Auditive Localization.
Head movements, an additional cue in Localization

vorgelegt von
Diplom-Physiker, Magister
Philip Mackensen
aus Berlin

Von der Fakultät I - Geisteswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Philosophie
– Dr. phil. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. W. Hendricks
Berichter: Prof. Dr. H. de la Motte
Berichter: Prof. Dr. S. Weinzierl

Tag der wissenschaftlichen Aussprache: 26. Januar 2004

Berlin 2004
D 83

To Liele

Acknowledgements

First of all, I would like to thank Prof. Dr. Manfred Krause, who unfortunately died in August 2003, just before this work was finished. Throughout all the years he supported me and encouraged me to finalize this work.

Another great encouragement came from Prof. Dr. Helga de la Motte. So, likewise, my sincere thanks for her intensive guidance and support — especially after the death of Prof. Krause. I am also thankful to Prof. Dr. Stefan Weinzierl for taking over Prof. Krause's part and who contributed to a successful finish of this work.

Additionally, a very special thanks to Prof. Dr. Georg Plenge for his support.

Also, I am indebted to the Institut für Rundfunktechnik, Munich, for giving me the chance to be part of the scientific research team and to carry out all the listening tests. In particular, I would like to thank Dr. Günther Theile for various fruitful and encouraging discussions and the textual supervision. I also thank Marcus Fruhmann, representing my former colleagues of the Audio Systems Engineering group at IRT.

Many thanks to all subjects who took part in the various listening tests.

In addition, I would express my gratitude to Attila Karamustafaoglu from Studer Professional Audio AG, Dr. Ulrich Horbach and Dr. Renato Pellegrini, for their support and the development of the BRS Processor.

Special thanks to Dr. Gebhard Rehm, who supported me intensively in the last stage of this work. I also would like to thank Dr. Marcus Dülk for his professional support.

I am particularly grateful to my wife, Andrea (Liele), who has always supported and encouraged me in every stage of work through her endurance and love.

Abstract

During the localization process, the hearing system simultaneously evaluates a multitude of cues. Investigating the process of localization, and thus trying to understand the hearing system, not only requires a knowledge of these different cues, but also an understanding of the relevance of *all* localization cues – especially with respect to dynamic (head movement) cues.

This work is divided into four chapters: the *classification of cues*, a *validation of the cue classification* together with an examination of the *entirety of cues*, our *experiments regarding the importance of head movements* and an new auralization method as an *application* of the results.

The *first* chapter presents a new classification scheme for localization cues. With this scheme we will describe and categorize various possible localization cues. This classification divides the localization cues into three main groups: a set of cues typical of the sound source itself, a group of cues relating to the environment where both sound source and listener are located in and, thirdly, cues originating from the listener's head, his body or his movement.

In the *second* chapter of this work, we demonstrate that various examples of localization experiments found in the literature are compatible with our cue classification. This can be regarded as verification of our categorization scheme. Furthermore, we discuss the importance of taking all localization cues into consideration. Wherever possible, we describe "pairs of experiments" with differing results or conclusions: One experiment takes the influence of a certain cue into consideration whereas the other experiment disregards this cue. In particular, we focus on experiments considering dynamic cues due to head movements an important but yet often disregarded cue.

The strong impact of head movements on localization is discussed in the *third* chapter. We performed localization experiments which focus on the exact reproduction of dynamic localization cues. For the first time, a head-tracked dummy-head system of such a high fidelity was used so that no localization failures could have been attributed to the system's inaccuracy. These experiments were carried out at IRT, the Institut für Rundfunktechnik in Munich, Germany, and contributed to the development of a new auralization method for the replication of a real listening room.

In the *fourth* chapter, we describe this new, data-based auralization method, the so-called Binaural Room Scanning (BRS). Our localization experiments at IRT contributed to the development of this method within the framework of a research project in cooperation with Studer Professional AG, Zurich, Switzerland. For the first time, an auralization method accounts for both head movements and the entirety of auditory localization cues. We portray the fundamental idea and technology behind the BRS method, and discuss future experiments and applications of this auralization technology.

Abstrakt (German Version)

Das Gehör wertet bei der Lokalisation eine Vielzahl von Lokalisationsmerkmalen aus. Um den Prozess des Lokalisierens und somit auch die Funktionsweise des Gehörs verstehen zu können, erfordert es einerseits die Kenntnis der verschiedenen Lokalisationsmerkmale, als auch das Bewußtsein, dass deren *Gesamtheit* relevant ist - insbesondere dynamische Lokalisationsmerkmale (Kopfbewegungen).

Diese Arbeit ist in vier Abschnitte untergliedert: der *Klassifizierung der Lokalisationsmerkmale*, dessen *Validierung* zusammen mit einer Betrachtung der *Gesamtheit aller Lokalisationsmerkmale*, unseren *Hörversuchen am Institut für Rundfunktechnik*, und einer *neuen Auralisationsmethode* als Anwendung der gefundenen Ergebnisse.

Der erste Abschnitt stellt ein neues Klassifikationsschema vor, das die Lokalisationsmerkmale gruppiert. Dabei unterteilt die Klassifizierung die Lokalisationsmerkmale in drei Hauptgruppen: einer Gruppe Schallquellen spezifischer Merkmale, einer Merkmalsgruppe, die sich auf die Umgebung bezieht, in der sich Schallquelle und Hörer befinden, und schließlich alle Merkmale, die individuell von Hörer zu Hörer verschieden sind.

Im zweiten Abschnitt dieser Arbeit wird anhand verschiedener Beispiele von Lokalisationsversuchen aus der Literatur aufgezeigt, dass dieses Klassifikationsschema damit verträglich ist und somit als verifiziert erachtet werden kann. Darüberhinaus wird erörtert, wie bedeutend es ist, die Gesamtheit aller Lokalisationsmerkmale zu betrachten. So möglich, werden paarweise Versuche angeführt, deren Ergebnisse jedoch differieren: Während bei einem Versuch der Einfluss eines Lokalisationsmerkmals betrachtet wurde, wurde er beim anderen Versuch nicht untersucht. Insbesondere werden die Versuche hervorgehoben, die ein wichtiges, aber oft nicht beachtetes Lokalisationsmerkmal untersuchen: Kopfbewegungen (dynamische Lokalisationsmerkmale).

Die überaus wichtige Bedeutung von Kopfbewegungen auf die Lokalisation wird im dritten Abschnitt behandelt. Dazu wurden Lokalisationsversuche durchgeführt, die insbesondere die exakte Reproduktion von dynamischen Lokalisationsmerkmalen zum Thema hatten. Zum ersten Mal wurde ein Kunstkopfsystem verwendet, das mittels eines Head-Tracker nachgeführt wurde, und eine bis dahin nicht erreichte Genauigkeit und Übertragungstreue aufwies, so dass sämtliche Lokalisationsfehler zumindest nicht auf Systemungenauigkeiten zurückgeführt werden konnten. Diese Versuche wurden am IRT, dem Institut für Rundfunktechnik in München, durchgeführt. Sie haben maßgeblich zur Entwicklung einer neuen Auralisationsmethode für die Nachbildung von realen (Ab-)Hörräumen beigetragen.

Im vierten und letzten Abschnitt wird diese neue, daten-basierte Auralisationsmethode, das sogenannte Binaural Room Scanning (BRS) beschrieben. Die Hörversuche am IRT waren eingebettet innerhalb eines Forschungsprojektes in Kooperation mit Studer Professional AG, Zürich, Schweiz, und beeinflussten die Entwicklung dieser Methode. Zum ersten Mal hat eine Auralisationsmethode sowohl Kopfbewegungen berücksichtigt, als auch der Gesamtheit aller Lokalisationsmerkmale Rechnung getragen. Die grundlegenden Ideen und Technologien hinter dem BRS-Verfahren werden vorgestellt, sowie zukünftige Versuche und Anwendungen dieser Auralisationstechnologie diskutiert.

Contents

List of Figures

Chapter 1

Introduction

The human being is endowed with at least six senses: the visual sense, the auditiv sense, the sense of smell, the sense of taste, the tactile sense and the equilibrium sense. And although it is possible to use several senses in order to localize an object, in most of the cases we only use the visual or the acoustical sense, or maybe a combination of both for this task. We rather rarely determine the location of an object by smelling (e. g., gas or odor), touching or feeling. Therefore, we will consider only the visual, the acoustical and the equilibrium sense[1] with respect to localizing an object.

The Visual World

In the visual world, several mechanisms allow to localize an object. Firstly, the *angular* displacement of the eyeball when focusing the object permits to ascertain its lateral position. The vertical position is determined in the same way. The brain registers both angular displacements, the lateral and the vertical, through the tactile information of the corresponding eye-muscles. This holds true for both eyes, the left as well as the right.

But there is also a further cue, provided by the combination of the left and the right eye. Because of the distance between the two eyes, i. e., their *lateral* displacement, an angle between their *optical axes* arises when the eyes focus an object. This (difference) angle helps to evaluate an object's position. This angle is especially a cue for the perception of distance.

Another, more crucial, optical cue for distance is the "sharpness" of the visual object's image – caused by the curvature of the eye's lens. When the eyes focus a visual object, its image is projected sharply onto the retina.

A further cue originates from the relationship between the object and its environment. The "known" size and dimensions of an object can lead to a distinct information about its position within the environment. For example, objects of identical size and

---

[1] The equilibrium sense is supported by the tactile sense that signalizes the tension of the neck-muscles using some receptors.

shape differ visually depending on their distance from the observer. The further away an object is placed, the smaller is the affected area on the retina (fovea), indicating thus its greater distance.

It can sometimes be helpful to actually turn the head in the direction of the object in order to face it straight ahead. Rotating the head induces dynamic visual cues, i. e., it alters the angle between the optical axes. For example, when turning the head the angluar displacement of the optical axes is greater in case of a close object than if an object is distant.

The Acoustic World

However, when we solely rely on the acoustical sense to localize a sound source, the situation is completely different and not as obvious as in the optical case, or at least more difficult to understand.

Bregman [16] illustrates the complexity of the localization[2] process in the acoustical world:

> *“ Imagine that your are on the edge of a lake and a friend challenges you to play a game.*
>
> *The game is this: your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide, and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the sides of the channel. As waves reach the side of the lake they travel up the channels and cause the two handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer series of questions:*
>
> *How many boats are there on the lake and where are they? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped suddenly into the lake? . . .*
>
> *Solving this problem seems impossible, but it is a strict analogy to the problem faced by our auditory systems. ”*

In contrast to the optical case, where the eyes can “move” and focus an object without turning the head, it is impossible to orientate the human pinnae towards a sound source. Consequently, it is not possible to use a correlation of angles to aid in localization (especially, in distance perception).

Similarly to the optical analogy, a lateral displacement between the left and the right ear causes interaural differences. For sound sources positioned outside the median plane, i. e., the vertical symmetry plane of the head, these interaural differences strongly contribute to localize the sound source. However, within the median plane no interaural differences exist (at least theoretically), and so other cues must be responsible for the distinction between frontal, above and rear sound sources [9, 8]. The spectral variations caused by the shape of the outer ear (pinna), head and the body are possible cues.

---

[2] The exact definition of *localization* is given in chapter 2.1.2.

In the visual localization process movements of both, the eyes and the head are required to enable a projection onto the fovea. Also in the acoustical case, head movements can have quite a strong impact on localization. This work aims at demonstrating the strong influence of this dynamic cue. A further goal is to emphasize the importance of taking into account the entirety of cues as opposed to only a single cue. This holds especially true when exploring the general functioning of the human hearing system through experiments.

We therefore put forward a classification scheme in order to categorize all cues relevant to the acoustic localization process. We try to verify this scheme by applying it to examples of localization experiments from the literature. Thereby, we demonstrate the influence of a single cue within the frame of the entirety of cues. This is followed by a detailed description of experiments focusing on head movements that were carried out at the Institut für Rundfunktechnik (IRT) in Munich, Germany. At the end of this work, an auralization method is described that takes into account the listener's head movements as well as the entirety of cues. It can be viewed quasi as an possible application of the results found.

Chapter 2

New Classification Scheme of Localization Cues

In this chapter, we will explain some *basic terms* in order define the term *localization.* Subsequently, we will introduce a new *classification scheme* that categorizes all possible localization cues.

2.1 Definition of Terms

At first blush, the meaning of "*to localize*" might seem obvious: to find, to locate, to determine the position. This is because of the term being part of the common language, for example in "*to localize the failure*". However, it is not necessarily clear what it means in the context of auditory localization. We will give an exact definition at the end of this chapter. Beforehand, some basic terms require explanation.

2.1.1 Basic terms

Throughout this work the terms *sound event, auditory event, space* and *position* have the meanings as defined below.

Sound Event

A *sound event* is a physical event resulting in an emission of sound waves. The German Standard DIN 1320 [28] defines the term *sound* as "*mechanical vibrations and waves of an elastic medium, particularly in the frequency range of human hearing (16 Hz to 20 kHz)*". For example, *sound* originates from the pushing or the pulling of a string, the plunging of a stone into water, or the movement of a loudspeaker membrane. All these events have in common that a measurable physical quantity, such as the density of the air, varies in time or other respects. Due to its physical character and measurability, a *sound event* occurs in the real physical world and is "induced" by a *sound source.*

Auditory Event

In contrast to the sound event existing in reality, the *auditory event* takes place only in the "imagination" of the listener, and it cannot be "proven" by applying measurement techniques. It is rather *perceived* or *felt*, either consciously or unconsciously. Sometimes an auditory event occurs even without a corresponding mechanical vibration or sound event. For example, a symptom of the so-called *tinnitus* disease is the perception of a "ringing" or a noise in the ears *without* any mechanical sound waves actually being present.

Sound Space and Auditory Space

The environment surrounding us is the three-dimensional, physical space. In this space, we can unambiguously describe every point by reference to *three* coordinates. The *length*, the *width* and the *height* jointly constitute the so-called *cartesian coordinate system.*

Another commonly used coordinate system is listener oriented, and its origin is the middle of the listener's head. This system is called *head-related* or *spherical coordinate system.* The three variables here are the *distance to the origin*, the *azimuth* or *azimuthal angle* and the *elevation* or *elevation angle.* It is often used in localization experiments.

Since all sound events take place in the real physical space, the latter can also be called *sound space.* Auditory events, however, occur in a space, which exists solely in the imagination of the listener. This "imaginary" space will correspondingly referred to as *auditory space.*

Sound Event Position and Auditory Event Position

The position of a sound event will be called *sound event position.* It is a point in the sound space and can be described by three coordinates. Similarly, the place where an auditory event is perceived will be called *auditory event position*, and it also can be unambiguously defined by three coordinates.

The relationship between *sound event*, *sound event position*; *auditory event* and its corresponding *auditory event position* will be referred to by the term *localization.* A closer definition of localization follows now.

2.1.2 Localization

In principle, *localization* can be understood as the relation between "corresponding" positions in the sound space and in the auditory space. The following examples may illustrate this statement.

- A listener can allocate a certain auditory event position to certain sound event positions. This situation is called a *good localization.*

- In case of the so-called *in-head localization*, the listener perceives the position of the auditory event to be inside his head, although the sound sources are located outside his head, e. g., when using headphones.
- Sometimes the listener does not perceive any specific, “well-defined” auditory event position. It is rather a *blurred* impression, an auditory event *region* — not a single spot. This situation will be called *localization blur.*
- The auditory percept might not have any specific position at all, not even a blurred region or an in-head localization.

At first, we shall cite two definitions of *localization* as found in the literature.

Blauert’s Definition

Blauert [11] defines *localization* as “*law or rule by which the location of an auditory event (e. g., its direction or distance) is related to a specific attribute or attributes of a sound event, or of another event that is in some way correlated with the auditory event*”.

The definition not only relates the auditory event position to the location of the sound source, but it also postulates the possibility of influencing the auditory event position by certain attributes of the sound source(s), as for example the spectral distribution.

Theile’s Definition

Theile [122] restricts the term *sound event* to that part of a sound that originates from a single sound source, and that determines or influences the position or shape of the corresponding auditory event(s). Presupposing this “restricted definition” of sound event, he defines *localization* as a “*law between the auditory event position outside the head and characteristic cues of one or several sound events*”.

This definition is stricter than Blauert’s definition of localization as it includes auditory events that are “evoked” by several sound sources such as the so-called *phantom source* in stereophony using loudspeakers.

To determine the auditory event position, i. e., for purposes of localization, several localization cues are necessary. Some of them can be attributed to the sound source alone. These localization cues will be discussed in detail in the following chapter.

Definition of Localization employed here

Throughout this document, the term *localization* refers to the law or rule between the auditory event position and characteristic *cues of the sound event(s)*, *cues of the environment* and also *cues typical of the listener* himself. This definition of localization obviously complements Theile’s definition in that it also takes cues of the environment and of the listener himself into account.

Having defined the term *localization* we turn to defining and classifying the different localization cues in the following chapter.

2.2 New Classification Scheme

This section introduces a new classification scheme and describes numerous different localization cues. This scheme divides the cues into three main groups: cues that are typical of the *sound source*, cues which are characteristic of the enclosing *environment*, and localization cues related to the *listener*.

2.2.1 Source Cues

The first group of localization cues are typical of the sound source: the *source cues*. They can be subdivided into *spectral cues*, *temporal cues* and *local cues*.

Spectral Cues

The spectrum of the sound source can have a strong influence on localization. A characteristic of a spectrum is its *spectral width*. Localization may depend on whether an emitted sound is a single sine tone, with a narrow width, or a complex sound with a broad spectral width.

Another option for classification is the distribution of spectral energy in the spectrum, the "shape" of the spectrum. This classification will be called *spectral distribution*. The distribution of spectral energy determines whether a bandwidth limited signal, characterized by its center frequency, is attributed rather to the lower or to the higher spectral region.

The third characteristic, the level of the sound, is closely related to the overall spectral energy. A "loud" sound source might be localized at a different point than a "quiet" one. We will analyze this cue later, and it is often associated with the distance between the sound source and the listener.

The *phase* is closely linked to the level cue. We will take this into account in form of a *complex* level cue, having a *real* (level) and an *imaginary* (phase) part.

Temporal Cues

Temporal cues are another class of source related localization cues. On the one hand, there is the *overall dynamic*, i. e., the "temporal shape", of the sound. There are sounds with so-called *onset transients*, i. e., much dynamical movement takes place in the very beginning of the sound [120]. Examples for this category are the sound of a steel guitar or a crash cymbal. Furthermore, there are sounds whose level is more or less static, e. g., a sine tone or the sound of a flute or an organ. Both classes of sounds might be localized differently.

On the other hand, the localization of a sound source may be influenced by the time a sound is emitted. The localization of a noise sound presented for only some few milliseconds can be different from that if the same sound is presented for some seconds. Hence, the *duration* of the sound can influence localization. Thus duration is another temporal cue.

Local Cues

The *local cues* make up the third subgroup of sound source cues. They relate to the *position* of the source[1]. Under *freefield conditions*[2] it is simply the relative position with respect to the listener. Distinct positions of the sound event can affect localization. For example, the sound event position may vary within the *horizontal plane* or the *median plane*, or with respect to its relative *distance* to the listener. The position of the source with respect to the environment also plays a role in all non-freefield conditions. This latter factor will be discussed in section 2.2.2.

Of course, the position of the sound source can change in time. And although this variation can be qualified also as a temporal cue, it will be regarded here as a further local cue: *alteration of the sound event position*. As in the optical case, small changes in the position of an object can aid in localization.

Apart from the source cues there are listener and environmental cues, the latter of which will be discussed in section 2.2.2 and 2.2.3.

2.2.2 Environment Related Cues

In addition to the source cues, certain cues relate to the environment "surrounding" both the sources and the listener. The *environmental cues* not only comprise the room with all its surfaces, but also the position of the sound source within the room itself. When a room encloses the listener and sound sources, all its surfaces, like walls, the floor or the ceiling, reflect the emitted sound, and so (by superposition of the sound waves) a dense "sound pattern" arrives at the ears of the listener.

The first sound arriving at the ears of the listener is so called the *direct sound*. The *early reflections* are perceptible only a few milliseconds (0.8 ms – 20 ms) after the direct sound. These reflections allegedly facilitate the localization process and they help to gain a knowledge of the size and the acoustical properties of the room [27, 71, 82]. Finally, the late *diffuse reflections*, also known as *reverb*, will reach the listener's ear.

If there are no such reflections, the listener and the sound source(s) must either be in an anechoic room or 'outside" in the freefield, e. g., a snow covered field or another area without reflecting surfaces. Therefore, the localization can still be totally different from a situation with room reflections.

All environmental cues belong to one of two sets of cues: the first set relates only to the *direct sound* of the source, the other set of cues comprises the rest of environmental "information" in form of *reflections*.

Direct Sound Cues

If the listener can actually see the sound source the sound waves can reach the listener directly. The source can be said to be "acoustically visible". This also holds true if an

---

[1] To be exact, an *absolute position* of the sound source does not exist in a listener centered coordinate system, and hence, it cannot be an attribute of the sound source itself. It is related to both the environment and to the listener.

[2] In a freefield condition no reflections from any surfaces arise. An anechoic room is a good approximation for a freefield condition.

acoustically permeable fabric (as used for loudspeaker covers or "curtains" in listening tests) hides the source optically.

Otherwise, any obstacle *between* the sound source and the listener will scatter, reflect or bend the sound waves and thus possibly distort localization. Therefore, *direct sound* is an environmental cue.

Reflection Cues

All reflections from surfaces, such as walls, the ceiling, the floor or other objects, can influence the perceived location as well. For example, the correct perception of distance strongly depends on the presence of reflections, as will be shown in the next chapter [94].

Additionally, the reflections also support the listener in acoustically determining the size and the dimensions of the room. A small room is different from a large hall in the surfaces' way of reflecting the sound, i. e., their temporal reflection pattern. But this difference and the impression of the *room-size* is only of secondary importance for localization.

2.2.3 Listener Cues

The group of *listener specific cues* constitutes the third and last main division of localization cues. This group subdivides into four classes, the *interaural cues*, the *HRTF cues*, i. e., cues due to the influences of head and pinnae, cues by *head movements* (*dynamic cues*) and the group of *informational cues*, i. e., all cues relating to the listener which allow him to gather additional information about the source.

Interaural Cues

A human being has two ears with only the outer ears (pinnae) being visible. In a natural listening situation both the left and the right ear, receive the sound emitted by a sound source.

If a listener receives a sound solely by a *single* ear, either because of using earplugs, headphones, or because of a unilateral deafness, this is a case of a *monaural presentation*. Similarly, if the listener uses headphones, and only one headphone reproduces the signal, this is a *monotic presentation* (Stumpf [118]). In both cases there are no *interaural cues*.

A *binaural presentation* requires both ears. Compared with the monaural presentation in this case a second ear signal, and thus possibly *interaural cues*, is available. These cues may be *interaural level differences*, *interaural phase differences* or *interaural time differences* or a combination of these. This also holds true for a *dichotic* presentation, characterized by the left and the right ear receiving different signals, as it is the case for headphone presentation. Here, each ear signal may vary individually without influencing the opposite signal.

Both ears receiving *identical* signals through headphones creates a mixed scenario. This case lacks interaural information, although both ears receive an acoustic signal. This situation is called a *diotic* presentation.

Whether *interaural cues* can be used for localization depends on how the listener receives the sound: monaurally or binaurally.

HRTF Cues

The *head-related transfer function* (*HRTF cue*) is closely related to the interaural cues. Various parts of the human body (head, shoulders and torso) and particularly the pinnae scatter, bend or reflect all incoming sound waves. The sound field is transformed linearly before reaching the eardrums of the listener.

Using a dummy head or synthesizing artificial HRTFs through a computer results in a different set of HRTFs. These distinct HRTFs can significantly alter localization. If the pinnae lack completely, e. g., using a spherical microphone (as described by Theile [126]), dramatic localization differences follow when presented over headphones.

Dynamic Cues (Head Movements)

A listener can move his body or his head in order to confirm the localization of a sound event. In the same manner as the position of the sound source plays a role in determining its location (or rather the location of the associated auditory event), the listener's position and head orientation can influence, ease or complicate localization.

Usually, the listener rarely makes *translational* movements in order to localize a sound source [3]. This may be due to the fact that obstacles reflect, bend or scatter sound waves. Thus, a small change of the listener's position normally does not greatly aid in localization and is therefore disregarded in most localization experiments. The experimental setup[4] rather fixes the listener's position.

On the other hand, *movements of the head* can have a strong impact on localization. Here, in principle only *rotations* about various axes matter. Head translations are difficult to carry out and thus do not play a role in localization. *Rotational* and *tipping* movements are the most important head movements. All cues relating to dynamic changes *caused by head movements* will be called *dynamic cues* or *head movement cues*. These head movement cues implies the usage of another cue to actually determine the orientation of the head or a change in orientation: the vestibular cue, also known as equilibrium sense. This cue belongs to the last group of cues, i. e., the non–acoustic cues.

---

[3] That is different from the optical case: If there is an obstacle between the observer and the object to be viewed, for example a source of light, the observer might step aside to gain free sight. In this case, a translational movement would help to localize the desired object.

[4] For a localization experiment involving the movement of the listener see for example Loomis [76]. There, all subjects were able to move freely in a room in order to find the virtual or real sound sources through approaching them.

Non–acoustic Cues

All listener cues that have not been mentioned so far, and which allow the listener to gather information about the source position, will be considered so-called *non–acoustical cue*. They all have in common that they are *not directly related* to acoustics.

The first non–acoustic cue is the *knowledge cue*. This cue in principle rests on a certain knowledge of (or at least familiarity with) the sound source, i. e., its spectral and temporal attributes.

The second cue is *optical information* about the sound source: Its visually perceived apparent position. This will be called *optical cue* and is very informative.

The *vestibular cue* is completely different from the previous two. In this case the head's orientation is "evaluated" by the spiral ganglia[5]. This cue is very important when considering head-movement cues as it is used to determine the orientation of the head (see section 2.2.3).

Obviously, the enumeration does not exhaust all cues available for localization. In the introduction, the *sense of smell*, of *taste*, or the *tactile* sense were mentioned. As already stated, however, we will not analyze the cues with the exception of the knowledge cue, the optical cue and the vestibular cue.

---

[5] The head orientation is also "perceived" by the tactile information of the receptors in the neck-muscles. But this "measurement" and analysis of the listener's head-orientation generally relates to the *dynamic cues*.

Chapter 3

Verification and Applicability of the Cue–Classification

In this chapter, we show that earlier experiments are compatible with the cue classification previously proposed. Sometimes however, these experiments fail to take into account the importance of the entirety of localization cues.

Most experiments documented in the literature demonstrate the influence of each (sub-)group of cues on localization. In some cases, "pairs of experiments", distinct with respect to one important cue, and therefore leading to different results, are documented.

3.1 Static Localization Cues

We begin with discussing experiments regarding *static cues*, i. e., all cues not involving head movements. *Dynamic cues*, a subgroup of listener cues, will be the subject of section 3.2.

3.1.1 Experiments relating to source cues

As stated in the previous section this group can be divided into three subgroups: *Spectral cues*, *temporal cues* and *positional* or *local cues*.

Influence of spectral cues

Especially if a sound is only received monaurally, its perceived position strongly depends on both the spectrum of the sound itself and the "deformations" in the spectrum superimposed by the pinnae. Because, "*a priori, a listener does not know whether a particular spectral structure is caused by location-dependent filtering or whether it is intrinsic to the source itself*"(Durlach & Colburn [29]). Thus, manipulations of the source spectrum can trigger changes in the perceived location [23, 92, 143, 146].

This section will give some examples of experiments analyzing the *spectral width cues*, *spectral distribution cues* and *sound level cues*.

Spectral width The *spectral width* of a source spectrum can influence localization. For example, a pure sine tone is "*notoriously difficult to localize*", as stated by Angell and Fite [1]. Numerous experiments investigated the localization blur of different signals [11, 33, 102, 130]. The localization blur often differs with the signal's changing spectral width. In particular, there is a difference between sinusoidal and narrowband signals on the one hand, and broadband signals (e. g., a white noise) on the other.

The localization of pure tones and signals with limited bandwidth can also differ from that of broadband signals when influenced by other "parameters". These parameters include *reflections*, *onset* and *duration* of the signal [52, 103, 104], the influence of the *outer ear* [38] or the *position* of the sound source [143, 146].

Spectral distribution Not only the spectral width of the signal may influence the perceived position of the sound source but also its distribution of spectral energy. For example, in case of a bandlimited signal, its center frequency can strongly affect localization.

