Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Binaural Technique

Basic Methods for Recording, Synthesis, and Reproduction

Hammershøi, Dorte; Møller, Henrik

Published in:
Communication Acoustics

Publication date:
2005

Link to publication from Aalborg University

Citation for published version (APA):
Hammershøi, D., & Møller, H. (2005). Binaural Technique: Basic Methods for Recording, Synthesis, and Reproduction. In J. Blauert (Ed.), *Communication Acoustics* (pp. 223-254). IEEE Computer Society Press.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: March 24, 2024

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 1

Binaural Technique – Basic Methods for Recording, Synthesis, and Reproduction

Dorte Hammershøi and Henrik Møller

Department of Acoustics, Aalborg University, Aalborg

Summary. The term "binaural technique" is used as a cover label here for methods of sound recording, synthesis and reproduction, where the signals in focus are the acoustic signals at the eardrums. If these are presented authentically to listeners, the listeners will obtain acoustic cues which are deemed sufficient for authentic auditory experience – including its spatial aspects. This chapter reviews the basic principles of binaural technique - putting a special focus on results of investigations which have been performed at Aalborg University. These basic principles form the foundation for current utilization of binaural technique at large. They include basic theory, investigations on sound transmission in the ear canal, measurements and post-processing of head-related transfer functions, HRTFs, transfer functions of headphones and their adequate equalization, and results from localization experiments in real life as well as with binaural recordings from real heads and artificial heads. Numerous applications to these methods exist. Some of them will be introduced exemplarily.

1 Introduction

Binaural technique starts from the concept that our auditory percepts are predominantly formed on the basis of only two inputs, namely the sound-pressure signals at our two eardrums. If these are recorded – literally – in the ears of listeners, and reproduced authentically when played back, then all acoustic cues are available to the listeners for forming authentic replicas of the original auditory percepts – including all spatial aspects.

The signals to the two ears, called binaural signals, may, instead of being picked up in the ears of listeners, also be synthesized by computer – once the sound transmission from the source to the listeners' ears has been mapped for all possible positions of the sound source. It can then be controlled by computer *where* the listener will hear the sound, and sound sources can be placed everywhere in the listeners' perceptual space. This technique is obvious for virtual-reality application, but many further areas of applications exist.

The definition of *binaural technique* may be extended more generally to *binaural technology* as follows.

> "Binaural technology is a body of methods that involves the acoustic input signals to both ears of the listener for achieving practical pur-

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 2

poses, for example, by recording, analyzing, synthesizing, processing, presenting, and evaluating such signals." [8]

This chapter presents methods for obtaining or generating binaural signals and reproducing them adequately. For a wider view on spatial hearing and binaural technology, see e. g., [9, 35].

1.1 Structure

With binaural recording the sound is recorded in the ear canals of listeners. For practical reasons the listener is often replaced by a manikin with the relevant acoustical properties similar to those of a human. The technique is therefore commonly referred to as "artificial-head" recording technique or "dummy head" recording technique. When binaural recordings are made, the spatial properties of the exposing sound field are transformed into two electrical signals representing the input to the left and right ears, respectively.

Play-back is most often done by headphones, as they offer complete channel separation and acoustical isolation from the surroundings, thus enabling optimal control of the reproduction situation. Authentic reproduction requires correct calibration of the complete recording and reproduction chain. Section 2 addresses principles of calibration from a theoretical point of view.

The headphone transfer function is often given very little attention. Often, headphones at hand are used without electronic compensation, disregarding the intended use and design goals of the headphones and their typically non-flat frequency responses. In Sect. 3 examples of headphone transfer functions and equalization filters are shown.

When an artificial head is used during recording – or a human different from the listener – then the recorded signals do not exactly correspond to the ear signals that the listener would have been exposed to in the recording situation. Localization performance with binaural recordings from humans and artificial heads is addressed in Sect. 4. This section also includes a comparison of head-related transfer functions, HRTFs, of some artificial heads and humans.

Binaural signals may alternatively be generated electronically by the use of filters representing HRTFs, in which case the method will be denoted "binaural synthesis". The success of binaural synthesis depends strongly on details of the procedures applied for determining and realising HRTFs, such as physical aspects of the measurement situation, post-processing of data, and implementation as digital filters. These issues are addressed in Sect. 5.

Finally, in Sect. 6, some of the numerous applications of binaural technology are reviewed. The list is in no way exhaustive, but offers some input for inspiration and useful links to players in the field. Applications of binaural technology are also described in other chapters of this book [10, 10, 19, 84].

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 3

2 Theory

2.1 Sound-Transmission Model

When the binaural signals are recorded – whether in the ears of a human or an artificial head – the spatial sound field is transformed into only two signals, one for each side. A prerequisite for the successful reproduction is thus that the full spatial information is maintained in these signals.

This issue has been frequently addressed in the literature, e. g., [37, 66–68, 99–101, 110], the latter containing a summary. It is generally agreed that there are other places in the ear canal than at the eardrum where the full spatial information is available. This has been demonstrated for points on the center line of the ear canal, including the point on the entrance plane. It has also been shown that the spatial information is maintained, even if the ear canal is blocked at the recording point or further towards the eardrum. This is of significant practical relevance, since larger microphones can then be used in human ear canals, and artificial heads can be made without ear canals. In addition, as will be seen in Sect. 2.3, individual effects from the ear-canal and eardrum impedance are removed, whereby the signals become more general with respect to the spatial information contained in it.

A sound-transmission model as introduced in [71] splits the sound transmission up into direction-dependent and direction-*in*dependent parts. The model, and a similar one for the reproduction situation – see Sect. 2.2, is used for determining the correct calibration of the complete recording and playback chain. Further, it enables assessment of other relevant properties, such as inter-individual variation, see Sect. 2.3. The model is depicted in Fig. 1[1]

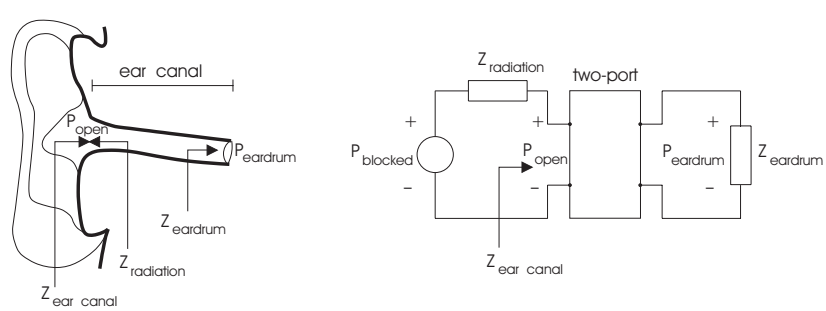

Fig. 1. Model of free-field sound transmission to the human external ear: Sketch of anatomy, *left*, and analogue model, *right* – adapted from [71]

[1] In the present chapter, capital letters denote terms given in the frequency domain. The symbols thus stand for complex functions of frequency. Time domain representations are denoted by the corresponding lower-case symbols. The connection between the two domains can, e. g., be given by the *Fourier* transform, $\mathcal{F}$, and its inverse, $\mathcal{F}^{-1}$, thus $P(f) = \mathcal{F}(p(t))$ and $p(t) = \mathcal{F}^{-1}(P(f))$

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 4

In the model, the complete sound field outside the ear canal – whatever the source and the sound field – is described by a *Thévenin* equivalent, consisting of the open-circuit sound pressure, P_{blocked}, and a generator impedance. The generator impedance is identical to the radiation impedance, $Z_{\text{radiation}}$, as seen from the ear canal into the free air. P_{blocked} does not exist during normal listening conditions, but if the ear canal is blocked to make the volume velocity zero, P_{blocked} can be found just outside the blockage. P_{eardrum} denotes the sound pressure at the eardrum and P_{open} denotes the sound pressure at the entrance to the ear canal. The ear canal itself is modelled by a two-port which is loaded by the eardrum impedance, Z_{eardrum}. The input impedance of the two-port as seen from the entrance of the canal, is denoted $Z_{\text{ear canal}}$. The total sound transmission from the free field to the eardrum can be described by the ratio of (i) the sound pressure at the eardrum and (ii) a reference sound pressure which is found at the position corresponding to the center of the subject's head with the subject absent, i.e.

$$\frac{P_{\text{eardrum}}}{P_{\text{ref}}}(\angle). \tag{1}$$

This ratio, $P_{\text{eardrum}}/P_{\text{ref}}$, depends on the angle of incidence of the sound wave, as indicated by the symbol $\angle$. It defines a head-related transfer function, usually abbreviated HRTF[2]. The sound transmission can be separated in consecutive parts in the following way,

$$\frac{P_{\text{eardrum}}}{P_{\text{ref}}}(\angle) = \frac{P_{\text{blocked}}}{P_{\text{ref}}}(\angle) \cdot \frac{P_{\text{open}}}{P_{\text{blocked}}} \cdot \frac{P_{\text{eardrum}}}{P_{\text{open}}}. \tag{2}$$

$P_{\text{eardrum}}/P_{\text{open}}$ describes the sound transmission along the ear canal and does not depend on the angle of incidence – due to one-dimensional wave propagation inside the ear canal. Nor does $P_{\text{open}}/P_{\text{blocked}}$ depend on the angle of incidence. This ratio denotes the pressure division at the entrance of the ear canal between $Z_{\text{ear canal}}$ and $Z_{\text{radiation}}$ as given in the equation below. In essence, only $P_{\text{blocked}}/P_{\text{ref}}$ depends on direction.

$$\frac{P_{\text{open}}}{P_{\text{blocked}}} = \frac{Z_{\text{ear canal}}}{Z_{\text{radiation}} + Z_{\text{ear canal}}}. \tag{3}$$

An example of the transfer functions for three directions of sound incidence which illustrates directional dependence and independence as discussed above is given in Fig. 2.

