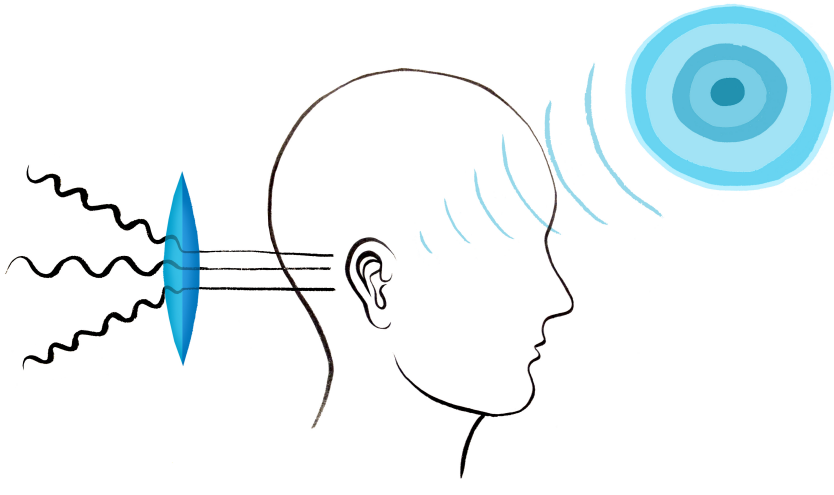
Technical University of Denmark

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 25

Johannes Käsbach

Characterizing apparent source width perception



Hearing Systems
Department of Electrical Engineering

Characterizing apparent source width perception

PhD thesis by
Johannes Käsbach

Preliminary version: August 2, 2016

Technical University of Denmark

2016

Cover illustration by Emma Ekstam.
Preprint version for the assessment committee.
Pagination will differ in the final published version.

This PhD dissertation is the result of a research project carried out at the Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark.

The project was partly financed by the CAHR consortium (2/3) and by the Technical University of Denmark (1/3).

Supervisors

Prof. Torsten Dau
Phd Tobias May
Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Abstract

Our hearing system helps us in forming a spatial impression of our surrounding, especially for sound sources that lie outside our visual field. We notice birds chirping in a tree, hear an airplane in the sky or a distant train passing by. The localization of sound sources is an intuitive concept to us, but have we ever thought about how large a sound source appears to us? We are indeed capable of associating a particular size with an acoustical object. Imagine an orchestra that is playing in a concert hall. The orchestra appears with a certain acoustical size, sometimes even larger than the orchestra's visual dimensions. This sensation is referred to as apparent source width. It is caused by room reflections and one can say that more reflections generate a larger apparent source width.
The apparent source width is an important perception in several aspects. It contributes to the perceived quality of concert halls where a large apparent source width is desired. In modern sound reproduction systems, it is attempted to reproduce a natural or convincing perception of apparent source width. Furthermore, apparent source width perception might also reflect a listener's ability to segregate sound sources in a way that large appearing sources might be more difficult to segregate. In analogy to the visual system, the auditory system might operate like an acoustical lens whose performance can be characterized by measurements of the apparent source width.
In this thesis, measurements of apparent source width perception are presented where the influence of various physical parameters was systematically tested. These parameters comprise, for example, the similarity between the two ear signals which is low in reverberant and high in dry listening conditions. Also effects of the frequency content and level of the presented sounds on apparent source width were investigated. Furthermore, the discrepancies in apparent source width perception in hearing-impaired compared to normal-hearing listeners were examined including the impact of hearing-aid signal processing. Finally, an auditory model was developed that predicts apparent source width in the presented measurement conditions. An effort was made to clarify the contribution of the various physical cues to apparent source width perception. The results might be helpful for the design and evaluation of acoustical spaces, sound reproduction systems and personal hearing devices such as hearing aids.

Resumé

Hørelsen hjælper os til at skabe en rummelig fornemmelse af vores omgivelser, især for lydkilder, der ligger udenfor synsfeltet. Vi bemærker fugle i et træ, høre et fly på himlen eller et tog i det fjerne. Lokaliseringen af lydkilder er for os et intuitivt koncept, men har vi nogensinde tænkt over, hvor stor lydkilden bliver opfattet? Det er nemlig muligt at associere en vis størrelse med et akustisk objekt. Forestil dig et orkester, som spiller i et koncerthus. Orkestret bliver opfattet med en bestemt akustisk størrelse, der nogle gang er større end orkestrets visuelle dimensioner. Denne opfattelse går under betegnelsen oplevet kilde bredde (apparent source width). Den er skabt af rummets refleksioner og generelt leder flere refleksioner til en større oplevet kilde bredde.
Den oplevede kilde bredde er en vigtig opfattelse set fra forskellige synspunkter. Den bidrager til den oplevede kvalitet af koncerthuse, hvor en stor oplevet kilde bredde er ønskelig. I højtaler systemer er det målet at gengive en naturlige og overbevisende oplevet kilde bredde. Den oplevede kilde bredde giver desuden oplysninger om, hvor god en person er til at afskille flere lydkilder fra hinanden. Eksempelvis kan det være mere krævende at afskille stort opfattede lydkilder fra hinanden. I analogi til det visuelle system, kan det tænkes at det auditoriske system fungerer som en akustisk linse, hvis egenskaber kan karakteriseres med målinger af den oplevede kilde bredde.
I denne afhandling præsenteres målinger af den oplevede kilde bredde, hvor indflydelsen af forskellige fysiske parameter systematisk er blevet testet i lytteforsøg. Disse parametre omfatter, for eksempel, ligheden mellem de to signaler fra hvert øre, som er lavt i områder med meget rumklang og højt i et lydabsorberende rum. Effekten af frekvensindhold og lydstyrken af de afspillede lyde på den oplevede kilde bredde blev også undersøgt. Forskellen i opfattelse af oplevet kilde bredde mellem normal hørende og hørehæmmede forsøgspersoner blev desuden undersøgt, inklusiv effekten af høreapparatets effektive signalbehandling. Slutteligt blev der udviklet en model, der forudser den oplevede kilde bredde i de testede lytteforsøg. De fysiske parameter blev relateret til opfattelsen af den oplevede kilde bredde og parametrenes bidrag blev kvantificeret. Resultaterne kan bruges til design og evalueringer af rumakustik, højtaler systemer og høreapparater.

Acknowledgments

I am very happy to finally have completed this long journey of my Phd. It was a time where I learned and experienced a lot, both scientifically and also for my personal life.
First of all, I would like to thank my two supervisors Torsten Dau and Tobias May who have been very supportive, inspiring, patient and also understanding during the sometimes unconventional course of my project. Further, I would like to thank all my colleagues at building 352. Especially, Marton Marschall who had been a good friend throughout and helped me as well with countless discussions and feedback on my scientific work. *Dankeschön* to Bastian Epp who helped me getting started on my project. Further, I say *tusind tak* to Alan Wiinberg for his motivating and optimistic feedback, especially during our paper revision. I appreciated as well all inspiring discussions in our 'spatial hearing group'. Thanks also to my office mates, that helped me out, cheered me up and joined me for good laughters. I am also grateful for the usage of 'the couch' for recovering power-napping in my colleagues' office ;) This was accompanied with refreshing chats 'from the couch' with Johannes Zaar, Gusztáv Lőcsei and Jens Cubick. My family deserves a huge hug since they always supported, motivated and advised me in all these years. In thoughts I am with my grandfather Johann Käsbach who would have been proud of my path as an engineer.
Thanks as well to all my friends in Copenhagen. I wave as well to all my friends that I made during my external stay in Montréal in the beginning of this project. I would like to express my appreciation for the support of Myriam Mahias during that time. *Merci beaucoup* to Catherine Guastavino whose research group at McGill university I could join during that stay. My other hand is waving at all the nice people that I met during my stay in San Francisco. *Oh thank you, cheers* to Manual Hahmann, who became a good friend during that time and who I enjoyed working with during a common project.

With this thesis I frame a long period, full of vivid activities. I find the completion of this project very rewarding. This period includes also my various musical activities, especially recordings which I made during my time in Montreal and which mean a lot to me. Thanks to all musicians at Rytmisk Centers Big Band and Oste(r)bæk's 4, especially Alfkil T. Wennermark, who I enjoyed very much playing music together with. *Vielen Dank, tusind tak, merci beaucoup and thank you all!*

Related publications

Journal papers

- Käsbach, J., Wiinberg, A., May, T., Jepsen, M. L. and Dau, T. (**2016**). "The effect of hearing-aid processing on apparent source width perception in normal-hearing and hearing-impaired listeners," Int. Journal of Audiology, *submitted*.

Conference papers

- Käsbach, J., Marschall, M., Epp, B. and Dau, T. (**2013**). "The relation between perceived apparent source width and interaural cross-correlation in sound reproduction spaces with low reverberation," Proc. of German Acoust. Association (DAGA), Merano, Italy.
- Käsbach, J., May, T., le Goff, N. and Dau, T. (**2014**). "The importance of binaural cues for the perception of apparent source width at different sound pressure levels," Proc. of German Acoust. Association (DAGA), Oldenburg, Germany.
- Käsbach, J., May, T., Oskarsdottir, G., Jeong, C. and Chang, J. (**2014**). "The effect of interaural-time-difference fluctuations on apparent source width," Proc. of Forum Acusticum, Krakow, Poland.
- Käsbach, J., Wiinberg, A., May, T., Jepsen, M. L. and Dau, T. (**2015**). "Apparent source width perception in normal-hearing, hearing-impaired and aided listeners," Proc. of German Acoust. Association (DAGA), Nürnberg, Germany.

- Käsbach, J., Hahmann, M., May, T. and Dau, T. (**2016**). "Assessing and modeling apparent source width perception," Proc. of German Acoust. Association (DAGA), Aachen, Germany.
- Käsbach, J., Hahmann, M., May, T. and Dau, T. (**2016**). "Assessing the contribution of binaural cues for apparent source width perception via a functional model," Proc. of the International Congress on Acoustics (ICA), Buenos Aires, Argentina.

Contents

1

General introduction

Modern audio reproduction systems cover a wide range of applications, such as entertainment and teleconferencing systems, virtual sound environments, as well as personal hearing devices, like headphones and hearing instruments. In many applications, there is a desire to correctly or plausibly reproduce a complex acoustic scene, which is, however, challenging to accomplish. Since human listeners are always the users of such systems, perceptual aspects of reproduced sound are essential to consider (Spors et al., 2013). This concerns, e.g., the perceived angle and distance of sound sources but also how the listener perceives the surrounding acoustical space. A room can be perceived as small and dry or as spacious and reverberant. A singer in a dry room is perceived as a point-like source whereas in a large room the voice might appear spatially expanded. The *apparent source width (ASW)* conveys this perception of the sound source's size in the horizontal dimension (Blauert, 1997).
This project investigates the manifold and diverse aspects of ASW perception. This involves considerations of the listening environment, sound perception by the listener as well as hearing devices that interact between the listener and the listening environment. A concert situation is an illustrative example of these three components. The ASW has been widely used as a quality measure in room acoustics (e.g. Ando, 2007). A concert hall is considered to have high acoustic quality when the playing orchestra appears spatially enlarged. In recordings and sound mixes of such an acoustic scene, sound engineers attempt to capture a natural representation of the orchestra's ASW. Furthermore, ASW provides information about the listener's hearing abilities. The individual instruments of the orchestra may be perceived as focussed by one listener, whereas another listener perceives them as spatially blurred. Similarly to the visual system, the auditory system might operate like an *acoustical lens* whose performance can be characterized by measurements of the ASW. Hence, the sensation of ASW may be

related to the listener's ability to discriminate multiple sound sources (Whitmer et al., 2012). Hearing-impaired listeners may have a distorted perception of the concert compared to normal-hearing listeners or experience difficulties, for example, in following a conversation in a crowded reception after the concert (Bronkhorst and Plomp, 1992). A perceptual characterization of ASW might, therefore, provide useful information regarding deficits in the listener's spatial hearing abilities. The information about such hearing deficits might be valuable in the fitting procedure of a hearing aid to a hearing-impaired listener. Ideally, the hearing aid should provide the listener with a natural perception of the acoustical environment such that a concert situation as illustrated here remains enjoyable.

