REFERENCES

[1] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localisation, rev. ed. (MIT-Press, Cambridge, MA, 1997)
[2] F. L. Wightman and D.J. Kistler, The Importance of Head Movementes for Localising Virtual Auditory Display Objects, in Proc. 1994 Conference on Auditory Displays
[3] D.R. Begault, E.M. Wenzel; Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized HRTFs on the Spatial Perception of a Virtual Speech Source; J. Audio Eng. Soc. Vol. 49, No.10; 2001
[4] S. Carlile, S. Delaney, A. Corderoy; The localistion of spectrally restricted sounds by human listeners; Hear. Res. Vol. 128, pp 175-189; 1999
[5] G. Theile, U. Felderhoff , P. Mackensen; Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale; 20. Tonmeistertagung Karlsruhe pp. 229 - 237; 1998
[6] P. Majdak P., M. Noisternig; Implementation kopfpositionsbezogener Binauraltechnik, Diplomarbeit 2002, Institut für Elektronische Musik und Aksutik, Graz
[7] A. Sontacchi, M. Noisternig, P. Majdak, R. Höldrich; An objective model of localisation in binaural sound reproduction systems, unpublished, proposed to the 21st AES Int. Conf., 2002
[8] M. A. Gerzon; General Metatheory of Auditory Localisation; Audio Eng. Soc. 92nd Conv. Vienna, preprint 3306; 1992
[9] A. Sontacchi; Neue Ansätze der Schallfeldreproduktion; Dissertation 2002, Institut für Elektronische Musik und Aksutik, Graz