In 1967/68, Blauert investigated the influence of the center frequency of third octave band noises on localization in the median plane [8, 9, 11]. The subjects perceived the sound in a darkened, anechoic chamber, and the sound had a duration of between 100 ms and 1 s. The presentation used loudspeakers positioned *in* the median plane (directly in front, above and behind the listener) as well as on the ear axis (left and right besides the ears). One experiment used headphones instead of the loudspeakers.

The subjects perceived the signal either in front, directly above the head, in the back, or even inside the head, depending only on the center frequency, but not at the actual position of the sound source (loudspeaker) itself. The difference between the actual loudspeaker position and the perceived position can be explained as follows: If the specific band noises coincide with frequency bands that are boosted in the HRTF for a specific location (*directional bands*), this location is the most probable for the auditory event.

For example, a low frequent 3rd octave band noise, e. g., with a center frequency of about 200 Hz, is perceived in the front. Altering the center frequency to 500 Hz or 8 kHz shifts the corresponding hearing event directly above the head. If one uses instead a center frequency of 1 kHz or its tenfold, the signal appears to come directly from behind. Again, about 2 kHz or 16 kHz center frequency will bring the impression back to the front.

An experiment on localization in the frontal median plane, carried out by Roffler and Butler [108], brought about similar results. Roffler and Butler used distinct pure tones as well as noise bands presented by loudspeakers in the median plane with different elevations around the horizontal plane. Hebrank and Wright [56] also found the localization of bandpassed noise to be independent of the sound source's original position.

Wightman and Kistler [145] proved the importance and dominance of *low-frequency* interaural time-difference cues in another experiment which demonstrates the strong influence of low frequencies on localization. Because of its close relation to *listener cues* we will describe this experiment in detail in section 3.1.3.

On the other hand, the high-frequency part is considered to be important for an exact localization because of both the influence of the head and ears (pinnae) [19, 21, 37, 50, 56, 92, 108] lead to evaluable interaural level differences.

Level Most localization experiments were carried out with a constant sound level. In the majority of the experiments investigating the influence of the sound level this cue did not have a strong impact on localization (especially: *direction*) [51, 104, 108]. It neither has an impact on the inertia of the human sound system [7], nor does it influence the perceived direction of the so-called *directional bands* [9].

However, Gardner carried out an experiment in 1968 in which he investigated the influence of the sound level on the perceived *distance* [36]. He used five loudspeakers arranged in two different rows in 0° and in 45° azimuthal direction and in various distances (from approximately 3 ft to 30 ft). He placed the loudspeakers in an anechoic chamber. Human speech, either spoken lively by a person or reproduced by the loudspeakers, served as a test signal. All in all, 20 subjects took part in this experiment. Their heads were fixed to avoid head movements. The subjects had to identify the loudspeakers, which were numbered in order to permit easy designation, according to the level's variation at the listener's ear.

In the case of *recorded* speech, only the level at the listener's ears determined the estimated location of the loudspeaker, regardless of its actual position. This also holds true for the loudspeaker array in the 45°–direction, as well as for horizontal loudspeaker arrays perpendicular to the median plane.

In the case of *live* speech, however, a human speaker had to shift between four positions: 3 ft, 10 ft, 20 ft and 30 ft, respectively. Four different types of voices were used: a *whispering* voice, a low or *confidential* voice, a *normal* and conversational tone of voice, and finally a *shouted* voice.

With the exception of whispering, solely the loudness level at the listener's ears determined the perceived distance of a live speaker. But in the case of the whispering voice some kind of "acoustic horizon" seemed to arise. The reason of this "acoustic horizon" could be the association between this type of speaking and the typical proximity of such a sound source (speaker).

Experiments relating to temporal cues

Whether the emitted sound has an onset-transient or not, can make a difference for localization. The sound's duration might also influence the perceived position of the sound source(s).

Some experiments clearly demonstrate the impact of onset-transient on localization. We will describe two of them briefly.

Franssen Effect One experiment was carried out by Franssen in 1959/60 [34, 11]. He used two loudspeakers in one room with the subject sitting in front of them at some distance. One of the loudspeakers emitted a pure sine tone with an exponential onset and decay envelope. The envelope of the second loudspeaker's signal was such that combining both signals would result in an envelope with rectangular shape. In other words, the second loudspeaker (mostly) transmitted the switching transients.

A subject listening to both loudspeakers had the impression of *only the second* loudspeaker emitting a sound. The listeners would only revise their erroneous assumption that the second loudspeaker transmitted the sustained part of the sound if the first loudspeaker was switched off. This effect has come to be known as *Franssen Effect*.

Localization in Rooms: Onset and Duration In 1986, Rakerd and Hartmann carried out the other experiment relevant in this context. They investigated the *localization of sound in rooms*, especially the effect of *onset and duration* [104].

They varied the onset time from 0 s (impulsive) to 5 s (no transients) and observed its impact on localization. A sine tone at a frequency of either 500 Hz or 2000 Hz served as test signal. One of twelve loudspeakers positioned in the horizontal plane emitted this pure tone. The tests were carried out both in a room with a "controllable" single reflection and in an anechoic room.

It turned out that localization was independent of reverberation for sounds with an onset transient. Without an onset transient, however, the reverberation of the room affected localization.

These two experiments proved onset transients in a sound to influence perception and thus help localization of the sound source.

Macpherson and Middlebrooks [80] as well as Hofman and van Opstal [57] showed the duration of a sound to influence localization. Their experiment is briefly described here:

The subjects had to localize noise bursts presented frontally. The noise bursts' perceived *elevation component* was observed. When their duration was reduced from 500 ms to 3 ms the auditory events became increasingly biased to the horizontal plane (i. e., "*compressed*"). And although the correlation between actual elevation and target elevation was high, it was less than 1:1. Likewise, when presenting 3 ms-noise trains, a similar compression was observed by increasing the periods of silence from 0 to 77 ms.

Experiments relating to local cues

The actual position of the sound source can strongly affect the whole localization performance. This connection is a well-known fact from every-day life experience. For example, sound sources in the horizontal plane are mostly easier to localized than sources in the median plane. A reason for this is the presence of interaural time-difference cues for sources outside the median plane. This kind of interaural time differences is easier to "evaluate" than interaural level differences, as in the case of median plane sources. Various experiments investigated this correlation [2, 11, 22, 66, 79, 84, 105].

While in the horizontal plane it is mainly interaural cues (time and level differences between left and right ear) that serve localization, localization of sources in the median plane is based on HRTF cues[2, 11, 54, 84].

Roffler and Butler, for example, proved the importance of pinnae cues for the localization of sounds in the median plane [108]. They flattened the pinnae using a flexible plexiglass band and thus prevented the outer ear from receiving any spectral cues. As expected, localization in the median plane decreased, whereas localization in the horizontal plane remained largely unaffected. Obviously, the sound sources' position exerted an influence on localization.

The source position can also influence the distance estimation of a sound source (see also section 3.1.2). However, this "positional effect" is not very strong and can only be observed in an anechoic environment [36, 38, 58, 94, 93].

Finally, the source can alter its position in time, i. e., the source itself can move. As a consequence, characteristic *listener cues* will change, as for example the *interaural cues* or *HRTF cues*. In principle, similar changes might occur if the listener moves himself or his head into the opposite direction (at least in the free-field situation). In this case, the only *additional* cue would be a *vestibular cue* (registering the movement) or the *tactile cue* (from the sensors in the muscles). These cues do not exist in case of a moving source and a fixed listener position. All dynamic cues will be considered in section 3.2.

3.1.2 Experiments relating to environmental cues

This section will describe some experiments dealing with environmental cues. Particular attention will be given to experiments exploring the influence of reflections, either in combination with a direct sound, or distinguishing between single reflection and a diffuse reverberation.

Direct Sound and reflections

The following experiment demonstrates that neglecting possible environmental influences of the environment may lead to completely different results.

The results of Gardner's experiment described in section 3.1.1 suggested a close relationship between the perceived distance and the sound level at the ears. But this experiment took place in an anechoic environment.

However, in 1992, Nielsen carried out several experiments on distance perception in different rooms [93, 94]. He thoroughly investigated the influence of various cues on distance perception[1]. Besides the role of spectral cues he explored the influence of the sound source's position, the loudness level and the environment itself.

The experimental environment was an anechoic room, a standard (IEC 268-13) listening room and a class room. Loudspeakers were placed at four distances: 1.0 m, 1.71 m, 2.92 m and 5.0 m. The loudness level varied between different runs, but within

---

[1] See also [58, 75] and the review of "distance experiments" in [93]

a single run, it was kept constant and independent of the loudspeaker distance. All loudspeakers were at ear height, and either in 0° or in 45° azimuthal direction.

The subjects (a total of 32 persons) had to indicate the apparent position of the perceived auditory event using a graphical interface and a computer mouse with a female voice serving as signal[2].

In the anechoic room the results were principally identical with Gardner's experiment, i. e., the estimated distance depends only on the loudness level at the listener's ears. It is independent of the sound source's actual distance. However, it would be wrong to conclude that *under any circumstance* a distant source will be perceived as distant, only because the loudness level at the listener's position decreases with increasing distance.

Using the standard listening room or the class room as the same experiment's environment led to a totally different result. In both rooms, early and late reflections existed. These reflections seemed to aid in correct perception of the distance. The estimated distance of the loudspeaker principally corresponded more or less to its real position *regardless* of the sound level at the listener's ears.

Sakamoto *et al.* [109] investigated the conditions for the "out-of-head"–perception of an auditory event. They found that the so-called *acoustical ratio*[3], i. e., the ratio between all the reflected sound and the direct sound, has to exceed a certain level for a sound event to be perceived out of head. This also seems to apply to the *distance perception* of an auditory event.

Results of an experiment regarding "in-head localization" and carried out by Toole [129] can be explained in the same way. Toole used an anechoic environment (the acoustical ratio was accordingly zero: $\mathrm{AR} = 0$) and placed loudspeakers symmetrically to the median plane. Interaural differences were avoided by using either headphones or pairwise symmetrical loudspeakers. This proved the lack of reflections be the reason for the subjects to perceive the sound inside their heads.

Nielsen's and Gardner's already cited experiments show that disregarding the influence of reflections (environmental cues) can lead to totally different results. On the one hand there is Gardner's result: It is solely the actual sound level at the listener's ears that determines the perceived distance. On the other hand there is Nielsen's contrary result, i. e., confining Gardner's result to an anechoic environment. Additionally, in every reflecting environment the perceived distance depends mainly on the ratio between reflected and direct sound.

Direction of Reflections

Not only the general existence of reflections might help in determining a sound source's distance but also the *direction* of the reflections.

This was investigated, for example, by Rakerd and Hartmann [52, 53, 103, 104]. In one experiment they altered the height of the ceiling, and thus "rearranged" the order of incoming reflections. The reflection from the floor came always first after

---

[2] preliminary tests also used noise or music, but all signals gave the same results.

[3] The exact definition of the acoustical ratio AR is: $\mathrm{AR} = \frac{\text{acoustic energy of reflected sound}}{\text{acoustic energy of direct sound}}$

the direct sound, whereas the reflection of the ceiling was arranged to be second (low ceiling condition). Altering the ceiling height changed the *direction* of the second reflection.

In another experiment Rakerd and Hartmann investigated the directive effect of the *first* reflection. This was accomplished by placing a single reflective surface inside a concert hall. The hall had a variable wall absorption and ceiling height. The floor, the ceiling or one of the walls alternately served as reflective surface.

The directional effect of the reflections can be summarized as follows: Early reflections stemming from the same azimuthal direction as the direct sound aid in localization, whereas early lateral reflections in reverberant rooms (e. g., concert halls) tend to complicate the localization process. Especially for tones with a slow-onset the directional impact of the first reflection on localization is much stronger than for impulsive tones [104].

Reverberation

Not only early reflections and their direction influence localization, but also reverberation (i. e., late and diffuse reflections). And although binaural hearing can "suppress" reverberation, as remarked by Koenig [70], it can nevertheless diminish the accuracy of localization.

An experiment by Giguère and Abel [43] shows the influence of reverberation on localization of frontal and lateral loudspeakers. These were positioned in a semi-arc around the listener (frontal: $\pm 15°, \pm 45°$ and $\pm 75°$ / lateral: 15°, 45°, 75°, 105°, 135° and 165°). A third-octave band noise of 500 ms duration served as test signal. Varying the carpet and wall absorption allowed creating two different environments with reverberation times T of about 150 ms and 1 s, respectively.

The results were that an increase in reverberation effected an overall decrease in localization accuracy. This effect is especially pronounced at low (center) frequencies.

Other environmental influences

Reflections are certainly the main localization cue provided by the environment. But other "cues", as for example "background noise", can also affect localization. It is less a *cue* than rather "disturbing noise" because its *absence* aids in localization.

The effect of (disturbing) noise on localization was investigated, for instance, by Good & Gilkey [45]. A broadband click-train signal had to be localized in both a quiet and a noisy environment. Therefor, a "masking" noise was located at the position 0° azimuth and 0° elevation. The variable parameter was the *Signal-to-Noise-Ratio* (SNR), subject to alteration in nine different steps. The signal itself originates from one of 239 spatial positions, covering the whole azimuthal range and an elevation area from -45° up to 90°.

It turned out that localization accuracy decreases if the SNR decreases correspondingly – nearly monotonically [45]: The louder the masking noise, the worse the localization of the pulse-train. However, azimuthal judgments (e. g. *left/right*) were less strongly influenced than those regarding the directions *up/down*, or *front/back*.

This is in accordance with the findings of Rakerd and Hartmann [104] mentioned above.

However, Kock and Koenig [69, 70] proved that using both ears (binaural perception, see section 3.1.3) can reduce the general influence of a background noise on localization.

3.1.3 Experiments relating to listener cues

In this section, we will give some examples of experiments regarding listener cues. Intentionally, we leave out experiments regarding the group of *dynamic cues* because we will investigate those in detail in section 3.2.

Experiments relating to interaural Cues

In this context, two experiments deserve emphasis. The first experiment by Hebrank & Wright explores the general influence of binaural perception. The other experiment, carried out by Wightman & Kistler, investigates conflicting interaural difference cues.

Monaural and Binaural Localization can depend on whether the sound is perceived *monaurally* or *binaurally*. For example, in contrast to a monaural or a *diotic* perception, a *dichotic* perception can reduce background noise [70]. Also, any effects due to small head movements are almost completely eliminated in monaural sound reception [19].

An experiment by Wightman and Kistler showed a different perception of real and virtual sound sources depending on the general existence of interaural cues. Localizing virtual sources monaurally yields different results than binaural localization of the same (virtual[4]) sources [146].

Generally, a monaural sound reception is rather unnatural. This is because "*when one ear is occluded, the resulting interaural difference cues skew the subject's perceptual space toward his open ear, preventing him from having the impressions of front, back, or elevation in the median plane, and causing him to err in his responses*" (Hebrank & Wright [55]).

However, there are situations in which the localization does not seem to depend on interaural cues. For instance, sound sources positioned in the symmetrical plane of the head, the median plane, do not evoke any interaural differences — at least theoretically. Therefore, any binaural disparity is considered to be irrelevant for median plane localization [111], which is thus in principle a monaural phenomenon [2, 11, 21, 37]. The influence of interaural cues on localization can in turn depend on the position of the sound source itself (cp. 3.1.1).

Localization is independent of interaural cues with respect to unfamiliar sound. Difficulties in localizing an unfamiliar sound arise similarly in both monaural and binaural perception. This was tested experimentally for example by Hebrank & Wright

---

[4] The term *virtual source* refers to sound sources that do not exist in reality but are "simulated" at that specific location.

[56]. They varied the spectral composition of a noise after each trial ("*scrambled*" or "*rippled*" noise) in order to prevent the auditory system from becoming familiar with the sound and "learning". This localization behavior assumes the existence of another cue: *memory* or *knowledge*. This cue will be discussed in section 3.1.3.

ITD and ILD In the early 20th century, Lord Rayleigh developed the so-called "*duplex theory*" of localization [117, 144]. He employed a simplified geometry of the head, and based his theory on the results of early psychophysical experiments and acoustic measurements.

According to this theory, localization of low frequency sounds depends on interaural time differences (ITDs), whereas localization of high frequencies rests on interaural level differences (ILDs). The physical size and dimensions of the head lead to an unambiguous ITD only at *low frequencies*. On the other hand, especially at *high frequencies* (greater than about 4 kHz), the pinnae influence the amplitude response and thus the ILD [111, 112, 113]. Detailed information about the various influences of torso, head and pinnae on the sound field at the ear drums is available in [41].

The ITDs may be used to determine the range of possible source positions, whereas the ILDs and the spectral cues can help to resolve localization ambiguities and to refine the position of the auditory event [83, 146]. In situations where ITDs and ILDs are in conflict, the ITD–cues dominate [145]. They seem to be more reliable than ILD–cues, apparently because of their higher consistency across frequency bands [73, 137, 145]. Wightman and Kistler demonstrated these correlations in an experiment. In 1992, they investigated *the dominant role of low-frequency interaural time differences in sound localization* [144, 145].

Sounds from 36 virtual directions in the horizontal plane were presented via headphones [142, 145]. For this purpose the interaural cues were processed as follows: While interaural level difference cues (ILD–cues) were normal and remained unchanged, all the interaural time difference cues (ITD–cues) were artificially set to correspond to a sound direction of 90°. Additionally, a high-pass filter with variable cut-off frequency was used to vary the source-typical localization cues.

A train of eight Gaussian noise bursts of 250 ms length followed by 300 ms of silence served as test signal. After each trial the spectrum was scrambled to avoid learning effects (see also section 3.1.3).

If both low and high frequencies existed in the source spectrum, the dominant ITD–cues strongly influenced the auditory perception, i. e., all auditory events were located at the side. If, however, the source spectrum was limited and lacked the lower frequencies, the ILD–cues determined the position of the auditory event. The ITD–cues fixed artificially had no influence.

This experiment clearly demonstrates that both interaural differences, ITDs and ILDs, and the spectral distribution of the source signal can have a strong influence on localization.

Experiments relating to HRTF cues

The spectral HRTF cues caused by the outer ears [37, 55, 56, 105] are the only cues available in monaural localization as well as for sound sources in the median plane.

A variety of experiments explored the way in which these spectral "deformations" of the HRTF influence localization and how manipulations of the spectrum alter the perceived position of a sound source [2, 29, 32, 81, 92, 108, 139, 140]. When we localize sources in the median plane, a boost of certain frequency bands can result in a change of perceived elevation [9, 12, 50, 134]. This has already been described in section 3.1.1

The use of non-individual HRTFs, e. g., a dummy-head or artificial HRTFs, can cause such a change in perception of source positions. Thus, using the "correct" HRTFs will have a strong influence on the monaural and median plane localization.

This was demonstrated, for example, in an experiment of Damaske & Wagner: If a listener perceives a sound through a dummy-head, i. e., foreign or artificial HRTFs, the perceived position of the sound source can strongly differ from that one localized if using "his own" HRTFs instead. In preparation of the listening experiment, the researchers recorded sounds in the median plane with a dummy head. If these recorded sounds were presented through headphones, the listeners perceived the sound "through the ears (HRTF) of the dummy head", i. e., not their own ears. The localization performance was shown to be distinctly inferior to that associated with free field listening where the subjects used their own ears [26].

In a similar experiment, Searle and his colleagues inserted microphones into the ear canal of the subjects in order to record the sounds, thus using the subject's *individual* HRTFs. By playing back these "individual recordings" via headphones the sources were localized at their original positions [111]. Therefore, it seems that when using individualized HRTFs, the localization is close to that of natural listening. A number of experiments [11, 17, 22, 38, 89, 86, 110, 119, 138, 141, 142, 146] focused on this influence of individualization of HRTFs on localization.

However, without pinnae (HRTFs) and with only interaural cues being available, a distinction between frontal and dorsal sound sources becomes more difficult [19, 47, 92, 96, 119, 138, 150]. This difficulty arises, for example, when using a sphere microphone as proposed by Theile [126]. Using such a sphere microphone for a binaural recording results in various front–back–inversions, as will be seen in more detail in section 4.4.1.

Similar acoustic conditions, i. e., the lack of HRTFs, prevailed in an experiment by Roffler and Butler [108]. They carefully arranged a plexiglass band with small holes at the positions of the ear canal entrances in order to flatten the subjects' pinnae. With his arrangement they found the subjects' ability of *median plane localization* to deteriorate sharply. Localization in the *horizontal plane*, however, was comparable to natural hearing – as long as front–back–inversions were ignored.

A special situation is the "pure" headphone-listening to *non-binaural* material[5]. The headphones eliminate the influence of the listener's own HRTF and thus lead to an acoustic perception that is *not* comparable to every-day listening: *in–head–localization*. Some do assert, however, that this kind of headphone-listening cannot lead to "normal" localization, but rather to a phenomenon called *lateralization* [64, 75].

---

[5] Non-binaural material shall be defined as sound material that was not recorded with a dummy head or in any kind processed using HRTFs

Nevertheless, Toole showed that an *in–head–localization* can also occur under freefield conditions. In this experiment, the listener perceived the sound as "inside the head", although he used his "own ears" to localize the sound [129]. Likewise, the lack of head-movements can lead to in–head–localization – in spite of the existence of HRTFs. This is a phenomenon occurring typically when listening to dummy-head recordings. We will discuss this phenomenon in the following main section.

Experiments relating to non-acoustic cues

Although there are various non-acoustic cues, in this subsection we will only comment upon and give examples for the *knowledge/memory* cue, the *optical* cue and the *vestibular* cue.

Knowledge & Memory To have any kind of knowledge about the sound source, i. e., being familiar with the source characteristics, can have an influence on localization. Especially in the monaural case, knowledge about the source is important because here the spectrum at the eardrums is the product of pinnae filtering and the source spectrum. The only way to "separate" these two components (and thus to "reconstruct" the position of the source) is by knowing about the source [2, 29, 55, 56, 105, 146]. Either this knowledge already exists or it can be acquired by training and learning. Generally, in listening situations *training* (and thus knowledge) can help to reduce errors and to improve localization [2, 50, 56]. Disregarding the influence of *learning* could lead to a different experimental result [37].

One way of using knowledge as a localization cue is to learn and memorize the spectral characteristics of the signal. This can be done either by repeating the signal, or by presenting it long enough for the listener to become familiar with it [101, 105]. The other possibility of taking advantage of the knowledge–cue is to use "*well-known*" sounds. Several localization experiments were based on using familiar sounds, e. g., the sound of a common music instrument such as a piano, or of a human voice.

In an experiment, Rakerd *et al.* [105] focused upon the difference in localization between familiar and unfamiliar sounds: They substituted a known (human) speaker by an unknown speaker and observed a drop in localization accuracy from about 90% to 50%.

Therefore, we must have come to understand the relationship between typical spectral manipulations of the HRTF on the one hand and associated positions of the sound source on the other hand from early childhood on and thus know it well[100]. As a consequence, when trying to localize a familiar sound with *foreign* HRTFs, such as that of a dummy head, localization errors might result which are caused by the *non-familiarity* with the these "localization-links" [87, 89].

However, we cannot only learn to establish and remember a link between HRTF and the source position, but also a relationship between certain sound characteristics of the source and positions of that sound source. In this context, a previously cited experiment (see section 3.1.1) by Gardner deserves to be mentioned.

A subject had to estimate the position of a human speaker. The speaker was asked to vary the loudness from *whispering* over *normally speaking* to *shouting*. The result relevant here was that all subjects *associated* a certain *type* of speech (whispering, talking, shouting) with a distinct distance. This again demonstrates how knowledge of the source spectrum can influence the perceived position.

That *knowledge of the spectrum* does influence the localization can also be shown by *scrambling* the spectrum from trial to trial. The listener each time hears a different spectrum, and hence it is impossible for him to familiarize himself with the spectral characteristics of the sound.

Using this scrambling technique most likely triggers a difference in monaural or binaural perception. When we perceive sources outside the median plane binaurally, localization is less dependent on the source characteristics [55]. In this case, a knowledge of the spectrum does not bring any further advantage, and scrambling does not influence the results greatly. But in case of monaural perception the results will deteriorate when using a scrambled spectrum [146, 143].

Visual & Vestibular Cue Closely related to the knowledge of the source position is the *visual cue*, i. e., *optical* information about where the sound source is located. If the loudspeakers are visible for the subject, the localization task will be reduced to a mere source identification [87]. Another effect is the “dragging” of the acoustic perception into optical vicinity. This is called the *proximity image effect* [38].

Another experiment deserves to be mentioned briefly here. This experiment was carried out by Wallach and focused on visual cues and *vestibular cues* [131, 133]. A subject was placed on a revolving chair surrounded by a curtain with vertical stripes that was prepared to rotate around the listener. Invisible to the subject a loudspeaker was placed on the other side of the curtain in front of him.

When the loudspeaker emitted a signal, it was localized correctly “in front”. Then the curtain rotated, and a visually induced “*ego-movement*” was created. Instantaneously, the listener had the impression of a sound originating directly above his head. The striking aspect of this result was that the acoustic perception was affected *solely* by an optical information, i. e., rotating stripes on the curtain, and *not* by any acoustic variations. Additionally, when the real loudspeaker was moved to the side of the listener, its perceived position descended slowly towards the horizontal plane.

Similar results were achieved when rotating the listener himself physically. His own physical rotation was registered by his vestibular organ and thus provided another localization cue, i. e., the *vestibular cue*.

But of course, as stated earlier, much more important is the equilibrium sense (vestibular cue) for detecting any changes in the orientation of the listener’s head as will be described in the next section about *Dynamic Localization Cues*.

3.2 Dynamic Localization Cues

This section will briefly describe a few experiments regarding *dynamic cues*. Such dynamic cues originate either from a movement of the sound source relative to the listener’s head, or a head movement, or a combination of both. Here, we will only take into account dynamic cues resulting from movements of the listener’s head.

3.2.1 The Origin of Motional Theories

First of all, we will give a short introduction to motional theories. The term *motional theories* refers to all localization theories describing the relationship between changes in the ear signals due to head movements and the corresponding positional change of the auditory event.

Von Hornborstel & Wertheimer

In 1920, von Hornborstel and Wertheimer suggested a theory that might be the "origin" of all *motional theories.* First of all, they disregarded the existence of head and pinnae. Therefore, no shadowing effect arises in this model. The ears were considered to be two points separated by 21 cm, lacking every spectral influence on the signal[6]. This all together constituted the simplest model of binaural hearing called *time difference theory* [60]..

Cone of confusion

Disregarding all shadowing effects caused by the head and the outer ears results in several localization ambiguities. For one, every point in the median plane is equally far away from the right "ear" as from the left "ear". Thus, the *difference* between those distances is zero, and so, correspondingly, is the time difference. Therefore, from a "time-difference point of view" *all* points in the median plane are equivalent.

Similarly, for each point *outside* the median plane there is an infinite number of points with an *identical* interaural time difference. If one considers only two dimensions, all of these equivalent points lie on a hyperbola. In three dimensions, this hyperbola becomes a hyperboloid, a conical shell that is commonly known as *cone of confusion.*

Van Soest

Van Soest was among the first to take into account the factor of *head movements.* Simply by turning the head it seems to be possible to determine the sound source — despite all ambiguities caused by the cone of confusion.

If a sound source is positioned in the median plane, both ears receive the identical sound at the same instant, i. e., there is *no interaural time difference.* Since this holds true for frontal *and* dorsal sources, the only way to discriminate these sources is by turning the head. For example, the head can be turned clockwise. In this case, the sound of a frontal source arrives earlier at the *left* ear than at the *right* ear. The opposite is true for a dorsal source, i. e., here the signal of the *left* ear lags behind that of the *right* ear.

---

[6] A more "advanced" model of the head is an (acoustically) *opaque*, rigid sphere of 17.5 cm diameter.

Van Soest assumed that we discriminate between frontal and dorsal sources by evaluating the *polarity* of the change in interaural time difference. However, he mistakenly ignored other cues that convey the direction of the head movement, as for example *visual*, *vestibular* or *tactile information* [115].

However, it is necessary to "*know*" the direction of the head movement in order to localize a sound source correctly. It is not hard to understand that a rotation of the head *clockwise* in case of a *frontal* sound source will produce the same change in interaural time difference, i. e., a "leading left ear", as a *dorsal* sound source in case of a head rotation *counterclockwise*.

Therefore, motional theories can be characterized as theories with multi-sensory input, or, as Blauert stated, "*heterosenory theories*" [11].

3.2.2 General Importance of Head Movements

The experiments described below are meant to demonstrate the general importance of head movements for localization.