The aim of this thesis was to characterize ASW perception. Several studies have investigated ASW perception either in concert halls (e.g. de Villiers Keet, 1968; Okano et al., 1998; Bradley, 2011), in loudspeaker systems (e.g. Plenge, 1972; Morimoto et al., 1995; Griesinger, 1999; Santala and Pulkki, 2011; Zotter and Frank, 2013) or in headphone presentations (e.g. Chernyak and Dubrovsky, 1968; Blauert and Lindemann, 1986a; Mason et al., 2005a). Listening in concert halls represents a natural listening situation. Room reflections as well as reflections from head- and torso are present leading to fluctuations of binaural cues which are important to consider for ASW perception (Blauert and Lindemann, 1986b). Furthermore, in the case of several concurrent sound sources, the sound of each sound source reaches both ears of the listener, leading to cross-talk. A high degree of realism is challenging to accomplish in headphone presentations which requires binaural simulations including tracking of the listener's head-movements. If binaural simulations are not employed, headphone presentations provide good experimental control, but the signals at the ears of the listener are separated and therefore deviate from natural listening situations. Recent studies have considered the dynamic behavior of binaural cues and their impact on spatial perception (e.g. Catic et al., 2013). However, the role of cross-talk and of dynamic fluctuations of binaural cues on ASW are not fully understood.
In the current thesis, the focus was on loudspeaker-based listening experiments. Specifically, stereo setups were used to elicit a wide range of ASW. This allowed for an easy placement of sound sources in a room reflecting an everyday listening situation. Furthermore, the loudspeaker setup also provided a simple test

environment for evaluating the influence of hearing aids which is not feasible with headphone-based presentation. This work investigated the physical cues that are available to a listener in such listening conditions and related them to the listener's perception of ASW. In particular, the influence of cross-talk, the temporal fluctuations of the binaural signals, as well as monaural measures like frequency content and sound pressure level of the presented sound sources were studied.

In comparison to normal-hearing listeners, spatial perception has been found to be distorted in hearing-impaired listeners (e.g. Van den Bogaert et al., 2006; Boyd et al., 2012; Whitmer et al., 2012; Cubick et al., 2014). Therefore, this work addressed deficits regarding ASW perception in the impaired auditory system. In this context, binaural listening is another important factor in aided listening. Many studies have investigated speech perception and localization performance when wearing hearing aids (e.g. Dillon, 2001), but other spatial attributes, such as ASW, have received less attention. Therefore, this thesis investigated the influence of hearing-aid processing on the perception of ASW in both normal-hearing and hearing-impaired listeners.

The prediction of ASW perception could be useful in the development process and evaluation of hearing devices, as well as in the context of room acoustics and loudspeaker systems in general. A reliable prediction might help to avoid otherwise time consuming listening tests. This requires, however, a thorough understanding of how listeners perceive the ASW of sound sources. Practical measures, such as the interaural coherence, describing the similarity between the two ear signals, have been suggested as predictor of ASW, but lack generalizability across various listening conditions (Bradley, 2011). In the present thesis, a binaural auditory model was used to analyze the role of the investigated physical cues in ASW perception in more detail. This allowed also for predictions of ASW perception in aided compared to unaided conditions.

Chapter 2 of this thesis comprises a summary of human spatial hearing with a specific focus on ASW perception. The binaural cues, interaural time and level differences, are introduced as important localization cues (Plack, 2005), and their dynamic variations in rooms are considered. The interaural coherence is introduced as a measure of the similarity between the two ear signals (Bernstein et al., 1999). A summary of auditory mechanisms encountered when

listening in rooms, e.g. the precedence effect (Blauert and Braasch, 2005), is provided. Relevant perceptual attributes, such as spaciousness (Bradley, 2011), listeners envelopment (Morimoto et al., 2007), distance perception (Zahorik et al., 2005) and externalization (Hartmann and Wittenberg, 1996) are described. Physical and perceptually-based sound reproduction methods are presented (Spors et al., 2013), with a focus on stereophonic loudspeaker reproduction as the experimental method used in this thesis.

Chapter 3 investigates the role of cross-talk, encountered in stereo setups (Blauert, 1997) due to the two extra pathways from each loudspeaker to the contra-lateral ear (Blauert, 1997), on the perception of ASW. The aim was to allow a comparison of ASW data obtained with headphones and with loudspeakers by switching the cross-talk paths on and off in the stimuli presented to the listeners over headphones.

In *Chapter 4* , the influence of three important physical parameters on ASW perception is studied: the interaural coherence (IC), the frequency-content and the sound pressure level. From headphone-based experiments (Blauert and Lindemann, 1986a) it is known that the ASW increases with decreasing interaural coherence. This relation is important because IC is commonly used as a predictor of ASW in room acoustics (Ando, 2007). The relation between IC and ASW is evaluated quantitatively for a stereo loudspeaker setup using band-limited Gaussian noise signals as used in the headphone experiments. Furthermore, it is tested whether the frequency dependence of ASW perception is similar to that found for headphone-based presentations (Blauert and Lindemann, 1986a). The sound pressure level, also mentioned in the literature (Okano et al., 1998) as an important contributor to ASW perception in concert halls, is investigated as well. Furthermore, the perceived vertical extent of the sound source, the apparent source height (ASH), is considered.

Chapter 5 focuses on investigating the contribution of interaural time differences (ITDs) on ASW. ITD fluctuations over time have been suggested as an important cue for ASW perception (van Dorp Schuitman et al., 2013). In an experimental study, binaural recordings are obtained in the listening position of a stereo setup, for each listener individually. A binaural synthesis approach is used to allow for the modification of the interaural time difference fluctuations of broadband Gaussian noise sources whereby the modifications of the interaural time differences are applied in different frequency bands to estimate their frequency-dependent contribution to ASW.

Chapter 6 addresses the consequences of hearing-impairment on ASW perception. Whitmer et al. (2012) and Whitmer et al. (2014) showed that hearing-impaired listeners indicated a roughly constant ASW, independent of the interaural coherence of the presented stimulus signal. Here, normal-hearing and hearing-impaired listeners rate ASW in stereo-loudspeaker setups. The role of hearing-aid processing on ASW is also examined. Specifically, the influence of wide dynamic-range compression is considered since it is commonly applied in hearing-aids to compensate for loudness recruitment in hearing-impaired listeners (Allen, 1996; Villchur, 1973).

In *Chapter 7* two models of binaural hearing, a functional model and a complex non-linear model, are employed to investigate the role of binaural cues on ASW in more detail. A functional model is developed to test the contribution of various binaural cues suggested in literature to predict ASW. The functional model uses either an analysis of the interaural time differences (van Dorp Schuitman et al., 2013), the IC (Okano et al., 1998) or a combination of interaural time and level differences (Blauert and Lindemann, 1986b; Mason et al., 2005b) at the output of the model. The complex model, as suggested by van Dorp Schuitman et al. (2013), includes level-sensitive components in addition to the analysis of the interaural time differences. The model predictions are correlated with experimental data described in Chapters 4 and 6 and one external data set to test the models' generalizability.

Chapter 8 provides a summary of the most important results obtained in this thesis. In an overall discussion, an attempt is made to link ASW perception to other related aspects of spatial perception such as sound source localization, discrimination and segregation. Finally, perspectives for future studies of ASW perception are outlined.

2

Background on spatial hearing

Our visual system helps us navigate through our natural environment. Our hearing system substantially supports the navigation process, especially for objects that occur outside our visual field. For instance, we notice a car that is passing by behind us or awareness can be drawn to distant, unseen sound sources. A fabulous feature of the auditory system is its ability to segregate multiple concurrent sound sources in space. Considering a crowded place, like a bar, where many people are chatting at the same time and also music is playing in the background, the auditory system is capable of following a conversation despite the interfering noise. The underlying auditory processes in such a *cocktail-party* like scenario are complex. This thesis provides some further insights in aspects of spatial hearing, specifically, the perception of apparent source width as a measure of spatial perception. This chapter presents relevant aspects of spatial hearing and introduces the concept of apparent source width perception.

2.1 Introducing apparent source width perception

Most sound sources occur in the horizontal plane relative to the listener. Here, human sound localization relies mainly on two binaural cues, namely *interaural time differences (ITDs)* and *interaural level differences (ILDs)*. Both cues are, however, not static but fluctuate over time. In anechoic conditions, this is due to reflections from the head and the torso of the listener. Therefore, the sound source is not perceived as a point source but has an increased expansion in the horizontal dimension, referred to as *apparent source width (ASW)*. In rooms, the fluctuations are more pronounced such that the correlation between the two ear signals decreases and leads to an increased ASW. The auditory system

is not fast enough (Siveke et al., 2007) to resolve the individual directions of the (early) reflections. Therefore, instead of perceiving multiple punctuate sound sources, the listener perceives one single fused sound object which, however, is locally blurred and elicits the perception of ASW (Blauert and Lindemann, 1986b, Blauert, 1997, Griesinger, 1999).

2.2 Localization in the horizontal plane under free-field conditions

The interaural time and level differences are the most relevant localization cues which also contribute to ASW perception. This section provides some more details on both, ITD and ILD cues. The ITDs range from -1000 to 1000 μs and resemble a monotonic increasing function of the azimuthal angle and are frequency-independent. For ongoing sounds, the ITDs are interpreted as *interaural phase differences* (IPDs). In contrast, ILDs are very small for frequencies below 1 kHz and become more pronounced with increasing frequency up to a magnitude of ±20 dB at high frequencies. This is due to the increasing shadowing effect of the head at high frequencies, where the wavelength becomes comparable to the head's dimensions. These purely physical considerations already suggest that ILDs are only relevant localization cues for the auditory system at high frequencies, at least in anechoic conditions. This consideration led to the duplex theory (Strutt, 1907; Macpherson, 2002) which states that human auditory localization is based on ITDs at low frequencies and on ILDs at high frequencies. For pure tones, ITDs can be exploited up to a cut-off frequency of around 1.5 kHz. The frequency range is limited by the inner-hair cell's capability of phase-locking to the signal's fine structure (Plack, 2005). The effective frequency range of ITDs is further limited to ca. 750 Hz by phase ambiguities in the physical waveform (or spatial aliasing due to the physical distance between the two ears), especially for ongoing sounds. At high frequencies, the auditory system can, however, track the signal's temporal envelope, especially for broadband signals which overcome the phase ambiguities and the frequency limitations due to the break down of phase locking. Therefore, envelope ITDs might complement ILDs at high frequencies as localization cues (Macpherson, 2002).

The human auditory system has remarkable discrimination abilities in terms

of the binaural cues. Just-noticible differences (JNDs) for ITDs measured in normal-hearing listeners (Musa-Shufani et al., 2006) are 30 μs at low (500 Hz) and 80 μs at high frequencies (4 kHz). ILD-JNDs are in the range of 1 dB at both frequencies. The *spatial resolution* of the auditory system can be accessed by measurements of the minimal audible angle (MAA) that is required to differentiate the perceived location of two sound sources (Mills, 1958). The MAA dependents on the signal type and its frequency content. For frontal sources, the MAA is as small as 1° (Mills, 1958) for pure tones and about 3° for broad band noises (Haustein and Schirmer, 1970). The MAA increases for lateral sources at ±90° to about 10° to 20°.
ITDs and ILDs fluctuate over time even in anechoic conditions due to reflections from torso, head and pinna. The auditory system exhibits temporal limitations in following the fluctuations of the binaural cues. This effect is called *binaural sluggishness* (Siveke et al., 2007). The location of a sound source can be tracked by the auditory system for ITD fluctuation frequencies of up to 2 to 3 Hz before the sound image becomes blurred and elicits ASW (Blauert, 1972, Blauert, 1997). This threshold is relatively low in comparison to the remarkable timing accuracy in ITD-JNDs (30 to 80 μs). ILD fluctuations can be followed for variation frequencies of about 79 to 85 Hz (Stellmack et al., 2005). This cutoff frequency is below the one of monaural amplitude modulation detection which is about 116 to 121 Hz (Stellmack et al., 2005).