The consequence of this is that both sound pressures at the eardrum, at the ear-canal entrance, blocked or open, or elsewhere on the center line of the ear canal, will have the full spatial information and may be used for binaural recording. Likewise, it suggests a wider definition of the head-related transfer

[2] Note that the term HRTF is commonly understood as standing for a set of two transfer functions, namely, one for the left and one for the right ear

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 5

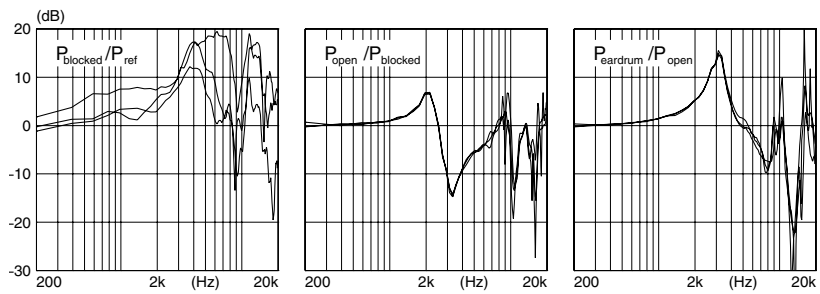

Fig. 2. $P_{\text{blocked}}/P_{\text{ref}}$, *left*, $P_{\text{open}}/P_{\text{blocked}}$, *center*, and $P_{\text{eardrum}}/P_{\text{open}}$, *right*. Measurements with sound coming from the front, from the left side and from the back are overlaid in each frame – left ear of Subject ML, from [37]

function, so that it covers all points in the ear canal that have the full spatial information and not only the point at the eardrum – as was denoted in (1). An HRTF, then, is, generally given by the ratio of the sound pressure at the ear of a listener and the reference sound pressure as follows:

$$HRTF(\angle) = \frac{P_{\text{ear}}}{P_{\text{ref}}}(\angle). \tag{4}$$

2.2 Calibration of the Complete Recording and Play-back Chain

In order to determine the full transmission of a binaural recording and play-back system, a model similar to that described in Sect. 2.1 for the free-field transmission is used for the play-back situation. The superscript "hp" is used in the following for signifying sound pressures that refer to the reproduction situation over headphones. Again, the input is described by a *Thévenin* generator, i. e. the blocked-entrance sound pressure, $P^{\text{hp}}_{\text{blocked}}$, and a generator impedance, $Z_{\text{headphone}}$. The complete transmission from voltage at the headphone terminals, $E_{\text{headphone}}$, to sound pressure at the eardrum, is then

$$\frac{P^{\text{hp}}_{\text{eardrum}}}{E_{\text{headphone}}} = \frac{P^{\text{hp}}_{\text{blocked}}}{E_{\text{headphone}}} \cdot \frac{P^{\text{hp}}_{\text{open}}}{P^{\text{hp}}_{\text{blocked}}} \cdot \frac{P^{\text{hp}}_{\text{eardrum}}}{P^{\text{hp}}_{\text{open}}}. \tag{5}$$

The sound transmission along the ear canal is the same whatever the source, thus $P^{\text{hp}}_{\text{eardrum}}/P^{\text{hp}}_{\text{open}} = P_{\text{eardrum}}/P_{\text{open}}$ holds. However, the *Thevénin* impedance is not the same. Consequently, the pressure division at the entrance to the ear canal differs between the two situations. The pressure division during reproduction is

$$\frac{P^{\text{hp}}_{\text{open}}}{P^{\text{hp}}_{\text{blocked}}} = \frac{Z_{\text{ear canal}}}{Z_{\text{headphone}} + Z_{\text{ear canal}}}. \tag{6}$$

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 6

If recording is made at the eardrum, M describes the transfer function of the recording microphone, and G_{eardrum} denotes the electrical gain from the output of the recording microphone to the input of the headphone, the total transmission from recording to reproduction can be written as

$$\frac{P_{\text{eardrum}}}{P_{\text{ref}}} \cdot M \cdot G_{\text{eardrum}} \cdot \frac{P^{\text{hp}}_{\text{eardrum}}}{E_{\text{headphone}}}. \tag{7}$$

The headphone produces the *correct* sound pressure at the eardrum when the transmission in (7) equals $P_{\text{eardrum}}/P_{\text{ref}}$, which means that the design target for the equalizer is given as follows – see [71] for calculations,

$$G_{\text{eardrum}} = \frac{1}{M \cdot (P^{\text{hp}}_{\text{eardrum}}/E_{\text{headphone}})}. \tag{8}$$

Similarly, when the recording is made at the entrance, open or blocked, respectively, the following equations apply,

$$G_{\text{open}} = \frac{1}{M \cdot (P^{\text{hp}}_{\text{open}}/E_{\text{headphone}})} \quad \text{and} \tag{9}$$

$$G_{\text{blocked}} = \frac{1}{M \cdot [P^{\text{hp}}_{\text{blocked}}/E_{\text{headphone}}]} \cdot \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}}. \tag{10}$$

The last term in (10) is the ratio of the pressure division during recording (3) and the pressure division during reproduction (6), which is denoted the pressure-division ratio, PDR. If the PDR is ignored – at least at this point, it can be seen that the exercise for finding the correct equalization is the same whatever the recording point is. In other words, the transfer function of the equalizer, G, shall equal the inverse of the microphone transfer function times the inverse of the headphone transfer function as measured at the same physical point in the ear where the recording was made.

Many of modern miniature and probe microphones have fairly flat frequency responses, in which case they may simply be represented by their overall sensitivities in the design target for G. Thus the main component of G is the inverse of the headphone transfer function measured at the position in the ear canal where the recording was made. For this reason, the equalizer is often referred to as the *headphone* equalizer, although, by definition, it serves to ensure that the entire reproduction chain is adequately calibrated. Headphone equalization will be further discussed in Sect. 3.

2.3 Inter-Individual Variation

In the previous sections it was inherently assumed that the recordings and the determination of the headphone transfer function were done with the

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 7

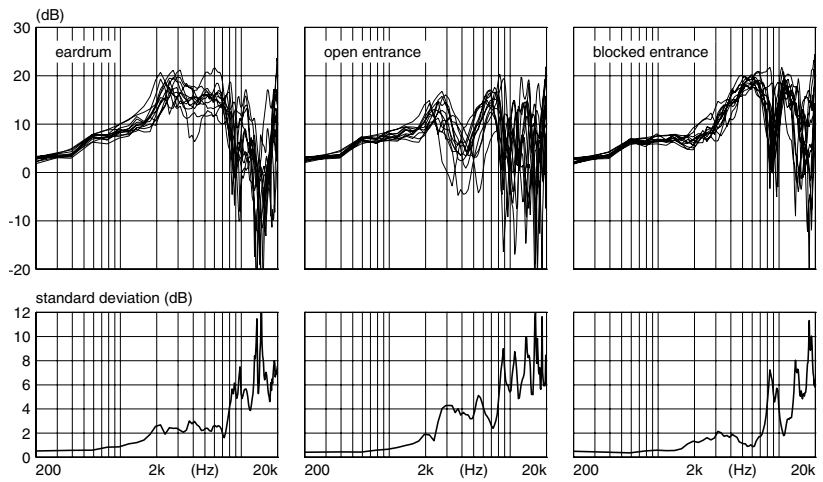

Fig. 3. Left-ear HRTFs for 12 subjects, measured at the eardrum, **left**, at the open entrance, **center**, and at the blocked entrance, **right**, to the ear canal – adapted from [37]. Sound from the left, standard deviations computed frequency by frequency in dB

same person, and that the listener was identical to this person. This is rarely the case, and since there are considerable differences in the shape of humans, including their outer ears, it must be anticipated that there will be individual differences in the binaural signals and in the headphone transfer functions. The consequence is that the binaural technique may result in correct signals only for those particular individuals that the recordings and the headphone-transfer-function measurement were made with. Thus, only these individuals may achieve the authentic listening experience.

The separation of free-field sound transmission into cascaded components, such as described in Sect. 2.1, paves the road to objective assessment of inter-individual variations. Data for inter-individual variation in sound transmissions have been presented previously, see, e. g., [9,30,37,39,76,97,110,111,117]. Comparisons as performed in [37, 76] provide the basis for the following discussion.

Examples of HRTFs measured at the eardrum, the open entrance and the blocked entrance are given in Fig. 3. If the general characteristics of the three types of HRTFs are compared, there are clear differences. This does not come by surprise, given the characteristics of the pressure division, $P_{\text{open}}/P_{\text{blocked}}$, and the transmission along the ear canal, $P_{\text{eardrum}}/P_{\text{open}}$, see e. g., Fig. 2. These elements are additively included in the transfer functions in the frames of Fig. 3 – in the order mentioned, i. e. from right to left. Individual characteristics of the elements will also add, but since the elements may be correlated, the variance will not necessarily sum up for each step in the transmission.

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 8

In the lower panels of Fig. 3, the standard deviations for the data in the top panels are shown. It can be seen that the standard deviation at the open entrance is higher than at the eardrum or the blocked entrance. An analysis of correlation [37] shows that this is due to a high negative correlation of the transmission along the ear canal, $P_{\text{eardrum}}/P_{\text{open}}$, and the pressure division at the entrance, $P_{\text{open}}/P_{\text{blocked}}$. This correlation does not surprise, since both terms are highly influenced by the acoustical properties of the ear canal and its termination.

It also becomes evident that the blocked-entrance HRTF has the lowest variation. This means that it is not only convenient from a practical point of view to record at the blocked entrance, but it also results in recordings which may offer a wider range of general applicability. An example of a convenient microphone technique for recording of blocked-entrance sound pressures is shown in Fig. 4.

Also the headphone transfer functions vary across subjects, as will be shown by examples in Sect. 3. The significance of inter-individual differences will thus also depend on the properties of these transfer functions, the procedures for determining them and the design of adequate equalization filters. An analysis – using terms from the model – is given in Møller et al. [77].

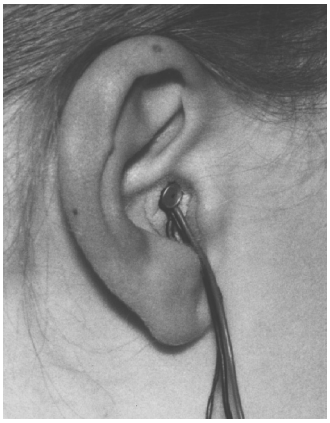

Fig. 4. Example of a microphone technique for blocked-entrance binaural recordings. A hole of suitable size is made in an EAR earplug – for instance by a soldering iron – and a Sennheiser KE 4-211-2 miniature microphone is inserted in the hole. The earplug, with the microphone inside, is then compressed and, once well compressed, earplug and microphone are inserted into the ear canal. Earplug and microphone should be held in position while the earplug decompresses, so that the microphone is ensured the correct position flush with the entrance – after [76]

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 9

3 Headphone Reproduction

As stated earlier, authentic reproduction of binaural recordings is aimed at through one-to-one transmission from the recorded sound pressure to the reproduced one. In the following, examples of headphone transfer functions, pressure divisions and equalization filters [73, 77] are given. For further examples check, e. g., [40, 55].