Young

In a series of experiments Young [150] investigated the effects of head movements on localization by means of a so-called *pseudophone*. The experimental setup consisted of two small funnels or trumpets made of hard-rubber, which he connected via rubber tubes directly to the entrance of the subject's ears. Thus, the funnels replaced the subject's own outer ears. Both funnels, identical in size or shape, were 17.3 cm apart and pointed with their openings in opposite directions.

The pseudophone was fixed, and so all of the listener's head movements were prevented from altering the binaural stimulus-pattern. As a consequence, the subjects were unable to localize correctly, and only *lateralization* was possible (*left*... *right*). Thus, it was impossible to discriminate between *up*... *down* or *front*... *back*.

However, when the pseudophone was attached directly directly to the head of the listener and thus allowed head movements, the "*localization with the pseudophone resembled the normal, unrestricted, tridimensional type and not the restricted type found when the 'pinnæ' are detached from the head*" [150]. Although Young did not make any systematic or quantitative observations, his findings are remarkable. Obviously head movements can compensate the lack of individual pinnae (at least within certain limits).

With respect to the magnitude of head movements, Young referred to Klemm [67] who reported the spatial blur for "click"-sounds in the frontal median plane to be in a dimension of 0.75° and 3°. A comprehensive overview about localization blurs can be found in Blauert [11].

This, however, means that the results of experiments using a simple chin- or head-rest to fix the head might be influenced by unconscious, small head movements and "*need to be examined critically from this angle*" [150].

De Boer & van Urk

In 1941 de Boer and van Urk used a *spherical* dummy head to record sound. Parallelly, the subject received this sound via headphones. During the presentation of the sound the dummy head was turned to the left and to the right [13].

Normally the problem with a *symmetrical* model of the head are "front-back-inversions"[7]. But if the subject moved the head in accordance with the dummy head's movement, localization in the horizontal plane was reliable. The subjects were able to determine whether the sound source was in front of or behind the dummy head. However, when they moved their head in the opposite direction, *front-back-reversals* occurred.

This again demonstrates the impact of head movements on localization. And as previously stated the head movements cue is strongly linked to the vestibular cue (to "feel" the direction of the head movement).

In particular, de Boer's and van Ulk's experiment, as well as similar experiments carried out by Klensch [68] and later by Jongkees and van de Veer [65] disproves van Soest's assumption that solely the *change* in binaural signal difference (but not the vestibular information) is necessary for correct localization.

Wallach's Theory

In 1938, Wallach put forward a theory based on head movements that totally disregarded the effect of the pinnae. He defined an angle, called *lateral angle*, that refers to the angular distance between the direction of the sound source and the aural axis. This angle exactly describes the cone of confusion because every point of the surface (determined by the binaural signals) possessed the same respective angular distance. For example, a lateral angle of magnitude 0° equals the aural axis itself. Every point in the median plane, on the other hand, has a lateral angle of 90°.

Moving the head results in *series of specific changes* of the lateral angle, which are *unique* for a given source location. Thus, every single position can be defined by a *sequence of lateral angles*. In his experiment, Wallach proved the perceived location of the auditory event to be *independent* of the sound source's actual position, as long as the characteristic changes in the lateral angle are presented in accordance with the head movements.

He used a circle of equidistant loudspeakers in the horizontal plane, all connected to a rotary switch. The switch was linked to the (rotational) movement of the listener's head. Music and human voices served as test signals. The one loudspeaker, whose position met the desired lateral angle, started to emit the signal. Subsequently, by rotating the head the signal was switched to the next loudspeaker. As a consequence, the auditory event was perceived at the "calculated" position — which not necessarily coincided with the loudspeaker's real position.

For example, when listening to the *frontal* loudspeaker and there being an angular distance between the loudspeaker twice the rotational angle of the listener head, the sound seemed to come from the rear [131, 132].

---

[7] These kind of localization errors will be demonstrated in more detail in section 4.4.1.

But not only *active* rotations of the head lead to those perceptions. As already described in section 3.1.3, *passive* movements of the head (using a swivel-chair) could also reach the desired result. In that case, the vestibular organ provided the localization cues. Similarly, optical cues (rotating a striped curtain) and a total absence of physical movements can influence the perceived position.

All these experiments by Wallach showed human localization to depend to a high degree on dynamic cues. Reproducing the typical binaural pattern in accordance with head movements can establish distinct locations of the auditory event — even if the position of the real sound source is different. The experiments showed the dynamic cues to dominate the pinnae cues because the physical direction (perceivable by the effects of the pinnae) and the perceived direction are widely different.

The fact that head rotations provide an immediate information as to whether a sound is in the frontal or dorsal hemisphere was confirmed by Burger. He investigated the front-back discrimination of the hearing system subject to head movements and pinnae. But he also remarked that "*covering one ear reduces discrimination to something just slightly better than wild guessing*" [19].

Likewise, Toole stated: "*The in-head localization was lost during the movement, but normally it was restored by a momentary return to the starting position*" [129]. This, however, contradicts Wallach's findings, namely that localization perception continues even after the head movements stopped.

Boerger & Fengler

When listening to mono or "plain" stereo, non-binaural material via headphones ("ordinary headphone listening"), normally *in-head localization* occurs. However, Boerger et *al.* showed that a "more natural" acoustic impression is possible, when taking head rotations (about the vertical axis) into consideration, and "distorting" the signals accordingly by means of a parametrical system.

They used a simple, electro-mechanical head tracker, and a parametrical system with a transfer function depending on the listener's head orientation only. This basic parametrical system led to an extracranial localization by taking into account the listener's head-movements, which obviously seem to be important for localization.

3.2.3 Different Kinds of Head Movements

The aim of this section is to answer the question if there is a *prominent* head movement. Therefore, a more detailed look on different *types* of head movements will be taken.

Wallach's theory neglected the existence of a pinna and assumed a symmetrical head. Consequently, the binaural cues determine only a *range* of possible source locations. This is commonly known as *cone of confusion* (also see 3.2.1).

To resolve these localization ambiguities *different* head movements are necessary. For example, when *turning* the heard, i. e., rotating the head horizontally, a frontal sound source can be distinguished from a source in the back. On the other hand, a *tilting* movement of the head around the frontal axis[8] permits a distinction between sound sources located above or below the horizontal plane.

Such *tipping* or *nodding*, however, does not have any influence within Wallach's theory because the nodding movement is a rotation about the aural axis. Therefore, this axis will not be displaced, and no change in interaural cues will result (changes due to the pinnae's form were neglected a priori).

Thus, in principle at least *two* head rotations around *different axes* are necessary to localize a sound source. But in real life, the head's revolution about a *fixed* axis is the exception to the rule. Normally, the rotational axis varies and this probably suffices to remove ambiguities which would otherwise result from an accurate rotation about a fixed axis.

Free head movements

In 1967, Thurlow, Runge and Mangels carried out the following experiments concerning the impact of *different* head movements on localization. They categorized the head movements in three aforementioned types: *pivoting*, *tipping* and *rotating* movements.

Their experimental setup consisted of 10 loudspeakers placed in an anechoic room (free-field condition), five of them reproduced only low frequencies and the other five only high frequencies. Two different band-passed noises were used each with a bandwidth of 500 Hz (low-band with a frequency range of an octave (500 Hz – 1000 Hz) and a high-band filter using a bandwidth of a third (7500 Hz – 8000 Hz)).

Blindfolded subjects' were asked to localize the different noises. They should denote their positions by pointing in the perceived direction. It was allowed to move the head to get aid in localization of the loudspeakers during the 5 s of the stimulus being presented. Their head movements were captured by a motion-picture camera, and a small lightweight frame mounted on their head facilitated the measurement of angular movements.

The results were that *all* different patterns and combinations of head movements types could be observed. Among these, the rotational movement exhibited the greatest amplitudes of all three movement-classes. The three most frequent combinations of head movements were a rotation combined with a tipping movement, followed by the combination consisting of rotation, tipping and pivoting, and finally head rotation alone.

When considering only movements with an amplitude greater than 10 degrees, the rotation led the table, followed by a combination of rotation and tipping. This again emphasizes the importance of head rotations on localization. However, strong inter-individual differences were observed regarding the maximal movements.

Thurlow *et al.* showed experiments that head rotations are the most frequent head movements. They occur either alone or in a combination with other head movements. However, as already observed by Wallach, in natural listening there are no dedicated head rotations about a stable axis. Moreover, the axis varies with time.

---

[8] The frontal axis is the axis perpendicular to both axes, the aural axis and the vertical axis.

Induced head movements

In another experiment, Thurlow and Runge “attached” an apparatus by means of a bite bar to the subject’s head. This apparatus was controlled by the experimenter so that he actually could *induce* head movements at a given instant. It could be rotated about the vertical axis and also about an horizontal axis to perform either a pivoting or a tipping movement. A motor powered the frame to rotate with an angular velocity of 19.8°/sec.

They repeated the afore described localization experiment (see 3.2.3) but with induced instead of free head movements, and compared the results in localization error with that in free motion [128]. In addition to the low- or high-pass filtered noise also filtered “click”-sounds were used. The signals were reproduced by a total of 14 loudspeakers, seven for the low-frequency range, and seven for the high frequencies.

It turned out that even induced head movements had a strong impact on localization accuracy in the azimuthal component if a *rotational* component was part of the movement. Even front-back-inversions were reduced similar to the experiment with free movements. This again stresses the importance of *head rotations* on localization.

3.2.4 Head Movements and Individual Binaural Cues

In the past, a lot of experiments investigated the importance of spectral pinnae cues for localization. Particularly in the case of an immobilized, fixed head the pinnae cues can be a prominent localization cue, especially in the median plane. Various authors report on HRTF-measurement techniques or the respective characteristics of HRTFs [6, 11, 20, 22, 32, 49, 81, 88, 114, 138].

Experiments proved the localization to be comparable to natural hearing only if the HRTFs employed to generate the binaural signals (dummy head or auralization) resemble the listener’s HRTFs [89]. This can be achieved, for example, by modelling the listener’s pinnae for a dummy head, or the exact measurement of the listener’s HRTF for an auralization. However, when allowing head movements, the dynamic cues are more prominent than the exact HRTFs.

In an experiment reported by Freedman and Fisher (1968), the subjects had to localize sound sources using either their own ears, artificial ears or with their pinnae occluded [32]. When their head was fixed they were able to localize the sound source much better with their own ears or with artificial pinnae than without pinnae at all (occluded ears). But when the listeners were allowed to move their heads freely, these differences in localization disappeared.

Already in 1931, Young had indicated that head movements to a high degree aid in localization, as was already described in section 3.2.2. He used small funnels connected with the ears of the listener by means of flexible rubber-tubes. The funnels were movable in accordance with the listener’s head.

Inanaga *et al.*

Boerger and Fengler used a parametrical system that simply added a time delay to the binaural signals according to the orientation of the listener’s head (see 3.2.2).

Inanaga *et al.* [61] conceived a more advanced headphone system. They measured the HRTF of a dummy head with loudspeakers in a normal (reverberant) listening room for different directions. These HRTFs were simplified and used for the convolution with the input signal to produce an auralization. The listener's head movements were registered with a head tracker, and the respective set of HRTFs was chosen.

As expected, localization was bad when head movements were not allowed, and frequently in-head localizations occurred. But when the subjects were allowed to rotate their heads, the sound image was perceived outside the head, and the signals seemed to come from their original directions. However, when the head tracker was switched off (leading to a listening situation with *fixed* dummy head) the spatial impression collapsed more or less instantaneously and in-head localization occurred again.

Loomis *et al.*

An interesting localization experiment was carried out by Loomis *et al.* in 1990. In most experiments, the subjects are restricted to a single position by seating them on a chair, surrounded by loudspeakers. In this experiment, however, the subjects were able to move their head freely and had to "home" the sources by actually walking to them.

The experiment took place in a large gymnasium (50 m long, 25 m wide and 15 m high) with 18 sound sources to be localized. In one trial, loudspeakers served as sound sources and the subjects used their own ears. In the other trial, the subjects were confronted with "virtual loudspeakers" via headphones, only simulating different localization cues[9].

For example, the *interaural time difference* (ITD) was approximated by Green's equation ($\vartheta =$ azimuth): $ITD = 257(\vartheta + \sin\vartheta)$. The signal was split into a lower and a higher frequency band. Only the HF-band ($f > 1800\,\mathrm{Hz}$) received an *interaural level difference* (ILD) that varied sinusoidally in accordance with the orientation of the listener's head. Additionally, pinnae shadowing effects and the influence of distance were simulated.

A head tracker, mounted on the headphones, and a camera registered the position of the subjects. The researchers used an update-rate of 72 Hz together with a latency time of 35 ms and an angular resolution of 1.4 degree, which seems to be small enough to not degrade the spatial impression[10].

When the subjects were confronted with the virtual loudspeakers via headphones, their localization performance due to head movements was similar to that of natural hearing. Dynamic cues seemed to dominate pinnae cues, or compensate the lack of them as assumed by Noble [95].

It should, however, be noted that an experiment by Müller and Bovet (1999) did not show a total compensation in all of their tested sound directions [91]. But they

---

[9] At that time a real-time convolution with HRTFs was not feasible because of the required processing power a 12 MHz 80268 computer was not capable of providing!

[10] These values of the auralization system are in accordance with Sandvad [110, 136] and the author's findings [66], which will be reported in section 4.

used *pure tones* and not speech signals or noise (bursts), as sounds to be localized. This might account for the different outcome of their experiments.

3.2.5 Head Movements and other cues

Two examples will be given here for the purpose of relating head movements to other localization cues like the effects of memory or conflicting cues.

Han – The Effects of Memory

Memory is often ignored as a localization cue [50, 90, 122, 125]. But it can nevertheless be a vital cue for a correct identification and localization of the source. Theile, for example, based his *association model* entirely on the relevance of memory [122]. Only if a listener *remembers* a sound or a class of sounds, he can later *associate* a perceived auditory event with this sound.

Memory effects can be supported by head movements. Han [50], for example, showed that (under certain conditions) source movements do not account for resolving front-back-inversions once the position and the sound source were identified and memorized.

Han used two loudspeakers, diametrally positioned in the median plane at 0 and 180 degree azimuthal angle, and in a distance of about 2 m. A subject was placed in between. The signal used was a noise-band between 800 and 1200 Hz. This part of the frequency band is most significant in front-back discrimination [11, 8].

The experiment consisted of four consecutive steps, and started with a period of silence in order to "clear the acoustic memory".

Initially (step 1), the *rear* loudspeaker emitted the noise signal and the subject had to turn his head to *identify* the loudspeaker and to ensure that he was hearing the rear loudspeaker behind him.

Then (step 2), the subject was told to face the frontal loudspeaker and keep the head still. Thereupon the signal was "*switched over*" from the rear to the frontal loudspeaker.

While the subject kept his head in the same position, the frontal loudspeaker was *moved perpendicularly* to the median plane, i. e., to the left and to the right (step 3). Astonishingly, the subject had the impression as if the *rear* loudspeaker (and not the frontal one!) were moving *behind* him.

Finally (step 4), the frontal loudspeaker was returned to its original position and the subject was allowed to *turn his head.* He identified the frontal loudspeaker to emit the noise.

This experiment demonstrated that *head* movements help to identify a sound source and to memorize its position (step 1). Even by changing the sound source (step 2) and executing *source* movements the perceived position remains unchanged (step 3). Only by *head* movements (step 4) the "true position" and the real source can be perceived correctly.

In a second experiment with the same setup, Han started with the *frontal* loudspeaker. In order to ensure again that the subject hear the frontal loudspeaker, he was instructed to move his head. In a second step, while facing the frontal loudspeaker, the signal was switched to the *rear* loudspeaker. After a few seconds it was *switched back* to the frontal speaker.

Despite the physical source's change (frontal speaker ⇒ rear speaker ⇒ frontal speaker) no *front-back-reversal* occurred. The subject only perceived the position of frontal speaker after memorizing (step 1). Even a short interruption (step 2) could not alter or falsify this impression.

These experiments showed the importance of memory for the localization and identification of sound sources in connection with head movements. Source movements, on the other hand, have a different impact on localization than head movements.

In this respect, it would be interesting to know if the previously described Franssen effect (see 3.1.1) could be avoided by using head movements to establish the location of the first loudspeaker before using the second one. For if the position of the loudspeaker emitting the non-transient part can be established and memorized, the memory might help to track its position.

Wenzel – The effects of conflicting cues

Wightman and Kistler have shown the dominance of ITD cues over ILD cues, when both being in conflict, as we described in the last section (3.1.3). They used *individualized* HRTFs for their experiment. Especially the lower frequency range in the test signal played a critical role. The ITD cues were seen to be dominant mainly because they are more or less consistent across frequency bands. Also, in the low frequency range the size of the head is of the same magnitude as the wavelengths and thus diffraction occurs. This diffraction prevents the existence of interaural level differences (ILDs). However, the experiment *disregarded* and excluded head movements.

To investigate the influence of dynamic cues on conflicting interaural cues Wenzel carried out a similar experiment three years later. This time the subjects were allowed to move their heads freely [135, 136]. Instead of individualized HRTFs she used non-individualized HRTFs for a dynamic (taking head movements into account) convolution. The subjects received the signals through headphones.

The interaural cues were either correctly or incorrectly (fixed at 0° azimuth and 0° elevation) correlated with the listener's head motion to generate conflicting cues. There were three different conditions:

Condition 1: Both, ITD and ILD cues were correlated correctly

Condition 2: ILDs were correlated correctly, ITDs were fixed

Condition 3: As in previous condition, but with the ILDs fixed

The experiment produced the following results: In contrast to the *static* condition (Wightman–experiment), where the ITDs dominated the conflicting ILDs, here the opposite seemed to be the case, i. e., the ILDs dominated conflicting ITDs. Head movements aided in resolving front-back-ambiguities, primarily when the ILDs were correctly correlated with the head movements. However, a few front-back-inversions still remained also if the ITDs were correct.

When looking at the front-back-reversals the following could be observed: As was to be expected, the lowest rate of front-back-inversions occurred when there were no conflicting interaural cues. But with the ITDs being fixed and the ILDs being normal (condition 2) there were *less* front-back-reversals than in the third condition (ITDs correct and the ILDs fixed). This result is in contrast to the findings of Wightman and Kistler [145]. It suggests that in case of head movements pinnae cues (ILDs) may have a stronger impact on localization than what might have been expected from the former results of Wallach [131, 132, 133] and Wightman & Kistler.

3.3 Discussion

We showed that all the aforementioned experiments "fit" into the proposed localization scheme. It is possible to classify various experiments according to the localization cue analyzed.

Since a totally different result may emerge from overlooking a localization cue, we emphasize the importance of taking into account all of these valuable cues. Our cue classification helps to identify the relevant cues that were sometimes (deliberately) omitted in some experiments. This was to reduce the influence on the experimental results to a *single* variable, as, for example, the influence of the environment on distance perception in Gardner [36] was neglected and only the influence of loudness was studied.

We will briefly discuss a few applications of these results concerning the perception of music. The first example illustrates the general inability to always localize sound sources sharply. Another example deals with the perception of distance.

Localization Blur and Envelopment

The localization blur, i. e., the inability of the human hearing system always to localize and "pinpoint" a sound source, seems to be the basis of the aural sensation of *envelopment*. The sound sources are not defined exactly in shape and position; moreover a kind of "acoustical aura" is created.

For example, the large reverberation times of churches were intentionally used by the composers to create a sensation of "surround sound" in their works. Architects, on the other hand, create concert halls in order to achieve an optimal envelopment.

In electro-acoustic transmission systems, there is a trend to use not only the conventional two stereophonic channels but also so-called surround sound channels. These systems aim at producing an ambient sound that seems to come from indefinite directions. If the human hearing system could always localize the sound sources (loudspeakers) without any blur, those systems could not work.

Reverberation as a mean to create “Spaciousness”

Nielsen proved the relationship between direct sound and reverberation to influence the perceived distance of a sound source (section 3.1.2). This cue has a stronger influence on localization than the perceived loudness of a sound, as was found by Gardner [36, 93].

In modern production of music it is sometimes desired to create an impression of depth and, thus, spaciousness, for certain signals, e. g., a solo-guitar. By usage of a so-called *reverb*, an electronic device that adds artificial first and late reflections (echoes) to the original signal, the balance between direct sound and reverberation can be varied.

Deliberately, a typical characteristic of the hearing system is used in order to artificially create the impression of depth and spaciousness in music production.

Chapter 4

Investigating the Importance of Head Movements quantitatively

This chapter describes our experiments carried out at the *Institut für Rundfunktechnik* (IRT) in Munich, Germany, between 1997 and 2000, in which we thoroughly investigated and analyzed the influence of head movements (rotations) on localization. These experiments were part of a research project in cooperation with *Studer Professional Audio AG*, Zurich, Switzerland, and contributed significantly to the development of a new auralization method described in section 5.

4.1 General Experimental Setup

We describe in this section the general experimental setup and the procedure we used to carry out these listening tests.

4.1.1 The Setup

All experiments involved a dummy head, mounted on a motor-driven turntable and controlled by a head tracker in connection with a personal computer. The dummy head was placed at the optimal listening point, the so-called "*sweet-spot*", in a standard surround sound loudspeaker setup (3/4 stereo format) according to ITU-R Rec. BS 775–1 [63]. The distance between dummy head and loudspeakers was exactly 3.0 meters.

Depending on the actual experiment, the whole setup was placed either in a *normal listening room* ("studio") according to the EBU Tech 3276 standard, or in an anechoic room. The latter was chosen to simulate rather unnatural *free field* conditions.

The *Neumann KU 100* served as a dummy head. It was connected to *STAX SR-Lambda Professional* headphones via headphone amplifier *STAX SRM-Monitor* (see

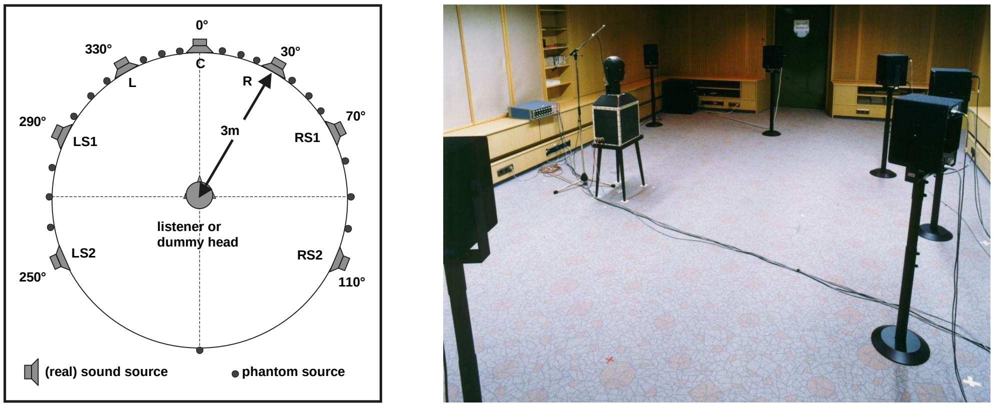

Fig. 4.1: Surround sound loudspeaker setup according to ITU-R Rec. BS 775–1 in the standard listening room at IRT

figure 4.2). Both the dummy head and the headphones were *diffuse-field equalized* [62].

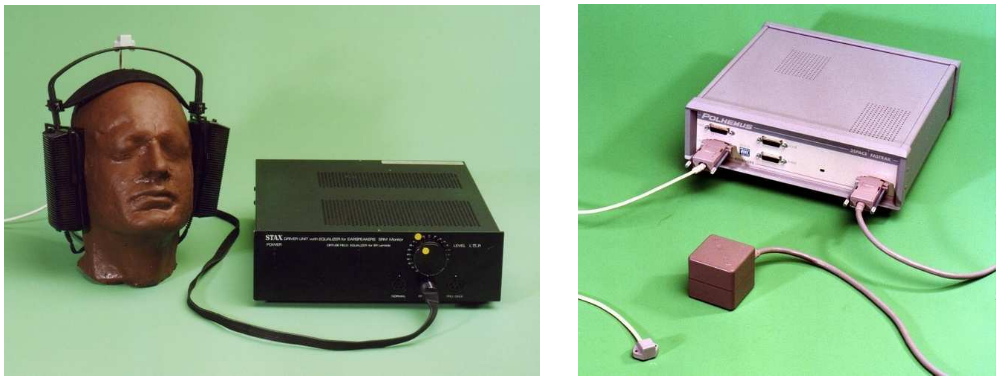

Fig. 4.2: STAX SR-Lambda Professional headphones with STAX SRM-Monitor (left) and Polhemus 3Space FasTrak head tracker (right)

The diffuse-field equalization is the optimal interface between the recording part and reproduction part as described by Theile [123, 124] and Larcher [74]. It avoids errors in the transmission path from dummy head to headphones, and thus ensures a true and faithful reproduction.

The idea to track the head movements of the listener and transmit these to a "recording device", e. g., a dummy head, is not entirely new. The literature documents previous experiments using a kind of mechanical or electro-mechanical head tracker [14, 15, 69, 70, 99, 150].

In the experiments at IRT, a precise *Polhemus 3Space FasTrak* head tracker (see figure 4.2) was used, with an angular resolution of 0.1 degree. It consisted of three components: a *receiver*, a *transmitter* and a *base unit*. In contrast to (electro-

)mechanical head trackers its mode of operation is based on an electro-magnetical principle. The transmitter sent out an electro-magnetical field (EMF), and the receiver detected its position within this EMF. This method recognized all six degrees of freedom: three rotations and three translations. The experiments at IRT only took rotations about the vertical axis into account. Only one experiment, described in section 4.3, registered a second head movement: the tipping or tilting movement (rotation about the ear-axis).

When connecting only one of four possible receivers to the main unit, the update-rate was 120 Hz (i. e., every 8.33 ms new positional data were sent), which at the same time was the maximal update-rate. The data were transferred to a PC via the serial interface (RS–232C) at 19,2 kBd in an ASCII-format using a nullmodem-cable.

The PC controlled the motor-driven turntable, likewise via RS–232 interface. The turntable and its control-interface was integrated in a wooden box in order to reduce the noise level caused by the motor-unit (*TR3* by *THOMA Filmtechnik*) by a total of 35 dB. The construction is depicted in figure 4.3. Details about this motor-driven turntable can be found in [107]. The angular accuracy of the whole system, consisting of head tracker and motor-unit, was one degree.

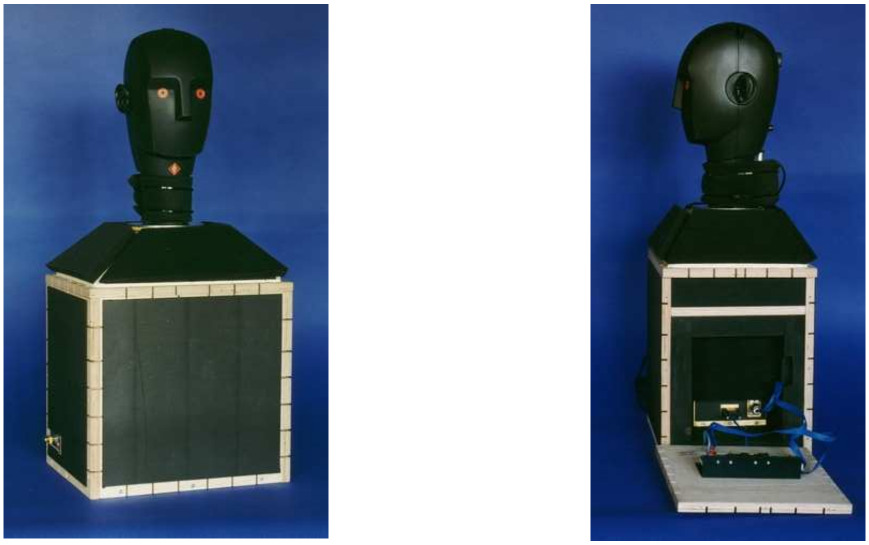

Fig. 4.3: Motor-unit for the dummy head inclusive control interface and wooden case (left: view from the front, right: view from the side)

4.1.2 Experimental Procedure

The task was the following in all experiments: A sound source (loudspeaker) or a phantom sound source was presented either to a dummy head or to the subjects directly. The echo-free ("dry") recording of a male voice (track 50 on the EBU SQAM-CD [30]) served as a signal. The sound was audible for a duration of nine seconds. The subjects had to denote the perceived position of the aural events graphically on paper.

There were 27 relevant locations, all of them in the horizontal plane. Apart from the positions of the seven real loudspeakers[1], further 20 positions of phantom-sources[2] had to be localized. The phantom sound sources were created by a level differences of +6 dB, 0 dB and -6 dB, respectively, between two adjacent loudspeakers. At the listening position the total of the sound pressure level was always 0 dB. All in all, 30 positions had to be localized, including three locations that were presented twice for cross-checking.