2.3 Localization and spatial impression in rooms

In contrast to anechoic conditions, room reflections cause pronounced fluctuations of the binaural cues. Figure 2.1 shows mean and standard deviation of the ITD and ILD fluctuations, respectively, in a classroom with a reverberation time of $T_{60} = 0.5$ s. The mean ITDs and ILDs (first row) are both reduced in their extrema and show more ripples as in anechoic conditions. Accordingly, their fluctuations (second row) vary up to 800 ms for ITDs throughout the entire frequency range and around 5 dB for ILDs, especially at low frequencies. This shows that, for the localization of sound sources in rooms, the contribution of ITDs and ILDs across frequency might be different than in anechoic conditions. Generally, the fluctuations of ITDs and ILDs over time are essential for

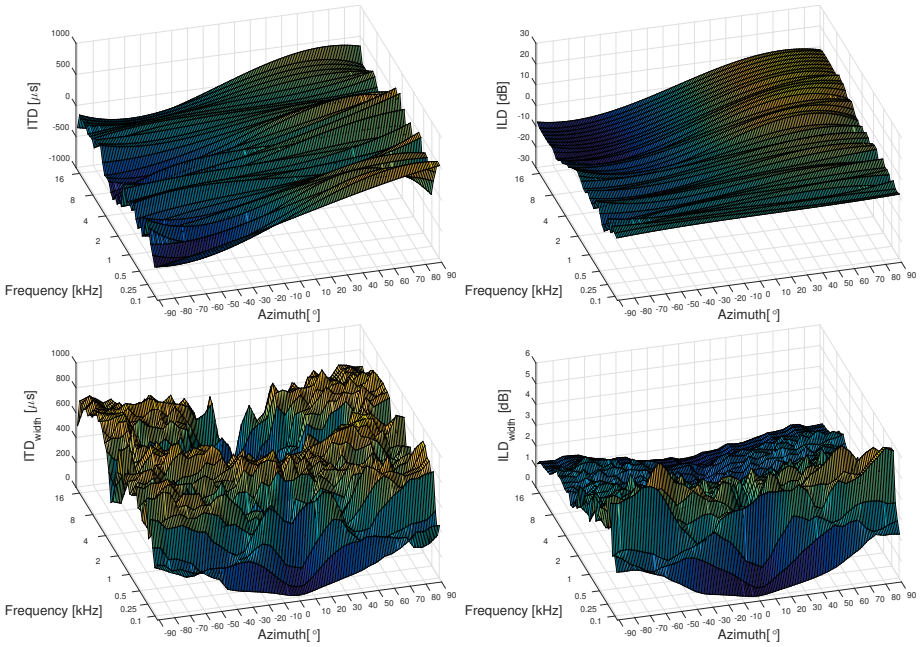


Figure 2.1: ITD (left panels) and ILD (right panels) maps across azimuth and frequency obtained from HATS measurements in a class-room. Third row: Mean ITDs and ILDs. Second row: Standard deviation of ITDs and ILDs. Both cues were extracted from simulations where for each source position a 1 s long white noise signal was convolved with a binaural room impulse response (BRIR) measured with a Cortex Instruments Mk.2 Head and Torso Simulator (HATS) (Hummersone, 2011)

the listener's *spatial impression* of the room (Blauert and Lindemann, 1986b and Blauert and Lindemann, 1986a).

The perceived location of a sound source in a room is primarily determined by the direct sound. The directional information of reflections arriving in the first 80 ms is suppressed by inhibition processes, known as the *precedence effect* or the *law of the first wavefront* (Blauert, 1997). For distinct reflections arriving later than 80 ms, *echoes* will be perceived by the listener. Despite the precedence effect, the direction of early reflections affect the localizability of sound sources. While early reflections from the frontal direction (including ceiling and floor reflections) improve the localization performance, lateral early reflections decrease the localizability of the sound (Hartmann, 1983). Mostly stationary broadband signals are affected but the localization performance generally improves with increasing spectral density of the sound source, i.e. a larger amount of harmonic frequency components per auditory filter (Hartmann, 1983).

In addition to affecting localization performance, room reflections in the precedence window cause an increased *loudness* perception, a *colored* frequency

spectrum, leading to a change in timbre, as well as an increased apparent source width (Blauert and Braasch, 2005). When multiple sound sources are present, e.g. an orchestra, the perceived width of the entire ensemble is described by the *ensemble width* (Rumsey, 2002). Besides the perceived expansion of the sound source in the horizontal dimension, also the perceived height and depth might increase. The sound source can even appear larger than its visual dimensions (Kuhl, 1978). In concert halls, the listener might even feel immersed into the auditory scene. The degree to which the listener feels immersed in the sound is described by the *listener envelopment (LEV)*. Both, the ASW and LEV are considered to compose the overall spatial impression (Bradley, 2011).

It has been claimed that listeners can distinguish the perception of ASW and LEV because they are assigned to two different perceptual streams, the *direct stream* and the *reverberant stream*, respectively (van Dorp Schuitman et al., 2013, Griesinger, 1999). The direct stream indicates sound-source related sensations and is associated with early arriving reflections at the listener's ears within the first 80 ms (Okano et al., 1998) to 150 ms (Griesinger, 1999). This is roughly congruent with the time window of the precedence effect. The reverberant stream represents room related perceptions, like for example, *reverberance* and LEV, and is affected by late reflections, arriving at the listener's ears after about 150 ms.

The perception of distance relies on intensity and high-frequency attenuation in free-field conditions, whereas in rooms, the most important cue is the *direct-to-reverberant (D/R) ratio*. It is defined as the energy of the direct sound and early reflections over the energy of the late reflections. With increasing distance, the level of the reflections increases and becomes dominating compared to the direct sound such that the D/R ratio decreases. In other words, the perception of distance might resemble the ratio of the (internal representation of the) direct and the reverberant stream. The volume and reverberation time of a room determine the *reverberation distance*, which describes the radius around a sound source where the $D/R = 1$, i.e. the energy of the direct sound and late reflections is equal. This is, for example, important for the installation of loudspeaker systems in reverberant environments, e.g. a church, to allow for a reasonable speech intelligibility.

Dynamic binaural cues also contribute to externalization perception, which has been defined as correct localization in terms of angle and distance of compact sound sources with small ASW (Hartmann and Wittenberg, 1996). External-

ization describes natural listening situations where sound sources appear in the real world outside of a listeners head, as opposed to listening situations with, e.g., headphones where sound sources are internalized, i.e. they appear inside the listener's head. Fluctuations of ILDs have been identified as major contributing cue to externalization (Catic et al., 2013).

2.4 Physical measures of apparent source width

In concert hall acoustics, ASW is an important qualitative perceptual attribute (Bradley, 2011). Mostly early lateral reflections are considered as the main contributors to the perception of ASW (Okano et al., 1998). Room reflections generate large fluctuations of the binaural cues, the ITDs and ILDs, and lead to a decrease of the coherence between the two ear signals. Therefore, the perception of ASW is inversely proportional to measurements of the *interaural cross-correlation (IACC)* (Ando, 2007, Schroeder et al., 1974) of early-arriving reflections ($\mathrm{IACC_E}$) which is defined below. In addition, the *early lateral energy fraction* ($\mathrm{LF_{early}}$) is often used as a measure of ASW perception (Bradley, 2011). However, Frank (2013) found that the $\mathrm{LF_E}$ is not suited for measurements in loudspeaker-based reproduction spaces. Therefore, it was not further considered in this thesis.

2.4.1 Definition of the interaural cross-correlation

de Vries et al. (2001) defined the IACC as the correlation between the left-ear signal, $p_l(t)$, and the right-ear signal, $p_r(t)$, normalised with their root-mean-square (RMS) values, calculated with a delay time interval of $|\tau| \leq 1$ ms and using a time window of $t_2 - t_1$:

$$\rho_{lr}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t) p_r(t+\tau) \mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} p_l^2(t) \mathrm{d}t \int_{t_1}^{t_2} p_r^2(t) \mathrm{d}t}} \tag{2.1}$$

Negative and positive values of the $\rho_{lr}(\tau)$ function might lead to different perceptual results. While negative values are associated with a change in distance perception, positive values are associated with a change in ASW perception (Plenge, 1972). Therefore, the *IACC coefficient* corresponds to the absolute maximum of the cross-correlation function $\rho_{lr}(\tau)$ (Blauert and Lindemann, 1986a):

$$\text{IACC} = \max|\rho_{lr}(\tau)| \tag{2.2}$$

In the short-time analysis of the $\rho_{lr}(\tau)$, the *interaural coherence (IC)* is calculated in each time frame as the maximum of the $\rho_{lr}(\tau)$ (Faller and Merimaa, 2004; de Vries et al., 2001):

$$\text{IC} = \max\rho_{lr}(\tau) \tag{2.3}$$

Both, IACC and IC, take values between zero and one. Throughout this thesis, the term IACC will be used as an abbreviation for interaural cross-correlation and the term IC will be used to refer to interaural coherence as a binaural cue (in addition to ITDs and ILDs). The IACC coefficient will be computed according to Eqn. 2.2 and the IC will be computed according to Eqn. 2.3.

2.4.2 Discrimination of IC values

The discrimination of two IC values in normal-hearing listeners dependents on the reference IC value. While for two fully decorrelated white noise signals (IC = 0) a difference of 0.4 can be discriminated, the required change at two fully correlated, i.e. identical, white noise signals (IC = 1) is only 0.04. These results are based on headphone measurements by Pollack and Trittipoe (1959). Using bandlimited noise signals with a center frequency of 500 Hz, Gabriel and Colburn (1981) found that for noise signals with IC = 1, the JNDs were even below 0.04 (broadband JND) when decreasing the bandwidth below 115 Hz. The opposite effect was found for noise signals with IC = 0, where a bandwidth

above 115 Hz led to the smallest JND of 0.4 (as in the broadband case) but the JND increased for smaller bandwidths.

2.5 Localization and spatial impression in loudspeaker-based reproduction systems

The goal of most loudspeaker-based reproduction systems is to obtain a *plausible*, i.e. convincing, representation of an acoustic scene for the listener. In the case that no difference to an internal or explicit reference is audible, the reproduction is *authentic* (Spors et al., 2013). In sound quality evaluations of stereophonic and multiple loudspeaker systems, it was found that timbral fidelity accounts for 70% of the total quality and that approximately 30% of the total quality are assigned to spatial fidelity (Rumsey et al., 2005). This demonstrates the importance of considering perceptual effects in the design of loudspeaker-based reproduction systems (Spors et al., 2013).

2.5.1 Stereophony

Stereophony is still one of the most common reproduction techniques used by sound engineers. It allows for a plausible spatial reconstruction (Spors et al., 2013) of e.g. an orchestra recording. Stereophony is based on the principle of *summing localization* (Blauert and Braasch, 2005), stating that two sounds arriving from the two loudspeakers within 1 ms are fused by the auditory system as long as they are coherent. This generates a single sound image, a *phantom source* or *auditory event*, in front of the listener. This allows, in principle, for a difference in distance of the listener to the loudspeakers of 34 cm (corresponding to 1 ms) to stay in the perceptual time span of summing localization. The optimal listening position, the *sweet spot*, is, however, at an equal distance to both loudspeakers which are typically placed at an opening angle of $\pm 30°$ with respect to the listener. At the sweet spot, frequency-dependent interference patterns between both loudspeaker signals will occur, generating spectral comb-filters (Spors et al., 2013). However, the coloration effects are only marginally audible due to the *binaural decoloration* abilities of the auditory system (Brüggen, 2001). In a listening situation with headphones, the inter-channel differences,

comprising time and level differences as well as the coherence, directly become the respective binaural cues, ITDs, ILDs and IC. In contrast, in a stereo setup *cross-talk* is present, i.e. two additional pathways between each loudspeaker and the contralateral ear. The frequency-dependent interference patterns at the listener's head, including cross-talk, imply that the inter-channel differences based on the loudspeaker signals are different from the binaural cues measured at the listener's ears.

The location of the phantom source on a path between the two loudspeakers can be adjusted by applying time or level differences to the loudspeaker signals (Lipshitz, 1986). Amplitude panning (applying a level difference) is commonly used which generates ITDs below 1.5 kHz at the listener's ears (Leakey, 1959). In comparison, time-delay panning (applying time differences) is considered as a less effective way to control the source location. The phantom source's ASW can be controlled by the opening angle of the loudspeakers, the listener's distance to the setup and the inter-channel differences between both loudspeaker signals, which affects the fluctuations of ITDs, ILDs and IC at the listener's ears (Frank, 2013). The reflections in the listening room will further modify the binaural cues. The control of the ASW for experimental or artistic purposes via the inter-channel differences is signal dependent. For a Gaussian noise signal, the coherence between the two channels can be adjusted between zero and unity with the asymmetric or symmetric two-generator method (Hartmann and Cho, 2011). For natural signals, like music and speech, stereophonic recordings can be applied that directly capture the inter-channel time and level differences. However, for monaural, anechoic recordings, so-called *decorrelation algorithms* (Zotter and Frank, 2013) or *pseudo-stereo* principles (Blauert, 1997) need to be exploited to adjust the inter-channel differences. A high fidelity decorrelation algorithm provides the desired change in ASW without causing severe coloration artifacts (Zotter and Frank, 2013). However, any kind of partial decorrelation of the two loudspeaker signals will affect the degree of interference at the listening position which will consequently change the timbre of the sound (Frank, 2013). The phantom source will be perceptually stable, unless the sound field results in incoherent signals at the listener's ears, leading to *image splits*, i.e. the perception of two sound sources instead of a single one (Blauert and Lindemann, 1986a). For time delays larger than 1 ms and up to 80 ms between both loudspeaker signals, perceptual effects assigned with the afore mentioned precedence region will occur, such as coloration and an increased ASW. Echoes

can be artificially generated for time delays larger than the echo-threshold of 80 ms (Blauert and Braasch, 2005).