3.1 Headphone Transfer Functions

In [73] the transfer characteristics of a selection of headphones were studied, mostly of Hi-Fi type – for examples see Fig. 5. It became obvious that none of the headphones had the flat frequency response required for authentic reproduction of binaural signals, compare Sect. 2.2, and that several of them have complicated structures that are difficult to compensate with low-order filters.

For frequencies above 7 kHz the structures are also highly individual, so that individual equalization has to be considered. It was also noted that the headphones have very different frequency responses, even those that are claimed to be designed according to the same design goal, e. g., free-field or diffuse-field – see Sect. 3.4. Further, the sensitivity of the headphones varied considerably between subjects for some headphones. This may be considered a minor problem, but the accurate calibration of level may be important for the generation of perceptually-authentic auditory scenes, particularly for scientific experiments. Thus, generally speaking, the transfer function of the headphones used for the reproduction of binaural signals must be known and compensated for.

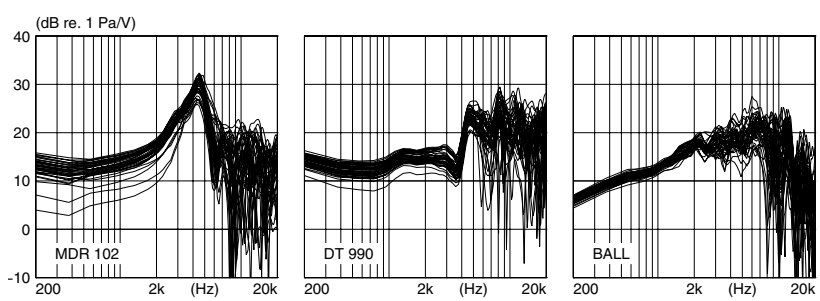

Fig. 5. Headphone transfer-function examples for Sony MDR-102, **left**, BeyerDynamic DT-990 headphones, **center**, and for a free-standing ball loudspeaker, **right**, measured at the blocked-ear-canal entrance of 40 human subjects [73]

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 10

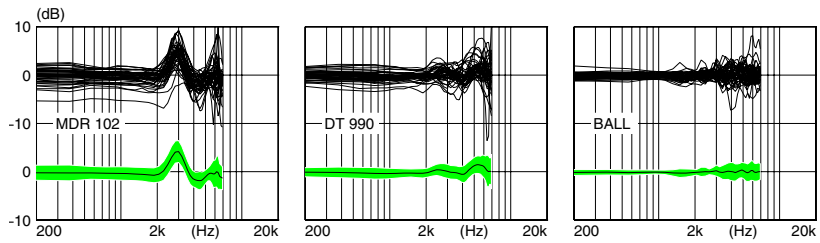

Fig. 6. Examples of pressure-division ratios, PDRs, for the Sony MDR 102, **left**, the BeyerDynamic DT 990 Professional headphones, **center**, and for a free-standing ball loudspeaker, **right**. Data from 40 human subjects [73], individual curves, **above**, mean values ± one standard deviation, **below**. Each PDR is computed on basis of four measurements, i. e. open and blocked entrance measured in the free field and open and blocked entrance measured with the headphone as the source. The validity of the results for higher frequencies is doubtful, thus no results are shown for frequencies above 7 kHz

3.2 Pressure Division

If the pressure division in the headphone situation equals the pressure division in a free-air situation, PDR reduces to unity – refer to Sect. 2.2. The term "free-air-equivalent coupling", FEC, was introduced in [73] to describe headphones for which this is the case. For headphones that have FEC properties, the equalization filter thus only needs to compensate for the headphone transfer function itself, even if recordings are made at the blocked ear canal. For headphones that do not have FEC properties, the equalization filter also needs to compensate for the PDR.

Results from [73], see the examples in Fig. 6, indicate that the pressure divisions in the free-air and headphone situations differ only slightly. Hence, the pressure division ratio can be assumed negligible for some headphones. Not surprisingly, this is the case for small, free-standing ball loudspeakers – Fig. 6, right frame – but also some of the more traditional headphones, including the BeyerDynamic DT 990 Professional – center frame – have pressure division ratios close to unity. Even for headphones for which pressure division ratios differ most clearly from unity, e. g., the Sony MDR-102 – left frame – deviations are, although significant in magnitude, limited to a rather narrow frequency range. In any case, the equalization needed to compensate for the pressure division ratio is much smaller than the one needed to compensate for the headphone transfer function itself. This is particularly relieving from a practical point of view, as determination of the PDR is complicated and requires a total of four measurements, i. e. in the free field, P_{open} and P_{blocked}, and with headphones on, $P^{\mathrm{hp}}_{\mathrm{open}}$ and $P^{\mathrm{hp}}_{\mathrm{blocked}}$ – see Sect. 2.2.

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 11

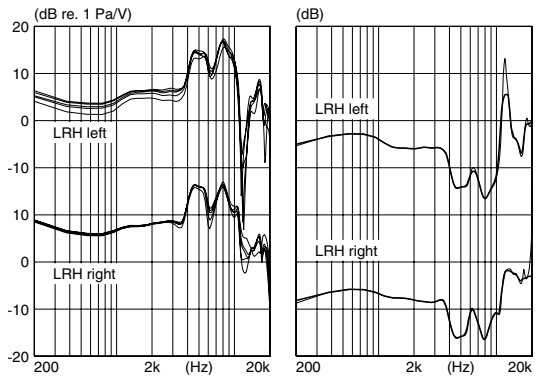

Fig. 7. $P^{\mathrm{hp}}_{\mathrm{blocked}}/E_{\mathrm{headphone}}$ for BeyerDynamic-DT-990-Professional headphone, **left**, subject LRH, five repetitions. 32nd-order-IIR equalization filter, **right** [77]. Target function depicted as *thin line*

3.3 Headphone Equalization

Headphone equalization requires a successful inversion, i. e. de-convolution, of the headphone transfer function – see 8, 9, 10. Numerous mathematical methods exist for this inversion, and – considering the variation in characteristics of headphone transfer functions – it may be that one method may be ideal for one headphone type, and another ideal for another type.

Whatever the method used for inverting the headphone transfer function, the problem generally becomes easier to handle if the transfer function to be inverted represents a minimum-phase system. Generally, all-pass sections do exist in headphone transfer functions, but they are often ignored in the design of the equalization filter – see [69] for further discussion.

An example of measurements of a headphone transfer function and a designed filter is shown in Fig. 7. A target function for the filter was derived from five measurements. For each measurement, the subjects themselves arranged the headphone for most comfortable fit. It can be seen that the variations between measurements are much less than the variations between subjects, compare with Fig. 5. The low variation in the repeated measurements means that an individual headphone filter can be reliably designed. The high variation in transfer functions across subjects means that it probably should be designed individually, at least for critical purposes.

It can further be seen that a very deep dip exists in the left-ear headphone transfer function around 12–13 kHz which, when inverted for the equalization filter, gives a strong peak. Since it *is* slightly dependent on the exact positioning of the headphone, it should also be conservatively compensated for. This is done by reducing the amplitude of the corresponding high peak in the equalization filter, so that excessive amplification is avoided.

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 12

3.4 Pre-Equalization of Binaural Recordings

The possibility of applying a standard equalization for the use with binaural recordings is vital for the wider applicability of the methods. It has been suggested to pre-filter or pre-equalize binaural recordings with a weighting function that corresponds to the inverse of the design goal for standard stereo headphones, i. e. either a free-field or a diffuse-field equalization. This would, in the ideal case, lead to compatibility of the binaural recordings with existing standard-stereo headphones. Unfortunately, the variation across headphones and the general lack of success in reaching these design goals have compromised this idea. See [74] for a comparison of headphone transfer functions with traditional design goals, and [9, 102] for discussions of design philosophies.

4 Performance with Binaural Recordings

The listening experience obtained with reproduced binaural signals has been discussed recurrently, among other things, questioning the fundamentals principles – particularly regarding the issue of whether the frontal direction can be properly reproduced with human-like transfer functions. This was the background for a series of localization experiments with binaural recordings using real [77, 78] and artificial [70, 79] heads. The results are summarized in the following.

4.1 Localization Experiments

A series of listening experiments were carried out in which the localization with binaural recordings was compared to real-life localization. A set-up including 19 loudspeakers at various directions and distances was made in a standard listening room built according to IEC 268–13 [43] – see [59] and Fig. 8. The subjects were asked to indicate on a tablet the loudspeaker where they perceived the sound. In other words, although only loudspeaker positions were possible responses, the subjects were asked to report on the position of "what they heard", i. e. the *auditory event*, and not in any way encouraged to speculate on sound-source positions.

The subjects listened in consecutive experimental series to (a) the stimulus sound played back in the real-life set-up, (b) their own binaural recordings, (c) binaural recordings from other humans and (d) binaural recordings from artificial heads. Responses which did not correspond to the stimulus position, were subdivided into four main *error* categories as inspired by the nature of our hearing, namely

– Distance error ... when the direction was correct but stimulus and response differed in distance. There were four loudspeakers in the front

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 13

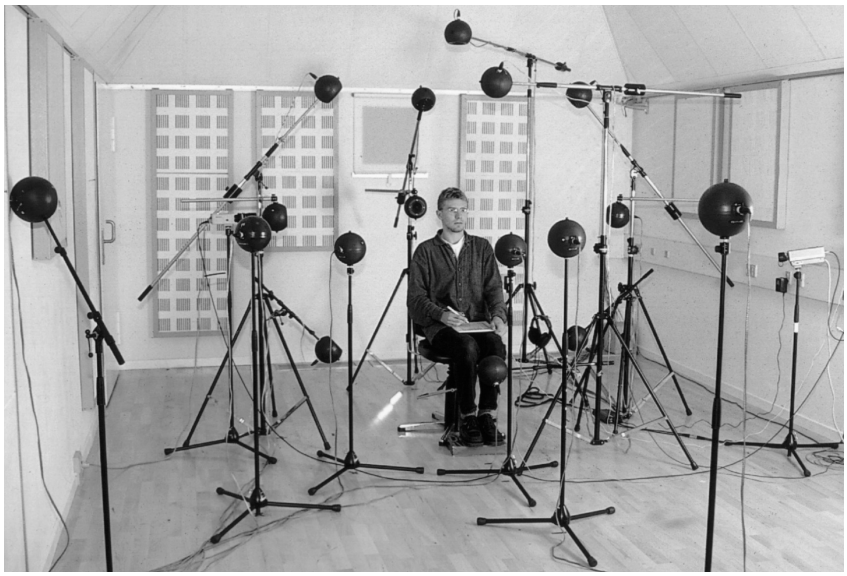

Fig. 8. Photo of set-up in standard listening room [70]

and three loudspeakers at 45° azimuth, right, that could be confused in distance.