In the first part of the listening test the sound was presented to the dummy head. The head tracker was deactivated so that the dummy head remained in a fixed position. The subject received the sound of the dummy head via the headphones. It sat in a separate darkened room, surrounded by an opaque curtain in a distance of about 1m. A dimmed light enabled the subject to mark positions on a prepared experimental sheet. The second part was almost identical with the first part apart from the rotational component of the listener's head being tracked and "passed on" to the motor-driven dummy head.

Finally, in the third part, the subject employed its "own" ears to listen to the sound. Therefor, the listener sat in the "original" room at the dummy head's position. It was allowed to move its head freely. The previously mentioned curtain (sound transmitting, opaque fabric) prevented the subject from actually seeing the loudspeakers and thus from using visual cues.

A short training session preceded each experiment. It consisted of ten localization tasks in order for the subjects to familiarize themselves with the (new) hearing situation.

4.1.3 Influence of the System's Latency Time

In preparation of the actual listening tests a preliminary experiment served to determine the system's maximal latency time and its effect on localization. This is the time between the head rotation was detected by the head tracker and tracking of the motor unit with the dummy head.

The update rate of the Polhemus FasTrak was 120 Hz, and the baud rates of the serial interfaces between head tracker, computer and motor-unit were adjusted in order to permit the transmission of the head tracker's positional data to happen in between two head tracker updates, i. e., within a time-window of 8.33 ms.

To determine the maximal latency time of the system, an *additional variable delay* was added to the system's inherent minimal latency time of 50 ms. Hence, the system's latency time varied from 50 ms up to 150 ms in 8.33 ms–steps (see Tab. 4.1):

For the sake of this experiment's preparation various sounds (music, speech and solo instruments) as well as different loudspeaker setups (mono, stereo, surround) were tested. The most critical condition was a castanets-sound (EBU SQAM-CD, track

---

[1] Real LS-positions: C, L &R, LS1 & RS1 and LS2 & RS2

[2] Phantom source positions, three each between the loudspeaker pairs: C-R & C–L, R–RS1 & L–LS1, RS1–RS2 & LS1–LS2, and a single centered phantom source between L–R and between LS2–RS2

Delay	0	8	17	25	33	42	50	58	67	75	83	92	100
Latency	50	58	67	75	83	92	100	108	117	125	133	142	150

(All values in ms)

Table 4.1: Values of the variable delay and the corresponding total latency time

27) using a single frontal loudspeaker. The loudspeaker (Klein + Hummel O108/TV) was placed in a distance of 2.5 m to the dummy head in the median plane.

The subject was placed in a different room, separated from the dummy head with the loudspeaker. Using a switch it could alter between two listening situations: In one situation the system's total latency time equaled the minimal latency time (50 ms) and in the other situation an additional delay was added. The subject was instructed to perform small head rotations about the vertical-axis to ease the perception of the difference. The experimenter changed the additional delay between each run.

It was the subject's task to report verbally the situation of either that included the additional delay time. At the five most critical latency times a cross-checking was performed. 17 expert listeners took part in the test. If they perceived a difference and named the correct situation, a "1" was denoted, otherwise a "0".

Results

Figure 4.4 displays the results together with the mean value of all 17 persons and a 95 %-confidence interval depending on the total latency time. An average value of 0.5 was assumed as the threshold that needs to be passed to allow a perception. Hence if the 95 %-confidence interval of the mean value exceeded this limit, the respective total latency time was counted to be perceptible.

Each (total) latency time falling short of 85 ms was ignored by the subjects. The transition between "*not perceptible*" and "*perceptible*" occurred in a range between 85 ms and 101 ms. For latency times exceeding 101 ms localization artifacts were perceived.

A *tracing-effects* of the corresponding auditory event was observed for very long latency times. When the subjects turned their head, firstly, the auditory event was "pulled" in the *same* direction as the head moved, but then suddenly the auditory event "fell back" to its original position (see figure 4.6). This effect intensified when the latency time increased far beyond T_{max}.

The result of this preliminary experiment was the following: The system's inherent latency time of 50 ms was not perceptible. The *maximal* total latency time of the system not being perceived was $T_{max} = 85$ ms. These findings are consistent with Sandvad's results. He found localization not to deteriorate significantly as long as the latency time falls short of 96 ms [110].

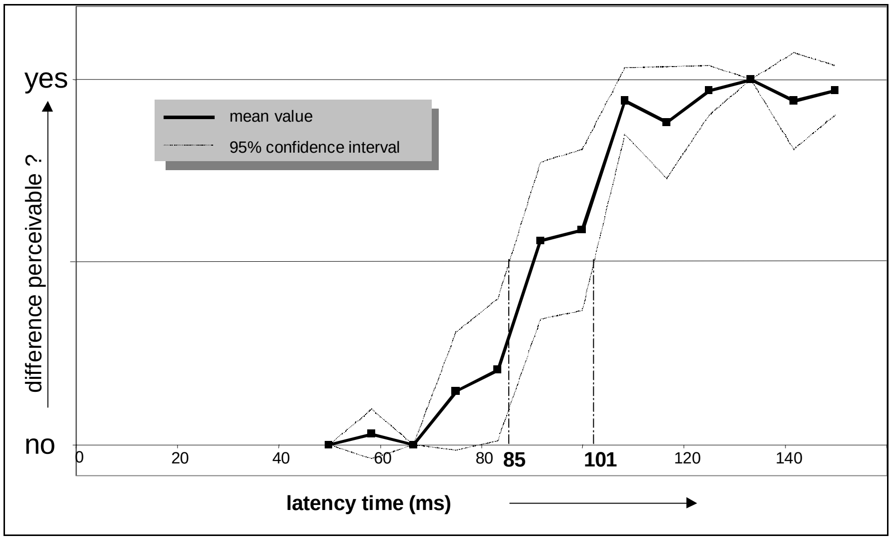

Fig. 4.4: Influence of latency time on localization. Increasing the system's total latency time leads to the perception of artifacts. This is denoted as the perceivable difference with respect to the case of minimal latency time. Only for latency times shorter than $T_{max} = 85\,\mathrm{ms}$ these artifacts are ignored.

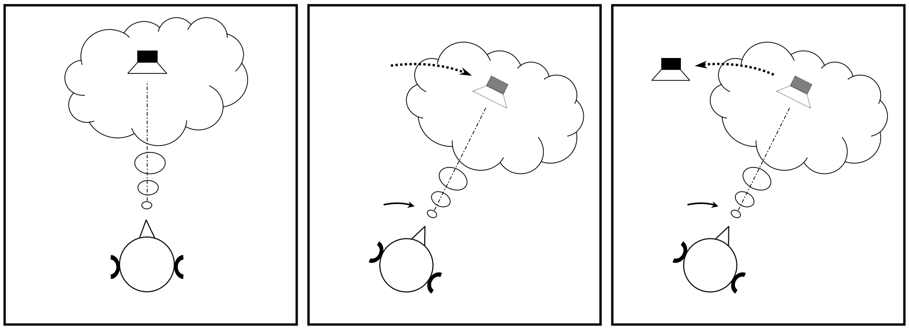

Fig. 4.5: Latency time greater than t_{max}

Fig. 4.6: For $t_{Latency} = 0$ the auditory event remains at its position (left). For latency times $0 < t_{Latency} < T_{max.}$ a small "dragging-effect" can be perceived (middle). Finally, audible artifacts occur for $t_{Latency} > T_{max.}$ (right). When moving the head the auditory event moves at first in the same direction as the head moves before returning to the original position (tracing-effect).

4.2 Horizontal Head Movements (Rotation)

Section 3.2 discussed several experiments involving dynamic cues. In one experiment Thurlow *et al.* investigated the role of various head movements [127]. They found head *rotations* about the vertical axis to be important for localization.

Throughout this document the terms *head rotations* or *rotations* refers only to the head rotation about the *vertical axis* (also known as *z-axis*). In contrast, the other revolutions of the head about the y- or x-axis will be called *pivoting* or *tipping*, respectively (in accordance with the definitions by Thurlow).

This section investigates in detail the influences of head rotations on localization under *natural* and *unnatural* hearing conditions.

4.2.1 Natural Conditions (Studio)

In this first experiment, the dummy head was placed in a *studio*, a standard listening room at IRT. The head tracking was disabled in the first trial. The subject thus had to use a fixed dummy head for purposes of localization. The subjects were also instructed to keep their heads still. During the second trial, the head tracking was activated, and the subjects were allowed to turn their head (though only in the horizontal plane). Finally, in the third session, the subjects themselves sat in the studio (instead of the dummy head) and were completely free to move their head. A total of 17 persons took part in this experiment.

Results

Figure 4.7 – 4.9 displays the results of this experiment. In these graphs the x-axis denotes the azimuthal angle of the *presented* sound event, and the y-axis the *perceived* azimuthal angle of the corresponding auditory event. The bold lines denote positions of the loudspeakers. The dashed lines refer to prominent phantom-sources (between L and R, or LS2 and RS2, respectively). In case of an ideal localization the perceived position of the auditory event is congruent with the actual position of the real sound source. Graphically such congruence results in a diagonal through the origin, starting from the lower-left corner and ending in the upper-right corner ("ideal diagonal").

Using a fixed dummy head resulted in numerous front-back-inversions, as depicted in figure 4.7 in form of bifurcations. These occurred mainly in the range from -30° to 30°, i. e., between the frontal left and right loudspeaker (L and R)[3]. The fairly big spread suggests a reliable localization with a fixed dummy head to be impossible.

In the second trial, where head tracking was enabled, these localization ambiguities vanished almost completely, although the HRTFs of the dummy head and that of the listener were clearly not identical, and no individualized HRTFs were used. The deviation from the "ideal diagonal" was fairly small. The accuracy of localization was comparable to that of the third session.

---

[3] The range -30° to 30° is denoted in the diagram as the ranges 330°–360° and 0°–30°

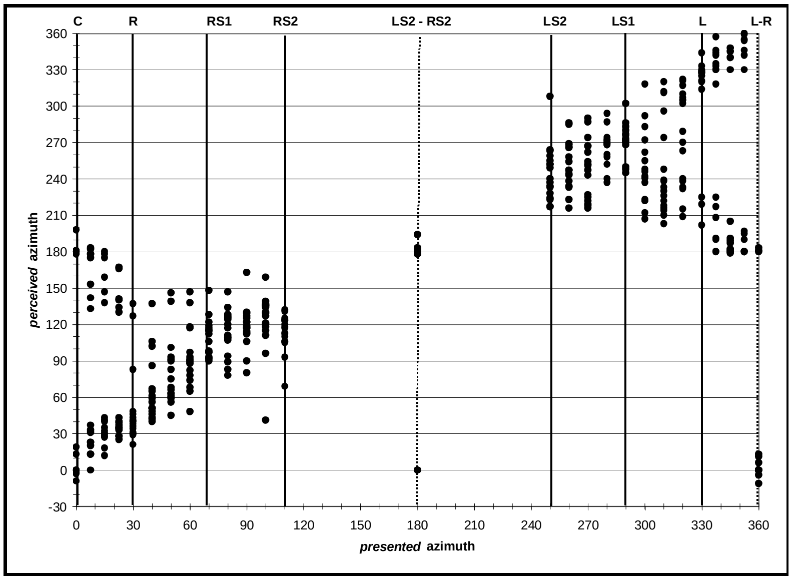

Fig. 4.7: Localization using a fixed dummy head without head tracking under natural listening conditions (IRT studio). A lot of front-back-inversions occur in the region between the first surround loudspeakers LS1 and RS1.

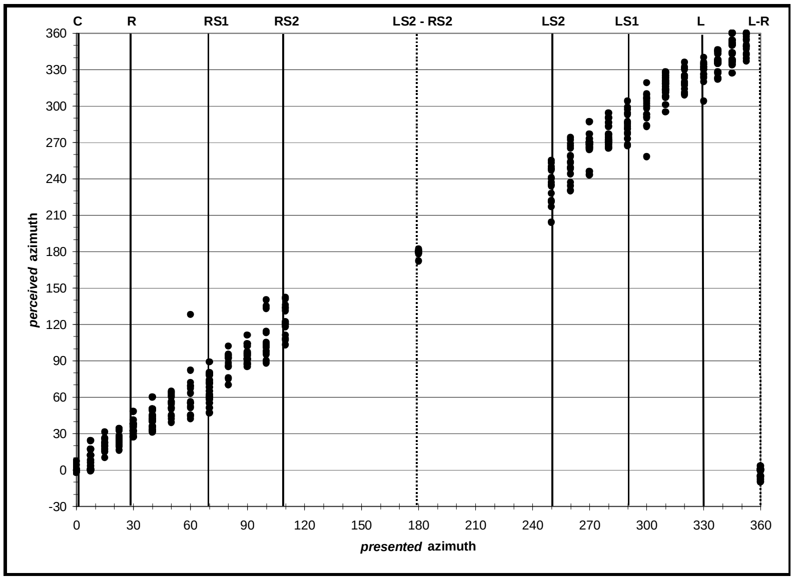

Fig. 4.8: Localization using a dummy head *with* head tracking in the studio (natural listening conditions). The various front-back-inversions of the fixed case do not exist anymore, and the localization resembles that of natural hearing.

In the third trial, when the listeners were allowed to use their "own ears" and to move their head freely, the exactitude of localization did not increase considerably (see fig. 4.9). The spreads were fairly small and comparable to that of the second trial.

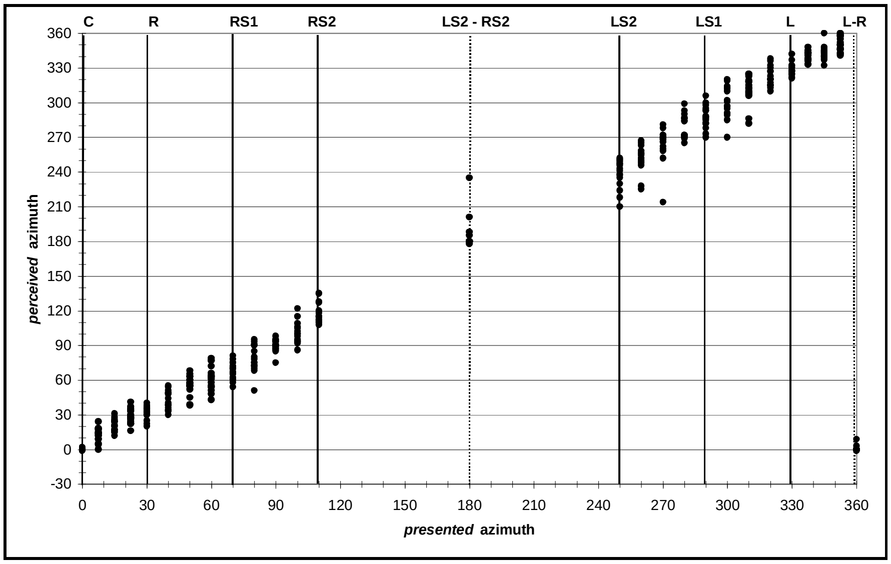

Fig. 4.9: Localization with the listeners being under natural listening conditions (IRT studio) using their "own ears". Localization is reliable without front-back-inversions.

The general result of this experiment can be summed up as follows: When using a dummy head whose movements are linked to the listener's head rotations localization is comparable to natural hearing. This confirms the importance of head rotations on localization.

4.2.2 Unnatural Conditions (Anechoic Room)

Having investigated the influences of head rotations on localization under natural listening conditions (studio), we repeated the same experiment under "unnatural conditions", in an anechoic room. Here, all reflections from walls, the floor, the ceiling etc. were absent, and the listener could only evaluate the direct sound.

As in the previous experiment three sessions took place: One session *without* head tracking, another one *with* head tracking, and a third session where the listener itself was placed in the anechoic chamber to localize the 30 sound events.

Results

Without head tracking a lot of front-back-inversions occurred (fig. 4.10), comparable to the results of the previous experiment. Here, even a stronger bias existed towards

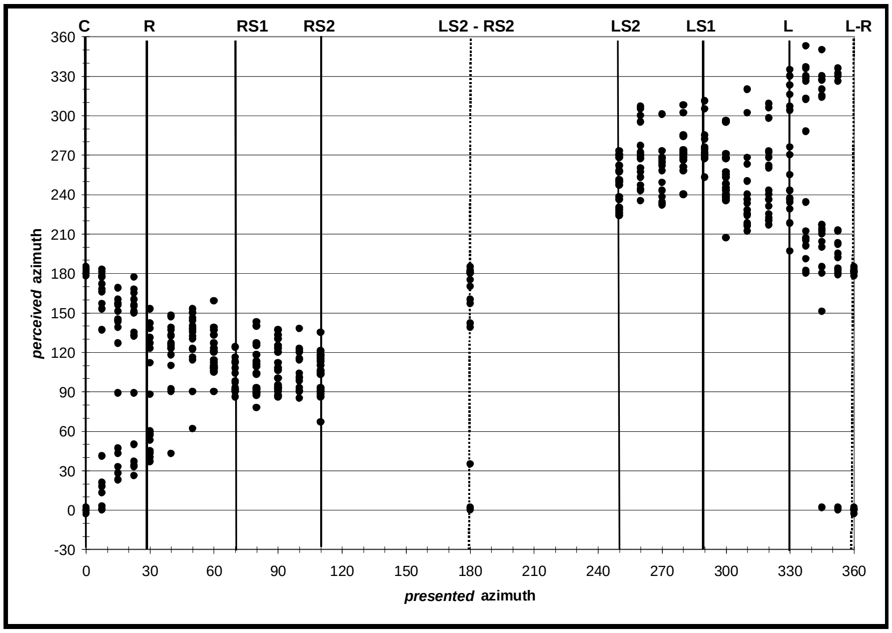

Fig. 4.10: Localization using a fixed dummy head without head tracking under unnatural listening conditions (anechoic room). The localization is bad — even worse than under "fixed" studio conditions.

inversions, depicted as bigger bifurcations pointing away from the "ideal diagonal". Also, the scattering increased, especially in the range of the frontal speaker, compared to natural conditions. As expected, the *fixed* dummy head did not allow a reliable localization under unnatural conditions.

Enabling the head tracking dramatically improved localization. Front-back-inversions occurred only sparsely, and the spreads diminished (fig. 4.11). However, these spreads still exceeded the results noticeable the natural case.

In the third session, i. e., listening via "one's own ears", the localization improved even further. This contrasts with the corresponding results of the experiment under natural conditions, where no improvements were observed. This shows the spread was further reduced, and the front-back-reversals almost disappeared (fig.¡ 4.12).

4.3 Vertical Head Movements (Tilting)

The previous section illustrated the strong impact of head rotations on localization in the horizontal plane. Under both natural and unnatural conditions, the accuracy of localization hardly differed between using the listener's HRTFs or the dummy head's HRTFs if the listener was allowed to move his head.

However, some of the subjects perceived slightly *elevated* auditory events when listening via the dummy head. Various authors using non-individualized HRTFs independently made similar observations [110, 119, 142]. Some assumed that these

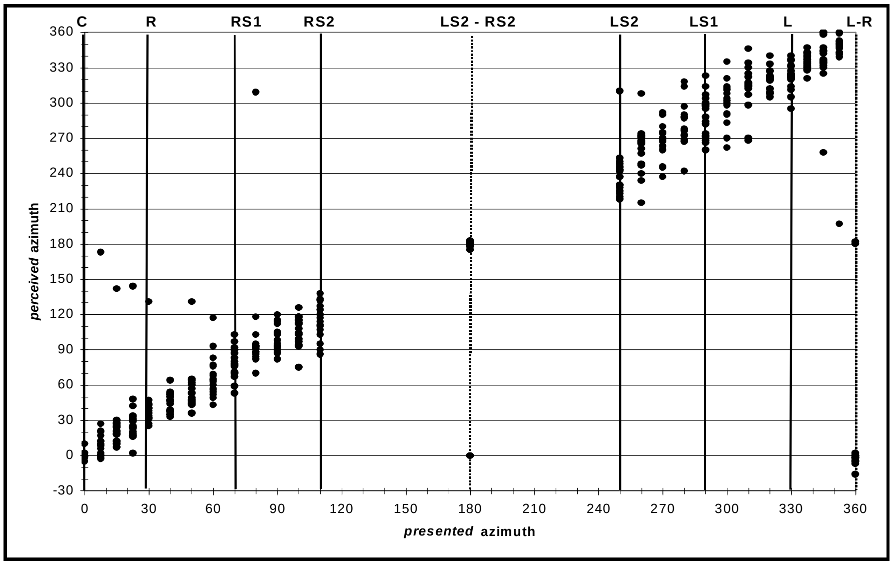

Fig. 4.11: Localization using a "dynamic" dummy head *with* head tracking under unnatural listening conditions (anechoic room). The localization is reliable — comparable to the "dynamic" studio condition.

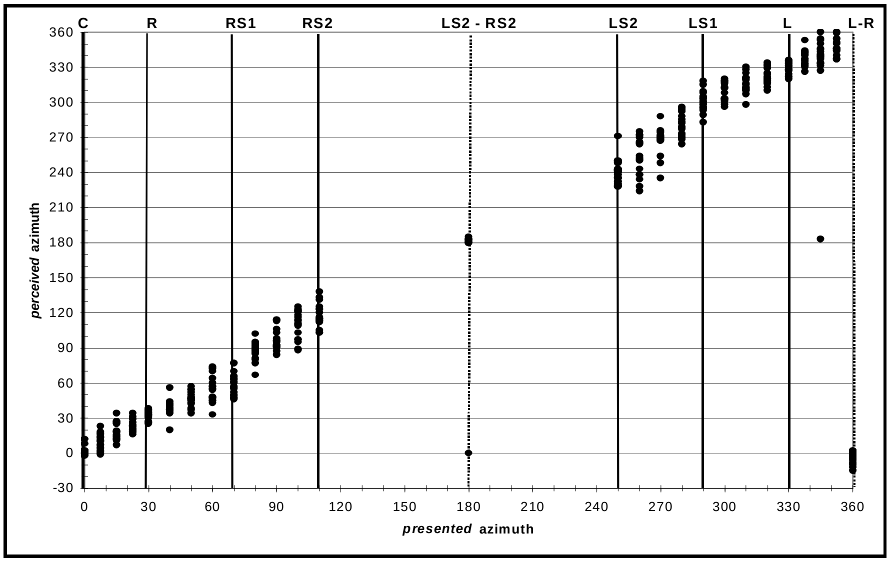

Fig. 4.12: Localization with the listeners being in the anechoic room (unnatural listening conditions) using their "own ears". Localization is fairly reliable, almost identical with studio conditions.

elevations stem from the difference of listener and dummy head HRTFs. These discrepancies in HRTF result in a different HF-pattern and that in turn could lead to an elevation [2, 56].

The lack of "vertical head tracking", i. e., no tracking of the vertical head movements, may also cause unwanted elevations. Therefore, an experiment was set up to investigate the impact of additional *vertical* head movements (tipping) on localization.

Nine loudspeakers with different elevation served as sound sources in and outside the median plane. They were positioned in a vertical arc in the median plane of the dummy head. The angular distance between the loudspeakers was 30° and the elevations ranged from -30° to 210° (loudspeaker positions: -30°, 0°, 30°, 60°, 90°, 120°, 150°, 180° and 210°). Further two loudspeakers were placed outside the median plane, one at 45° azimuthal angle, and the other one at 135°. Both were elevated at 45° vertical angle.

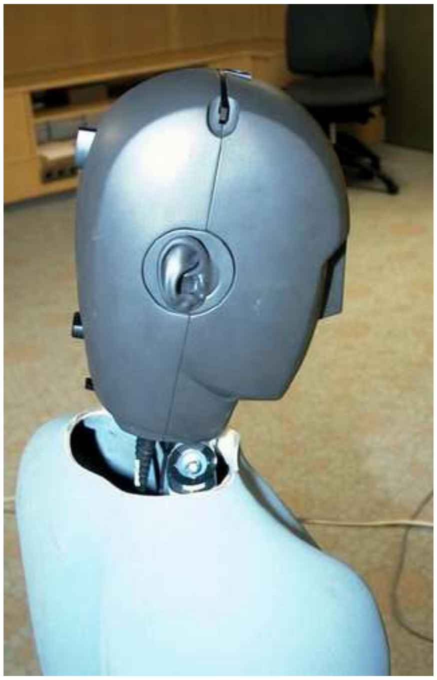

Fig. 4.13: Dummy head with an additional hinge for the vertical movements (tipping) mounted on a torso.

Since it was *mechanically* impossible to control the horizontal *and* the vertical orientation of the dummy head at the same time, a "trick" was used to perform these two rotations independently: An "*acoustical clone*" of the whole experimental situation was generated electronically by using the *BRS-System* (described in section 5.2). This auralization method did not run into any mechanical limits and permitted an *exact reproduction* of the desired experimental setup. The BRS-System was capable of controlling the *rotation* within a range between -42° and 42° azimuthal angle and the *tipping* in a range of -20° to 15° elevation angle.

This localization experiment consisted of two parts: In the first part, the additional tilting movement was disabled, thus allowing only horizontal rotations. In the second run,the additional degree of freedom was taken into account by evaluating the particular tipping data of the head tracker.

All in all, 20 source positions were tested (each twice) with 21 subjects taking part in this experiment. Again, the male voice from the SQAM-CD (track 50) served as a signal. The subject's task was to indicate both horizontal *and* vertical position of the perceived auditory event, i. e., azimuth and elevation angle, by marking the corresponding positions on a paper sheet. Full circles (with the head being the center) denoted the horizontal or the median plane, respectively [121]. All subjects had a general difficulty in correctly marking the perceived position on the sheet [79].

Results - Median Plane

Figure 4.14 displays the localization performance *without* additional tilting movement and with only the head tracking of (horizontal) rotations being enabled. All elevations for every subject are comprised in the plot. The dispersion seems to be fairly large (about ± 30°), except for an elevation angle of 90° in case of a sound source straight above the head. Two cluster points stand out: Either the source was perceived correctly, or it was perceived inside the head (denoted as an elevation angle of 0°). All in all, vertical localization did not seem to be very reliable.

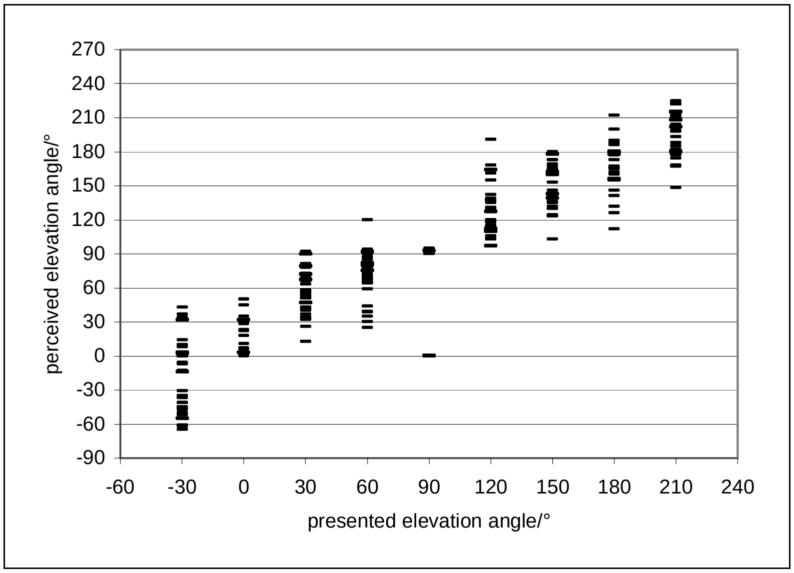

Fig. 4.14: Distribution of hearing event elevations over all subjects without head tracking the tilting movement. The spread is fairly large. Two cluster points exist only for sources directly above the head. Either the sources were perceived correctly (90° elevation angle) or inside the head (denoted as an 0°-elevation).

What would happen if the vertical component of head movements (tilting) were tracked additionally? Would localization improve drastically as was the case when switching from a fixed dummy head to a “rotational” dummy head in the experiments previously described? Or would localization performance remain more or less the same as predicted by Wallach’s theory (see 3.2.2)?

In fact, the latter seemed to be the case, as depicted in figure 4.15. The localization did not improve at all compared to the previous run, in which only the rotational movements had been tracked. While allowing head movements in the horizontal plane has an enormous impact on localization with respect to horizontal sound sources, the same is not true for the vertical positions of sound sources. The spread remains more or less identical. For the 90° elevation angle even the opposite seems to be true: A spread is noticeable, not merely two cluster points.