2.5.2 Virtual sound environments

In auditory research, virtual sound environments (VSEs) are increasingly used to simulate and study complex auditory scenes related to the cocktail-party problem. Such an environment is composed of a large number of loudspeakers and has the aim of accurately reproducing the complete auditory scene, preferably in an authentic manner. To assess the overall quality and artifacts of VSEs in general, attribute catalogues have been designed covering timbral as well as spatial fidelity aspects (Lindau et al., 2014; Spors et al., 2013). Especially, coloration artifacts and localization errors have been investigated (Spors et al., 2013). ASW perception has also been considered as an important spatial attribute (Santala and Pulkki, 2011, Frank, 2013).
Basically, two reproduction approaches are distinguished in VSE designs, sound-field synthesis (SFS) techniques and reproduction methods that are based on psychoacoustic knowledge. To the first class belong methods like higher order ambisonics (HOA) and wave field synthesis (WFS) that aim at a physical reproduction of the sound field in the center of the loudspeaker array. In contrast to, e.g., a simple stereo setup, their sweet spot is increased to a listening area. However, due to the limited number of loudspeakers in practical realizable setups, artifacts, such as spatial aliasing, occur. Further, both systems use correlated loudspeaker signals, such that the increased number of loudspeakers introduces further coloration artifacts.
The second class of reproduction methods is based on psychoacoustic knowledge. These methods aim at a perceptually convincing reproduction without considering a physically accurate sound field reconstruction. To this class belong, e.g., directional audio coding (DirAC) (Pulkki, 2007) and the spatial decomposition method (SDM) (Tervo et al., 2013). DirAC is based on a decomposition of the sound field into diffuse and non-diffuse components estimated with the short-time Fourier transform (STFT). The non-diffuse components, such as the direct sound, are played back using vector-based amplitude panning (VBAP), which is an extension of stereo to a triplet of loudspeakers (Pulkki, 2007). The diffuse part uses all loudspeakers for playback and aims at a correct spatial impression, especially an authentic ASW perception. A simple, nonetheless

very effective, approach is considered in the SDM method. It estimates the direction of each sound wave in short-time windows which are shifted sample by sample. Each window is panned to the according nearest loudspeaker. Both systems, DirAC and SDM, aim at a high timbral fidelity by reducing coloration artifacts (Pulkki, 2007, Tervo et al., 2013).

2.6 Impact of hearing impairment on spatial hearing

The consequences of hearing impairment (HI) are highly individual. Besides a higher absolute threshold, commonly at higher frequencies, HI listeners suffer from loudness recruitment (Steinberg, 1937; Moore et al., 1996) and impairment of their spatial perception (Van den Bogaert et al., 2006; Boyd et al., 2012; Whitmer et al., 2012; Cubick et al., 2014). Especially, in multi-talker environments, HI listeners often have difficulties to segregate sound sources which impairs the intelligibility of a target talker (Bronkhorst and Plomp, 1992; Peissig and Kollmeier, 1997). This might be linked to a reduced localization performance (Van den Bogaert et al., 2006), impaired distance (Cubick et al., 2014) and externalization perception (Boyd et al., 2012) in HI listeners. Furthermore, some focus has recently been directed to the reduced sensitivity in ASW perception in HI listeners (Whitmer et al., 2012, Whitmer et al., 2014).

2.7 Choice of methods in this thesis

In psychoacoustic experiments, there is always a trade-off between realism and controllability of the perceptual attribute that is sought to be measured. On the one hand, headphone experiments offer the highest level of control of the ear signals, but the obtained results might not easily translate from the laboratory to real listening scenarios. On the other hand, measurements in virtual sound environments allow for a plausible representation of the acoustic scene, but exhibit challenges in isolating a single perceptual attribute from others. Regarding the measurements of ASW perception, both approaches have been successfully used in the literature (Pollack and Trittipoe, 1959, Blauert and Lindemann, 1986a, Plenge, 1972, Favrot and Buchholz, 2010). Specifically, a drawback of headphone experiments is that the sound image may be internalized, i.e.

perceived inside the head if not individualized head-related transfer function (HRTF) are used. In the case of an internalized sound image, the measurements of ASW might not correspond to measurements where the sound is externalized as in realistic listening situations. In addition, headphone experiments do not allow for the testing of hearing devices such as hearing aids. Therefore, it was decided to use a loudspeaker-based reproduction for the ASW measurements. In preliminary listening tests, a VSE was used in combination with the LoRA toolbox and the room-modeling software ODEON (Favrot and Buchholz, 2010). The amount of early reflections was used to control the IC at the listener's ears and, consequently, the ASW. This was achieved by truncating the room impulse-response that was convolved with the source signal. In this scenario, also the height and depth of the sound source changed and reverberation was audible as well. Thus, for the studies presented in this thesis, an intermediate step was chosen, using a stereo loudspeaker setup. The simplified setup was considered in an attempt to have a better control over ASW, by simultaneously ensuring an externalized sound image.

3

The influence of cross-talk on apparent source width [a]

Abstract

The interaural cross-correlation (IACC) has been proposed as an objective measure of the apparent source width (ASW). This relation has been well-established in headphone-based experiments and IACC is commonly used to estimate ASW in reverberant spaces. However, in low-reverberation environments of sound reproduction systems, this relation is less clear. The present study investigates such a case in detail. The ASW of a band-limited noise signal with varying inter-loudspeaker cross-correlation was subjectively evaluated for a typical stereo-setup in a low-reverberation playback room. Consistent with results from earlier studies, the perceived ASW was found to increase monotonically with decreasing inter-loudspeaker correlation of the noise. The IACC, evaluated at the ear canals of a dummy head, however, predicted a saturation of the ASW from medium to low inter-loudspeaker correlation coefficients. These discrepancies resulted from the interaural cross-talk in the stereo-loudspeaker setup, leading to an ambiguity in the IACC calculation. Such ambiguities were absent in headphone presentation and have a negligible effect in reverberant environments. The results provide constraints for the applicability of the IACC as a measure of ASW in sound reproduction and synthesis systems, and in terms of computational models for the estimation of ASW in room acoustics.

[a] This chapter is based on Käsbach et al. (2013) (Proc. of DAGA, Merano, Italy, 2013).

3.1 Introduction

Apparent source width (ASW) is a perceptual attribute that describes the perceived width of a sound image. The interaural cross-correlation (IACC) is commonly used in room acoustics as an objective measure for ASW. Early reflections in a room cause a decorrelation of the two ear signals, i.e., a reduction of IACC, which leads to a larger ASW. In a dichotic listening condition using headphones, the ASW changes when varying the correlation between the two channels. Blauert and Lindemann (1986a), showed for bandlimited noise signals that the ASW increases with decreasing inter-channel correlation (IC). For a given bandwidth, the ASW was found to increase with decreasing center frequency of the noise while keeping the IC constant. In such conditions, the sound image is internalised, i.e. perceived inside the head. Since ASW relates to the spatial extent of a sound source, this measure becomes more meaningful for externalised sources, such as produced by two loudspeakers. A phantom source is then perceived in the middle of the two loudspeakers (Plenge, 1972). In contrast to the headphone (HP) presentation, the loudspeaker (LS) presentation produces an interaural cross-talk (CT) referring to the cross paths between left LS and right ear and vice versa. The present study investigates the influence of the interaural cross-talk on the perception of ASW in rooms with low reverberation that are typically used for sound reproduction systems and virtual sound environments.

3.2 Method

Three experiments were performed: 1) LS presentation in a listening room; 2) HP presentation of the same listening room using head and torso simulator (HATS) recordings and 3) HP presentation of an anechoic listening condition using a HRTF database.

Apparatus

The experiments were conducted in the Immersive Presence Lab at CIRMMT, McGill University, with a reverberation time of $T_{60} = 0.2$ s. A typical stereo-setup

with an opening angle of $\alpha = 30°$ was installed using Genelec 8030A loudspeakers. The listener was seated at a distance of $d = 2.3$ m (Experiment 1). Room impulse responses (IR) were measured for each loudspeaker in the same room with a B&K head and torso simulator (HATS) of Type 4100 placed at the listener's position. The signals were convolved with the IRs and presented to the listener via headphones. The resulting stimuli were then either presented including or excluding the interaural cross-talk (Experiment 2). For the anechoic listening condition, the CIPIC HRTF database (CIPIC, 2004) was used (Experiment 3).

Stimuli and experimental procedure

The stimuli were bandlimited noises with center frequencies of $f_c = 0.25$kHz or 1 kHz and a bandwidth of $\Delta f = 1.5$ octaves. The signals were generated from a 2-dimensional multivariate normal distribution where each dimension corresponded to one loudspeaker channel. The bandpass filter was a digital 4th-order Butterworth filter with 24 dB/octave roll-off. The signals were amplitude modulated with a modulation depth of 60% and a modulation frequency of 8 Hz. All stimuli were presented with constant spectral density of 35dB/Hz (at IC = 0) and had a duration of 4s. The IACC was calculated according to Eqn. 2.2 for the same stimuli as presented to the listeners, i.e. the binaural room impulse responses, measured with the HATS, were convolved with the noise signals. The integration time window $t_2 - t_1$ of Eqn. 2.1 was chosen to be the duration of the stimuli.

The experiments were separately tested using a Multi Stimulus test with hidden reference and anchor (MUSHRA) (ITU-R BS.1534-1, 2003), excluding anchor and reference. In Experiment 1, the ASW was measured for three inter-channel correlation (IC) values of 0, 0.6 and 1 at the two center frequencies of 0.25 and 1 kHz. In Experiments 2 and 3, the two conditions with and without interaural cross-talk were considered for the same noise signals and ICs. Subjects were asked to rate the stimuli relative to each other in terms of ASW on a scale from 0 (narrow) to 100 % (wide). In addition, the task was to identify the narrowest and the widest stimulus with 0 and 100, respectively. The stimuli were presented in random order and repeated six times. Thirteen normal-hearing subjects participated in Experiments 1 and 2, and six of these participated in Experiment 3. The results were analysed with a 3-way ANOVA with the null hypothesis that

all stimuli produced the same ASW. The three factors were stimuli, subjects and repetitions. The post hoc analysis were based on Least Significant Difference (LSD) bars. Differences were considered significant for $p < 0.05$.

3.3 Results

The results of the three experiments are shown in Figures 3.1 to 3.3. The measured ASW scores (open symbols) are indicated as median values and 25th and 75th percentiles for the three considered IC values (0, 0.6 and 1). The corresponding interaural cross-correlation coefficients were calculated and are represented as 1-IACC. They are indicated as filled symbols connected by solid lines in the figures for comparison with the measured ASW scores.

The results from Experiment 1 are shown in Figure 1. The measured ASW scores were similar for the two noise bands centered at 0.25 kHz (open circles) and 1 kHz (open squares). At both frequencies, ASW decreased with increasing IC. At both frequencies, a comparable dynamic range (of $0.8-0.9$) was obtained, representing the difference between the largest and smallest values. In contrast to the data, the calculated IACC values were different for the two bands. Larger IACC values were produced at 1 kHz than at 0.25 kHz. Furthermore, a reduced dynamic range of only 0.2 was obtained for the low-frequency band, compared to the high- frequency band with a dynamic range of 0.4. The statistical analysis revealed no significant differences for the two frequency bands for IC $= 0$ and 0.6, but the difference for IC $= 1$ was significant.

Figure 3.2 shows the measured data from Experiment 2 obtained in the conditions with cross-talk (open squares) and without cross-talk (open circles) using the HRTFs recorded in the CIRMMT lab. The results for 0.25 kHz are shown in the top panel and the results for 1 kHz are represented in the lower panel. The data show that, in the case of cross-talk, lower ASW scores were obtained than in the absence of cross-talk, for all specified IC values. The statistical analysis revealed that there was a significant difference between the two conditions at both frequencies. The median of the ASW scores was found to be shifted downwards by the same amount of about 0.4 in both conditions. Thus, the dynamic range for the ASW values as a function of IC was similar with and without cross-talk which was the case for both frequency bands. The only exception was the

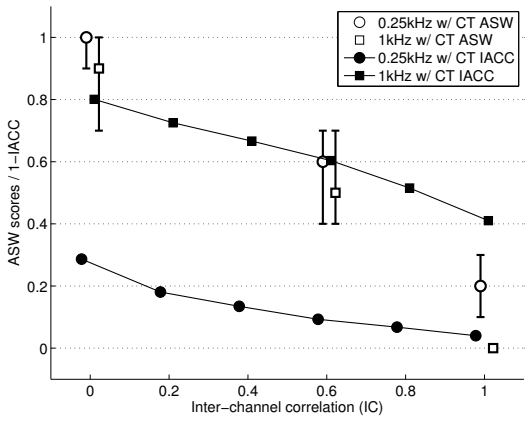


Figure 3.1: ASW scores and calculated 1-IACC values as a function of the inter-channel correlation in Experiment 1. Two frequency bands centered at $f_c = 0.25$ and 1 kHz were considered in the CIRMMT lab, presented via two loudspeakers. The subjective results are represented by their median and 25th and 75th percentiles.

high-frequency band for IC = 1 where a similar ASW score was observed with and without cross-talk. In contrast, the IACC predictions, indicated by the filled symbols connected by solid lines, resulted in a shallower curve in the condition with cross-talk (filled squares) than in the condition without cross-talk (filled circles). At 1 kHz, this resulted in an intersection of the two IACC curves at IC = 0.4.