- Median-plane errors ... when stimulus and response differed in direction but were both within the median-plane. There were ten source positions in the median plane.
- Within-cone errors ... confusions between source positions that were on "cones" with the same inter-aural time difference. There were two such source positions on the left side, 45° elevated up and down, and five on the right, 45° azimuth in the horizontal plane at three different distances, and 45° elevated up and down.
- Out-of-cone errors ... all other mistakes, indicating severe directional errors.

Directional errors were considered more *severe* than distance errors, thus, if a response represented both a distance *and* directional error, it was counted as a directional error only.

The results of these localization experiments do not depict spatial hearing in every detail or give a full story of human sound localization. They do, however, provide a fairly representative performance map for common and natural listening situations. Authentic reproduction would result in a performance map comparable to the performance map for real-life listening. This, indeed, held true for the experiments with the subjects' own recordings, whereas a statistically-significant degradation was seen for all experiments with reproduction of non-individual and artificial-head recordings.

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 14

4.2 Non-individual Human Recordings

In [78] a panel of 20 listeners participated in localization experiments including binaural recordings from a panel of 30 human *recording heads*. There was considerable variation in the percentage of errors obtained with the different sets of non-individual recordings. The heads were subsequently "ranked" according to the number of median-plane errors, and the head which rendered least median-plane errors in total was "selected". This procedure will always as a result point to a "best" head, even if the low number of errors occurred by chance. An additional experiment was therefore carried out to verify that the recordings from the "selected" head consistently rendered the lower number of errors. The result of this experiment confirmed that it actually did so. This recording head thus incorporates the most common and/or salient features required for good sound localization. Consequently, it was denoted a "typical" human head.

Also others have studied the performance with non-individual binaural signals, see, e. g., [3,5,6,22,49,50,61,80,105–108,112], and other philosophies have been put forward for the selection of "typical" or "optimal" heads. One suggestion is that the heads of persons who localize well themselves, "good localizers", will also provide the best cues for other listeners. This idea can be assessed from Fig. 9. If the ears or head of a good localizer would also provide good cues for other listeners, then the points should follow a monotically increasing trend. Such pattern is not seen – thus this hypothesis is not supported by the current data.

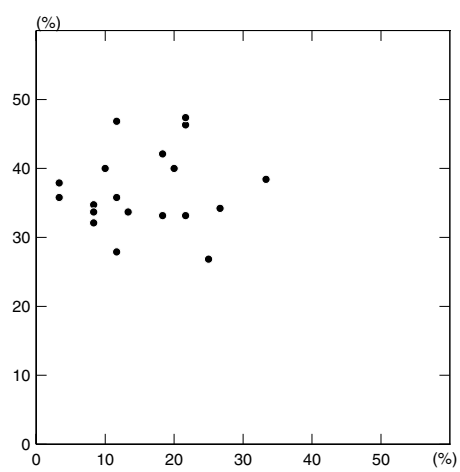

Fig. 9. Average median-plane errors of all listener of the panel, listening to the recording made from specific listeners, *ordinate*, plotted against the specific listeners' own performance in real life, *abcissa* [78]

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 15

4.3 Artificial Heads

The performance of artificial heads was first investigated in the advent of artificial-head technique, e. g., [16, 27, 28, 30, 31, 33, 61, 62, 86–88, 103, 114] – reviewed in [79]. All investigations pointed at imperfect performances, but the differences in experimental conditions, way of specifying and presenting results, computing statistics – only few have employed statistical comparisons at all, and, particularly, the very limited number of subjects in most of these investigations, disable further comparison.

A series of listening tests similar to the localization experiments reported in Sect. 4.1 and 4.2 have been carried out with recordings from artificial heads as well. The same group of listeners participated and the experimental design was identical in every other way. The results, median-plane errors only, of both human and artificial recording heads are shown in Fig. 10.

Figure 10 shows the average percentage of median-plane errors for a given recording head, human as well as artificial. The mentioned variation in error percentages for the human recording heads is clearly seen. The human

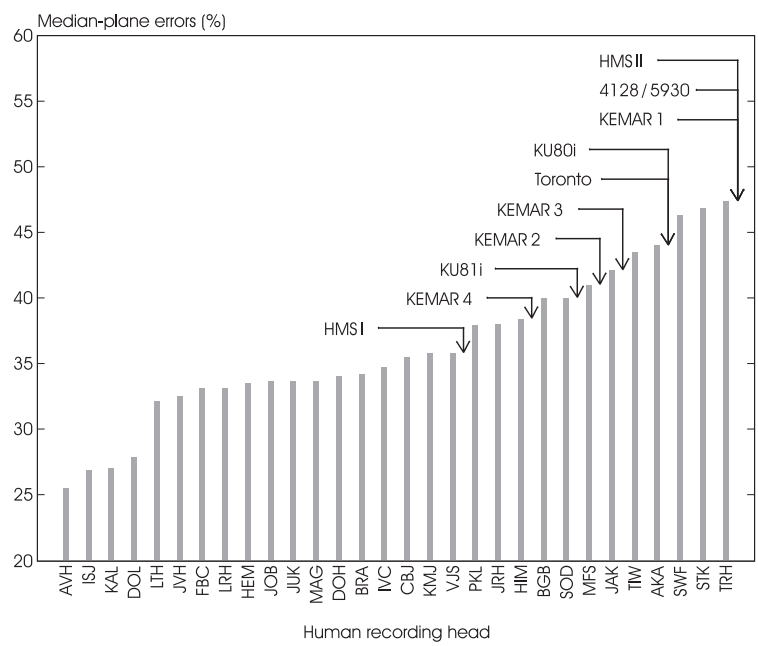

Fig. 10. Average median-plane error percentages for a panel of 20 listeners for the recording head indicated on the abscissa. Arrows indicate the performance with artificial head recordings: KEMAR X from Knowles Electronics with ear 'X', i. e. four types of ears included, KU80 and KU81i, HMS I and HMS II from Head Acoustics, 4128/5930 from Brüel and Kjær, types 4128 and 5930, respectively, and the TORONTO experimental head from Toronto University – adapted from [79]

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 16

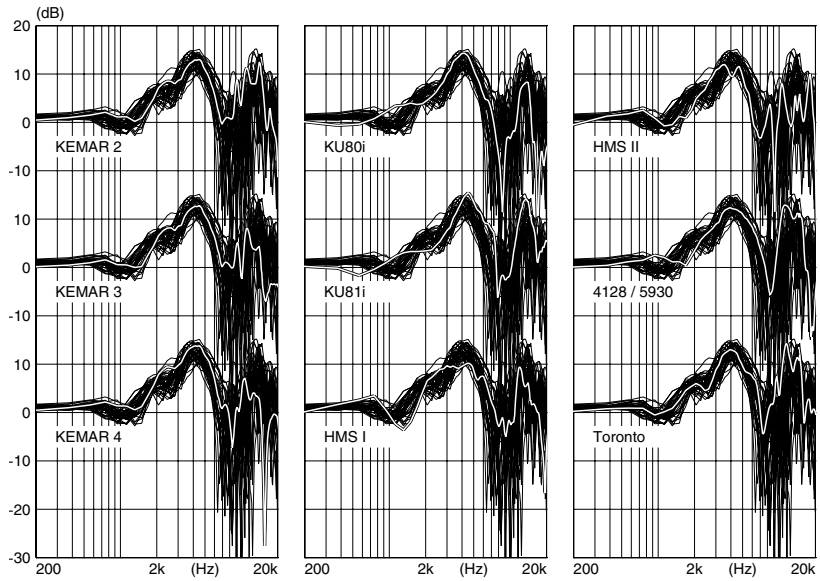

Fig. 11. Blocked-entrance HRTFs for frontal sound incidence for 40 humans, *thin lines* and manikins, *white line*, [36]

recording head, AVH, was the one that gave the lowest median-plane error percentage, i. e. 26%. The performances for several artificial heads are indicated with arrows inserted in between the columns for humans. It is seen that even the better artificial heads compare only to the human recording heads in the poorer end of the scale, namely, 60% of the human recording heads are better. This is disappointing as artificial heads should in principle represent the "average" or "typical" listener.

The differences between human and artificial heads may be assessed objectively by comparison of their HRTFs. Frontal HRTFs for the artificial heads are given in Fig. 11. It can be seen that the human HRTFs are well grouped, though with higher variations the higher the frequency is, as can also be seen in Fig. 3. It can further be seen that the HRTFs of the artificial heads do generally not well represent the human HRTFs. The artificial-head HRTFs deviate in several cases considerable from the "average" or "typical" structure, and often do not even fall within the range of humans. This is disappointing, since the general design goal for the artificial heads is to replicate humans.

The deviations in Fig. 11 do, however, not directly predict the ranking shown in Fig. 10. The Toronto head has a very human-like transfer function, but it has a relatively high percentage of median-plane errors even so. The HMS I, on the contrary, has a fairly non-human-like transfer function, but has a lower percentage of median-plane errors, at least when compared to the

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 17

other artificial heads. If more directions are inspected – not shown here, it is seen that whereas the front direction may reveal certain types of shortcomings of the heads, other directions reveal other types of shortcomings that result in other types of localization errors [70, 79]. Generally, both the results from the listening experiments and the analysis of the transfer functions indicate that the differences between artificial heads and humans are larger than differences between humans.

Successive experiments were carried out including more artificial heads, among others an in-house experimental recording head, VALDEMAR. These experiments were carried out over several years with a varying group of listeners, thus requiring more caution when comparing across investigations. It was shown that, generally, the various experimental series gave results of good validity, and that artificial heads provide the better performances the more human-like they are [70].