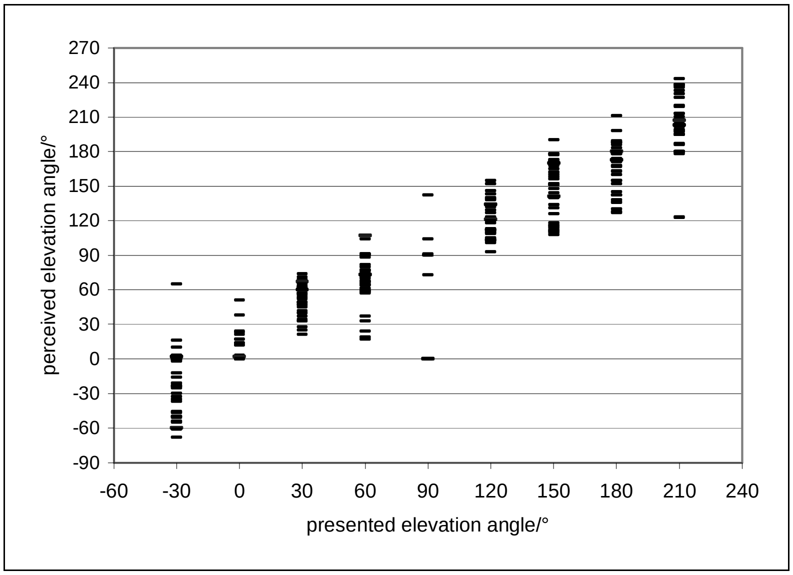

Fig. 4.15: Distribution of hearing event elevations over all subjects *with* head tracking the tilting movement. The results are not very much different from the “non-tipping” case.

The median values of both sessions are plotted in figure 4.16 to illustrate the strength of the additional vertical head movement’s influence in detail. Only a small improvement compared to the first run (rotation only) occurs in the region between -30° and 60°. Nevertheless, the loudspeakers at 0°, 30° and 60° seem to be perceived at an unnatural elevation. This situation is explained in detail in figures 4.17 and 4.18.

An elevation is apparent with respect to all frontal sources (fig. 4.17). In the case of two loudspeakers (-30° and 0°), the impact of additional vertical head movements on localization is stronger than for the rest, although the interquartiles do not confirm this finding. The interquartiles are nearly constant in all positions. This allows the conclusion that the additional tilting movement does not improve the localization of

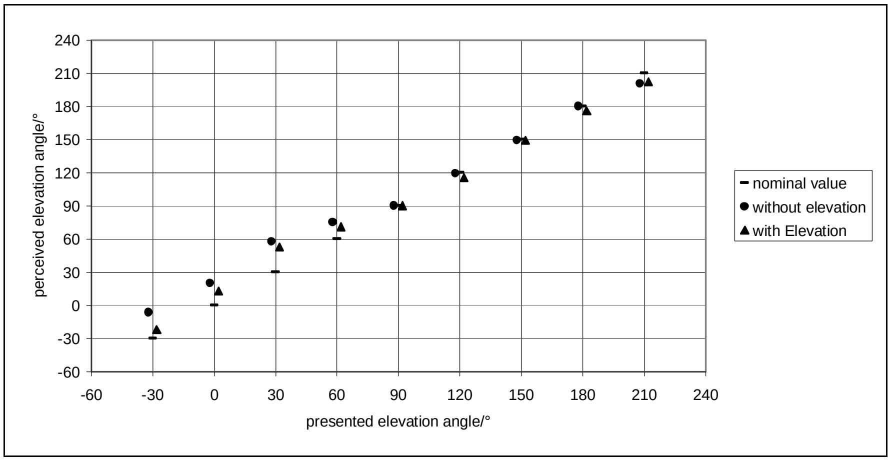

Fig. 4.16: Median values of auditory event elevations in the entire region -30° . . . 210°. The additional tilting movement does not aid very much in localization. Its influence seems to be slightly stronger in the frontal hemisphere (-30° . . . 90°) than in the rear hemisphere.

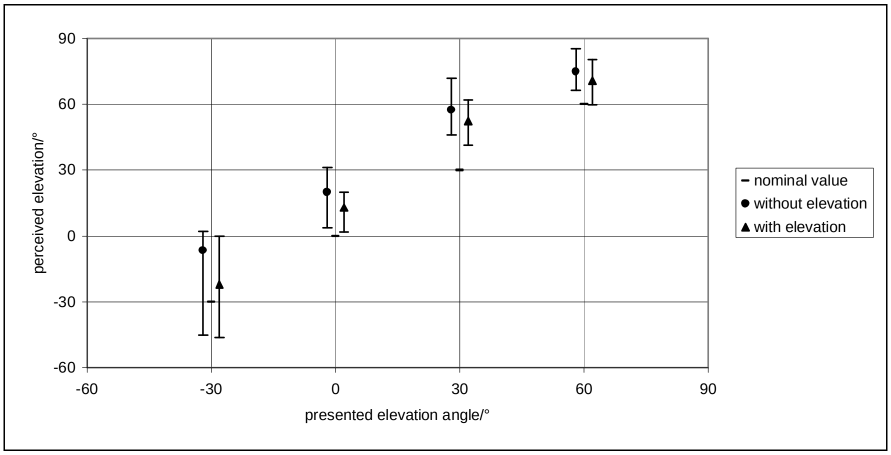

Fig. 4.17: Median and interquartiles of *frontal* auditory event elevations (-30° . . . 60°). The additional tilting movement reduces the perceived elevations.

frontal sources.

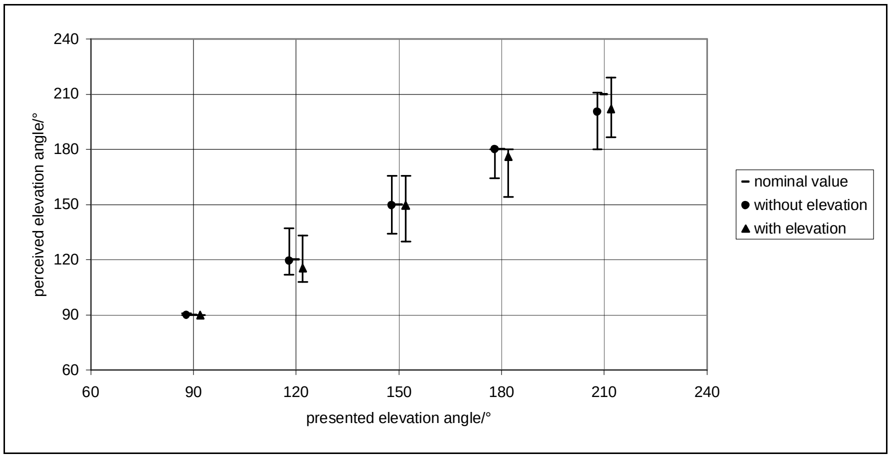

Fig. 4.18: Median and interquartiles of *rear* auditory event elevations (90° ... 210°). The additional tilting movement reduces the perceived elevation in the direction of the sound source's nominal elevation.

For rear sound sources (fig. 4.18) a reliable localization performance is indicated by the median values. In general, almost all rear loudspeakers were perceived correctly even without tilting being allowed. Only the sound source at 210° is perceived slightly elevated[4]. As fas as frontal sources are concerned, taking the tilting into account does *not* improve localization.

Results - Off-Median Plane

In principle, the same applies to the two loudspeakers at 45° and 135° azimuthal angle (outside the median plane). Again, allowing additional tilting does not improve localization (fig. 4.19). In case of the frontal speaker the scattering is even smaller without tipping-movements.

Even when repeating the same listening test with the standard surround sound loudspeaker setup [63], where additional loudspeakers are positioned 10° above and below the nominal (horizontal) loudspeakers, the result remains the same: The localization does not improve when allowing vertical head movements [66]. Nevertheless, it deserves to be emphasized that most of the subjects perceived even the real loudspeaker slightly elevated, although they listened with their own ears.

However, it was remarkable to discover that vertical head movements seem to influence *horizontal* localization. Figure 4.20 displays the distribution of the perceived positions. Here, a tilting movement seems to slightly influence *horizontal* localization.

---

[4] Elevation of rear sources: Auditory events in the rear being perceived elevated are characterized by a *smaller* elevation angle than the nominal angle of the sound source!

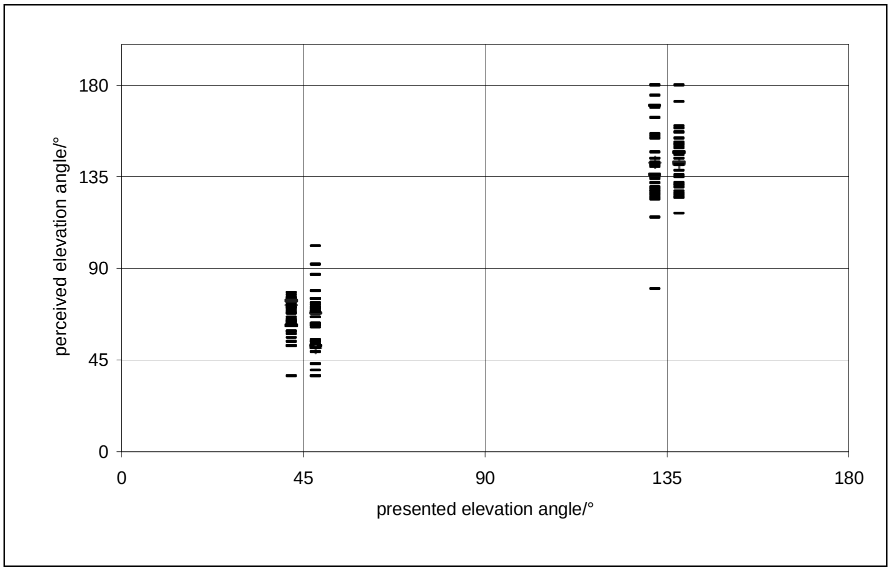

Fig. 4.19: Distribution of perceived *elevation* angles of elevated sound sources outside the median plane (left without tilting / right with tilting)

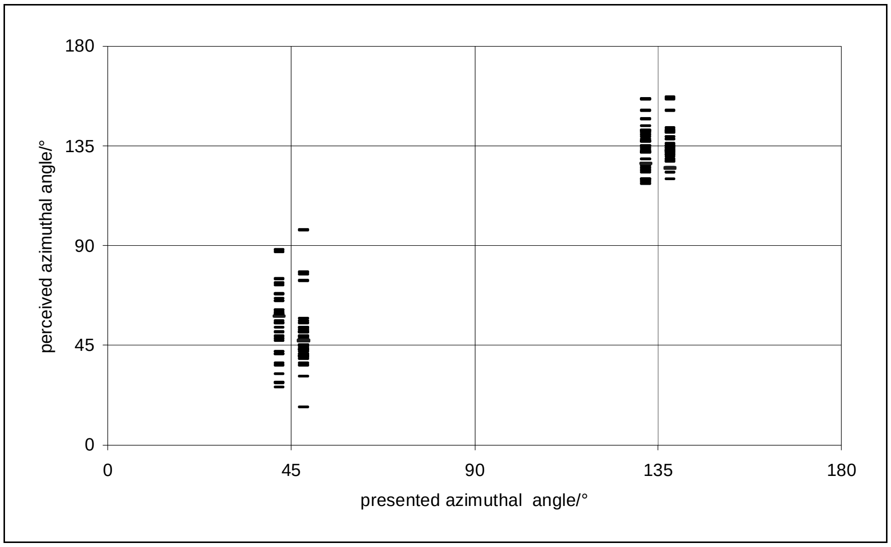

Fig. 4.20: Distribution of perceived *azimuthal* angles of elevated sound sources outside the median plane (left without tilting / right with tilting)

4.4 Importance of HRTF Cues

The experiments demonstrate that the anechoic room, that cuts off all reflections, does not seriously alter localization performance as long as head tracking is enabled.

This raises the question of the consequences of all "normal" spectral cues of the HRTF (especially the pinnae) being absent. Furthermore: How do different HRTFs (dummy heads) contribute to localization performance? The following two sections are dedicated to exploiting the importance of "individual"[5] binaural cues.

4.4.1 Absence of HRTF Cues

In this experiment, an *artificial* listening situation was produced [31, 66, 78]. Instead of using the dummy head, as had been done in all previous experiments, the *Schoeps KFM 6*, a *sphere microphone* according to Theile [126] was used. This device can substitute a kind of an abstract minimalistic model of the head. Any individual features, especially the pinnae, are absent. The experiment took place in the same IRT studio as in the first experiment.

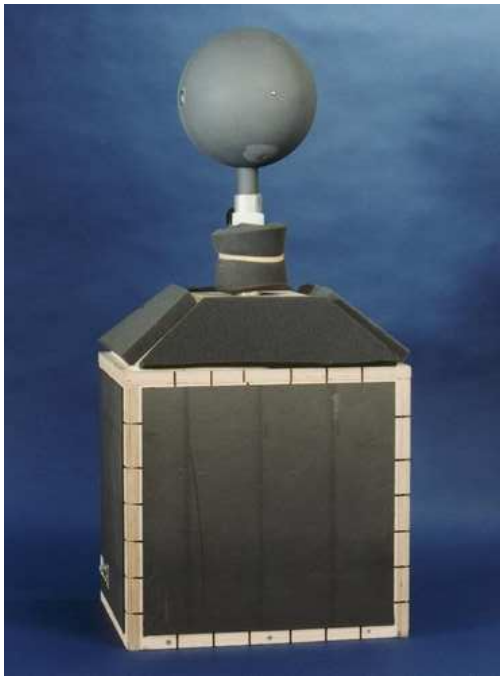

Fig. 4.21: Sphere microphone mounted on the motor-unit.

This time only two sessions were carried out: One without head tracking and the other one with head tracking enabled. A total of 14 subjects took part in the exper-

---

[5] Here, the term "individual" does not refer to individualized HRTFs as for example in [89] and [138], but to the spectral pinnae cues.

iment. The subjects had to indicate both the perceived azimuth and the elevation (see also section 4.3).

Results

As depicted in figure 4.22, a poor localization performance can be observed in the static case. The numerous front-back-inversions are visible as distinct bifurcations with a large spreading.

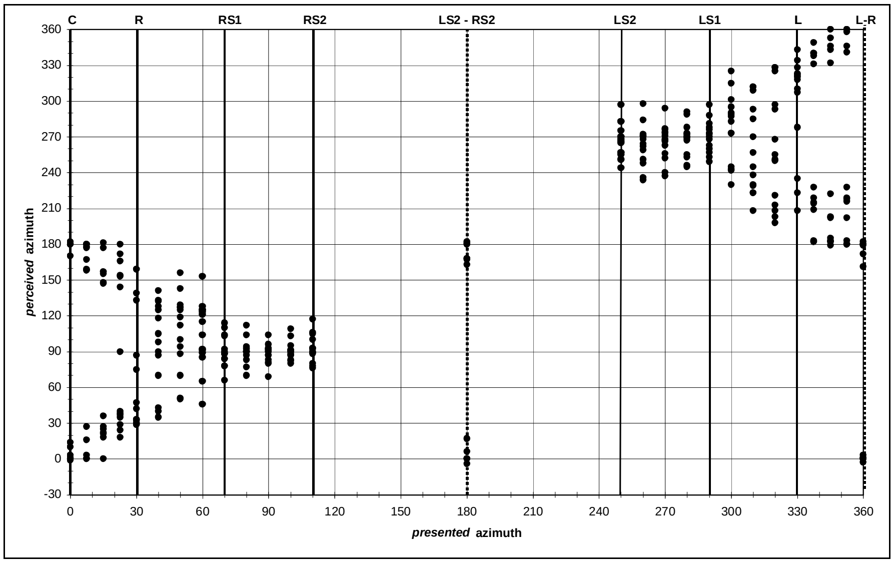

Fig. 4.22: Localization when using a fixed sphere microphone without head tracking in the studio. A reliable localization is not possible. Numerous front- back-inversions occur.

With head tracking enabled, however, most of these front-back-reversals disappeared with only a few occurring infrequently (fig. 4.23). And although all typical spectral cues considered to be important in the static case were missing, localization without front-back-reversals in the horizontal plane was possible when tracking the head movements.

We did not expect using a sphere microphone to result in a similar localization accuracy as a dummy head with head tracking allowed. In case of the sphere microphone scattering was slightly larger and a few strong deviations occurred. Again, these results stress once more the importance of dynamic localization cues.

However, most of the subjects perceived elevated auditory events when the sphere microphone was used (see fig. 4.24). These elevations occurred to a lesser extent when using a dummy head. This effect was particularly pronounced in the region between the frontal loudspeakers if the sphere microphone was fixed and the head tracking disabled.

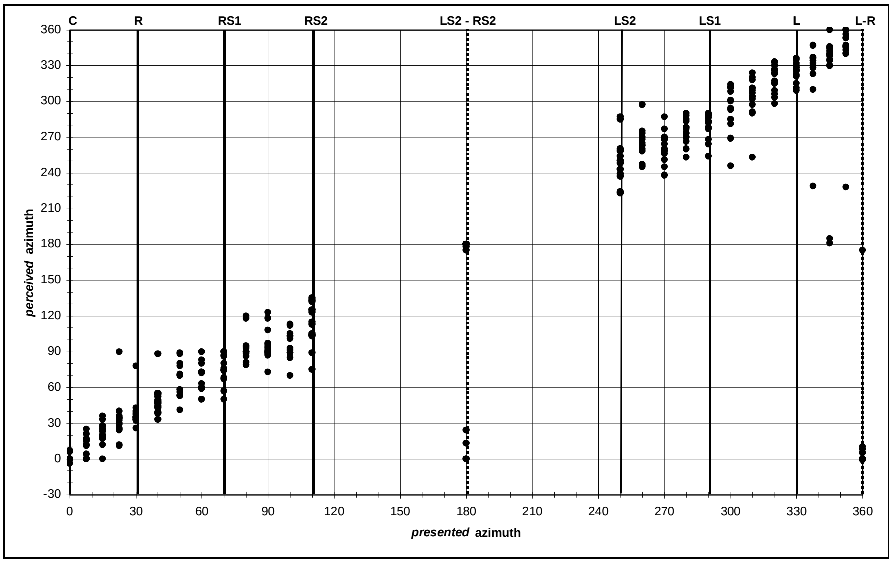

Fig. 4.23: Localization when using a head tracked sphere microphone in the studio. In contrast to the fixed condition, localization is very reliable. Although there are no HRTF-cues, only a few front-back-inversions occur.

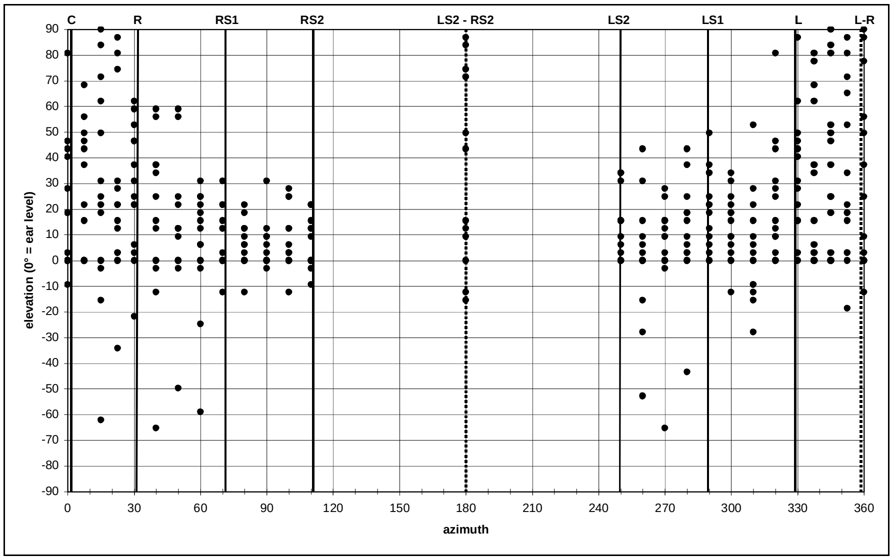

Fig. 4.24: When using the sphere microphone subjects perceived unwanted elevations in both conditions (static and "dynamic"). Here: Static condition (= *fixed* sphere microphone)

4.4.2 Influence of different HRTFs

When determining the azimuth of (horizontal) sound sources, the dynamic cues caused by head rotations dominate the cues by HRTF and pinnae. This was proven in the last section. Furthermore, additional vertical head movements (tipping) do not reduce unwanted elevations (see section 4.3).

In preliminary tests it was attempted to reduce elevations by boosting or attenuating certain frequency bands (see also [2, 8, 11]). It turned out, however, to be impossible to control the elevation of the auditory event by applying such an equalization technique [66].

This section explores the influence of different HRTFs on elevation. For this purpose, this experiment uses six dummy heads which are common in the field of acoustics research. The following six dummy-head systems (K1 ...K6) were used, partly in combination with a replica of a torso:

K1: KU 100 (Neumann)

K2: Manikin MK1 (with torso, Neutrik-Cortex)

K3: KU 81 (Neumann)

K4: HMS III (with torso, HEAD acoustics)

K5: HUGO (with torso, Institut für Technische Akustik, RWTH Aachen)

K6: KU 100 with Cortex-Torso

To provide for high-fidelity reproduction, diffuse-field equalization was again chosen as interface between the dummy heads and the headphones. Therefore, the *diffuse-field transmission coefficient* of the headphone (*STAX SR Lambda*) and of the dummy head systems have already been determined and were equalized accordingly in order to obtain a plain transmission coefficient [106].

This exact alignment of the diffuse-field transmission coefficient was necessary to eliminate or at least reduce possible artifacts, such as unwanted elevations. These elevations stem from a difference between diffuse-field transmission coefficients of the dummy head and that of the headphone, thus from a deviation of the optimal one-to-one transmission.

Since the different torso replica did not allow a mechanical, motor-driven control of the dummy head(s), the BRS-System was used again (as described in the previous section).

Experimental Setup

This experiment was designed to investigate the influence on localization of HRTFs and the vertical position of the sound source. Influences on the azimuthal component were disregarded. There are only small differences in horizontal localization regardless of whether one uses a dummy head with head tracking or natural hearing. This was proven in section 4.2.

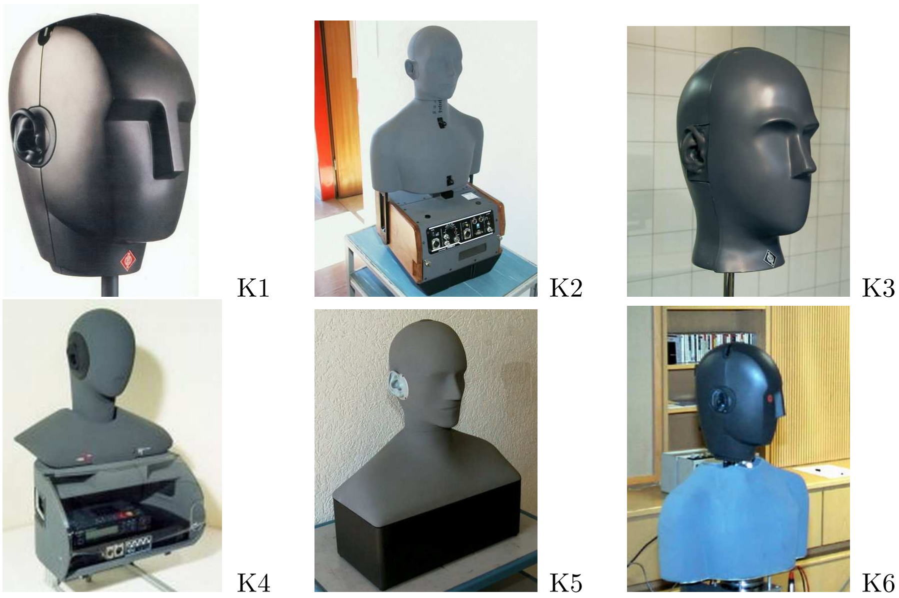

Fig. 4.25: Different dummy heads were used to investigate their individual influence on localization (elevation) and tonal impression. The dummy heads are (from upper left to bottom right): K1 = KU100 (Neumann), K2 = Manikin MK1 (with torso, Neutrik-Cortex), K3 = KU 81 (Neumann), K4 = HMS III (with torso, HEAD acoustics), K5 = HUGO (with torso, Institut für Technische Akustik, RWTH Aachen), K6 = KU 100 with Cortex-Torso

The same loudspeakers as in the previous experiments (Klein & Hummel O108/TV) served as sound sources. They were placed in the "studio" in the standard surround sound loudspeaker setup with only the three frontal speakers being used. The loudspeaker were mounted in a height of 130 cm. The distance between loudspeakers and dummy head (sweet spot) was 3 m.

Real sources and phantom-sources were used. The phantom sources were created by weighting the volume of two adjacent loudspeakers at the ratios 2:1, 1:1 or 1:2, respectively. Thus, the nominal positions of the real sources were 0° and ±30° azimuthal angle, and for the phantom sources ±7.5°, ±15° and ±22.5° azimuthal angle.

To eliminate any optical cues and to aid in "marking the apparent position" of an auditory event, a rectangular, sound transparent linen was placed between listener and loudspeakers (see fig. 4.26). This linen was divided into squares, each 15 cm × 15 cm, by means of spanned, black colored threads. The distance between two squares was about 6° (measured from the sweet spot).

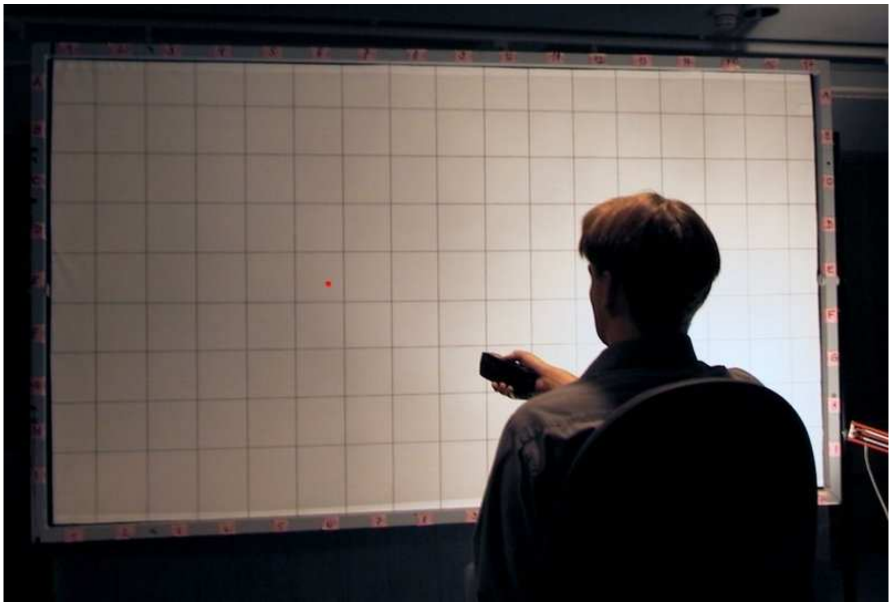

Fig. 4.26: A sound transparent linen with a grid serving as a graphical user interface. The subject used a laser to mark the direction of the perceived auditory event.

Each square was identified by means of two coordinates — similar to the nomenclature of a chess-board: The columns were assigned numbers from 1 to 17 starting with the leftmost, and the rows were given letters, beginning with *A* for the top row and ending with *I* at the bottom row. To mark the perceived position of the auditory event, the listener had to point at the corresponding square using a laser pointer. The corresponding letter-number-combination was taken down designated, e. g., "*F*3".

18 subjects took part in the experiment. A pink noise pulse train consisting of five pulses with a length of 2 ms alternating with pauses of the same duration served

as a test signal. The noise signal was chosen because its high-frequency components are essential for localization of elevated sources [2, 11].

The subjects had to localize the sources using the BRS-System in the first session. Therefore, the whole experimental setup was "scanned" using the various dummy heads (see also section 5.2). In the second session the subjects listened with "their own ears" to the real loudspeakers in the studio.

Results

Figure 4.27 depicts the differences in elevation between real and virtual sound sources. Its display integrates the mean value of 18 subjects and the corresponding 95%-confidence interval. The figure shows that there is no significant difference in elevation between the various dummy heads. On average, all dummy head systems showed a small elevation of about 7°.