Figure 3.3 shows the results from Experiment 3 for the two conditions with and without cross-talk in the anechoic listening environment. The results were similar to those obtained in Experiment 2 and the statistical analysis revealed that there was a significant difference between the two conditions at both frequencies. Lower ASW scores were obtained when the cross-talk was present. The difference between the two conditions was, however, slightly smaller than in Experiment 2. The IACC predictions for the 1-kHz band provided similar values for ICs between 0.2 and 0.6 and decreasing values for outside this IC range. This is in contrast to the measured ASW values that showed a monotonic decrease with increasing IC.

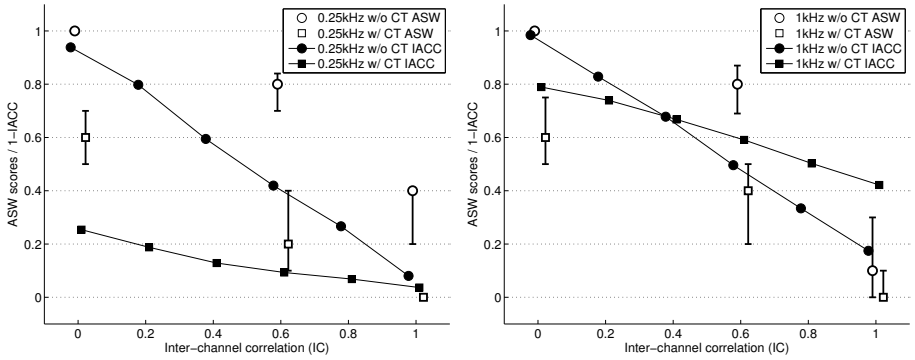


Figure 3.2: ASW scores and 1-IACC predictions as a function of IC in Experiment 2. The two conditions with and without cross-talk were presented via headphones in the CIRMMT lab using HATS recordings, for two frequency bands. The measured data are represented by their medians and 25th and 75th percentiles.

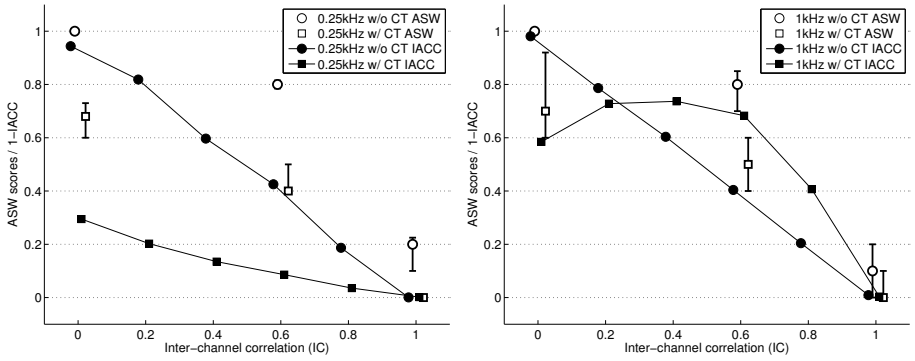


Figure 3.3: ASW scores and 1-IACC results as a function of IC in Experiment 3. The two conditions with and without cross-talk were presented via headphones in an anechoic listening environment using the CIPIC database (CIPIC, 2004), for two frequency bands. The measured data are represented by their medians and 25th and 75th percentiles.

3.4 Discussion

The results demonstrated that the IACC predicted a reduced dynamic range of values in the cross-talk conditions whereas the measured ASW scores showed a similar dynamic range in the conditions with and without cross-talk and a downward shift in the cross-talk condition. This discrepancy from the data can be explained by considering the effect of cross-talk on the cross-correlation function $\rho_{lr}(\tau)$ from Eqn. (2.1) as illustrated in Figure 3.4 (left panel). $\rho_{lr}(\tau)$ is shown for a broadband signal for five inter-channel correlation values. Besides the peak at zero lag which varies with IC, two additional constant peaks occur

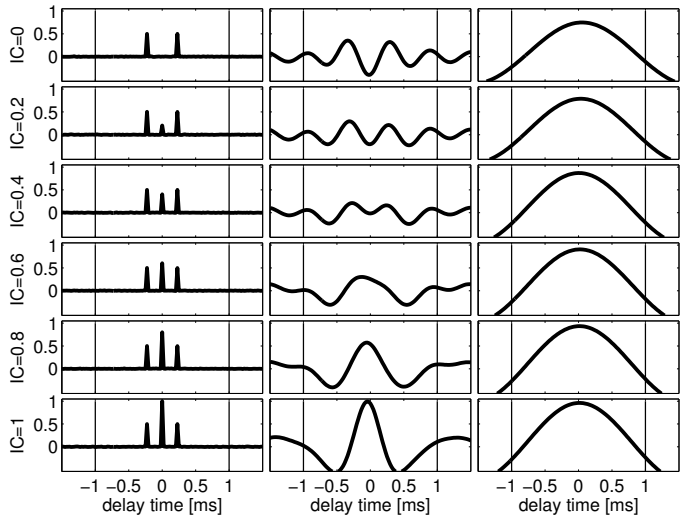


Figure 3.4: The normalized cross-correlation function $\rho_{lr}(\tau)$ (2.1) in the conditions with cross-talk, for five values of IC and three stimulus conditions. Left: for a broadband signal; middle: for the noise band centered at 1 kHz from Experiment 3; and right: for the noise band centered at 0.25 kHz from Experiment 2. The maximal delay time interval of $|\tau| \leq 1$ms is indicated by the vertical lines.

at time lags corresponding to the angular offset of the loudspeakers from the median plane. Due to this offset, the signal from the left LS will reach the left ear before the right ear, and vice versa for the right LS. For low correlation values between the two LS signals (from IC = 0 to IC = 0.4), the peaks caused by the delay are larger than the actual specified IC value (peak at zero lag). According to Eqn. (2.1), the IACC coefficient corresponds to the maximum of $\rho_{lr}(\tau)$, which is equal to the value of the side peaks unless the peak at zero lag becomes larger. A monotonically decreasing 1-IACC with increasing IC is thus only resulting in the absence of cross-talk. When bandlimiting the signal, the peakwidths increase and interfere with each other. The middle panel of Figure 3.4 shows $\rho_{lr}(\tau)$ for the 1-kHz frequency band considered in Experiment 3. The pattern shows a minimum at IC = 0 for a delay time of 0ms and a roughly constant maximum value (defined by the two side peaks) for ICs between 0.2 and 0.6, consistent with the corresponding 1-IACC function shown in Figure 3.3 (filled squares in the lower panel). For the low-frequency band at 0.25 kHz (right panel of Figure 3.4) from Experiment 2, the width of the peaks increase such that they are not distinguishable from each other and result in one large peak. This explains the reduced range 1-IACC values shown in Figure 3.2 (filled squares in the upper panel).

The outlined differences between data and predictions based on the IACC suggest that other cues might be available to a listener. More advanced models, such as the one proposed in van Dorp Schuitman (2011), suggest that the stimulus energy below 500 Hz and the variation of interaural time differences above 500 Hz provide important cues for the perception of ASW. The analysis of these cues was outside the scope of the present study. However, the listeners mentioned a difference in timbre between stimuli with different ASW. A stimulus with a "brighter" timbre was often associated with a narrower sound source, such as in conditions with high IC values, particularly in the headphone-based Experiments 2 and 3. Furthermore, image splitting was reported which refers to the perception of two or multiple sources (Blauert and Lindemann, 1986a), in contrast to one fused sound image. While in the LS-based experiment (Experiment 1) only one listener reported image splitting, six listeners in Experiment 2 and two listeners in Experiment 3 reported this effect. In case of image splitting, subjects based their judgement on the overall ASW as they reported after the experimental procedure. Most subjects mentioned that the task of evaluating ASW was easiest for the LS presentation in Experiment 1. Nevertheless, it should be noted that the measured data showed remarkably stable results obtained with the chosen method.

3.5 Summary and Conclusion

A stereo set-up was used to study the perception of ASW for partially correlated bandlimited noise signals between the two loudspeakers and to test the performance of the IACC that is commonly used to predict ASW. The following results were obtained: 1) In a listening environment with low reverberation, ASW was found to be frequency independent for ICs of 0 and 0.6, in contrast to the case of dichotic HP presentations as investigated in Blauert and Lindemann (1986a), whereas a slight frequency effect was observed for IC = 1; 2) With and without cross-talk, the ASW decreased monotonically with increasing IC. The cross-talk generally caused a smaller ASW than the conditions without cross-talk, but the dynamic range of the ASW ratings was preserved. This was the case both for the listening room with a low reverberation time and the anechoic listening condition and was found to be independent of frequency. In the absence of cross-talk, the predictions based on the IACC were consistent with the measured

data. However, in the presence of cross-talk, the IACC did not describe the data correctly. The change of ASW with IC was strongly underestimated at 0.25 kHz. At 1 kHz, the IACC based prediction did not account for the monotonic decrease of ASW with increasing IC. The discrepancies are caused by the interferences in the cross-correlation function introduced by the cross-talk delays. More advanced predictors are needed to correctly account for the measured data.

Acknowledgments

The author would like to thank Bastian Epp and Marton Marschall for their significant contributions to this research.

4

The influence of interaural coherence, frequency and sound pressure level on apparent source width[a]

Abstract

In concert halls, sound sources are perceived as spatially extended whereby the apparent source width (ASW) describes the extend in horizontal direction. It is mainly caused by a decorrelation of the two ear signals, i.e. the ASW increases with decreasing the interaural coherence (IC). Blauert and Lindemann (1986a) showed in headphone-based experiments that for a constant IC, low frequencies are perceived with a larger ASW than high frequencies. Further, Okano et al. (1998) claimed for concert hall acoustics that, with the increase of the monaural sound pressure level (SPL), the ASW increases as well. The objective of this study was to measure the ASW as a function of the IC and the frequency content by using noise signals presented in a stereo setup. The dependency on the sound pressure level, i.e. the salience of the auditory cues at different SPLs was thereby taken into account. The obtained results confirmed the findings in the literature and will be important for testing auditory models that aim at reliably measuring ASW.

[a] The results of this chapter were presented in Käsbach et al. (2014a) (Proc. of DAGA, Oldenburg, Germany, 2014).

4.1 Introduction

Regarding room perception, apparent source width (ASW) is an essential measure that describes the perceived spatial extent of a sound source. Mainly three important factors have been mentioned that contribute to this percept: (i) The degree of correlation of the two ear signals (IC_{ears}), whereby a decorrelated signal causes a large ASW, (ii) the frequency-content, i.e. low-frequency sounds are perceived as being wider than high-frequency sounds for a fixed IC_{ears} value (Blauert and Lindemann, 1986b) and (iii) the sound pressure level (SPL) at low frequencies, i.e. ASW increases with increasing SPL (Okano et al., 1998). However, the exact contributions of these individual cues to the complex percept are still unclear. This study presents a psychoacoustic evaluation where the three parameters IC_{ears}, frequency-content and sound pressure level were varied in order to quantify their influence on ASW under controlled conditions. Besides the horizontal dimension, potentially also the height and depth of the sound source can be perceived as expanded. In addition to the ASW, also the apparent source height (ASH) was measured.

4.2 Method

4.2.1 Loudspeaker setup

The experiment was conducted in the Spacelab at the Technical University of Denmark, with a reverberation time of $T_{60} = 0.2$ s. A loudspeaker-ring with 11 loudspeakers was placed in the horizontal plane, ranging from $-75°$ to $75°$ with a spacing of $15°$ to each other (see Fig. 4.1). All loudspeakers were of type Dynaudio BM6. Only the two loudspeakers at $\alpha = 30°$ (typical stereo-setup) were used to play back the stimuli. One additional loudspeaker at $0°$ was used for a reference condition. The listener was seated at a distance of $d = 1.8$ m to the loudspeaker setup. For further analysis of the presented signals, binaural room impulse responses (BRIRs) were measured with a B&K head and torso simulator (HATS) of Type 4100, placed at the listener's position, and were convolved with the signals.