5 Binaural Synthesis

A human recording head, or alternatively an artificial head, serves to perform the alteration of the sound field that the listener would have caused in the recording situation. When placed in a free sound field, the listener will obstruct an incoming sound wave, and ears, head and body will cause a linear filtering of the sound signal. This filtering is completely and uniquely described by the HRTFs.

Figure 12 shows the HRTF in the time and frequency domains for a single subject with sound coming from the side. It is obvious in both domains that the sound reaching the ear closest to the sound source, left ear, is stronger than the sound reaching the ear opposite to the sound source, right ear. In the frequency-domain representation it can also be seen that the sound transmission to the left ear exceeds 0 dB, which is mainly due to pressure

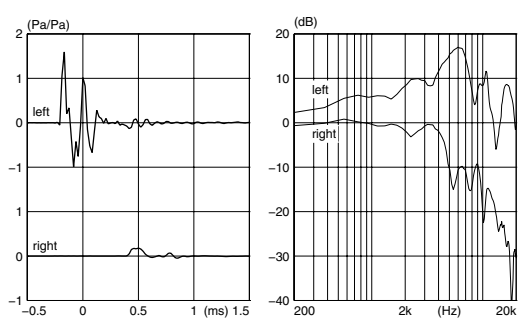

Fig. 12. Head-related transfer function, HRTF, in the time domain, **left**, and in the frequency domain, **right** – measured at the blocked entrance with sound from the left [76] – data from subject LRH

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 18

build-up and diffraction around ear, head and body. The frequency response for the right-ear sound transmission resembles the characteristics of a low-pass filter – due to the shadowing effect of the head. In the time domain it can be seen that the sound at the closest ear arrives some hundred microseconds earlier than the sound at the opposite ear.

It is the direction-dependent characteristics in the sound signals reaching our two ears which enable us to localize the sound source by our hearing. A very strong localization cue is the arrival-time difference between left and right ear signals, which *positions* the auditory event laterally on a given spatial cone of equal inter-aural time differences, also when other cues are ambiguous – thus the term "cone of confusion".

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, ILD/IID as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.

5.1 Measurement

The ideal HRTF describes the transformation from a freely-propagating plane wave to sound pressures that occur when a listener *enters* the sound field and, thus, obstructs wave propagation. In the laboratory, plane waves exist only in approximation. When a source is set up for measurements in the anechoic chamber, it will be within a limited distance of the subject, and the wave that hits the subject will thus not be exactly plane, but marginally spherical.

The an-echoic chamber also presents an approximation in itself, since the absorption at the boundaries will fail to work at low frequencies, and possibly – depending on design and material – also for higher frequencies. Platforms, chairs, etc., that are needed for sources and subjects, also cause unintended reflections and, thus, are a source of measurement errors. Hence, it is preferable to minimize the number of loudspeakers present in the room, and to reduce the amount of equipment to a minimum needed carrying, supporting and/or fixating the subjects. One solution is to use an arc with one or more loudspeakers, and to move the arc [57, 58] or the listener [76]. An example is shown in Fig. 13.

In recent years, a number of measurement techniques have been developed that return the transfer function in the time-domain representation, i. e. for the present application, as head-related impulse responses, HRIRs. A particularly successive technique is based on a pseudo-random noise stimulus, known as a maximum-length sequence, e. g., [18, 90, 97, 116]. This technique enables a time gating of the measured impulse where later parts of the impulse can be rejected. In this way it is, e. g., possible to "gate out" an unavoidable reflection from an imperfectly absorbing wall, or from the platform carrying the subject.

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 19

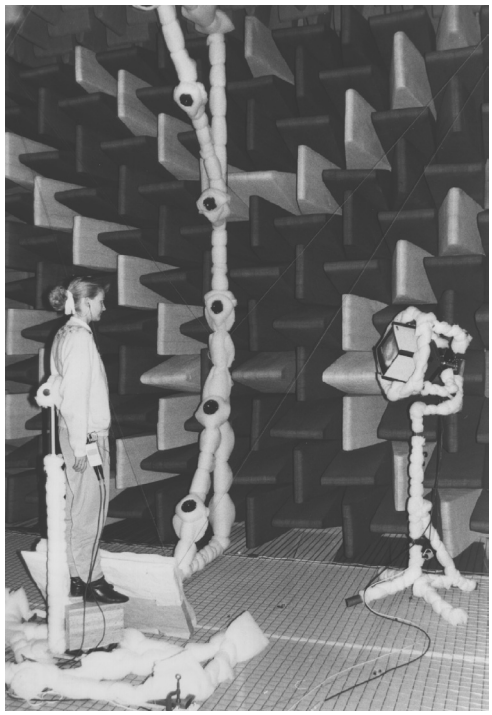

Fig. 13. Set-up for measurement of full-sphere HRTFs. Eight loudspeakers are placed in an arc, and the subject is standing on a platform supported with a back-rest. The platform can be turned in discrete steps to render source directions for the whole sphere. The monitor shows the subject from above and enables the subject to center head position accurately [76]. For photo-technical reasons the monitor is closer to the subject than during measurements

Raw measurement of P_{ear} and P_{ref} will include the transfer functions of all elements of the transmission chain, e. g., power amplifier, loudspeaker, microphone, microphone amplifier, anti-aliasing filters. Thus, a prerequisite for cutting away unwanted reflections from, e. g., supporting platforms, is that the head-related impulse response convolved with the impulse responses of amplifier(s) and loudspeaker does not overlap in time with the unwanted reflections. This can lead to a peculiar contradiction of compromises, since, e. g., a good loudspeaker response at low frequencies is desirable for a good signal-to-noise ratio, SNR, but a such loudspeaker will often have a long impulse response, and will thus limit the possibility of gating out reflections. Also, band-limiting filters with steep slopes of their transfer function can result in long impulse responses.

HRTFs are computed by a complex division in the frequency domain – see 4. For the highest frequencies, i. e. above the cut-off frequency of the anti-aliasing filters and other band-limiting elements, the result will be based on

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 20

one set of invalid data divided by another set of invalid data. This results in a time-aliased impulse response with random high-frequency, i. e. "ringing" components which must be removed by appropriate low-pass filtering. For the lowest frequencies it is tempting to do the same, but as will be demonstrated in the following, it is important to control the low-frequency characteristics in a more direct way.

5.2 Low-Frequency Control

The two measurements used for the calculation of an HRTF are always carried out with equipment that has a lower limiting frequency. Thus, even if the measurements return values at DC, these values are not the results of true measurements, but they simply reflect the off-set-voltage properties of the amplifiers and A/D converters involved. As a consequence, the complex division carried out to obtain the HRTF gives a wrong and more or less random value at DC. One might think that this is not important as long as the signals do not have DC components. However, an incorrect DC value has consequences for the processing of signals in a wide frequency range as it will be seen in the following.

Fortunately, it is easy to correct for the DC value. At low frequencies, the presence of a person does not alter a propagating sound wave and an HRTF will thus asymptotically approach unity gain, i. e. 0 dB, when the frequency decreases toward DC. As a consequence, the calculated – and wrong – DC value of the HRTF can simply be replaced by unity in the frequency domain, just before taking the inverse *Fourier* transform to obtain the HRIR. Alternatively, the individual taps in the HRIR filter can be adjusted accordingly.

Figure 14 shows the consequence of slightly incorrect DC values. In the time-domain representation the errors certainly look modest, but in the frequency-domain representation it is seen that the magnitude is severely affected not only at DC, but also at frequencies well within the range of

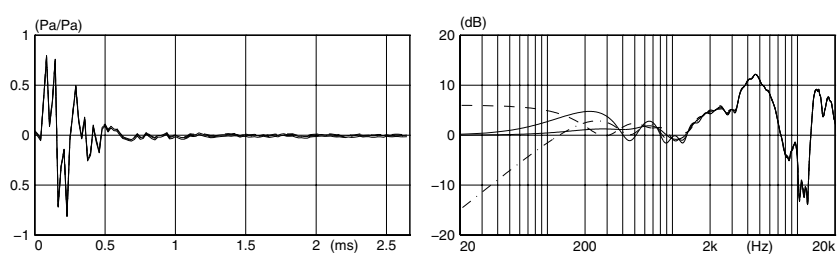

Fig. 14. Blocked-entrance HRTF in the time domain, 128 points, **left**, and in the frequency domain, **right**. Transfer function for a DC value corrected to unity, *medium solid line*, for a slightly-too-high DC value, *dashed line*, for a DC value of zero, *dash-dotted line*, and for a slightly-too-low DC value, *thin solid line* [72]

.

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 21

hearing. At the frequencies, which equal an integer times the frequency resolution of the original HRIR, in this case 375 Hz, all curves coincide. The concerning observation is the behaviour of the amplitude response between these points. These bumps are several dB in magnitude, and can be clearly audible as miscolouration of the lowest frequencies. If, by misfortune, the DC-value is slightly too low in the HRIR for the one ear and slightly too high in the HRIR for the opposite ear, then a large difference can exist. This does not normally occur in real life, and can give a quite unpleasant feeling of sub- or super-pressure in the ears. Hence, in the authors' experience, it is generally required to correct the DC-value to ensure a good sound quality.

5.3 Filter Length and Order

It appears from visual inspection that the length of the HRIR itself is on the order of 1 ms or slightly more, e. g., Fig. 12 and 14. The HRIR can be implemented directly as an FIR filter with an order corresponding to the number of taps. The possibility of reducing the order for minimizing computational resources has repeatedly been investigated, e. g., [3, 4, 14, 15, 38, 41, 46, 64, 93, 94] – for a review see [42].