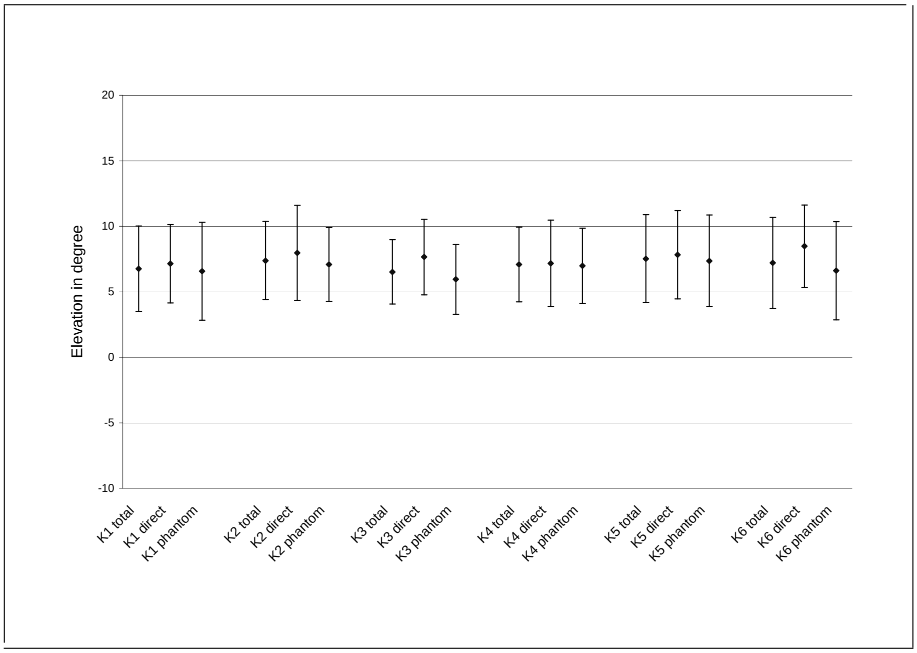

Fig. 4.27: Differences in elevation between real and virtual sound sources using six different dummy head systems K1 . . . K6 (as described previously: K1 = KU 100, K2 = Manikin MK1 with Cortex-torso, K3 = KU 81, K4 = HMS with HEAD acoustic-torso, K5 = HUGO with torso and K6 = KU 100 with Cortex-torso). No significant differences are observed between the various dummy head systems.

Nevertheless, a few elevations occurred which clearly deviate from the mean elevation. No significant differences are noticeable between various dummy heads. However, no significant difference could be established between real sources and phantom sources. It was thus possible to average out all positions per dummy head.

4.5 Discussion

The described experiments have shown the importance of head movements, in particular head rotations, on localization. Consequently, head movements have to be taken into account if a listening situation is to be reproduced in an authentic or convincing way. As a first example of applying these results in practice we will now discuss the acceptance of dummy head stereophony. A few comments on the general question of whether such reproduction has to be authentic or merely convincing follows. In this context, the effect of better localization on the *perception of music* will be emphasized.

Acceptance of Dummy Head Stereophony

When listening to a "normal" stereophonic recording using headphones, the auditory percept normally resides *inside the listener's head*, somewhere on the aural axis between the ears, and thus makes localization impossible. It is considered to be a *lateralization.* The position of the auditory event is influenced only by the interaural level or time differences, but not by any spectral characteristics [18, 64, 100, 116, 148, 149].

Using a dummy head to record the sound binaurally allows avoiding lateralization and granting the listener the sensation of a true localization. For this purpose, the dummy head is placed at an optimal listening position and remains fixed. However, the dummy head stereophony (binaural recordings) mainly for two reasons did not become widely accepted: The first reason was that headphones had to be used because a presentation with loudspeakers (a so-called transauralization) did not really work very well [5, 11, 24, 40, 39, 48, 85, 123]. Secondly, mainly for sound sources in the median plane either *front-back-inversions* occurred or the aural events were perceived *inside* the head [42, 87, 89, 138].

If there had been a possibility of using a "dynamic" dummy head system that would have been capable of accounting for the listener's head movements (as used for the IRT-listening tests), the situation of binaural stereophony might have been different: Simply by rotating the head a little bit to the sides, a natural localization would result effortlessly and almost automatically, and the sensation of "actually being there" would immediately be felt.

Two issues remain to be considered: The first one relates to the "*tonal*" perception of the sound. In general, it can be assumed that the listener's and dummy head's HRTFs are different. The tonal perception is thus likely to be different when listening with the dummy head's HRTFs. However, the listener might prefer the brilliant localization of such a system and therefore accept some tonal differences.

The other issue is the realization of such a dynamic dummy head system. The problem is that the dummy head can only follow the head movements of a *single* listener. Whenever the listener moves his head, the dummy head will turn accordingly and a change in the binaural signals will result. Therefore, all head movements have to take place at the very moment of perception and cannot be recorded. This in turn limits the possible range of such an application: It only could be used by a *single* listener at the specific time the sound event (e. g., a concert) takes place, and a "dynamic" recording is *not* possible.

To sum up, on the one hand a *dynamic* dummy head system could greatly increase the acceptance of binaural perception, but on the other hand, a reasonable realization of such a system is not possible, especially not as a "recording tool".

Artistic Perception vs. Localization

The previous section discussed the importance of a reliable binaural localization for the general acceptance of a (dynamic) dummy head system. This part will address the question if a reliable or even a natural localization really is *necessary* (if not indispensable) for the *artistic perception* of music.

Various authors investigated the influence of different electro-acoustic transmission systems on localization and perception [3, 4, 10, 72, 97, 77]. Experimental parameters were, for example, the number of transmission channels (from mono to multichannel surround sound), the way of reproduction (headphones or loudspeakers), the position of the loudspeakers in the room, the size and type of room used, etc..

All these investigations focused on "basic" aspects of perception, e. g., localization, timbre, spaciousness. However, none of them investigated the possible impacts on the *artistic* component of music. For example, the localization using a single-channel transmission, e. g., a monophonic portable radio, is certainly by no means comparable to that of a surround sound recording, let alone the actual listening to a concert. But to what degree does that influence the *artistic appreciation* of music?

The analogy in the optical case is the presentation of a movie either by a small black-and-white television or in a cinema. Surely, the overall impression of a movie shown in a cinema is probably "better". A small b/w-television cannot capture one's attention in quite the same way nor produce the same emotional sensations.

Artistic perception seems to be similar in the acoustic world. Listening to the "favorite" music will activate the same emotions, regardless of, say, the localization of the individual instruments being excellent or not. The longtime existence of stereophonic recordings seems to support this hypothesis. Since stereophony has an inherently limited capacity to produce a convincing acoustic replica of the original situation, it makes reliable localization impossible. Nevertheless, the artistic appreciation seems to be mainly unaffected.

Another question is whether we accept a transmission system to deliver an *authentic* or merely a *convincing* and *plausible* reproduction. Strictly speaking, using not the listener's HRTFs, but "merely" the dummy head's HRTFs, cannot lead to an authentic reproduction. However, in most cases when using the *dynamic* dummy head system these differences were not perceived to be disturbing. Here, a *convincing* (albeit not totally authentic), reproduction of the original acoustic situation was sufficient for a reliable localization.

Chapter 5

Auralization Methods and the Entirety of Localization Cues

When summarizing the results of the previous chapters, two main issues stand out: Firstly, the reproduction of a listening situation requires to consider head movements. Secondly, it is always necessary to take the entirety of localization cues into account.

We describe a new auralization method enriched by respecting these two main findings: The data-based auralization, called Binaural Room Scanning (BRS). In a cooperation between Studer Professional Audio AG, Zurich, Switzerland, and IRT the BRS Processor was developed. Our localization experiments at IRT contributed to the research project and were the basis for this development.

5.1 Auralization Methods

The aim of an *auralization* is to produce a convincing *virtual acoustic environment* which results in a spatial 3-dimensional perception. Commonly, headphones are used to reproduce such an virtual acoustic environment. Although theoretically a pair of loudspeakers (or maybe even more loudspeakers in another setup than the common stereo setup) may also be capable of performing this task, certain problems arise when trying to put this into practice, namely the exact cross-talk compensation [5, 25, 40, 39, 147]. Such an auralization using loudspeakers for reproduction is called *transauralization*.

Generally, each auralization is based on the following assumption[1]: When confronting a subject with the *identical*[2] binaural signals, that would have also been perceived at the time of recording, the same acoustic perception will result. Hence, the same hearing sensation will occur if either the subject is actually present at the recording location or the same acoustic signals are presented via headphones.

---

[1] However, this assumption is not exactly true, as for example cognitive effects may occur, but nevertheless this assumption can be seen as a working hypothesis.

[2] Here the term *identical* does not necessarily means an identity between left and right signal.

A simple system consisting of a dummy head and a headphone can be considered a basic auralization system. Such system with a *fixed* dummy head "captures" the original sound field at the position of the dummy head, and reproduces the "binaural representation" of that particular sound field by means of headphones. Additionally, the signals delivered by the dummy head can be recorded and therefore be reproduced.

However, head movements are an important component in localization as proven in the previous chapters. Therefore, localizing with a *fixed* (dummy) head is an *unnatural* listening situation and thus often results in localization errors. But if the listeners head movements are taken into account such a listening situation resembles more natural hearing, and often localization errors are reduced.

The following sections introduce different methods for creating and reproducing a virtual acoustic environment. These auralization methods take head movements into account.

5.1.1 Model Based Auralization Methods

The optical analogue to a model based auralization method is the *construction* of a "picture" by using geometrical elements. In acoustics a model based auralization method *synthesizes* the virtual acoustic environment from scratch.

In principle, the synthesis is done in four basic steps: The first step consists of *recreating the sound field* at the desired listening position, e. g., the *sweet spot* by analyzing the characteristic acoustic properties of the room. For each sound source the direct sound, a certain number of reflections from surfaces like walls, the floor or the ceiling, and the room-modes at the listening position have to be calculated [5]. The more reflections are calculated the better and more convincing the overall result gets, but also the higher the computational requirements are.

In a second step, the *binaural impulse response* will be determined using a database of previously measured and stored HRTFs. Thus, this step models the properties of the human outer ear (pinnae) and head.

The third step consists of *convolving* the source signals with the calculated binaural impulse response. It simulates the acoustic path between the sound source and the entrance to the ear-channel (concha). This convolution has to be carried out for all of the various sound sources and each reflection (mirror sources) with the appropriate "directional" HRTF. However, depending on the sources and their sound emission characteristics, the number of reflections, the geometry of the room etc., the required computing-power increases rapidly.

Insufficient processing power available today is the reason for adding some "artificial" reverberation instead of calculating a very high number of (discrete) reflections (fig. 5.1). This is one of the limitations of a model based auralization method [5].

The final and fourth step consists of *reproducing* the calculated binaural signals through headphones.

In order to create a *realistic* and *convincing* virtual acoustic environment, it is essential to take the listener's head movements into account, as has been shown in the previous chapter. In turn, this requires that all aforementioned steps be repeated at least 60 times per second to produce a convincing virtual acoustic image [110, 136].

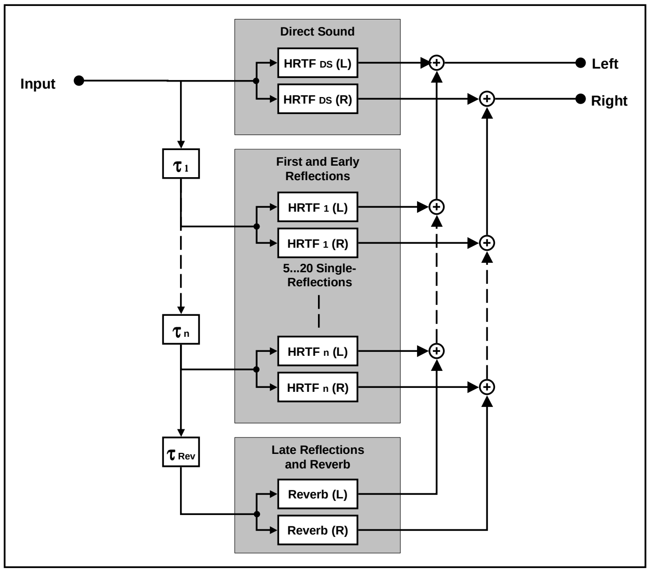

Fig. 5.1: A conventional model-based auralization system. Separate convolutions have to be carried out for the *direct sound*, *each (first) reflection* and the *reverb*. These calculations have to be executed for every new orientation of the listener's head.

Thus a new set of binaural impulse responses has to be calculated for every new orientation of the listener's head. Therefore, it seems to be rather impossible to synthesize a non–virtual, real acoustic environment (e. g., the 3/2 loudspeaker surround sound setup of the IRT-listening tests in part 4) with such accuracy that listeners would not notice any difference between the original and the virtual environment.

The processing-power needed for such a dynamic real-time production is currently beyond today's possibilities. And every simplification, e. g., shorter impulse responses or fewer reflections, would make the impression less convincing.

Nevertheless, a great advantage of model based auralization systems is their absolute control of sources, room and listener position. At least in principle, everything can be modelled and accurately controlled – be it "jumping" sound sources, an artificial (unnatural) reflection of the walls, or the change of the listener's position inside the room.

5.1.2 Data-Based Auralization Methods

As stated in the previous section, the processing-power needed for the *real-time*, dynamic[3] reproduction of a convincing acoustic environment is currently beyond today's capabilities. This is mainly due to the vast number of reflections required to be imitated in order to reach authenticity. To achieve such an authentic auralization, a different kind of auralization system is therefore necessary.

---

[3] dynamic = taking head movements into account

A *data-based auralization system* is capable of reproducing *dynamically* such an authentic acoustic environment in real-time.

The main idea of a *data-based* auralization system is to "*capture*" the acoustics of a real listening room by measuring the relevant data, and storing these *measured* room data in a database. This measurement of the room-acoustics (an acoustic "snap-shot") produces an *acoustic clone*" of the original environment. No model-based synthesis technique can achieve a comparable accuracy. The data-base technique carries: the disadvantage of having to be auralized for each situation (e. g., each position of the sound source) and thus requiring a *physical rearrangement* of the acoustic setup.

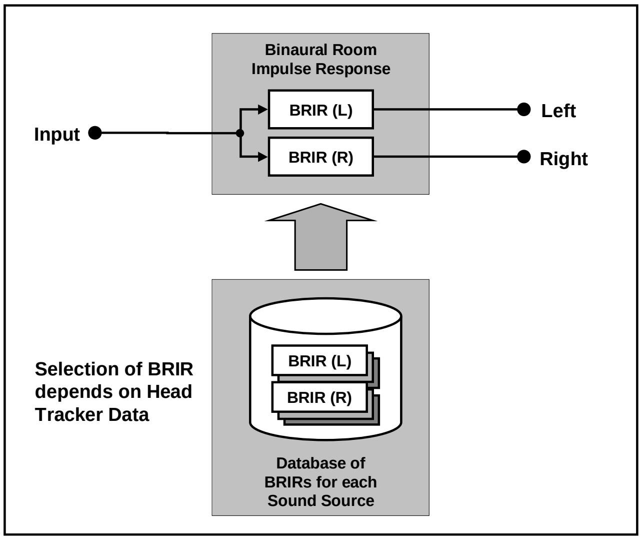
Binaural Room Impulse Response
Input
BRIR (L)
BRIR (R)
Left
Right
Selection of BRIR depends on Head Tracker Data
BRIR (L)
BRIR (R)
Database of BRIRs for each Sound Source

Fig. 5.2: A data-based auralization system. The Binaural Room Impulse Response (BRIR) contains the environment's *complete* acoustic information (= *entirety of localization cues*). As a consequence, a single convolution has to be carried out. Depending on the orientation of the listener's head the corresponding set of BRIRs is selected.

If one compares the model-based auralization methods with the construction of a picture, the analogy to this measurement- and data-based auralization technique is a *photography*. Thus, in order to obtain a different picture, either the camera or the objects would have to be moved physically.

Any data-based auralization method basically requires three steps: First, the room's acoustics at the listening position has to be *captured binaurally*. A loudspeaker at a fixed position serves as a sound source and the impulse response of the room is measured using the dummy head and the MLS-technique[4].

This measurement is taken for different orientations of the dummy head. Within a certain angular range covered, e. g., from -90° to 90°, the dummy head is rotated step

---

[4] MLS = *maximum length sequence*

by step and afterwards a measurement is taken. A database then stores each measured binaural room impulse response. A *complete set* of binaural impulse responses will be stored for every loudspeaker in a standard 3/2-surround sound loudspeaker setup.

In a second step, each input signal (normally fed to the loudspeakers) is *convolved* with its respective binaural impulse responses stored in the database. Here, the advantage of *measuring* the impulse response becomes evident: The measured impulse response contains *all* relevant (linear) acoustical properties of both the room and the loudspeaker(s).

Instead of calculating a vast amount of reflections to achieve a more or less realistic reproduction (at the expense of high processing-power), an "acoustic photography", i. e., the highest possible realistic image, is used for the convolution. The processing-power needed for a dynamic auralization at an update rate of at least 60 Hz nowadays is already available because for every (virtual) loudspeaker to be auralized only two convolutions are necessary.

The final step consists of delivering the binaural signals to the listener by headphones. It is strongly recommended, however, that both the dummy head (used for the measurement process) and the headphones are diffuse-field equalized, as described in section 4.1.1. For any headphone a diffuse-field equalization can be realized as a preceding filter.

This auralization method results in an *acoustic clone* of the real environment with the only shortcoming of the loudspeakers at *fixed* positions serving as sound sources. The *data-based auralization method* allows capturing and reproducing of all, in reality existing listening situations using loudspeakers. On the other hand, this is one of the "disadvantages" of the data-based approach because *only existing* listening rooms can be auralized.

5.1.3 Hybrid Auralization Methods

The previous two sections introduced two different auralization methods, i. e., the model-based auralization method and the data-based approach. The following remarks are meant to sketch the respective advantages of both systems combined in a hybrid auralization method (see fig. 5.3).

The model-based auralization method has the major advantage of allowing the control of every single parameter. In principle, this permits creating, for example, unnatural virtual environments (such as an environment with only two single first reflections and no reverberation), or environments before actually being built. It suffers, however, from requiring enormous processing-power for the real-time calculation of all parameters necessary to produce a convincing acoustic environment.

This disadvantage can be compensated by taking advantage of a data-based system, i. e., the use of stored binaural impulse responses for calculation. These impulse responses can either be previously *measured* or *rendered* upon the condition that only loudspeakers serve as sound sources, and that the positions of the sources as well as the listening position be fixed.

Without the aural environments being amenable to be reproduced in real time, such systems can be extremely complex. The computed final result can be stored

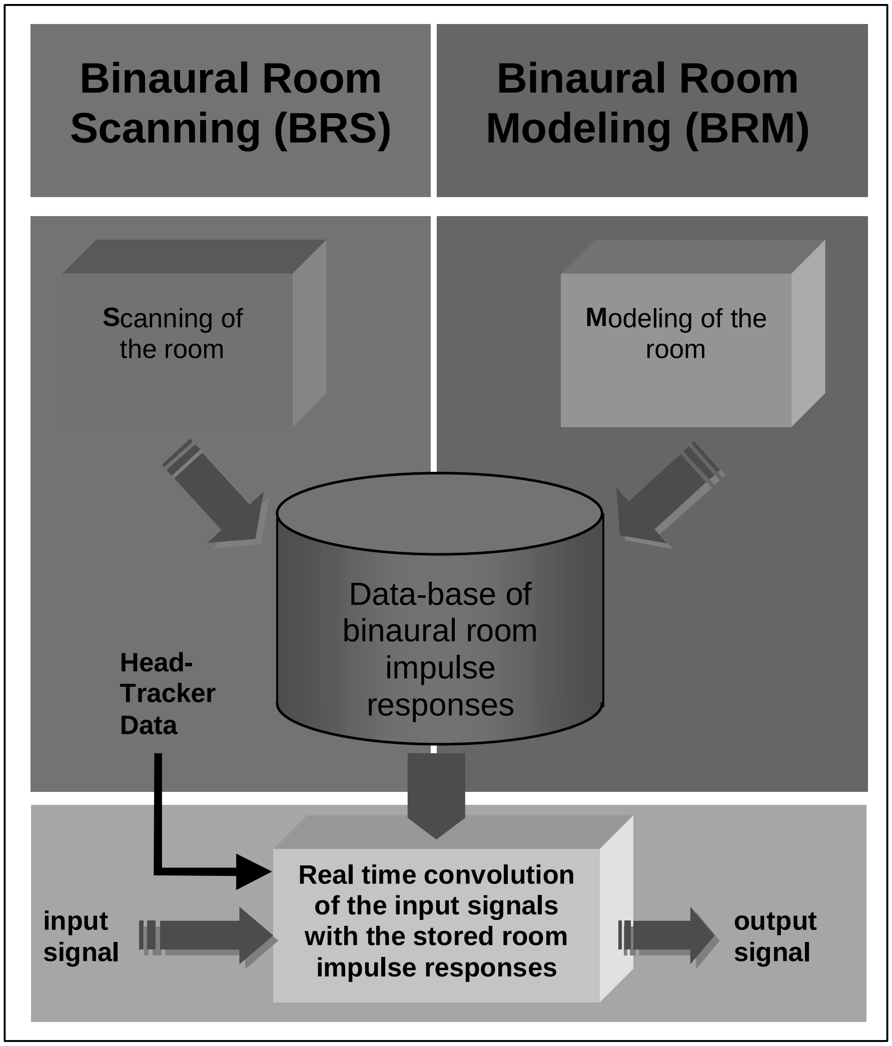
Binaural Room Scanning (BRS)
Binaural Room Modeling (BRM)
Scanning of the room
Modeling of the room
Data-base of binaural room impulse responses
Head-Tracker Data
Real time convolution of the input signals with the stored room impulse responses
input signal
output signal

Fig. 5.3: A hybrid auralization system combines the advantages of both auralization methods. In detail all rooms can be rendered using a flexible model based approached. The computed data are subsequently stored in a database. These previously rendered binaural room impulse responses can be used for the auralization.

as a pair of *calculated* binaural impulse responses in the database. The calculation then has to be repeated for all relevant head orientations resulting in a set of binaural room impulse responses for each source. A previous measurement of the listener's HRTF (used for the computing of the stored binaural room impulse responses) can even improve the auralization's fidelity.

For purposes of auralization the input signals are simply convolved with the respective previously calculated impulse responses (stored in the data-base) depending on the actual orientation of the listener's head.

Pellegrini [98] compares model- and data-based auralization methods. First applications of such a system in practice can be found in [46].

5.2 Binaural Room Scanning

Our localization experiments at IRT were the basis for the development and realization of the Binaural Room Scanning Processor (BRS Processor). In this chapter we will describe the data-based BRS auralization method in more detail including considerations about the system design.

5.2.1 System Design Considerations

Before describing how the data-based auralization concept, namely the BRS Processor, can be put into practice, a few remarks on some specific, underlying psychoacoustics experiments are warranted. These experiments were carried out at IRT and their results were used to ensure both a high fidelity of the reproduced auralized environments as well as the optimal exploitation of processing-power.

Latency, Update-Rate and Spatial Resolution

In natural hearing, we immediately perceive a rotation of the head because of the information from the vestibular organ (as well as by the tactile receptors in the muscles) and the relevant changes in the acoustic signals at the ears.

However, using an auralization system (electro-mechanical or electronic device with head tracker) will cause a delay between the listener's head movements and the transmission of the resulting changes through headphones. This delay, i. e. the system's latency time, is unwanted yet inherent in the system. It can influence the auditory perception and the fidelity of the auralized environment.

Sandvad and Wenzel [110, 136, 137] have previously investigated the influence of the system's total latency time on localization. They concluded that for producing real time auralization the auralization system has to meet the following requirements: The *total latency* time has to be below 91 ms, the *update rate* has to be at least 60 Hz and a *minimal spatial resolution* of about 2° is required.

The experiment at IRT which focused on latency (described in section 4.1.3) produced similar the results, i. e., a *maximal latency time* of 81 ms using an *update rate* of 120 Hz and a *spatial resolution* of 5° (due to the step-motor). The head tracker system itself had a latency time of $t_{headtracker} = 8.3\,\mathrm{ms}$ and an angular resolution of 0.1°.

Since the electronic auralization system (BRS Processor) itself has a shorter latency time ($t_{BRS} < 6\,\mathrm{ms}$), the *system's total latency time* is in the order of magnitude of about $t_{total} = t_{headtracker} + t_{BRS} \simeq 15\,\mathrm{ms}$. Thus does not exceed the upper limit of 85 ms and fulfils the requirements previously mentioned.

However, with a non-mechanical auralization system the spatial resolution can be increased by interpolating between adjacent binaural room impulse responses. Section 5.2.2, and in more detail [59], provide a short overview over this interpolation algorithm used in the BRS Processor.

Partially Dynamic Convolution

In a data-based auralization system, a pair of binaural room impulse responses has to be stored in a database for each relevant orientation of the listener's head. Afterwards, depending on the actual head orientation, the corresponding impulse response is chosen in a table-lookup process and is subsequently convolved with the input signal.

If the listener turns his head while receiving an input signal, this signal's onset needs to be convolved with *a different* binaural room impulse response pair than the binaural room impulse response that is used for the convolution with the later part of the same signal. This effect is particularly pronounced if the environment to be auralized has a long room impulse response.

If only the first part of the binaural room impulse response depends on the orientation of the listener's head, this might influence localization. Fruhmann *et al.* [35, 79] investigated this effect.

They created a "modified" impulse response combining a first part being dependent on head orientation and a second part being independent of head orientation.

The first part (*dynamic part*) was the "normal" and natural (i. e., head orientation dependent) impulse response up to a time t_{dyn}. The second *static part*, i. e., starting from t_{dyn} till the end of the impulse response, always consisted of the same part of the 0°-impulse response regardless of the listener's head actual orientation.

However, because of the systems design (memory available for the data-base), the *maximal length* of a binaural room impulse response had to be limited. This, of course, limits the number of listening rooms fit for auralization.

Therefore the studio listening room at the *Bavaria Film Studios*, Munich, was chosen in order to allow setting this limit. This studio's base dimensions of $11\,\mathrm{m} \times 15\,\mathrm{m}$, a height of 6 m, and a volume of about $890\,\mathrm{m}^3$ (resulting in a reverberation time of $\mathrm{T}_R = 820\,\mathrm{ms}$ at frequencies around 510 Hz) ensured that most of the common listening rooms are capable of being auralized by this system.

Despite the length of the original room impulse response being 300 ms, an impulse response with a shortened length of only 85 ms was finally implemented. It turned out that all differences between the real listening situation (with the original length) and

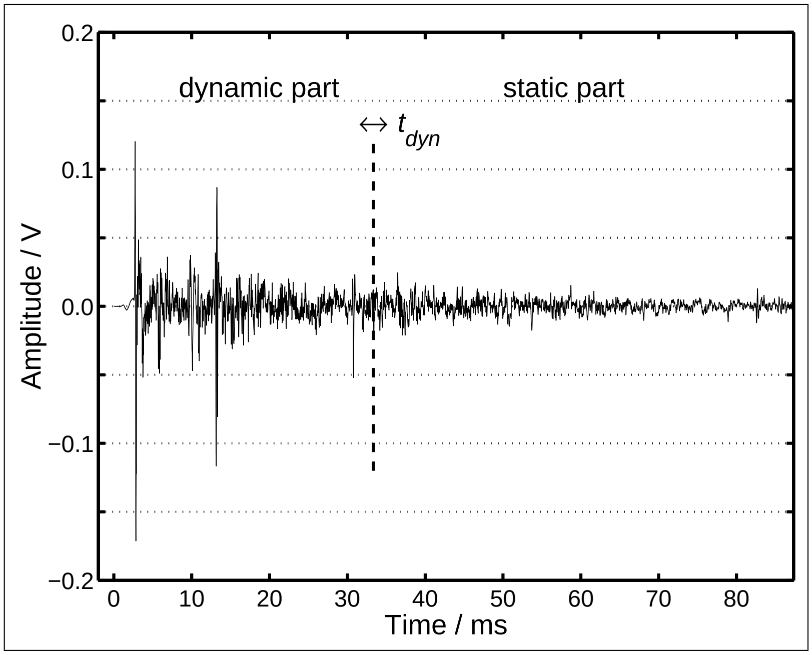

Fig. 5.4: Impulse response with a dynamic part (t $< t_{dyn}$) and a static part (t $> t_{dyn}$).

the shortened impulse response were perceivable only for expert listeners in the case of extremely critical signals, as for example click-sounds. This is due to the decline of the amplitude in the original impulse response being below the 30 dB-threshold after 85 ms.

The listening situation did not meet the standard surround sound setup (acc. to ITU–Rec. BS 770–1) because the listening room at the *Bavaria Film Studios* is designed to be in conformity with a cinema. Again, a diffuse-field equalized KU 100 dummy head, in combination with an *Audio Precision System ONE*, measured the binaural room impulse responses.