4.2.2 Stimuli

The source signal was either Gaussian white noise or band-pass (BP) filtered Gaussian noise with a bandwidth of 2 octaves at a center frequency of 0.25 kHz, 1 kHz or 4 kHz. In addition, an 8 kHz high-pass (HP) filtered Gaussian noise signal was used. The bandpass and highpass filters were both digital Butterworth filters of eighth and fourth order, respectively. The two loudspeaker signals had a duration of 2 s and were generated from a 2-dimensional normal distribution where the covariance was adjusted to yield inter-channel correlation values (IC_{LS}) of 0, 0.3, 0.6, 0.8 and 1. Due to the influence of the head, cross-talk and the listening room, the corresponding interaural cross-correlation values (IC_{ears}) at the ears of the HATS deviated from IC_{LS} and are listed in Table 4.1. The stimuli were presented at three sound pressure levels of 50, 60 and 70 dB within limits of ±2 dB variations among the IC_{LS} values.

Table 4.1: Measured IC_{ears} values for corresponding IC_{LS} values at all center frequencies (cf).

	IC_{LS}				
cf [kHz]	0	0.3	0.6	0.8	1
0.25 (BP)	0.63	0.75	0.83	0.89	0.93
1 (BP)	0.30	0.41	0.57	0.67	0.77
4 (BP)	0.27	0.32	0.53	0.66	0.80
white	0.27	0.30	0.49	0.62	0.79
8 (HP)	0.30	0.29	0.44	0.58	0.73

4.2.3 Procedure

In order to measure ASW, listeners were asked to indicate the apparent opening angle in degrees on a horizontal scale with a 5° resolution of each presented stimulus via a touchscreen as indicated in Fig. 4.1. At the same time, the listener had also to indicate the ASH in cm on a vertical scale with a 10 cm resolution, which was centered at the midpoint between the tweeter and bass-driver of the center loudspeaker. Since measurements with decorrelated signals can lead to image splits, which is the perception of two separate sound images instead of a single fused one, listeners had to indicate such event for the presented stimulus by marking a check box. The stimuli were presented in random order and repeated three times. A reference was available to the listener which was

a broadband noise signal at 45 dB SPL presented over the loudspeaker in the center. This produced a very narrow source with an opening angle of about 0°. Both, stimulus and reference, could be played back as often as desired by the listener. Thirteen normal-hearing listeners participated (leading to 39 responses per stimulus), of which seven evaluated the additional condition with the highpass filtered noise signal (leading to 21 responses per stimulus). All listeners were made familiar with the evaluation procedure in a short training and the entire evaluation procedure lasted about one hour per subject.

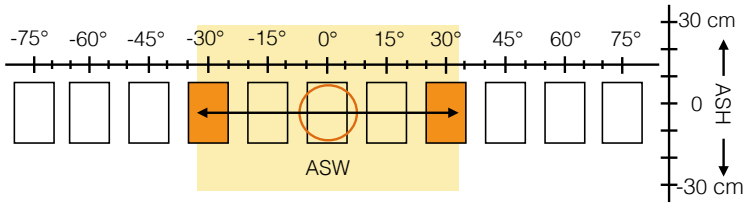


Figure 4.1: Sketch of the experimental set-up. The loudspeakers at ±30° generate a phantom source at 0°. Subjects were asked to indicate ASW in degree and ASH in cm on the given scales.

4.2.4 Statistical analysis

For the statistical analysis, the obtained data was fitted with a linear mixed-effects model. The data-fit comprised the three fixed factors, the IC_{LS}, the source signal and the SPL, and the two random factors, participants and repetitions. The overall ASW, given as the difference between the left and right boundary, was used as response variable. An additional linear mixed-effects model was fitted to the data using ASH as response variable. A post-hoc analysis was performed using pairwise comparisons with Bonferroni corrections, with a correction factor of $c = \sum_{l=1}^{L-1} l$, where c represents the number of comparisons between the L levels of each considered fixed factor in the mixed model (which will be specified in the results section together with $p_{posthoc}$). The occurrence of image splits was simply counted per stimulus across all presentations and listeners.

4.3 Results

The psychoacoustic data are shown in Fig. 4.2 for the three sound pressure levels 50 dB (bottom panel), 60 dB (middle panel) and 70 dB (top panel). The ordinate shows the measured ASW (in degrees) as a function of the measured IC_{ears} values obtained with the HATS on the abscissa. Shown are the mean values with their standard deviations.

4.3.1 ASW as a function of IC_{LS} and frequency

The data at 50 dB SPL (bottom panel) shows that ASW increases monotonically, from ±10° to ±35°, with decreasing IC_{ears} for all tested frequencies, represented by the different colors, as expected from literature. It is important to note that even for stimuli with frequency components above 2 kHz, the listeners were able to discriminate ASW. In this graphical representation of the IC_{ears} values (the exact values can be read in Table 4.1) it is prominent that low frequencies ($IC_{ears} = 0.6...1$), as opposed to high frequencies ($IC_{ears} = 0.3...0.8$), provide higher IC_{ears} values and a smaller dynamic range of IC_{ears}. This is due to the larger wavelength of low frequencies which defract around the listener's head and cause the two ear signals to be similar. On the other hand, high frequencies have a shorter wavelength compared to the head's dimensions and get reflected such that they cannot be fully correlated (see Lindevald and Benade, 1986). Besides the physical consequences due to frequency, two frequency-dependent effects on ASW can be observed: First, considering a fixed IC_{ears}, for instance at $IC_{ears} \approx 0.6$, ASW decreases with increasing center frequency. At 50 dB SPL (bottom panel), the ASW is ±35° at 0.25 kHz and reduces to ±15° at 8 kHz. Second, the dynamic range of ASW, representing the difference between the largest and smallest values, also decreases with center frequency. Interestingly, the broadband signal provided similar results as the high frequency stimuli. This indicates that the high-frequency content is dominating the overall ASW in this case. The statistical analysis revealed the two fixed factors, IC_{LS} ($F(4, 48) = 113.6$, $p < 0.001$) and source signal ($F(4, 42) = 97.2$, $p < 0.001$) to be significant, as well as an interaction between both factors ($F(16, 2436) = 21$, $p < 0.001$). The post-hoc analysis supported the finding that the listeners could differentiate the ASWs for the presented IC_{LS} ($c_{IC_{LS}} = 10$, $p_{posthoc} < 0.01$).

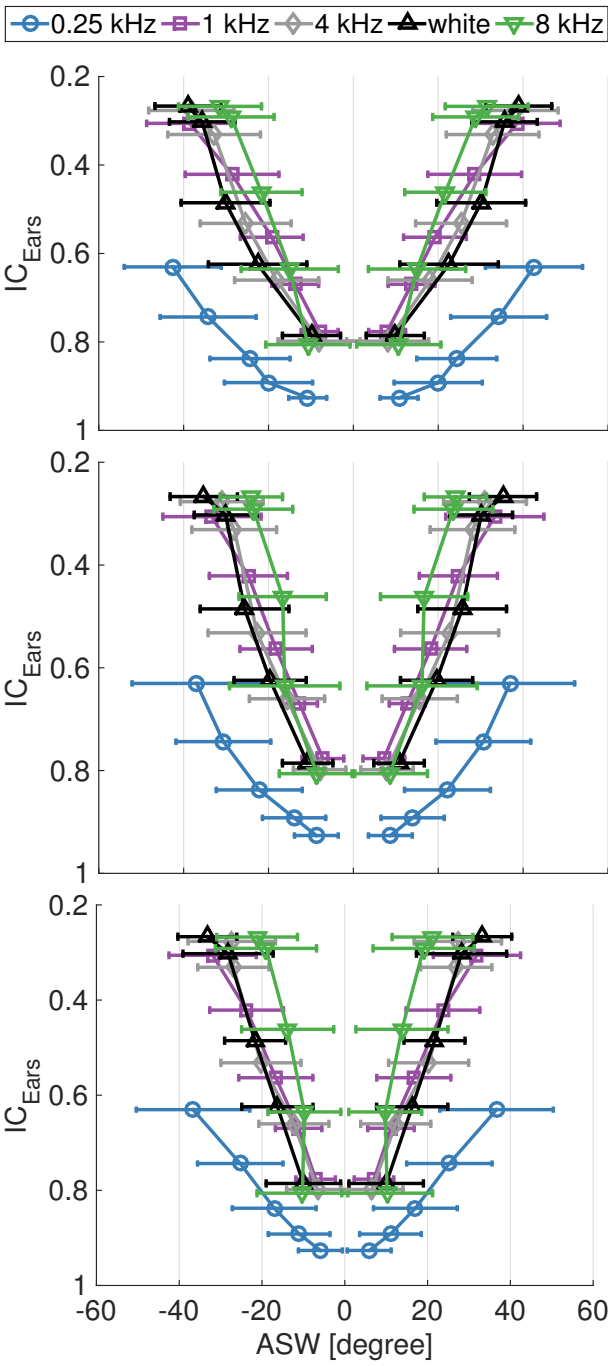


Figure 4.2: ASW data measured in degrees as a function of IC_{ears} for three different SPLs: 50 dB (bottom), 60 dB (middle) and 70 dB (top). Shown are the mean values and standard deviation for the Gaussian white noise, the BP filtered noise signals with center frequencies $f_c = 0.25$, 1 and 4 kHz and the HP filtered noise signal at 8 kHz.

The post-hoc analysis for the source signals showed (non-significant) similarities between the signals at 1 kHz and 4 kHz ($c_{signal} = 10$, $p_{posthoc} = 0.84$), between the white noise and the signal at 0.25 kHz ($c_{signal} = 10$, $p_{posthoc} = 0.08$), and between the white noise and the signal at 4 kHz ($c_{signal} = 10$, $p_{posthoc} = 0.06$).

4.3.2 ASW as a function of the SPL

Increasing the sound pressure level to 70 dB (top panel) resulted in a larger ASW ($c_{SPL} = 3$, $p_{posthoc} < 0.001$). This is easier to see in Fig. 4.3 where the ASW data is plotted across the different SPLs for each source signal, separately. At stimuli with $f_c = 0.25$ kHz (bottom panel), all ASW results are equally shifted outwards by about 10°, such that the dynamic range of ASW is preserved. For the broadband signal (middle panel) and the HP filtered stimulus (top panel), the higher SPL increased the ASW by about 10° at large ASW values, whereas the smallest ASW value was not affected, such that the dynamic range of ASW was increased. These results suggest that the increase of SPL affects ASW in a broad frequency range, which complements the findings in Okano et al. (1998) where this effect was only observed at low frequencies. The fixed factor SPL ($F(2, 25) = 78.8$, $p < 0.001$) was found to be significant, as well as its interaction with IC_{LS} ($F(8, 2427) = 6.8$, $p < 0.001$) and its interaction with the source signal ($F(8, 2408) = 8$, $p < 0.001$). The interaction between all three fixed factors was not significant ($F(32, 2427) = 1.6$, $p = 0.017$). Furthermore, significant effects were found for the random factor participants ($\chi^2 = 20.7$, $p < 0.001$), as well as its interaction with the repetition factor ($\chi^2 = 373.6$, $p < 0.001$), an interaction with the IC_{LS} $\chi^2 = 322$, $p < 0.001$), an interaction with the source signal ($\chi^2 = 23.4$, $p < 0.001$), and an interaction with the SPL ($\chi^2 = 15.2$, $p < 0.001$).

4.3.3 ASH measurements

Measurements of the apparent source height (ASH) varied from 0 to 40 cm (not shown here). In contrast to the ASW measurements, the ASH was independent of the IC_{LS} values. This was confirmed by the statistical analysis that revealed a small effect size of IC_{LS} ($F(4, 2476) = 4$, $p < 0.01$). The changes in ASH were rather dependent on the SPL ($F(2, 25) = 45.4$, $p < 0.001$).