Figure 15 shows the probability of detection in a three-alternative forced choice listening experiment, where finite-impulse-response filters, FIR filters, with reduced order were compared with the original HRIR, i. e. 256-point FIR. The sampling frequency was 48 kHz, so 48 taps correspond to a filter duration of 1 ms. From Fig. 15 it can be seen that for high filter orders, i. e. down to 72 taps, the difference is undetectable. For shorter filters the differences are heard, though still with a low detection at 48 taps and even at 24 taps. Thus, depending on the application at hand, filters of these or-

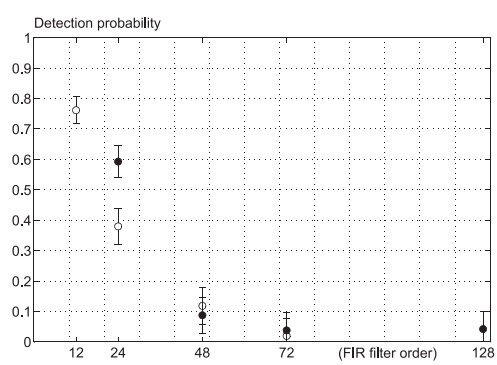

Fig. 15. Detection probability of reduced-order FIR implementations, pooled data for 15 subjects and 17 directions covering the whole sphere. Individual synthesis and individual headphone calibration was employed with speech, *open circles*, and noise signals, *solid circles* [94]

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 22

ders may in many cases work satisfactorily. In general, with current signal processing technology, FIR filters of 72 taps are considered relatively short.

The results match well with what one would conclude from visual inspection of HRIRs and from considering the physical dimensions of humans. Nevertheless, it is not an uncommon understanding that the length needs to be considerably longer, and that the quality of the filter is better the more taps are included. One reason for disagreements is the possible lack of low-frequency control. The longer the filter is, the lower is the frequency at which the "bumps" occur. This means that the errors can be indirectly shifted to such low frequencies that they become inaudible. An easily overlooked fact is that similar errors are likely to occur when the transfer functions are measured using frequency-domain methods or when the filtering is carried out in the frequency domain.

The possible advantages of using alternative, possibly computationally more efficient filter representations, e. g., infinite-impulse-response filters, IIR filters, has also been investigated, e. g., [3,13,15,38,41,44,51,52,64,91,93,94]. None of the results points to a convincingly superior computational performance. This may – at least partly – be due to the fact that IIR filters are often superior for describing low frequency spectral details, which are virtually non-existent in HRTFs.

It is a pre-requisite for many IIR-filter-design methods that the system has minimum-phase properties. This is, however, not the case for HRTFs, e. g., [76]. All HRTFs include a linear-phase component, i. e. pure delay, which is vital for maintaining the correct inter-aural time difference. HRTFs do also typically have one or more all-pass sections. Yet, it was demonstrated [69] that, without any audibly consequences, the all-pass sections may be replaced by an additional pure delay – determined as the low-frequency phase delay or group delay of the all-pass sections. Though the use of IIR-filter design may not be computationally favourable, the possibility of using minimum-phase approximations has many other practical advantages.

5.4 Performance of Binaural Synthesis

There have been numerous investigations testing or using binaural synthesis, e. g., [3,5,6,49,50,80,105–108,112]. Most have been made in an-echoic chambers with the subjects in fixed positions, thus simulating a single sound wave with a given direction, i. e. static free-field simulation. There may be several reasons for this. Technically, all that is required for a free-field simulation is the HRTFs, and the appropriate headphone equalizations. Second, if a simulation of more realistic environment should be made, a sound-transmission-analysis program is required. Such a program will also have its approximations and imperfections, which will influence and possibly deteriorate the overall result. Third, the first-arriving sound wave has a leading significance, when it comes to directional hearing, the so called "precedence effect", see e. g., [9] for a review. Thus, the static free-field simulation is in a sense a

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 23

relevant and critical test of the success in rendering authentic *directional* perception.

A general conclusion from these investigations is that, when properly implemented, the binaural synthesis is successful in giving a near-real-life performance. The results have contradicted some of the myths associated with binaural synthesis, e. g., that the technique cannot render frontal localization or inevitably leads to inside-the-head locatedness. In the studies where these effects have been observed, often non-individual synthesis has been used, or the studies have had such diverting objectives that details of the synthesis technique have been given very little attention.

The binaural synthesis can, however, be considerably improved by making use of a single very important difference between recording and synthesis of binaural signals, namely, the possibility in the synthesis to change the HRTFs interactively when the direction of the sound incidence changes due to listeners' head movements. It is well known that in real life head turns are often used unconsciously, e. g., to resolve front/back confusions, e. g., [8].

With binaural recording – and static synthesis, the listener has to sit still to have an authentic listening experience, since a recording cannot respond to later dynamical changes, such as the listener's head movements. This is not an impediment of principle nature for the synthesis technique. The position and orientation of the listener can be tracked by so called head trackers, e. g., magnetic devices, and the synthesis can be made responsive to these changes by continuously updating signal processing.

The importance of correct updating of binaural synthesis in response to head-movements has been understood for years. More recently, a number of investigations, e. g., [7, 20, 92, 113] have demonstrated the proficiency of dynamic synthesis. Localization errors such as front/back confusion are almost non-existent, and cone-of-confusion errors are well resolved in general, also for non-individual recordings.

The very different properties of the listening experience in dynamic simulation calls for considerably revised experimental schemes for scientific assessments of the methods. Localization experiments are still fundamental in such assessments, but the continuously-changing signal processing can lead to artifacts, e. g., "clicks", that degrade the quality of the sound, without severely impeding localization.

6 Applications

In the following, the principles of some *classical* application areas of binaural technique, mostly synthesis, are introduced – see also [10, 19, 84], this volume.

Room Simulation

Room-simulation systems perform numerical analyses of the acoustical environment and, with the use of binaural synthesis, the analysis results may

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 24

be rendered audible. Room-simulation systems are usually developed for assistance in architectural design and acoustical adjustment of concert halls, auditoria, theaters etc. Re-arrangements of walls, orchestras or audience can be listened to before alterations are made in the real world. Listening conditions at critical places in the concert hall and the speech intelligibility of the auditorium can be checked by actual listening. The simulation systems cannot only be used to suggest physical changes of the room, but also, by applying superposition, multiple sound sources can be simulated in the computer model and the system can be used for adjusting any multi-channel public address or electro-acoustical enhancement system prior to the costly set-up. The room-simulation system can also be used in various contexts where standardized listening conditions are needed. This could be used for instance for the evaluation of different loudspeakers. This way the loudspeaker producer can avoid building a rather costly listening environment. It can also be used in mobile monitoring environments for studio recording and mixing that lack space for a loudspeaker set-up. The potential of room-simulation systems has been, and is, a driving force in the research and development of the binaural-synthesis techniques. In later years the term binaural "auralization" has been used in general for rendering such numerical predictions audible – see, e. g., [2, 26, 53, 56, 60, 63, 81, 104, 115].

Binaural Mixing Consoles

A common way of making recordings in studios is to record each instrument or voice in its own channel, and then afterwards mix these down to the intended stereo perspective, using time and intensity differences. With a binaural mixing console such multi-channel recordings are transferred into binaural signals in the mixing phase. The binaural mixing console can thus be regarded an *electronic artificial head*. If an artificial head is used for the recording, then the physical set-up in the recording studio defines the position of the sound source for the play-back situation. These recordings, as well as ordinary recordings processed by means of the binaural mixing console, can be mixed together to binaural signals. Yet, the success of using binaural recordings in audio has been limited, possibly due to the lack of standardized methods for the processing procedures – for examples, see, e. g., [32, 34, 45, 48, 95].

Surround Sound by Headphones

The use of multi-channel sound-reproduction systems is increasing, e. g., in HDTV and DVD formats. The compatibility with headphone play-back is not trivial. Using binaural synthesis, the five channels can be transformed into a two-channel signal for headphone play-back. The idea is not at all new, e. g., [11], and has also been proposed for headphone play-back of traditional stereo recordings, where the same compatibility problem exists. Today, two approaches are seen. One that strives to simulate the optimal surroundings,

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 25

and one that strives to simulate the end-user multi-channel set-up and environment, whatever the properties and qualities of this – for examples, see, e. g., [1, 65, 83, 89, 98].

Communication Systems

Binaural technique is often proposed for communication systems, e. g., [24, 25, 47, 109], for situations where the listener must pay attention to more than one communication channel at a time. The idea is to spatially arrange the communication channels around the listeners and, in this way, enable them to distinguish between the different channels as in daily life. The advantage of spatially arranged channels originates from the fact that our hearing can focus on one out of several sound sources in noisy environments with the sounds being heard at different directions, the so-called "cocktail-party effect". Implementations of this kind are, e. g., seen within military research institutions to be used between pilots and air control towers, e. g., [107]. More general kinds of tele-conferencing systems also exist.

3-D Auditory Displays

The idea of 3-D auditory displays is to create spatially arranged auditory icons for situations where it is either natural to have audible information passed or where, for instance, the visual information channel is not available or is fully occupied, e. g., [17, 21, 82, 85]. The sound of a fuel tank drying out could carry an important message for a pilot, and the sound of a printer could be a way to transfer the graphical interfaces of modern computers into something blinds can use. Auditory icons – sometimes called "earcons" – can also be used for transfer of information which has no origin in the auditive world. Examples are radar or ultrasonic information, where the distance information is passed and used, for instance, as a navigation tool for blind persons, or the association of sounds to stock exchange information, e. g., high- and low-pitched sounds for rates going up and down – e. g., [23]. Obviously, earcons serve a purpose even when they are not arranged spatially, yet, but 3-D auditory displays offer an environment where the user may effectively monitor more activities at the same time.

Virtual Reality

In Virtual-Reality applications, e. g., [12, 29, 84, 96] the ultimate objective is immersive simulation of a non-existing environment. The distinction between VR systems and other sound simulation systems can sometimes be difficult to see. VR systems aim at providing stimuli to a persons' senses, which are perceptually plausible to such an extent that these persons develop a persistent experience of actually being somewhere else – so called sense of "presence". The perceptual world that they are experiencing this way may only exist as a computer model which the computer uses to control various kinds of actuators

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 26

that stimulate multiple modalities – besides the auditory one, one or more of the following, i. e. visual, tactile, olfactory, gustatory and proprioceptive ones.

Acknowledgement

This chapter reviews a series of studies carried out at the Department of Acoustics of Aalborg University, where all members of the staff have contributed to an inspiring atmosphere. In particular, *S. B. Nielsen, P. Ch. Frandsen, M. F. Sørensen, C. B. Larsen* – former *Jensen, J. Sandvad, F. Christensen, S. K. Olesen, P. Minnaar*, and *J. Plogsties* are thanked for their collaboration in different phases of the work. Various sources of financial support are also acknowledged, particularly the Danish Technical Research Council, the European Commission, the Danmon Research Fund, the *C. W. Obel*'s Fund, the National Agency of Trade and Industry, and the national Center for IT Research.