This dummy head was at 8 m distance of the frontal center speaker, which was aligned with the left and right loudspeakers. It had an equal distance to the side walls, i. e., in the middle between the side walls. Each of the five loudspeakers' room impulse response was binaurally measured in 15 different orientations of the dummy head: from -42° to 42° azimuthal angle in 2°–steps.

The BRS Processor served as an auralization tool in the listening test, enabling a direct A/B-comparison between the "original" and the "modified" impulse response. In case of the original impulse response, the head movements throughout the whole duration of the impulse response were taken into account. In contrast to the modification, only the first part up to t_{dyn} was changed according to the head movements of the listener. The latter part was, as already mentioned, cross-faded to the impulse response of 0°-orientation. t_{dyn} varied between 7 ms and 43 ms in steps of 4 ms.

12 expert listeners participated in the experiment. They were instructed to move

their head in order to better perceive the difference between the two alternatives. They verbally reported the difference. A dry recorded female voice (EBU SQAM-CD track 49) served as test signal. The experiment aimed at finding the threshold of t_{dyn} at which a noticeable difference in auralization arises.

Results

The subjects reported differences depending on the boundary value t_{dyn}. At small values of t_{dyn} (7 – 11 ms), front-back-inversions and in-head localizations occurred comparable to the case of a static, fixed dummy head. Increasing t_{dyn} led to the perception of additional echoes, which in turn resulted in an incorrect impression of the auralized room's size. Around the threshold value of t_{dyn} the subjects perceived a small bass enhancement.

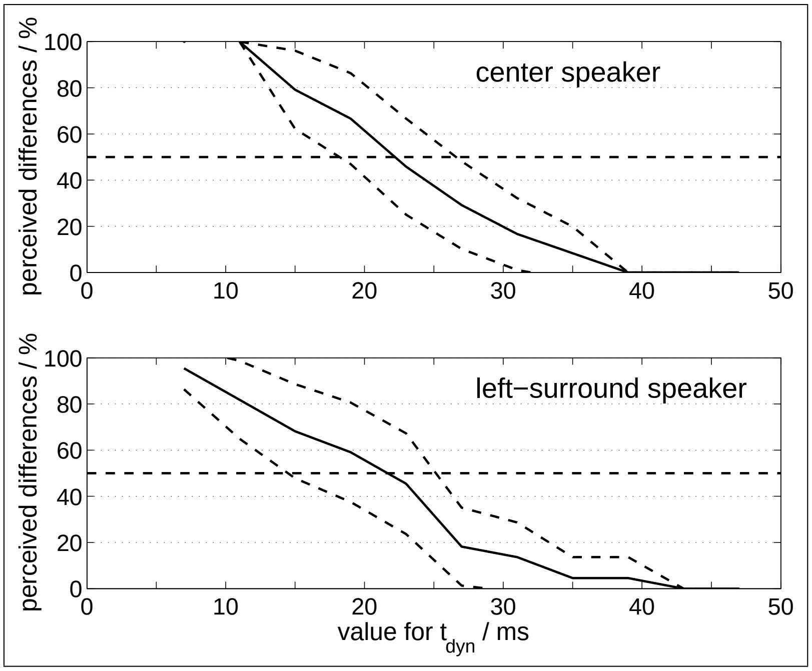

Fig. 5.5: Results for the center loudspeaker (upper panel) and the left surround speaker (lower panel). The perceived differences are displayed in percentages (solid line). Smaller differences are evoked with increasing dynamic cues due to a higher value for t_{dyn}. All results are shown as an arithmetic mean and 95% confidence intervals over 12 subjects.

In the graphic, the value 1 denotes a clear detection of a difference, whereas the value 0 means that the subjects perceived no difference at all. As figure 5.5 shows, the threshold was about 22 ms for the frontal speaker, and 27 ms for the surround speaker. At the threshold itself the probability between perceiving or not a difference was 50%.

The figures display the mean value and the 95% confidence interval (shown as dashed lines). The frontal loudspeaker's value of $t_{dyn} = 22\,\mathrm{ms}$ may possibly be due to a temporal coincidence with the first early reflection. This reflection reaches the listener's ear within the first 20 ms according to Hartmann [52, 53].

An important result of this experiment was that head movements have to be taken into account only for the first 30 ms in rooms with a reverberation time smaller than 300 ms. For the later part of the binaural room impulse response it is sufficient to use the fixed part of the 0°-orientation impulse response for convolution — regardless of head orientation.

5.2.2 The BRS Processor

The *BRS Processor* was created by *Studer Professional Audio AG* in close cooperation with our group at IRT and served as the first application of the data-based auralization method. It allows auralizing a room with up to five independent loudspeakers. The BRS Processor is a 19" device that integrates the whole signal processing board (memory, DSPs). Figure 5.6 presents some views of the BRS Processor.

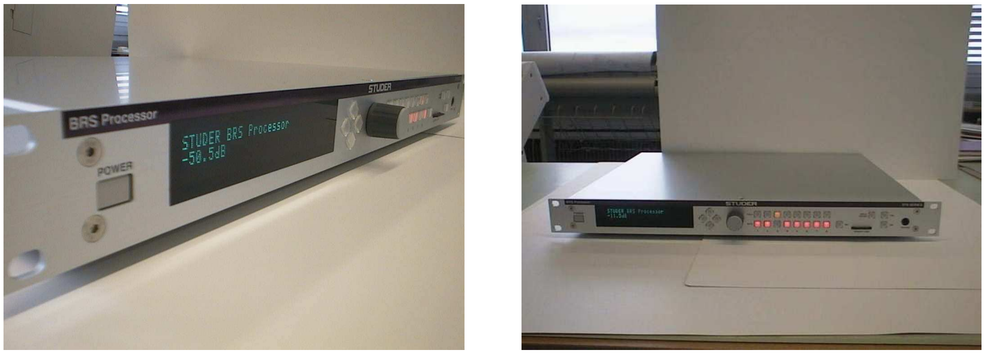

Fig. 5.6: The BRS Processor by Studer Professional AG, Switzerland

Block Processing & Interpolation

The convolutions are identical for all input channels, and they are implemented as "fast convolutions" in the frequency domain. The length of input- and output-buffer was kept short in order to keep the system's latency time below 6 ms. The complex products of the signal spectrum within one period are repeatedly convolved. Subsequently, the small delayed parts of the room spectrum are added up to enable the processing of long room impulse responses. The procedure corresponds to the basic functioning of a FIR-filter. Figure 5.7 denotes this a complex convolution. A more detailed description of the block processing can be found in [66, 59].

As described above, the room impulse response was measured binaurally by means of a diffuse-field equalized dummy head in combination with the MLS-technique of an *Audio Precision System ONE*. This "recording process" allows to take an "acoustic

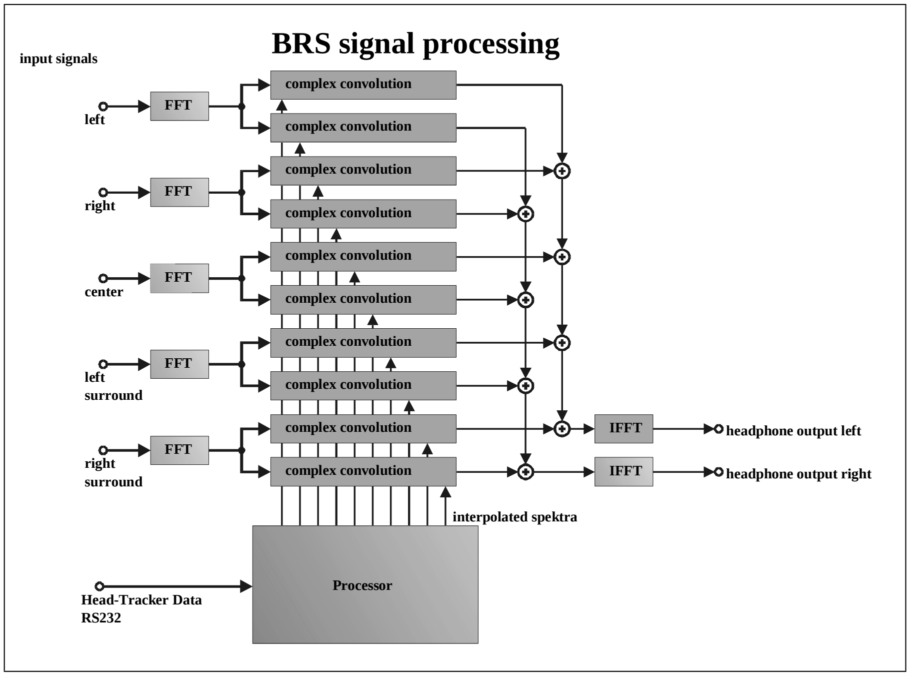

Fig. 5.7: Schematic signal processing in the BRS Processor.

snapshot" and thus to capture the room acoustics. The dummy head was mounted on a computer controlled and motor-driven turntable (described in section 4.1.1). A measurement was taken for each loudspeaker and for each of the dummy head's orientations (from -42° to 42° azimuthal angle in 6°-steps). This amounted to a total of 15 binaural room impulse responses (each of 85 ms duration) per loudspeaker. These responses were transmitted to the BRS Processor and stored internally.

Since a binaural room impulse response is only stored at every 6°, these data need to be interpolated depending on the actual position of the listener's head. This avoids unwanted artifacts, as for example noise, that would otherwise result from the "coarse" 6°-steps. Therefore, new, optimized algorithms in the frequency domain were developed enabling an interpolation without adding non-linear distortions and thus ensuring a 24 bit signal quality [66, 59].

The head-tracker was connected to the BRS Processor by means of a standardized RS–232 interface. The same interface was used to transmit the room data for internal storage.

The BRS Processor, with its underlying data-based auralization method controlled by the listener's head movements, ensured high fidelity and accuracy. The listener hears "through the ears" of a *virtual* dummy head that is positioned at the optimal listening position in the virtual listening room with its virtual loudspeakers.

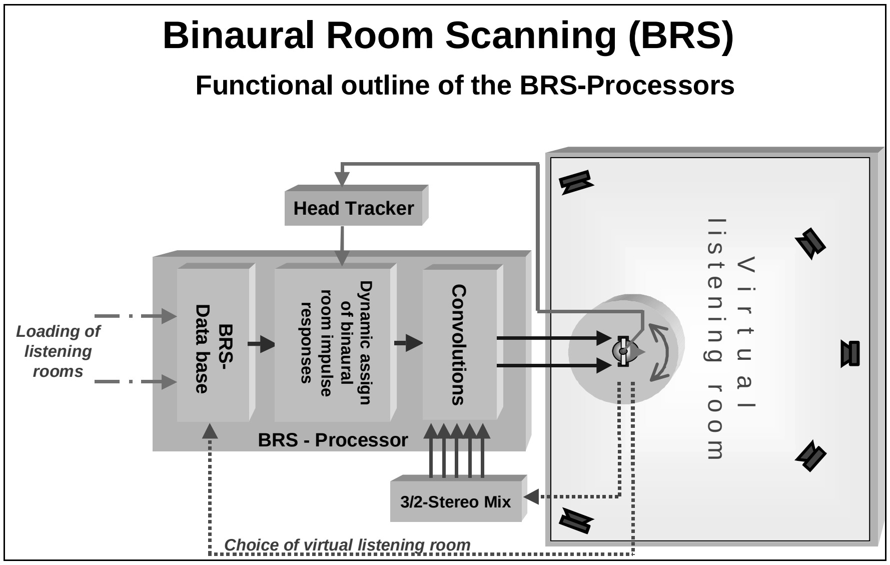
Binaural Room Scanning (BRS)
Functional outline of the BRS-Processors
Head Tracker
Loading of listening rooms
BRS-Data base
Dynamic assign of binaural room impulse responses
Convolutions
BRS - Processor
3/2-Stereo Mix
Choice of virtual listening room
Virtual listening room

Fig. 5.8: Schematic view of the rendering with the BRS Processor. Depending on the orientation of the listener's head the respective BRIR is selected from the database and used for the convolution.

Advantages of the BRS Technology

This new data-based Binaural Room Scanning technology allows to *acoustically clone* an existing listening situation with up to five loudspeakers.

Using the BRS Processor it is possible to simulate the complete acoustics of certain[5] *existing* surround sound production studio, including the loudspeakers and all other characteristic parameters, in a quality that can be called *authentical*. This emulation can take place, for example, in a broadcasting van, although the physical dimensions and the acoustic properties of the van itself would never allow to actually reproduce such a listening situation [66].

Another advantage of this system is to actually have *several sweet spots* when using several BRS-Processors at the same time. This provides the producer and the sound engineer parallelly with the optimal listening situation.

Furthermore, the system allows realizing a music production in a "standardized" listening situation, or "testing" the same music production in different listening environments such as a normal listening room, a car, etc.

First practical experiences - Listening Tests

A demonstration of the BRS Processor took place at the BBC with listening tests, in which approximately 30 — 40 BBC personnel participated. Additionally, some

---

[5] The only condition is that the reverberation time of the listening room is smaller than 300 ms (see also section "*Partially Dynamic Convolution*").

engineers volunteered to use the BRS Processor in a broadcasting van for a production on location.

The BBC Maida Vale Studio 5 was binaurally measured with a standard 3/2 surround sound listening situation. The actual loudspeakers used to measure the room were not available for an A/B-comparison test. So they used the Studer A5 loudspeaker as frontal loudspeakers and the A3 loudspeaker as rear speakers. The Studer loudspeakers were of even higher quality than those used for measurement.

Speech from a multichannel test DVD, as well as both classical and pop music in a standard stereo format, served as test signal. Furthermore, a surround recording of the Eagles' "Hotel California" was used to demonstrate the surround capability of the BRS Processor. Additionally, all listeners were invited to provide material they were familiar with.

At first, the listeners had to identify the real loudspeakers by listening to the speech sound in order to acquaint them with the room and the acoustic environment. Subsequently, they were asked to put on the headphones and to identify the virtual speaker. During this task the real speakers were muted.

Results

All participants perceived the voice to originate from the corresponding loudspeaker. Localization was particularly reliable for the rear speakers. Most of the listeners were astonished about the BRS-System's fidelity of reproduction. Some tonal differences were noted, but these could also be attributed to the difference between the loudspeakers used for the measurement and those in the listening test (Studer A5/A9).

However, the center speaker appeared to move slightly in the opposite direction when moving the head, as if to *compensate* for the listener's head motion. For example, when moving the head to the left the center appeared to be shifted a little bit to the right.

With respect to "normal listening" to music, i. e., classical and pop music, most listeners were surprised by the difference of quality between "normal headphone-listening", i. e., without the BRS Processor, and the listening to the virtual speakers using the BRS Processor. All participants preferred listening to the BRS Processor because it was much easier for the ear to listen to it for longer periods. Only the tonal differences due to the deviation in loudspeakers used, as mentioned above, were a subject of common criticism.

First practical experiences - OB truck production

Two BBC sound engineers also tested the BRS-System in an OB truck on location. The first location was Westminster Cathedral, where a live stereo broadcast of a classical concert took place. The other situation was a production during MetAid at Wembley Stadium.

Results

Using the BRS Processor in an OB van produced a few problems unseen in the studio situation. The electromagnetic head tracker systems (Polhemus FasTrak) could not be used because of an interference with all the real loudspeakers in the OB truck. An ultrasonic head tracker from *Logitech* replaced the *Polhemus*. But due to insufficient range and stability the ultrasonic system did not produce such perfect localization as the *Polhemus*.

On the other hand, visual discrepancies required the engineers to make a "mental adjustment": The visual environment (a close and narrow OB van interieur) interfered with the virtual acoustic environment reproduced by the BRS Processor, i. e., the virtual Maida Vale Studio 5. Also, tonal differences between the *measured* loudspeakers and the loudspeakers the engineers were used to contributed to the difficulty of hearing the "correct sound".

But in general, the engineers experienced the BRS Processor to be a vast improvement on OB-truck monitoring. This was particularly true if the head tracking issues could be adjusted and if the scanning process was supported by the "correct" loudspeakers.

As an overall result of the first practical experiences the following can be summarized:

- The general judgment on the BRS-System was extremely positive.
- The audio quality and the localization ability of the system is already high enough for professional use.
- A head tracker system needs to be used that does not cause any interference in an OB-van
- A built-in possibility to store internally up to eight different rooms is more than adequate.

5.3 Discussion

The data-based Binaural Room Scanning is the first auralization method that takes into account both head movements and the entirety of (acoustic) localization cues. The acoustical scanning "captures" the whole listening situation and, thus, each acoustical cue. As a consequence, the entirety of cues is reproduced in the replay process and a convincing and authentic impression of the reproduced virtual environment results.

In this section we will discuss a few proposals and directions for possible future work. They are closely related to the BRS Processor. On the one hand, we propose a method to optimize the reproduction of a data-based auralization system, on the other hand, we suggest some applications of the BRS technology.

Reduction of unwanted Elevations

When using the dynamic dummy head system (see chapter 4), some subjects reported elevations of sources that originally were placed in the horizontal plane. These elevations occurred mainly with respect to sound sources in the frontal region. Similar observations were made when using the data-based BRS Processor (see sections 5.1.2 and 5.2.2).

Applying an individualized equalization as described in [66] was meant to reduce these unwanted elevations. But these spectral manipulations did not help to obtain the desired "flat horizontal" localization.

However, Wallach [133] showed that changes in the lateral angle ($\Delta\psi$) caused by such a head rotation (alterations of the aural axis $\Delta\beta$) depend on the *elevation* (ϑ) of the sound source (see also section 3.2.2). Mathematically, the relation between the displacement β of the aural axis, the lateral angle ψ and the elevation ϑ of the source can be expressed as: $\sin(90^\circ - \psi) = \sin\beta \cdot \cos\vartheta$. For example, $\Delta\beta$ equals 0° for sound sources straight above the head ($\vartheta = 90^\circ$). On the other hand, for sound sources in the horizontal plane ($\vartheta = 0^\circ$) these differences are equal in magnitude, and differ only in their sign ($\Delta\psi = \pm\,\Delta\beta$).

That means the other way round that if the rotation of the listener's head were *not* transmitted in a 1:1 relation to the dummy head, the perceived elevation of the auditory event would be influenced. Varying this "rotational ratio" between listener and dummy head for the frontal sources might reduce the unwanted elevations: Something worth to be investigated in a future experiment.

Another idea with respect to reducing unwanted elevations is the implementation of a *tilting* movement. According to Wallach, a pivoting movement might have a greater influence on localization than the tilting movement. This assumption was tested in an experiment at IRT, described in section 4.3. Tilting the head displaces the aural axis in a similar fashion as a rotation around the vertical axis. The tilting movement, on the other hand, does not displace the aural axis at all. Therefore, the use of tilting movements to help reducing unwanted elevations might be the subject of another experiment.

Influence of Head Movements on Monaural Perception

The dynamic dummy head system easily allows to investigate the influence of head movements on monaural localization. It simply requires interrupting the left or the right transmission path between the dummy head and the headphones. When doing so, no acoustical information of the original sound field ever arrives at the other ear.

In contrast, if only occluding one of the listener's ears (for example by means of putty) and placing the listener himself into the original sound field, there is always a residual risk of sound arriving at the masked ear. Such sound reaching the "deaf" ear might, however, distort the result [146]. Using a dynamic dummy head system or the BRS Processor permits to completely ignore this source of failure.

A possible experiment could investigate the impact of dynamic cues on monaural localization compared with binaural localization. Dynamic cues could originate from head rotations or maybe from a tilting movement.

Comparisons using the BRS Processor

Switching almost instantaneously between different listening situations, e. g., loudspeaker setups, is one of the BRS Processor's advantages. These situations could be easily compared with each other without leaving the room or physically altering the setup — a process that always causes a break between comparisons.

Olive *et al.* [97] used the *static* binaural technology (fixed dummy head) in a listening experiment. They compared different loudspeaker setups without burdening the subject's small acoustic memory by changing the setup or moving the subjects from one room to another. However, if one additionally allows head movements, a more natural localization is possible and thus the comparison between different listening situations is more meaningful. Such head movements can be taken into account when using the BRS Processor.

A surround sound listening situation of a cinema could be instantaneously compared to the "home cinema" atmosphere, or even to a "car hifi" situation. Also, the "compatibility" of a surround mix with the normal 2-way stereo or even mono reproduction could be checked and assessed using either a data-based auralization system or several dynamic dummy head systems.

To sum up, a *dynamic* dummy head system or the *data-based* auralization method in form of the BRS Processor enables a researcher to carry out different types of listening tests and simultaneously to take into consideration the entirety of all localization cues.

Chapter 6

Summary and Conclusions

The last section reviews the novel contribution of this work and emphasizes its main aspects.

This paper addresses four different topics: A *classification scheme* of localization cues, an *applicability* of this cue-classification combined with an emphasis on taking into account the *totality of localization cues*, a quantitative analysis of the *importance of head movements*, and finally, an *application of our results* in form of a new auralization method [1].

Classification Scheme of Localization Cues

This work dealt with *localization cues*, in general, and with *head movements* in particular. A *new classification scheme* was proposed to distinguish the localization cues. It divides the cues into three main categories: cues directly relating to attributes of the source (*source cues*), cues originating from the surrounding environment (*environmental cues*), and finally a group of listener related cues (*listener cues*).

These main groups can be further divided into *subgroups.* Within the main group of *source cues* we can distinguish between *spectral cues*, *temporal cues* and *local cues.* The group of spectral cues comprises the influences of the *bandwidth* (narrow or wide) and of the *spectral distribution* (low or high frequencies). The temporal cues take into account the *dynamic behavior* of the sound (dynamic or static) as well as its *duration* (short or long). And the *relative position* of the source with respect to the listener characterizes the local cues. However, positional changes of either, source or listener, are disregarded here.

The main group of *environmental cues* consists of cues due to *direct sound* and cues originating from *reflections.* In a strict sense, reflections can be subdivided into *early reflections* and *reverberation.*

---

[1] This auralization method is restricted to existing listening rooms with a reverberation time of less than 300 ms (see also section "*Partially Dynamic Convolution*").

Finally, in the last main group, the *listener cues*, we can draw lines between four subgroups: the *interaural cues*, the *HRTF cues*, *head movement cues* and *non-acoustical cues.* Interaural cues are localization cues resulting from the *monaural* or *binaural* reception of sound. The influence of using the "*own HRTFs*", *foreign HRTFs* or even *no HRTFs* at all characterizes the HRTF cues. All *dynamic cues* caused by the *movement of the listener's head* belong to the group of head movement cues. These cues in general are *head rotations* around varying axes — due to anatomical reasons all *translations* of the head can be neglected. Finally, *optical cues*, *vestibular cues* and *knowledge cues* are the most prominent localization cues that are not directly related to acoustics, and thus come under the heading of *non-acoustical cues.*

This new cue classification permits arranging various experiments with regard to their localization cues used.

Application of the Cue Classification

A number of authors conducted several experiments in oder to understand the hearing system. These experiments could be arranged into groups according to the cue-classification and, thus, verifying the scheme. In most experiments, one or two localization cues were varied, whereas the rest was kept constant — if not even completely disregarded. This rested on the assumption that a separate and independent observation of different cues be permissible. And although all of those results are by no means wrong in principle, they have to be interpreted taking into account the respective prevailing conditions.

This work showed, by citing several experiments and assigning them to different cue groups, that disregarding a cue can lead to completely different results (including in-head localization). Wherever possible, the experiments were presented in "pairs", i. e., in one experiment a cue was disregarded, whereas in the other one it was taken into account. Therefore, wherever a single localization cue is concerned, a potential influence of all other localization cues always should be kept in mind in order to avoid drawing a false conclusion.

Gardner's experiment, cited in section 3.1.1, is an example for such a "false" conclusion. He used an anechoic room as an environment in order to determine the relevant cues for distance estimation, and concluded the sound level to be most relevant cue. In contrast, a series of experiments by Nielsen (see 3.1.2) proved the environment (room) to have the most important influence. Depending solely on the relation between direct sound and reflections the sound source will be perceived as close or distant.

However, certain non-acoustical cues as for instance the *memory* can also influence perception: A "*whispering*" is always associated with a close distance, the opposite is true for "*shouting*".

It is thus important to always take into account the possible influence of other localization cues before one "deduces" a certain functioning of the hearing system.

Quantitative Approval of the Importance of Head Movements

Head movements are an important cue yet often disregarded in experiments. These dynamic cues can have a strong influence on localization. They are especially a mean to resemble localization ambiguities, as for example, front-back-inversions.

Therefor, we thoroughly investigated quantitatively the impact of head movements on localization in various listening tests that had been carried out at IRT between 1997 and 2000. All of these experiments contributed to research project with Studer Professional Audio AG, Zurich, Switzerland.

A diffuse field equalized dummy head (Neumann KU 100) was used for these experiments in combination with likewise diffuse field equalized headphones (STAX SR Lambda Pro). The dummy head was mounted on a computer-controlled and motor-driven turntable, that in turn was controlled by a Polhemus FasTrak head tracker. By using a head tracker the dummy head was capable of following the listener's head rotations.

At first, the *fidelity of the whole system* was very high because the systems's inherent inaccuracy was minimized (latency time, angular accuracy of head tracker and step motor). An optimal reproduction was ensured at a *total latency time* shorter than 85 ms resulting in an aural sensation of "being there".

Nevertheless, it is not possible to exclude localization failures originating from the system itself even if such a high fidelity of the system was achieved. In future, the system's accuracy could be optimized, e. g., by a further reduction of the latency time or the usage of individualized HRTFs (outer ears).

When the listening tests were carried out *without* taking the listener's head movements into considerations e. g., with a *fixed* dummy head, the localization was not very reliable. A lot of front-back-inversions occurred, and the sources in the median plane were often perceived inside the head. This bad performance was nearly identical in different environments (listening room and anechoic chamber) and therefore independent of reflections being present.

If, however, head movements were allowed, the localization was more or less *independent of HRTF cues*, and resembled that of natural hearing. This held true for "foreign HRTFs" (dummy head) as well as for a total lack of HRTFs such as in the case of a sphere microphone.

To be more precise, only head *rotations* around the *vertical* axis, but not around the ear axis (tilting) improved localization. These findings were in accordance with a theory proposed by Wallach (see 3.2.2).

The experiments showed the enormous importance of dynamic cues caused by head movements (mainly head rotations) on localization as they aid to resolve localization ambiguities. All situations with a fixed (dummy) head are not a natural, and often result in localization errors.

BRS - A data-based Auralization Method

A new auralization method, the *data-based* auralization called *Binaural Room Scanning* (BRS) was introduced as a first application of the overall result, i. e., the strong impact of head movements and a general importance of *all* localization cues. Our experiments at IRT were the basis for the development of the BRS Processor.

The idea behind Binaural Room Scanning is the use of a dummy head for taking an acoustical "snapshot" (*acoustical clone*) of the environment to be auralized[2]. For this purpose, the dummy head is placed at the optimal listening position, e. g., the sweet spot, and a (binaural) measurement of the room impulse response is taken (e. g., by using the MLS-method) for every loudspeaker. This measurement has to be repeated for the dummy head's *different orientations*. All measurements are then stored in a *database*.

A head tracker evaluates the orientation of the listener's head for the auralization. Depending on this orientation, the respective binaural room impulse response is selected from the database and used for the convolution with the input signals, normally fed to the loudspeakers. The result is a not only a convincing, but also an *authentic reproduction* of an *existing* acoustic environment. *Without* head movements this auralization method would merely resemble a dummy head recording with a *fixed* dummy head, including all its inherent localization disadvantages (e. g., in-head localization for median-plane sources, etc.).

Conclusions

What overall result can be drawn? A head that is kept still, or a fixed dummy head, are an *unnatural* listening situation. And this work supports the findings of such unnatural limitations leading to localization errors. Or the other way round: Head rotations (movements) are of enormous importance for localization as they resemble more a natural listening situation, thus reducing localization errors. Also, it is important to take all localization cues into account to obtain an optimal localization and thus an authentic reproduction.

---

[2] Regarding the class of the listening rooms that can be auralized using the BRS-Processor the same limitations apply (reverberation time less than 300 ms) as described in chapter 5, section "*Partially Dynamic Convolution*".

Bibliography

[1] J. Angell and W. Fite. The monaural localization of sound. *Psychol.Rev.*, 8:225–246, 1901.

[2] F. Asano, Y. Suzuki, and T. Sone. Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.*, 88(1):159–168, 1990.

[3] S. Bech. Perception of timbre of reproduces sound in small rooms: Influence of room and loudspeaker position. *J. Audio Eng. Soc.*, 42(12):999–1007, 1994.