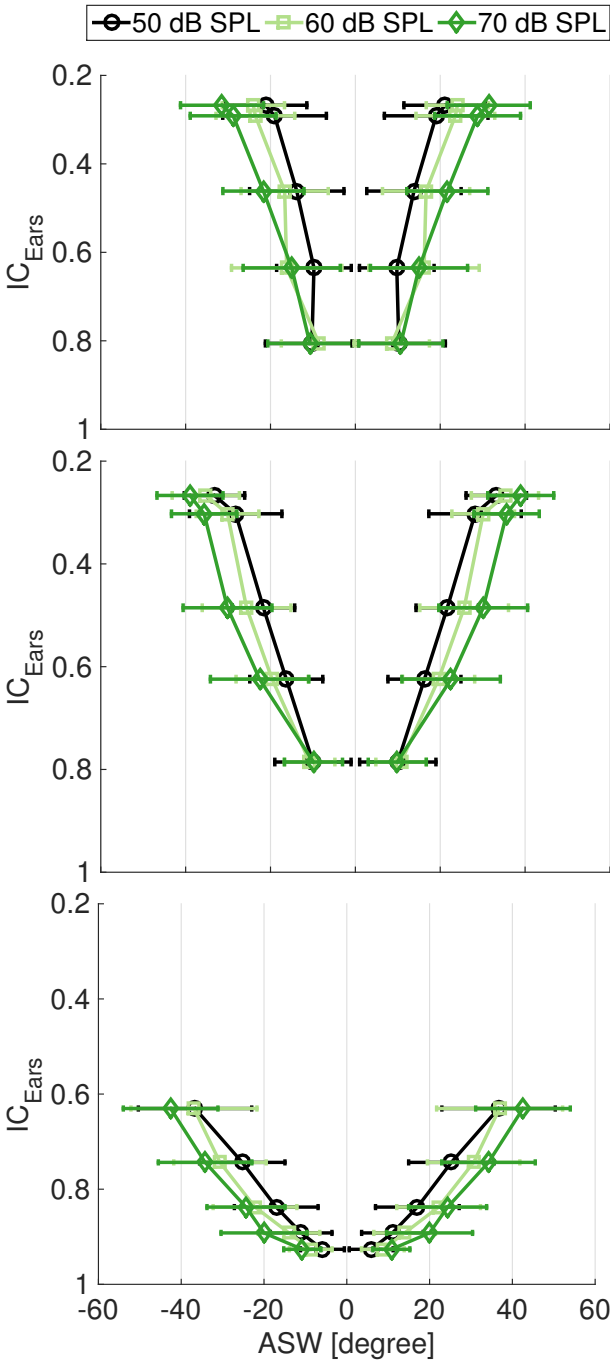


Figure 4.3: ASW data measured in degrees as a function of IC_{ears} for three source signals: 0.25kHz (bottom), Gaussian white noise (middle) and HP 8 kHz (top). Shown are the mean values and standard deviation for the the three measured sound pressure levels.

Similar to the ASW measurements, the ASH increased with increasing the SPL ($c_{SPL} = 3$, $p_{posthoc} = 0.01$). Also the source signal ($F(4, 42) = 24.5$, $p < 0.001$) had a similar effect on ASH, such that ASH was increased for the white noise and the BP filtered signal at 0.25 kHz as opposed to the high frequency stimuli ($c_{signal} = 10$, $p_{posthoc} < 0.02$). Besides an interaction between IC_{LS} and the source signal ($F(16, 2476) = 3.7$, $p < 0.001$), interactions were found between IC_{LS} and the SPL ($F(8, 2476) = 2.9$, $p < 0.01$), the source signal and the SPL ($F(8, 2482) = 9.1$, $p < 0.001$), as well as an interaction term between all three fixed factors ($F(32, 2476) = 4.1$, $p < 0.001$). Furthermore, significant effects were found for the random factor participants ($\chi^2 = 16$, $p < 0.001$), as well as its interaction with the source signal ($\chi^2 = 135.4$, $p < 0.001$), its interaction with the SPL ($\chi^2 = 58.9$, $p < 0.001$) and its interaction with the repetition factor ($\chi^2 = 435.6$, $p < 0.001$).

4.3.4 Occurrence of image splits

At an SPL of 70 dB, image splits occurred most often at $IC_{LS} = 0$ with maximal 11 out of 39 responses, which decreased to only 2 occurrences at $IC_{LS} = 1$. The stimuli at 1 and 4 kHz were hereby mostly prone to split images than the low frequency and white noise stimulus. Interestingly, at 50 dB SPL the frequency-dependent effect of image splits vanished and maximal 7 out of 39 responses were counted at $IC_{LS} = 0$. For the HP stimulus at 8 kHz, only 2 out of 21 responses revealed an image split which was independent of the IC_{LS} and the SPL.

4.4 Discussion

In the current study, findings in literature regarding ASW perception were confirmed for a stereo setup in a room with quasi anechoic conditions. An increase in sound pressure level, obviously leads to an extent of the apparent area of the sound source, i.e. besides the ASW also the ASH increases. This is in line with findings by Perrott and Buell (1982) where expansions in the volume of a sound source due to an increased SPL were found for headphone experiments. This might be understood by considering the loudness as a perceptual weighting that increases the salience of the binaural cues. Also, the broader auditory filters

at higher SPLs might contribute here. The current findings on a larger ASW for low-frequency content than for high-frequency content for a fixed IC_{ears} confirms those in Blauert and Lindemann (1986a) for headphone-based experiments. They explained this effect by an earlier study (Blauert, 1978), where the lateralization decreased with increasing the center frequency of the bandpass filtered noises, even though the interaural time-differences were kept constant. At low frequencies, the localization uncertainty raises due to the large wavelength, which might cause the increase in lateralization and in ASW. It was further shown here, that IC_{ears} exploited higher values at low frequencies. The auditory system is more sensitive to changes at high correlation values (Pollack and Trittipoe, 1959) which might in turn lead to a larger ASW. Further, it is important to note that the internal representation of IC_{ears} values differs from their physical values. Low IC_{ears} values will be higher in the internal representation mainly due to inner-hair cell transduction, which is commonly modeled as half-wave rectification and low-pass filtering (Faller and Merimaa, 2004). This will be further explored in Chapter 7 about modelling ASW perception. Finally, cognitive effects might also play a role, since in nature low frequency sounds are most likely produced by large objects or animals (Perrott and Buell, 1982). In chapter 3, the frequency-dependent effects regarding ASW could not be shown. The results had been measured using a MUSHRA test procedure. Since the current findings support findings in literature, the usage of MUSHRA tests in connection with the investigation of spatial impression is not recommended for further research.

4.5 Summary and Conclusion

The psychoacoustic data showed ASW as a function of IC_{ears}, frequency and sound pressure level. Listeners were able to discriminate ASW even for stimuli with a frequency content above 2 kHz. Increasing the sound pressure level showed a frequency-dependent increase in ASW. Future studies will further investigate the role of the different auditory processes on ASW, also including more complex stimuli like e.g. speech and music.

5

The impact of interaural time differences on apparent source width [a]

Abstract

For the perception of spaciousness, the temporal fluctuations of the interaural time differences (ITDs) and interaural level differences (ILDs) provide important binaural cues. One major characteristic of spatial perception is apparent source width (ASW), which describes the perceived width of a sound image. The temporal fluctuations of the binaural cues cause the signals at a listeners' ears to be decorrelated. Therefore, ASW has traditionally been measured by using the interaural cross-correlation (IACC). In particular, ITD fluctuations (below 2 kHz) have been suggested to be the dominant cue for the perception of ASW. However, the contribution of the ITD statistics on the percept of ASW has not yet been clarified. In the present study, the impact of ITD fluctuations in different frequency bands on the perceived ASW was investigated. In a psychoacoustic evaluation, a source signal was convolved with individual binaural room impulse responses (BRIRs) and presented to the listener via headphones. The obtained signals were passed through a gammatone filterbank with an analysis and synthesis stage which enabled the modification of the ITD fluctuation statistics in individual frequency bands. The ITD fluctuations of broadband noise stimuli were compressed while the effect of this compression on the ILD statistics was kept minimal. The IACC was kept the same for stim-

[a] This chapter is based on Käsbach et al. (2014b) (Proc. of Forum Acusticum, Krakow, Poland, 2014).

uli with compression below 2 kHz and for the uncompressed noise which should lead to the same ASW percept in the two conditions. However, the psychoacoustic data showed a reduced ASW for the modified signals, particularly in conditions with an applied compression around 1 kHz. In contrast, above 2 kHz, the compression had no effect on ASW, whereas the IACC increased. The results suggest that the broadband IACC can be a misleading objective measure of ASW and that ITD fluctuations around 1 kHz are crucial for ASW perception.

5.1 Introduction

The apparent source width (ASW) is an important perceptual measure of the space surrounding a listener. It has a long tradtition in room acoustics where it is employed as an attribute to describe the complex perception of rooms and their perceived quality (Bradley, 2011). In contrast to anechoic conditions, where the signals at the ears of a listener are highly correlated, the reflections in a room alter the left and the right ear signals separately which effectively reduces the correlation between the ear signals. Therefore, the interaural cross-correlation (IACC) was established as a measure that is inversely proportional to ASW (de Vries et al., 2001). This decorrelation, which is in particular caused by lateral reflections (Blauert and Lindemann, 1986b), results in fluctuations of binaural cues, namely interaural time differences (ITDs) and interaural level differences (ILDs), both of which are contributing to the perception of ASW. Moreover, in van Dorp Schuitman et al. (2013) the fluctuations of ITDs below 2 kHz were used in a complex, nonlinear auditory model to predict ASW. However, the frequency-specific contribution of these fluctuations has not yet been clarified. Therefore, the aim of this study is to investigate the importance of ITD fluctuations in different frequency regions for the percept of ASW.

In Catic et al. (2013) the contributions of ILD and ITD fluctuations to the externalisation of sound sources, i.e. the perception of a sound source outside the head with a clear perception of distance, was studied. This was achieved by compressing the fluctuations of either ILDs or ITDs in different frequency regions and analysing the degree of externalisation for the manipulated signals in a listening test. In the present study, this method was applied to modify

instantaneous ITD fluctuations of binaural signals. Specifically, the ITD fluctuations of a sound source with a wide ASW characteristic were compressed in different frequency bands. The influence of the applied manipulation on ASW was evaluated in a listening test where listeners had to rate the perceived ASW compared to a reference condition without ITD compression.

5.2 Method

5.2.1 Apparatus and stimuli

The listening experiment was conducted in the IEC listening room (properties follow the ISO 268-13 standard) at the Technical University of Denmark with a volume of $V = 98.3\,\mathrm{m}^3$ and a reverberation time of $T_{60} = 0.4\,\mathrm{s}$. A standard stereo loudspeaker setup with an opening angle of $\alpha = 30°$ was placed at a distance of $d = 2.1\,\mathrm{m}$ in front of the listener position, located in the center of the room. Both loudspeakers were of type Dynaudio BM6.

As in Catic et al. (2013) the stimuli were created from individual binaural room impulse responses (BRIRs) and were played back to the listener via headphones during the experiment. The BRIRs were measured by inserting miniature microphones of type Sonion 8002 with a closed fitting in the listener's ear canals. Logarithmic sweeps with a duration of 5 s and 10 repetitions played back from each loudspeaker were recorded with a sampling rate of $f_s = 44.1\,\mathrm{kHz}$. To obtain the BRIRs the inverse sweep spectrum was applied to the measured responses in the frequency domain. The resulting BRIRs were windowed with a 500 ms $\cos^2$ window in order to remove nonlinear components from the responses. Headphone impulse responses were measured in the same fashion, applying 2 s long logarithmic sweeps, by placing headphones of type Sennheiser HD-580 on the ears of the listener.
The stimuli were generated by convolving the signal $x(n)$ with the corresponding BRIR, $h(n)$, for each listener individually, where n indicates the discrete time index. Including cross-talk of the loudspeaker setup, the left and right ear

signals, $s_L(n)$ and $s_R(n)$, were calculated as

$$\begin{aligned} s_L(n) &= x_1(n) * h_{L1}(n) + x_2(n) * h_{L2}(n) \\ s_R(n) &= x_1(n) * h_{R1}(n) + x_2(n) * h_{R2}(n) \end{aligned} \tag{5.1}$$

where the first suffix of $h(n)$ indicates the ear channel (L and R) and the second the loudspeaker channel (1 and 2). The signals, $y_L(n)$ and $y_R(n)$, presented to a listener were obtained by filtering $s_L(n)$ and $s_R(n)$ with the inverse headphone impulse responses, $g_L(n)$ and $g_R(n)$, to account for the influence of the headphones using

$$\begin{aligned} y_L(n) &= s_L(n) * g_L(n) \\ y_R(n) &= s_R(n) * g_R(n). \end{aligned} \tag{5.2}$$

5.2.2 ITD compression

The signal $x(n)$ was white noise that was separately generated for both loudspeaker-channels, resulting in decorrelated signals between the loudspeakers. ITD compression was applied for frequency regions corresponding to approximately 2 octaves centered around $f_c = 0.25$, 1 and 4 kHz. The instantaneous ITD fluctuations were compressed by using the method described in Catic et al. (2013) that is schematically illustrated in Figure 5.1.

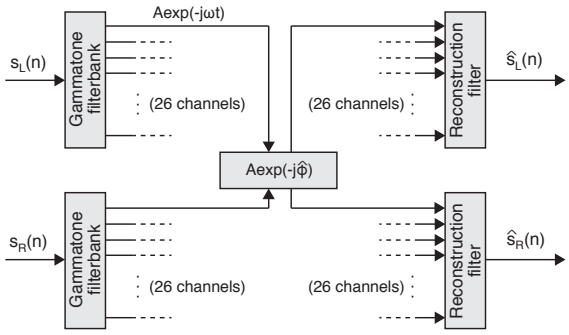


Figure 5.1: Schematic plot of the signal processing for the ITD compression. The input signals were synthesized by a fourth-order gammatone filterbank with 26 channels, each with a bandwidth of 0.5 ERB. After sample-wise assigning a new phase $\hat{\Phi}$ to the complex output of the desired channels, the signal was reconstructed by applying a synthesis filterbank.