References

1. Aarts R (2003) Applications of DSP for Sound Reproduction Improvement. 23rd Audio Engr Soc Int Conf, Copenhagen, Denmark, paper 4
2. Ahnert W, Feistel R (1993) EARS auralization software. J Audio Engr Soc 41:894–904
3. Asano F, Suzuki Y, Sone T (1990) Role of spectral cues in median plane localization. J Acoust Soc Amer 88:159–168
4. Begault D R (1991) Challenges to the successful implementation of 3-D sound. J Audio Engr Soc 39:864–870
5. Begault D R (1992) Perceptual effects of synthetic reverberation on three-dimensional audio systems. J Audio Engr Soc 44:895–904
6. Begault D R, Wenzel E M (1991) Headphone localization of speech stimuli. Proc Human Factors Soc, 35th Ann Meet, Santa Monica, CA 82–86
7. Begault D R, Wenzel E M (2001) Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J Audio Engr Soc 49:904–916
8. Blauert J (1997a) An introduction to binaural technology. In: Gilkey R H, Anderson T R (eds) Binaural and spatial hearing in real and auditory environments 593–609, Lawrence Erlbaum, Mahwah NJ
9. Blauert J (1997b) Spatial hearing: the psychophysics of human sound localization. 2nd rev edn. MIT Press, Cambridge MA
10. Blauert J (2005) Analysis and synthesis of auditory scenes. Chap 1 this vol
11. Blauert J, Laws P (1973) True simulation of loudspeaker sound reproduction while using headphones. Acustica 29:273–277
12. Blauert J, Lehnert H, Sahrhage J, Strauss H (2000) An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation. ACUSTICA/acta acustica 86:94–102

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 27

13. Blommer M A (1996) Pole-zero modeling and principical component analysis of head-related transfer functions. Doct diss. University of Michigan, Ann Arbor, Michigan
14. Blommer M A, Wakefield G H (1994) On the design of pole-zero approximations using logarithmic error measure. IEEE Trans Sig Processg 42:3245–3248
15. Blommer M A, Wakefield G H (1997) Pole-zero approximations for head-related transfer functions using logarithmic error criterion. IEEE Trans Speech Audio Processg 5:278–287
16. Boerger G, Laws P, Blauert J (1977) Stereophone Kopfhörerwidergabe mit Steuerung bestimmter Übertragungsfaktoren durch Kopfdrehbewegungen (Stereophonic reproduction by earphones with control of special transfer functions through head movements). Acustica 39:22–26
17. Bolia R S, D'Angelo W R, McKinley R L (1999) Aurally aided visual search in three-dimensional space. Hum Fact 41:664–669
18. Borish J, Angell J B (1983) An efficient algorithm for measuring the impulse response using pseudorandom noise. J Audio Engr Soc 31:478–488
19. Braasch J (2005) Modelling of binaural hearing. Chap 4 this vol
20. Bronkhorst A M (1995) Localization of real and virtual sound sources. J Acoust Soc Amer 98:2542–2553
21. Bronkhorst A W, Veltman J A H, van Breda L (1996) Application of a three-dimensional auditory display in a flight task. Hum Fact 38:23–33
22. Butler R A, Belendiuk K (1977) Spectral cues utilized in the localization of sound on the median sagittal plane. J Acoust Soc Amer 61:1264–1269
23. Carlile S, Cohen M (Eds) (2002) Proceedings of the 8th International Conference on Auditory Display. Advanced Telecomm Res Inst (ATR), Kyoto, Japan (www.icad.org)
24. Cohen M, Koizumi N (1992) Exocentric control of audio imaging in binaural telecommunication. IEICE Trans Fund Electr E75A(2):164–170
25. Cohen M (2003) The internet chair. Int J Hum-Comput Interact 15:297–311
26. Dalenbäck B-I, Kleiner M, Svensson P (1993) Audibility of changes in geometric shape, source directivity, and absorptive treatment – experiments in auralization. J Audio Engr Soc 41:905–913
27. Damaske P, Wagener B (1969) Richtungshörenversuche über einen nachgebildeten Kopf (Directional hearing tests by aid of a dummy head). Acustica 21:30–35
28. Damaske P, Mellert V (1969/70) Ein Verfahren zur richtungstreuen Schallabbildung des oberen Halbraumes über zwei Lautsprecher (Sound reproduction of the upper semi-space with directional fidelity using two loudspeakers). Acustica 22:153–162
29. Djelani T, Pörschmann C, Sahrhage J, Blauert J (2000) An interactive virtual-environment generator for psychoacoustic research. II: Collection of head-related impulse responses and evaluation of auditory localization. ACUSTICA/acta acustica 86:1046–1053
30. Genuit K (1984) Ein Modell zur Beschreibung von Aussenohrübertragungseigenschaften (A model for description of outer-ear transfer functions). Doct diss. Techn Hochsch Aachen, Aachen
31. Genuit K (1988) Simulation des Freifeldes über Kopfhörer zur Untersuchung des Räumliche Hörens und der Sprachverständlichkeit (Simulation of the free field using headphones for investigation of the spatial hearing and speech intelligibility). Audiol Akust 6:202-221

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 28

32. Genuit K, Gierlich H-W, Künzli U (1992) Improved Possibilities of Binaural Recording and Playback Techniques. 92nd Audio Engr Soc Conv, Amsterdam, The Netherlands, preprint 3332. Abstr in: J Audio Engr Soc 40:444
33. Genuit K, Platte H-J (1981) Untersuchungen zur Realisation einer richtungstreuen Übertragung mit elektroakustischen Mitteln. Fortschr Akust, DAGA'81, Berlin, Germany 629–632
34. Gierlich, H-W, Genuit K (1987) Processing Artificial-Head Recordings. 82nd Audio Engr Soc Conv, London, UK, preprint 2460. Abstr in: J Audio Engr Soc 35:389–390
35. Gilkey R H, Anderson T R (1997) Binaural and spatial hearing in real and virtual environments. Lawrence Erlbaum, Mahwah N J
36. Hammershøi D, Møller H (1992) Artificial heads for free field recording: How well do they simulate real heads? Proc. 14th Int Congr Acoust, ICA, Beijing, paper H6–7
37. Hammershøi D, Møller H (1996) Sound transmission to and within the human ear canal. J Acoust Soc Amer 100:408–427
38. Hartung K, Raab A (1996) Efficient modelling of head-related transfer functions. Acta Acustica 82:suppl 1:88
39. Hellström P A, Axelsson A (1993) Miniature microphone probe tube measurements in the external auditory canal. J Acoust Soc Amer 93:907–919
40. Hirahara T (2004) Physical characteristics of headphones used in psychophysical experiments. Acoust Sci & Tech 4:276–285
41. Huopaniemi J, Karjalainen M (1997) Review of digital filter design and implementation methods for 3-D sound. 102nd Conv Audio Engr Soc, Munich, Germany, preprint 4461. Abstr in: J Audio Engr Soc 45:413
42. Huopaniemi J, Zacharov N, Karjalainen M (1999) Objective and subjective evaluation of head-related transfer function filter design. J Audio Engr Soc 47:218–240
43. IEC/TR 60268-13 (1981) Listening tests on loudspeakers. Publication 268-13, Int Electrotechn Comm, IEC, Geneva
44. Jenison R L (1995) A spherical basis function neural network for pole-zero modeling of head-related transfer functions. Proc IEEE worksh appl sig processg audio and acoust, New Paltz NY, IEEE catalogue number 95TH8144
45. Jot J-M, Wardle S, Larcher V (1998) Approaches to Binaural Synthesis. 105th Audio Engr Soc Conv, San Francisco, CA, preprint 4861. Abstr in: J Audio Engr Soc 46:1038
46. Jot J M, Warusfer O, Larcher V (1995) Digital signal processing issues in the context of binaural and transaural stereophony. 98th Conv Audio Engr Soc, Paris, France, preprint 3860. Abstr in: J Audio Engr Soc 43:396
47. Kang S H, Kim S H (1996) Realistic audio teleconferencing using binaural and auralization techniques. ETRI J 18:41–51
48. Karamustafaoglu A, Spikofski G (2001) Binaural Room Scanning and Binaural Room Modeling. 19th Audio Engr Soc Int Conf, Schloss Elmau, Germany, paper 1880
49. Kawaura J, Suzuki Y, Asano F, Sone T (1989) Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear (in Japanese). J Acoust Soc Jpn (J) 45:756–766
50. Kawaura J, Suzuki Y, Asano F, Sone T (1991) Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. J Acoust Soc Jpn (E) 12(5):203–216