[4] S. Bech. Timbral aspects of reproduced sound in small rooms. I. *J. Acoust. Soc. Am.*, 3:1717–1726, 1995.

[5] D. R. Begault. *3-D Sound for Virtual Reality and Multimedia.* AP Professional, 1994. ISBN 0-12-084735-3.

[6] G. Behler. Versuch zur Richtungswahrnehmung in der Medianebene [Experiment on the perception of direction in the median plane]. In *Fortschritte der Akustik – DAGA*, pages 483–486, 1985.

[7] J. Blauert. Ein Beitrag zur Trägheit des Richtungshörens in der Horizontalebene [On the lag of sound localization in the horizontal plane]. *Acustica*, 20:200–206, 1968.

[8] J. Blauert. Sound localization in the median plane. *Acustica*, 22:205–213, 1969.

[9] J. Blauert. *Untersuchungen zum Richtungshören in der Medianebene bei fixiertem Kopf [Investigations of directional hearing in the median plane with the head immobilized].* Ph.D. thesis, Technische Hochschule Aachen, 1969.

[10] J. Blauert. Vergleich unterschiedlicher Systeme zur originalgetreuen elektroakustischen Übertragung [A comparison of different systems for true electroacoustic transmission of sound images]. *Rundfunktechnische Mitteilungen*, 18:222–227, 1974.

[11] J. Blauert. *Spacial Hearing – The psychophysics of human sound localization.* MIT Press, Cambridge MA, 2nd revised edition, 1996. ISBN 0-262-02413-6.

[12] P. J. Bloom. Creating source elevation illusions by spectral manipulation. *J. Audio Eng. Soc.*, 25(9):560–565, 1977.

[13] K. d. Boer and A. Urk. Some particulars of directional hearing. *Philips Tech. Rev.*, 6:359–364, 1941.

[14] G. Boerger and R. Fengler. Characteristics of a parametric earphone reproduction system. In *8.th International Congress on Acoustics, London*, 1974. 702.

[15] G. Boerger, P. Laws, and J. Blauert. Stereophone Kopfhörerwiedergabe mit Steuerung bestimmter Übertragungsfaktoren durch Kopfdrehungen [Stereophonic reproduction by earphones with control of special transfer functions through head movements]. *Acustica*, 39:22–26, 1977.

[16] A. Bregman. *Auditory Scene Analysis.* The MIT Press, Cambridge, MA, 1990.

[17] A. W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.*, 98(5):2542–2553, 1995.

[18] T. Buell, C. Trahiotis, and L. Bernstein. Lateralization of bands of noise as a function of combinations of interaural intensitive differences, interaural temporal differences, and bandwidth. *J. Acoust. Soc. Am.*, 95(3):1482–1489, 1994.

[19] J. Burger. Front-back discrimination of the hearing system. *Acustica*, 8:301–302, 1958.

[20] M. D. Burkhard. *Binaural Measurements and Applications*, chapter 30, pages 665–681. In Gilkey and Anderson [44], 1995. ISBN 0-8058-1654-2.

[21] R. A. Butler. *Spatial Referents of Stimulus Frequencies: Their Role in Sound Localization*, chapter 5, pages 99–115. In Gilkey and Anderson [44], 1997. ISBN 0-8058-1654-2.

[22] R. A. Butler and K. Belendiuk. Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.*, 61(5):1264–1269, 1977.

[23] R. A. Butler and R. Flannery. The spatial attributes of stimulus frequency and their role in monaural localization of sound in the horizontal plane. *Percept. Psychophys.*, 28:165–173, 1980.

[24] D. H. Cooper and J. L. Bauck. Prospects for transaural recording. *J. Audio Eng. Soc.*, 37(1/2):3–19, 1989.

[25] D. Damaske and V. Mellert. Ein Verfahren zur richtungstreuen Schallabbildung des oberen Halbraumes über zwei Lautsprecher [A procedure for generating directionally accurate sound images in the upper half-space using two loudspeakers]. *Acustica*, 22:154–162, 1969/70.

[26] P. v. Damaske and B. Wagener. Richtungshörversuche über einen nachgebildeten Kopf [Localization tests by using a head-replica]. *Acustica*, 21:30–35, 1969.

[27] M. Dickreiter. *Handbuch der Tonstudiotechnik [Handook of sound engineering].* K. G. Saur, München, 5 edition, 1987.

[28] DIN 1320. Allgemeine Benennungen in der Akustik [Common terms used in acoustics]. Beuth-Vertrieb, Berlin, 1959.

[29] N. Durlach and H. Colburn. *Binaural Phenomena.* 1978.

[30] EBU. Sound quality assessment material recordings for subjective test. EBU SQAM CD. *EBU Tech. 3253-E*, 1988.

[31] U. Felderhoff, P. Mackensen, and G. Theile. Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale [Stability of localisation vs. distorted reproduction of binaural cues]. In *20. Tonmeistertagung, Karlsruhe*, 1998. 229–237.

[32] H. Fisher and S. Freedman. The role of the pinna in auditory localization. *J. Aud. Res.*, 8:15–26, 1968.

[33] A. Ford. A dynamic auditory localization: I. the binaural intensity limen. *J. Acoust. Soc. Am.*, 13:367–372, 1942.

[34] N. V. Franssen. *Some considerations of the mechanism of directional hearing.* Ph.D. thesis, Institute of Technology, Delft, 1960.

[35] M. Fruhmann, P. Mackensen, and G. Theile. Reduction of dynamic cues in auralized binaural signals. *acta acustica*, 88:443–445, 2002.

[36] M. B. Gardner. Distance estimation of 0 or apparent 0-oriented speech signals in anechoic space. *J. Acoust. Soc. Am.*, 45(1):47–53, 1969.

[37] M. B. Gardner. Some monaural and binaural facets of median plane localization. *J. Acoust. Soc. Am.*, 54(6):1489–1495, 1973.

[38] M. B. Gardner and R. S. Gardner. Problem of localization in the median plane: Effect of pinnae cavity occlusion. *J. Acoust. Soc. Am.*, 53(2):400–408, 1973.

[39] W. G. Gardner. *3-D Audio Using Loudspeakers.* Ph.D. thesis, MIT Media Lab., 1997.

[40] W. G. Gardner. *3-D Audio Using Loudspeakers.* Kluwer Academic Publishers, Norwell, MA., 1998.

[41] K. Genuit. *Ein Modell zur Beschreibung von Aussenohrübertragungseigenschaften [A model for description of the transfer characterstics of the pinnae].* Ph.D. dissertation, Technische Hochschule, Aachen, 1984.

[42] H. W. Gierlich. The application of binaural technology. *Applied Acoustics*, 36:219–243, 1992.

[43] C. Giguère and S. M. Abel. Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay. *J. Acoust. Soc. Am.*, 94(2):769–776, 1993.

[44] R. H. Gilkey and T. R. Anderson, editors. *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum Associates, Publishers, Mahwah NJ, 1997. ISBN 0-8058-1654-2.

[45] M. D. Good and R. H. Gilkey. Sound localization in noise: The effect of signal-to-noise ratio. *J. Acoust. Soc. Am.*, 99(2):1108–1117, 1996.

[46] S. Goossens, R. Stumpner, and H. Lamparter. High-quality-auralisation of the sound field in production rooms. In *22. Tonmeistertagung, Hannover, Proceedings*, 2002. CD-Rom.

[47] T. Gotoh. *Can the acoustic head-related transfer function explain every phenomenon in sound localization?*, pages 244–249. 1979.

[48] D. Griesinger. Equalization and spatial equalization of dummy-head recordings for loudspeaker reproduction. *J. Audio Eng. Soc.*, 37(1/2):40–50, 1989.

[49] H. Han. Measuring a dummy head in search of pinna cues. In *90. AES Convention, Paris*, 1991. preprint 3066.

[50] H. Han. On the relation between directional bands and head movements. In *92. AES Convention, Vienna*, 1992. preprint 3293.

[51] W. Hartmann and B. Rakerd. Auditory spectral discrimination and the localization of clicks in the sagittal plane. *J. Acoust. Soc. Am.*, 94:2083–2092, 1993.

[52] W. M. Hartmann. Localization of sounds in rooms. *J. Acoust. Soc. Am.*, 74(5):1380–1391, 1983.

[53] W. M. Hartmann. *Listening in a Room and the Precedence Effect*, chapter 10, pages 191–210. In Gilkey and Anderson [44], 1994. ISBN 0-8058-1654-2.

[54] K. Hartung, M. Mioyshi, M. Bodden, and J. Blauert. Merkmale der Vorne-Hinten-Lokalisation in der Horizontalebene unterhalb von 2 khz [Cues concerning the front-back-localization in the horizontal plane below 2 khz. In *Fortschritte der Akustik - DAGA*, pages 837–839, 1993.

[55] J. Hebrank and D. Wright. Are two ears necessary for localization of sound sources on the median plane? *J. Acoust. Soc. Am.*, 56(3):935–938, 1974.

[56] J. Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, 56(6):1829–1834, 1974.

[57] P. Hofman and J. V. Opstal. Spectro-temporal factors in two-dimensional human sound localization. *J. Acoust. Soc. Am.*, 103:2634–2648, 1998.

[58] R. Holt and W. Thurlow. Subject orientation and judgment of distance of a sound source. *J. Acoust. Soc. Am.*, 46(6):1584–1585, 1969.

[59] U. Horbach and R. Pellegrini. Design of positional filters for 3d audio rendering. In *105. AES Convention, San Francisco*, 1998. preprint 4798.

[60] E. Hornborstel and M. Wertheimer. Über die Wahrnehmung der Schallrichtung [On the perception of the direction of sound]. Sitzungsbericht Akad. Wiss. Berlin, 1920.

[61] K. Inanaga, Y. Yamada, and H. Koizumi. Headphone system with out–of-head localization applying dynamic hrtf (head related transfer function). In *98. AES Convention, Paris*, 1995. Preprint 4011.

[62] ITU–Recommandation ITU-R BS.708. *Determination of the Electro-Acoustical Properties of Studio Monitor Headphones*, 1990.

[63] ITU–Recommandation ITU-R BS.775-1. *Multichannel Stereophonic Sound System with and without accompanying picture*, 1994.

[64] L. A. Jeffress and R. W. Taylor. Lateralization vs. localization. *J. Acoust. Soc. Am.*, 33(4):482–483, 1961.

[65] L. B. W. Jongkees and van de Veer, R. A. On directional sound localization in unilateral deafness and its explanation. *Acta Oto-laryngol.*, 49:119–131, 1958.

[66] A. Karamustafaoglu, U. Horbach, R. Pellegrini, P. Mackensen, and G. Theile. Design and applications of a data-based auralization system for surround sound. In *106. AES Convention, Munich*, 1999. preprint 4976.

[67] O. Klemm. Untersuchungen über die Lokalisation von Schallreizen iii: Über den anteil des beidohrigen hörens [Investigations on the localization of sound stimuli iii: On what is contributed by two- eared hearing]. *Arch. ges. Psychol.*, 38:71–114, 1918.

[68] H. Klensch. Beitrag zur Frage der Lokalisation des Schalles im Raum [a contribution to the study of the localization of sound in space]. *Pflügers Arch*, 250:492–500, 1948.

[69] W. Kock. Binaural localization and masking. *J. Acoust. Soc. Am.*, 22(6):801–804, 1950.

[70] W. Koenig. Subjective effects in binaural hearing. *J. Acoust. Soc. Am.*, 22(1):61–62, 1950.

[71] W. Kraak and G. Schommartz, editors. *Angewandte Akustik*, volume 4. Verlag Technik, Berlin, 1990.

[72] W. Kuhl and R. Plantz. Kopfbezogene Stereophonie und andere Arten der Schallübertagung im Vergleich mit dem natürlichen Hören [Artificial-head stereophony and other methos of sound transmission compared with natural hearing]. *Rundfunktechnische Mitteilungen*, 19:120–132, 1975.

[73] G. F. Kuhn. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, 62(1):157–166, 1977.

[74] V. Larcher, G. Vandernoot, and J.-M. Jot. Equaization methods in binaural technology. In *105. AES Convention, San Francisco*, 1998. Preprint 4858.

[75] P. Laws. Entfernungshören und das Problem der Im–Kopf–Lokalisiertheit von Hörereignissen [Auditory distance perception and the problem of "‘in–head localization"’ of sound images]. *Acustica*, 29(5):243–259, 1973.

[76] J. M. Loomis, C. Herbert, and J. G. Cicinelli. Active localization of virtual sounds. *J. Acoust. Soc. Am.*, 88(4):1757–1764, 1990.

[77] P. Mackensen. Naturgetreue elektroakustische Übertragung als Ziel technischer und künstlerischer Bemühungen (?) [Natural electroacoustical transmission as a goal of technical and artistic effords (?)]. Diploma thesis, Technische Universität Berlin, 1997.

[78] P. Mackensen. *Von der spontanen Kopfdrehung zum Virtuellen Mehrkanal Abhörraum [From Spontaneous Head Movements to a Virtual Multichannel Listening Room]*, pages 171–182. Wissenschaft und Technik Verlag, Berlin, 1999.

[79] P. Mackensen, M. Fruhmann, M. Thanner, G. Theile, U. Horbach, and A. Karamustafaoglu. Head-tracker based auralization systems: Additional consideration of vertical head movements. In *108. AES Convention, Paris*, 2000. preprint 5135.

[80] E. A. Macpherson and J. C. Middlebrooks. Localization of brief sounds: Effects of level and background noise. *J. Acoust. Soc. Am.*, 108(4):1834–1849, 2000.

[81] S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.*, 61(6):1567–1576, 1977.

[82] J. Meyer. *Akustik und musikalische Aufführungspraxis.* Verlag Das Musikinstrument, Frankfurt/M, 3rd edition, 1980. ISBN.

[83] J. Middlebrooks and D. Green. Directional dependence of interaural envelope delays. *J. Acoust. Soc. Am.*, 87:2149–2162, 1990.

[84] J. C. Middlebrooks. Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.*, 92(5):2607–2624, 1992.

[85] H. Møller. Reproduction of artificial-head recordings through loudspeakers. *J. Audio Eng. Soc.*, 37(1/2):30–33, 1989.

[86] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen. Using a typical human subject for binaural recordings. In *100. AES Convention, Copenhagen*, 1996. preprint 4157.

[87] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen. Evaluation of artificial heads in listening test. In *102. AES Convention, Munich*, 1997. preprint 4404.

[88] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen. Head-related transfer functions of humans subjects. *J. Audio Eng. Soc.*, 43(5):300–321, 1995.

[89] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi. Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc.*, 44(6):451–469, 1996.

[90] B. C. Moore. *An Introduction to the Psychology of Hearing.* Academic Press, third edition, 1989. ISBN 0-12-505623-0.

[91] B. S. Müller and P. Bovet. Role of pinnae and head movements in localizing pure tones. *Swiss J. of Psychol.*, 58(3):170–179, 1999.

[92] A. D. Musicant and R. A. Butler. The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am.*, 75(4):1195–1200, 1984.

[93] S. H. Nielsen. *Distance Perception in Hearing.* Ph.D. thesis, Aalborg University, 1991. ISBN 87-7307-447-0.

[94] S. H. Nielsen. Auditory distance perception. In *92. AES Convention, Vienna*, 1992. preprint 3307.

[95] W. Noble. Earmuffs, exploratory head movements, and horizontal and vertical sound localization. *Journal of Auditory Research*, 21:1–12, 1981.

[96] S. Oldfield and S. Parker. Acuity of sound localization: a topography of auditory space. I. Normal hearing conditions. *Perception*, 13:581–600, 1984.

[97] S. E. Olive, P. L. Schuck, S. L. Sally, and M. E. Bonneville. The effects of loudspeaker placement on listener preference ratings. *J. Audio Eng. Soc.*, 42(9):651–669, 1994.

[98] R. S. Pellegrini. Comparison of data- and model-based simulation algorithms for auditory virtual environments. In *106. AES Convention, Munich*, 1999. preprint 4953.

[99] G. Plenge. Über einige Probleme bei der elektroakustischen Vermittlung von Höreindrücken aus Räumen. In *DAGA Stuttgart, Berichtsheft VDI Verlag Düsseldorf*, page 154 oder NOTE, 1972.

[100] G. Plenge. On the difference between localization and lateratlization. *J. Acoust. Soc. Am.*, 56:944–951, 1974.

[101] G. Plenge and G. Brunschen. A priori knowledge of the signal when determining the direction of speech in the median plane. In *7th Int. Congress on Acoustics*, 1971. Proceedings, 19, H10).

[102] C. Pratt. The spatial character of high and low tones. *J. Exp. Psychol.*, 13:278–285, 1930.

[103] B. Rakerd and W. M. Hartmann. Localization of sounds in rooms II: The effects of a single reflecting surface. *J. Acoust. Soc. Am.*, 78(2):524–533, 1985.

[104] B. Rakerd and W. M. Hartmann. Localization of sounds in rooms III: Onset and duration effects. *J. Acoust. Soc. Am.*, 80(6):1695–1706, 1986.

[105] B. Rakerd, W. M. Hartmann, and T. L. McCaskey. Identification and localization of sound sources in the median sagittal plane. *J. Acoust. Soc. Am.*, 106(5):2812–2820, 1999.

[106] B. Rathbone, M. Fruhmann, G. Spikofski, P. Mackensen, and G. Theile. Untersuchungen zur Optimierung des BRS-Verfahrens (Binaural Room Scanning) [Investigations on optimization of the brs-system (binaural room scanning)]. In *21. Tonmeistertagung, Hannover, Proceedings*, 2000. 92–106.

[107] K. Reichenauer. Einfluß von Kopfbewegungen auf die Lokalisation in der Horizontalebene [Impact of head movements on the localization in the horizontal plane]. Diploma thesis, Fachhochschule München, 1998.

[108] S. Roffler and R. A. Butler. Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.*, 46(3):1255–1259, 1967.

[109] N. Sakamoto, T. Gotoh, and Y. Kimura. On "out-of-head localization" in headphone listening. *J. Acoust. Soc. Am.*, 24(9):710–716, 1976.

[110] J. Sandvad. Dynamic aspects of auditory virtual environments. In *100. AES Convention, Copenhagen*, 1996. preprint 4226.

[111] C. Searle, L. Braida, D. Cuddy, and M. Davis. Binaural pinna disparity: Another auditory localization cue. *J. Acoust. Soc. Am.*, 57(2):448–455, 1975.

[112] E. A. Shaw. Earcanal pressure generated by a free field. *J. Acoust. Soc. Am.*, 39(3):465–470, 1965.

[113] E. A. Shaw. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.*, 56(6):1848–1861, 1974.

[114] E. A. Shaw. *Acoustical Features of the Human External Ear*, chapter 2, pages 25–47. In Gilkey and Anderson [44], 1994. ISBN 0-8058-1654-2.

[115] L. v. Soest. Richtungshooren bij sinusvormige geluidstrillingen [Directional of sinusoidal sound waves]. *Physica*, 9:271–282, 1929.

[116] T. Sone, M. Ebata, and T. Nimura. On the difference between localization and lateralization. In *Rept. 6th ICA, A-3-6, pp. A29–A32*, 1968.

[117] J. L. R. Strutt. On our perception of sound direction. *Philosphy Magazine*, pages 214–232, 1907.

[118] C. Stumpf. Differenztöne und Konsonanz [Difference tones and consonance. *Z. Psychol.*, 39:269–283, 1905.

[119] C.-J. Tan and W.-S. Gan. Direct concha exitation for the introduction of individualized hearing cues. *J. Audio Eng. Soc.*, 48(7/8):642–653, 2000.

[120] C. Taylor. *Exploring Music*. IOP Publishing Ltd., 1992. ISBN 0-7503-0213-5.

[121] M. Thanner. Einfluß von Kopfbewegungen auf die Lokalisation in der Medianebene [Impact of head movements on the localization in the median plane]. Diploma thesis, Fachhochschule München, 1999.

[122] G. Theile. *Über die Lokalisation im überlagerten Schallfeld [On Localization in a superimposed sound field]*. Ph.D. thesis, Technische Universtität Berlin, 1980.

[123] G. Theile. Zur Theorie der optimalen Wiedergabe von stereophonen Signalen über Lautsprecher und Kopfhörer [On the theory of the optimum reproduction of stereophonic signals by way of loudspeakers and headsets]. *Rundfunktechnische Mitteilungen*, 25(4):155–170, 1981.

[124] G. Theile. Untersuchungen zur optimalen Kopfhörerentzerrung [Examination of the optimal equalization for headphones]. In *FASE/DAGA '82 - Fortschritte der Akustik, Proceedings*, pages 1247–1253, 1982.

[125] G. Theile. Die Phantomschallquelle – ein Ergebnis der zweidimensionalen Reizverarbeitung im Gehör [The phantom source – a result of two-dimensional stimulus processing in the hearing system]. *VDT-Informationen*, Jan/Feb, 1984.

[126] G. Theile. Das Kugelflächenmikrofon [The sphere microphone]. In *14. Tonmeistertagung, Proceedings*, 1986. 277–293.

[127] W. R. Thurlow, J. W. Mangels, and P. S. Runge. Head movements during sound localization. *J. Acoust. Soc. Am.*, 42(2):489–493, 1967.

[128] W. R. Thurlow and P. S. Runge. Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.*, 42(2):480–488, 1967.

[129] F. E. Toole. In–head localization of acoustic images. *J. Acoust. Soc. Am.*, 44(4):943–949, 1969.

[130] O. Trimble. Localization of sound in the anterior and posterior and vertical dimensions of auditory space. *Birt. J. Psyhol.*, 24:320–334, 1934.

[131] H. Wallach. Über die Wahrnehmung der Schallrichtung [On the perception of the direction of sound]. *Psycholog. Forschung*, 22:238–266, 1938.

[132] H. Wallach. On sound localization. *Acoustica*, 10(4):270–274, 1939.

[133] H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *J. of Exp. Psychology*, 27(4):339–368, 1940.

[134] A. Watkins. *The monaural perception of azimuth*, pages 194–206. 1979.

[135] E. Wenzel. The relative contribution of interaural time and magnitude cues to dynamic sound localization. In *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995. New Paltz, NY, Oct. 15–18, IEEE Press.

[136] E. M. Wenzel. What perception implies about implementation of interactive virtual acoustic environments. In *101. AES Convention, Los Angeles*, 1996. preprint 4353.

[137] E. M. Wenzel. Analysis of the role of update rate and system latency in interactive virtual acoustic environments. In *103. AES Convention, New York*, 1997. preprint 4633.

[138] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1):111–123, 1993.

[139] R. Wettschurek. Die absoluten Unterschiedsschwellen der Richtungswahrnehmung in der Medianebene beim natürlichen Hören, sowie beim Hören über ein Kunstkopf-Übertragungssystem [The absolute difference limen of directional perception in the median plane under conditions of both, natural hearing and hearing with artificial– head–system]. *Acustica*, 28(4):197–208, 1973.

[140] F. Wiener and D. Ross. The pressure distribution in the auditory canal in a aprogressive sound field. *J. Acoust. Soc. Am.*, 18:401, 1946.

[141] F. L. Wightman and K. D. J. Headphone simulation of free-field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am.*, 85(2):858–867, 1989.

[142] F. L. Wightman and K. D. J. Headphone simulation of free-field listening. ii: Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878, 1989.

[143] F. L. Wightman and K. D. J. *Factors Affecting the Relative Salience of Sound Localization Cues*, chapter 1, pages 1–23. In Gilkey and Anderson [44], 1997. ISBN 0-8058-1654-2.

[144] F. L. Wightman and D. J. Kistler. Hearing in three dimensions: Sound localization. In *AES 8th Int. Conference, Proceedings*, 1990.

[145] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, 91(3):1648–1661, 1991.

[146] F. L. Wightman and D. J. Kistler. Monaural sound localization revisited. *J. Acoust. Soc. Am.*, 101(2):1050–1063, 1997.

[147] H. Wilkens, G. Plenge, and R. Kürer. Wiedergabe von kopfbezogenen stereophonen Signalen durch Lautsprecher [The reproduction of head-related stereophonic signals over loudspeakers]. In *AES Convention, Cologne*, 1971.

[148] W. Yost. Lateralization of pulsed sinusoids based on interaural onset, ongoing, and offset temporal differences. *J. Acoust. Soc. Am.*, 61:190–194, 1977.

[149] W. Yost and E. Hafter. *Lateralization*, chapter 3, pages 49–84. 1987.

[150] P. T. Young. The role of head movements in auditory localization. *J. of Exp. Psychology*, 4:95–124, 1931.

Publications and Preprints

- **Philip Mackensen**, Klaus Reichenauer, Günther Theile (1998): Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören [Impact of the spontaneous head rotations on the localization in binaural hearing], 20. Tonmeistertagung, Karlsruhe, Proceedings p. 218 228
- Uwe Felderhoff, **Philip Mackensen**, Günther Theile (1998): Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale [Stability of localisation vs distorted reproduction of binaural cues], 20. Tonmeistertagung, Karlsruhe, Proceedings p. 229 237
- Attila Karamustafaoglu, Ulrich Horbach, Renato Pellegrini, **Philip Mackensen**, Günther Theile (1999): Design and Applications of a Data-based Auralisation System for Surround Sound, 106. AES Convention, Munich, preprint 4976
- Günther Theile, Markus Fruhmann, **Philip Mackensen**, Gerhard Spikofski, Ulrich Horbach, Attila Karamustafaoglu (1999): Der virtuelle Surround Sound Abhörraum - Theorie und Praxis [The Virtual Surround Sound Listening Room - Theory and Practice], ITG Congress, Köln, ITG-Fachbericht 158, p. 15 - 21
- **Philip Mackensen**, Uwe Felderhoff, Günther Theile, Ulrich Horbach, Renato Pellegrini (1999): Binaural Room Scanning A new Tool for Acoustic and Psychoacoustic Research, 137th Meeting of the Acoustical Society of America, Tagungsband (CD-ROM)
- **Philip Mackensen** (1999): Von der spontanen Kopfdrehung zum Virtuellen Mehrkanal Abhörraum [From spontaneous head movements to a virtual multichannel listening room], erschienen in Impulse und Antworten, Feiten et. al. (Hrsg.), Wissenschaft und Technik Verlag, Berlin, 1999, 171 182
- **Philip Mackensen** (2000): Gedanken zur Gesamtheit aller Lokalisationsmerkmale [Considerations on the Totality of Localization Cues], 21. Tonmeistertagung, Hannover, Proceedings p. 239 - 250

- Birgit Rathbone, Markus Fruhmann, Gerhard Spikofski, **Philip Mackensen**, Günther Theile (2000): Untersuchungen zur Optimierung des BRS-Verfahrens (Binaural Room Scanning) [Investigations on Optimization of the BRS-System (Binaural Room Scanning)], 21. Tonmeistertagung, Hannover, Proceedings p. 92 - 106

- **Philip Mackensen**, Markus Fruhmann, Matthias Thanner, Günther Theile, Attila Karamustafaoglu, Ulrich Horbach (2000): Design and Applications of a Data-based Auralisation System for Surround Sound, 108 AES Convention, Paris, Preprint 5135

- Markus Fruhmann, **Philip Mackensen**, Günther Theile (2002): Reduction of dynamic cues in auralized binaural signals, ACUSTICA/acta acustica, Vol. 88 (2002), Nr. 3. Hirzel-Verlag, Stuttgart, p. 443 - 445

- **Philip Mackensen** (2002): Ein neuartiges Modell für HRTFs [A new model for HRTFs], 22. Tonmeistertagung, Hannover, Proceedings on CD-Rom only

Curriculum Vitae

Philip Mackensen

Personal Data	
Nationality	German
21.10.1968	born in Phu-Choung, Vietnam
June 1988	*A–Level (Abitur)*
	Evangelisches Gymnasium zum Grauen Kloster (Berlin, Germany)
Education	
Oct. 1997	*Master of Arts in Communication Sciences*
April 97 - Oct. 97	Communication Science at the Technical University Berlin, Germany
March 1997	*Diploma in Physics*
Oct. 92 - March 97	Physics & Man-Maschine-Communication at the Technical University Munich, Germany
Oct. 91 - Aug. 92	Physics at the University of Edinburgh, Scotland
Oct. 88 - Sept. 92	Physics & Communication Science at the Technical University Berlin, Germany
Professional Experience	
since May 2001	Consultant in the IT & Media-Business at T-Systems International GmbH, Media&Broadcast (Deutsche Telekom AG), Berlin, Germany
Nov. 1997 – April 2001	Research associate at the Institut für Rundfunktechnik (IRT) in Munich, Germany