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 29

51. Kendall G S, Martens W L (1984) Simulating the cues of spatial hearing in natural environments. Proc Int Comp Music Conf, Paris, France, 111–125
52. Kendall G S, Rodgers A P (1982) The simulation of three-dimensional localization cues for headphone listening. Proc Int Comp Music Conf 225–243
53. Kleiner M, Dalenbäck B-I, Svensson P (1993) Auralization – an overview. J Audio Engr Soc 41:861–875
54. Kulkarni A, Colburn H S (1995) Efficient finite-impulse-response filter models of head-related transfer function. J Acoust Soc Amer 97:3278
55. Kulkarni A, Colburn S (2000) Variability in the characterization of the headphone transfer function. J Acoust Soc Amer 107:1071–1074
56. Kuttruff K H (1993) Auralization of impulse responses modelled on the basis of ray-tracing results. J Audio Engr Soc 41:876–880
57. Langendijk E H A, Bronkhorst A W (2000) Fidelity of three- dimensional-sound reproduction using a virtual auditory display. J Acoust Soc Amer 107:528–537
58. Langendijk E H A, Bronkhorst A W (2002) Contribution of spectral cues to human sound localization. J Acoust Soc Amer 112:1583–1596
59. Langvad B, Møller H, Budzynski G (1989) Testing a new listening room. Arch Acoust 14:45–60
60. Larsson P, Vastfjall D, Kleiner M (2002) Better Presence and Performance in Virtual Environments by Improved Binaural Sound Rendering. 22nd Audio Engr Soc Int Conf, Espoo Finland, paper 228
61. Laws P. Platte H-J (1975) Spezielle Experimente zur kopfbezogenen Stereophonie (Special experiments regarding head-related stereophony). Fortschr Akust, DAGA'75, Braunschweig, Germany 365–368
62. Laws P, Platte H-J (1978) Ein spezielles Konzept zur Realiserung eines Kunstkopfes für die kopfbezogenes stereofone Aufnahmetechnik (The design of a special artificial head for head-related stereophony). Rundfunktechn Mitt 22:28–31
63. Lehnert H, Blauert J (1993) Principles of binaural room simulation. Appl Acoust 36:259–291
64. Mackenzie J, Huopaniemi J, Välimäki V, Kale I (1997) Low order modelling of head-related transfer functions using balanced model truncation. IEEE Sig Processg Let 4:39–41
65. McKeeg A, McGrath D S (1997) Using Auralization Techniques to Render 5.1 Surround to Binaural and Transaural Playback. 102nd Audio Engr Soc Conv, Munich Germany, preprint 4458. Abstr in: J Audio Engr Soc 54:412
66. Mehrgardt S (1975) Die Übertragungsfunktion des menschlichen Aussenohres: Richtungsäbhangigkeit und genauere Bestimmung durch komplexe Strukturmittelung (The transfer function of the human external ear: dependence on direction and accurate determination by means of complex structural averaging). Fortschr Akust, DAGA'75, Braunschweig, Germany 357–361
67. Mehrgardt S, Mellert V (1977) Transformation characteristics of the external human ear. J Acoust Soc Amer 61:1567–1576
68. Middlebrooks J C, Makous J C, Green D M (1989) Directional sensitivity of sound-pressure levels in the human ear canal. J Acoust Soc Amer 86:89–108
69. Minnaar P, Plogsties J, Olesen S K, Christensen F, Møller H (200x) The audibility of all-pass components in binaural synthesis and reproduction. In preparation

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 30

70. Minnaar P, Olesen S K, Christensen F, Møller H (2000) Localization with binaural recordings from artificial and human heads. J Audio Engr Soc 49:323–336
71. Møller H (1992) Fundamentals of binaural technology. Appl Acoust 36:171–218
72. Møller H, Hammershøi D, Jensen C B, Sørensen M F (1995) Method of generating binaural signals with filter sets – uses head-related transfer functions with short time domain descriptions and low differences between individuals. Patent, PCT/DK 95/00089, publication No. WO 95/23493
73. Møller H, Hammershøi D, Jensen C B, Sørensen M F (1995) Transfer characteristics of headphones measured on human ears. J Audio Engr Soc 43:203–217
74. Møller H, Jensen C B, Hammershøi D, Sørensen M F (1995) Design criteria for headphones. J Audio Engr Soc 43:218–232
75. Møller H, Minnaar P, Plogsties J, Olesen S K, Christensen F (200x) The audibility of all-pass sections in electro-acoustical transfer functions. In preparation
76. Møller H, Sørensen M F, Hammershøi D, Jensen C B (1995) Head-related transfer functions measured on human subjects. J Audio Engr Soc 43:300–321
77. Møller H, Sørensen M F, Jensen C B, Hammershøi D (1996) Binaural technique: Do we need individual recordings? J Audio Engr Soc 44:451–469
78. Møller H, Jensen C B, Hammershøi D, Sørensen M F (1996) Using a typical human subject for binaural recording. 100th Conv Audio Engr Soc, Copenhagen, Denmark, preprint 4157. Abstr in: J Audio Engr Soc 44:632
79. Møller H, Hammershøi D, Jensen C B, Sørensen M F (1999) Evaluation of artificial heads in listening tests. J Audio Engr Soc 47:83–100
80. Morimoto M, Ando Y (1980) On the simulation of sound localization. J Acoust Soc Jpn (E) 1:167–174
81. Naylor G (1993) Computer modeling and auralization of sound fields in rooms. Appl Acoust 38:89–92
82. Nelson W T, Hettinger L J, Cunningham J A, Brickman B J, Haas M W, McKinley R L (1998) Effects of localized auditory information on visual target detection performance using a helmet-mounted display. Hum Fact 40:452–460
83. Nielsen S (2001) Multichannel Signal Processing Tools - differences to multiple single channel processing. 19th Audio Engr Soc Int Conf, Schloss Elmau, Germany, paper 1901
84. Novo P (2005) Auditory virtual environments. Chap 11 this vol
85. Perrott D R, Cisneros J, McKinley R L, D'Angelo W R (1996) Aurally aided visual search under virtual and free-field listening conditions. Hum Fact 38:702–715
86. Platte H J (1977) Zur Bedeutung der Aussenohrübertragungseigenschaften für den Nachrichtenempfänger "menschliches Gehör" (On the relevance of the transmissions characteristics for the information receiver "human audition"). Doc diss. Techn Hochsch Aachen, Aachen
87. Platte H-J, Laws P (1976) Die Vorneortung bei der kopfbezogenen Stereophonie (Frontal localization in head-related stereophony). Radio Mentor Electron 42:97–10
88. Poulsen T (1978) Hörvergleich unterschiedlicher Kunstkopfsysteme (Auditory comparison of different dummy-head systems). Rundfunktechn Mitt 22:211–214

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 31

89. Richter G (1992) BAP Binaural Audio Processor. 92nd Audio Engr Soc Conv, Amsterdam, The Netherlands, preprint 3323. Abstr in: J Audio Engr Soc 40:444
90. Rife D D, Vanderkooy J (1989) Transfer-function measurements with maximum-length sequences. J Audio Engr Soc 37:419–444
91. Ryan C, Furlong D (1995) Effects of headphone placement on headphone equalization for binaural reproduction. 98nd Conv Audio Engr Soc, Paris, France, preprint 4009. Abstr in: J Audio Engr Soc 43:401
92. Sandvad J (1995) Dynamic aspects of auditory virtual environments. 100th Audio Engr Soc Conv, Copenhagen, Denmark, preprint 4226. Abstr in: J Audio Engr Soc 44:644
93. Sandvad J, Hammershøi D (1994) Binaural auralization: comparison of FIR and IIR filter representation of HIRs. 96th Audio Engr Soc Conv, Amsterdam, The Netherlands, preprint 3862. Abstr in: J Audio Engr Soc 42:395
94. Sandvad J, Hammershøi D (1994) What is the most efficient way of representing HTF filters? Proc. NORSIG'94, Ålesund, Norway, 174–178
95. Sapp M, Kleber J (2000) Universal Signal Processing System. ACUSTICA/Acta Acustica 86:185–188
96. Savioja L, Huopaniemi J, Lokki T, Vaananen R (1999) Creating interactive virtual acoustic environments. J Audio Engr Soc 47:675–705
97. Schmitz A, Vorländer M (1990) Messung von Aussenohrstossantworten mit Maximalfolgen-Hadamard-Transformation und deren Anwendung bei Inversionsversuchen (Measurement of external-ear impulse responses with maximum-length-sequence-Hadamard transformation and their application in inversion experiments). Acustica 71:257–268
98. Schobben D, Aarts R (2002) Three-dimensional Headphone Sound Reproduction Based on Active Noise Cancellation. 113th Audio Engr Soc Conv, Los Angeles, CA, preprint 5713. Abstr in: J Audio Engr Soc 50:977
99. Shaw E A G (1971) Acoustic response of external ear with progressive wave source. J Acoust Soc Amer 51:S150 and personal communication
100. Shaw E A G (1973) Acoustic response of external ear replica at various angles of incidence. 86th Ann Meet Acoust Soc Amer, and personal communication
101. Shaw E A G (1982) External ear response and sound localization. In Gatehouse W (ed): Localization of sound: theory and applications 30–41, Amphora, Groton CN
102. Theile G (1986) On the standardization of the frequency response of high-quality studio headphones. J Audio Engr Soc 34:956–969
103. Theile G, Spikofski G (1984) Vergleich zweier Kunstkopfsysteme unter berücksichtung verschiedener Anwendungsbereiche (Comparisons of two dummy head systems with due regard to different fields of applications). Fortschr Akust DAGA'84, Darmstadt, Germany, 223–226
104. Torres R R, Kleiner M, Dalenbäck B-I (2002) Audibility of "diffusion" in room acoustics auralization: An initial investigation. ACUSTICA/acta acustica 86:919–927
105. Wenzel E M (1992) Localization in virtual acoustic displays. Presence 1:80–107
106. Wenzel E M, Arruda M, Kistler D J, Wightman F L (1993) Localization using nonindividualized head-related transfer functions. J Acoust Soc Amer 94:111–123

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 32

107. Wenzel E M, Wightman F L, Foster S H (1988) A virtual display for conveying three-dimensional acoustic information. Proc Human Factors Soc 32nd Ann Meet, Anaheim, CA, 86–90
108. Wenzel E M, Wightman F L, Kistler D J (1991) Localization with non-individualized acoustic display cues. ACM Conf Comp-Hum Interact, New Orleans, LA, 351–359
109. West J E, Blauert J, MacLean D J (1992) Teleconferencing system using head-related signals. Appl Acoust 36:327–333
110. Wiener F M, Ross D A (1946) The pressure distribution in the auditory canal in a progressive sound field. J Acoust Soc Amer 18:401–408
111. Wightman F L, Kistler D J (1989) Headphone simulation of free-field listening. I: stimulus synthesis. J Acoust Soc Amer 85:858–567
112. Wightman F L, Kistler D J (1989) Headphone simulation of free-field listening. II: Psychophysical validation. J Acoust Soc Amer 85:868–878
113. Wightman F L, Kistler D J (1999) Resolution of front-back ambiguity in spatial hearing by listener and source movement. J Acoust Soc Amer 105:2841–2853
114. Wilkens H (1972) Kopfbezügliche Stereophonie – ein Hilfsmittel für Vergleich und Beurteiling vershciedener Raumeindrücke (Head-related stereophonie – an aid for the comparison and critical examination of different room effects). Acustica 26:213–221
115. Xiang N, Blauert J (1993) Binaural Scale Modelling for Auralization and Prediction of Acoustics in Auditoriums. Appl Acoust 38:267–290
116. Xiang N, Genuit K (1996) Characteristic maximum-length sequences for the interleaved sampling method. Acustica 82:905–907
117. Yamaguchi Z, Sushi N (1956) Real ear response of receivers. J Acoust Soc Jpn 12:8–13

Eight kHz, LLC
Exhibit 2019
IPR2023-01003
Page 33