The signals $s_L(n)$ and $s_R(n)$ were analysed by a fourth-order gammatone filterbank with a reduced bandwidth of 0.5 ERB for center frequencies ranging from 131 to 13563 Hz. To avoid overlap between the filterbank channels, only every third filter of the original gammatone filterbank was used for analysis and reconstruction. The instantaneous ITD compression in the desired frequency region, corresponding to six gammatone filters, was then achieved by assigning an average phase to the left and right complex output of the respective channel as

$$\hat{\Phi}_L(n) = \hat{\Phi}_R(n) = (\Phi_L(n) + \Phi_R(n))/2. \tag{5.3}$$

The resulting signals, $\hat{s}_L(n)$ and $\hat{s}_R(n)$, were transformed back to the time domain by applying the corresponding reconstruction filterbank (Hohmann, 2002).
The effect of the ITD compression is illustrated in Figure 5.2 for four signals, for which the relative occurance of ITDs per frequency channel are shown in form of histograms.

The binaural signals were processed by a fourth order gammatone filterbank with 1-ERB wide filters. Subsequently, ITDs for each frequency channel were computed for frames of 20 ms with 50 % overlap. Note that the specifications of this analysis gammatone filterbank are different from the one used for applying the ITD compression. It is expected that the ITDs show strong fluctuations across all frequency channels, due to the use of broadband, stochastic signals that are decorrelated. Furthermore, these fluctuations are emphasised by presenting lateral source positions via the stereo-setup. As can be seen in the top left panel of Figure 5.2, the reference condition without ITD compression produced indeed a uniform distribution of ITDs across frequencies. The manipulated signals, seen in the successive panels, show an accumulation of ITDs, i.e. reduced ITD fluctuations, in the frequency region where the ITD compression was applied (indicated by the dashed lines).
In addition, the influence of the applied ITD compression on the ILD fluctuations was analysed by calculating the frequency-dependent correlation coefficient between distributions of the ILD fluctuations for the reference condition and the signals with ITD compression. The results for one listener are displayed in Figure 5.3.

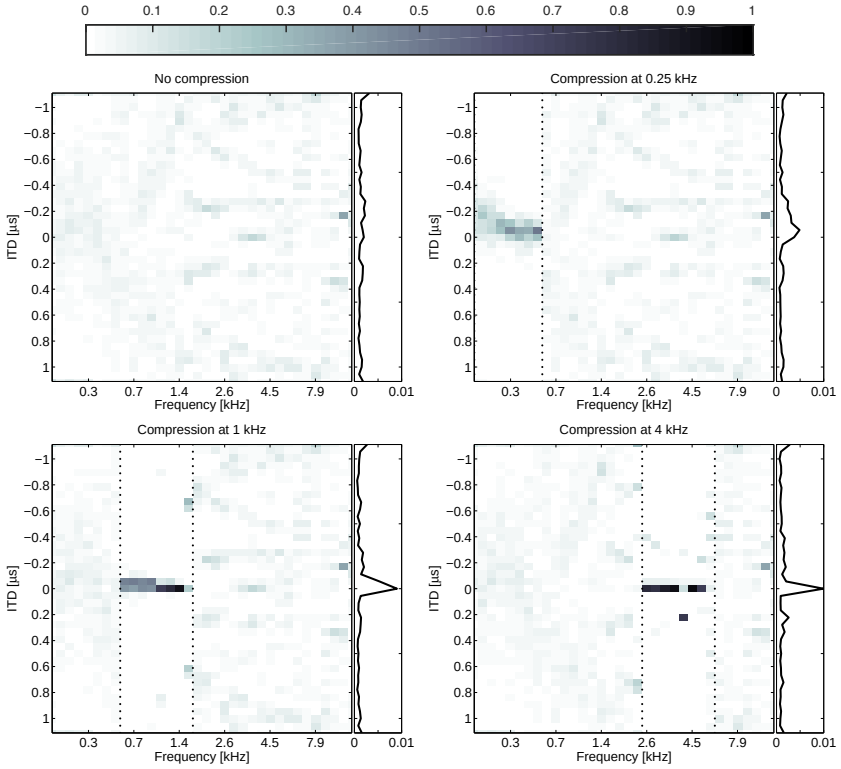


Figure 5.2: Relative occurrences of ITDs per frequency channel shown in form of histograms. These results were obtained from a time-frequency analysis of ITDs for each binaural stimulus of one listener. From top left to bottom right panel: Reference signal (without ITD compression) and the signals with ITD compression for a bandwidth of approximately 2 octaves around the center frequencies $f_c = 0.25$, 1 and 4 kHz. A dark color represents a high relative occurance of ITDs. The wide distribution of ITDs over the entire range of ±1 ms becomes concentrated around 0 ms in the frequency regions with ITD compression (indicated by the dashed lines). To the right of each panel the accumulated distribution across all frequency channels (center of gravity) is shown.

It can be seen that the correlation of ILD statistics between the unprocessed and processed signals was between 0.95 and 1 for most of the frequency channels. The lowest correlation values were obtained around the frequency regions where the ITD compression was applied. It can be concluded from these results that the applied ITD compression had a minimal influence on the ILD distributions.

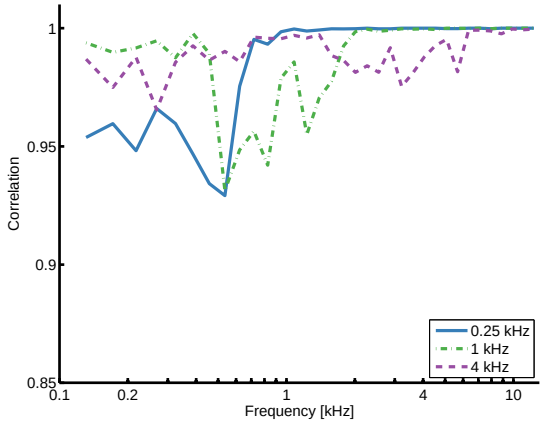


Figure 5.3: Correlation of ILD statistics. Histograms of ILDs were obtained in the same fashion as described for the ITDs in Figure 5.2. The correlation coefficient between ILD histograms of the reference signal and the ITD compressed signals per frequency channel are shown on a logarithmic frequency axis.

5.3 Experimental procedure

In the experiment, a modified Multiple Stimulus with Reference and Anchor (MUSHRA) procedure was used. It was decided to perform the experiment without an anchor since only the perceptual difference from the reference condition was of interest here. The listeners were asked to rate the four types of stimuli, three signals with ITD compression and the reference, in terms of apparent source width on a scale from 0 (narrow) to 100 % (wide). The total duration of one stimulus was 1.4 s. The stimuli were randomised for each presentation and listeners were instructed to identify the hidden reference by assigning a score of 100 %. The stimuli and the reference could be played back as often as desired by the listener. In total, five normal hearing listeners participated and the test was repeated four times per listener. All listeners completed an initial training phase prior to the experiment.

5.4 Results and Discussion

The results of the listening experiment are shown in Figure 5.4. On the x-axis, the experimental conditions are indicated by the center frequencies of the ITD compressed frequency regions. On the left y-axis, the obtained ASW data (rep-

resented by filled '□') is shown as mean value and standard error across all listeners and repetitions. The lower the assigned score in ASW the narrower the sound source was perceived. It can be seen that the perceived ASW decreased when the ITD compression was applied. While ITD compression at 0.25 kHz reduced the measured ASW by 50 %, the largest reduction of ASW was obtained for the ITD compressed signal at 1 kHz. For the stimulus with ITD compression at 4 kHz, no difference compared to the reference condition was found. The effect of the applied ITD compression at different center frequencies on the ASW data was analysed through a one-way repeated measures analysis of variance (ANOVA) [$F(3,12) = 39.98$, $p < 0.001$]. The results of a *post hoc* (Tuckey's test) comparison revealed that the described differences between the reference condition and the stimuli at 0.25 and 1 kHz were significant ($p < 0.001$), whereas no significant difference was found between reference condition and the stimulus at 4 kHz.

On the right y-axis of Figure 5.4, the measured interaural cross-correlation (IACC) values (represented by filled 'o') between left and right ear signal, $s_{\mathrm{L}}(n)$ and $s_{\mathrm{R}}(n)$, are shown. They were obtained for each listener separately (acc. to Eqn. 2.2 for the entire signal duration) and the mean and standard deviation across all listeners are plotted. Note that the scale in the plot is reversed to meet the general assumption that low correlation values correspond to a wide ASW. The IACC-based analysis, however, does not reflect the psychoacoustic data: It can be seen that for the first three experimental conditions the IACC has a constant value of around 0.2 indicating that the ITD compression did not alter the IACC value from the reference condition for these signals. In contrast, these conditions are perceived with a decreasing ASW. Similary, the stimulus at 4 kHz is perceived as unchanged compared to the reference condition whereas IACC increases to almost 0.4 which would suggest a narrower ASW.

On the same axis, the $\mathrm{IACC}_{\mathrm{E3}}$ value (represented by '□') is shown which is a frequency-averaged IACC over the octave bands 500 Hz, 1 kHz and 2 kHz as defined in Okano et al. (1998). By definition, only the early part (first 80 ms) of an impulse response is considered in the calculation. However, this upper integration limit was neglected here and $\mathrm{IACC}_{\mathrm{E3}}$ was calculated between the left and right ear signal, $s_{\mathrm{L}}(n)$ and $s_{\mathrm{R}}(n)$, for their entire duration. In contrast to the IACC, the $\mathrm{IACC}_{\mathrm{E3}}$ is more affected by the applied ITD compression. It follows the trend of the ASW data but with a reduced sensitivity compared to the data. At the reference condition, $\mathrm{IACC}_{\mathrm{E3}}$ has a value of 0.25 which decreases

at 0.25 kHz to around 0.35 and has the lowest value at 1 kHz of around 0.5. At 4 kHz the measured $IACC_{E3}$ equals the one obtained at the reference condition.

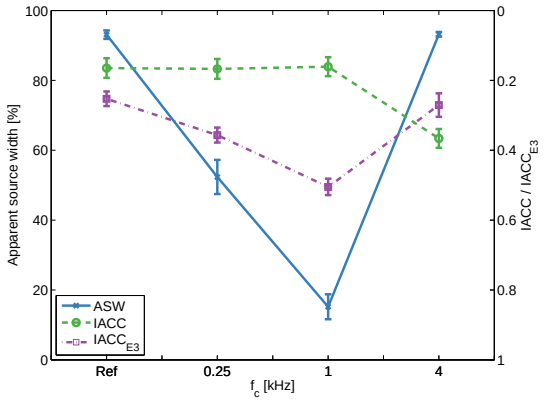


Figure 5.4: Measured ASW data and IACC values as a function of the experimental conditions. On the left y-axis, the ASW data (represented by filled '□') is shown as the mean and the standard error of the MUSHRA-test results across 5 listeners where 0 and 100 % corresponds to a narrow and wide ASW, respectively. On the right y-axis, the measured IACC values (represented by filled 'o') between the left and right ear signals are shown as mean and standard deviation across the 5 listeners. Note that the scale is reversed. The experimental conditions on the x-axis comprise the signals with ITD compression for a bandwidth of approximately 2 octaves around the center frequencies $f_c = 0.25$, 1 and 4 kHz and the reference signal Ref.

From the experimental results it becomes clear that for the perception of ASW the auditory system is most sensitive to ITD fluctuations for frequencies around 1 kHz. While at lower frequencies (around 0.25 kHz) the influence of ITD fluctuations on ASW is reduced, no contribution can be identified at higher frequencies (around 4 kHz). This indicates an upper frequency limit up to which ITD fluctuations are contributing to the perception of ASW. Because the ITD compression was applied with a bandwidth of approximately 2 octaves around the center frequencies, the frequency of 2 kHz is suggested as such limiting frequency.
In Blauert and Lindemann (1986a) and Käsbach et al. (2014a) (see Ch. 4) it was shown that listeners are still able to discriminate ASW for high frequency stimuli with a frequency content above 2 kHz. These findings are further supported by Blauert et al. (1986c) and highlight the importance of all frequency regions for the perception of the more general term spaciousness. Presumably, at higher frequencies the auditory system is likely to rely on ILD fluctuations as a cue for the perception of ASW. This needs to be further investigated though.
The results of this experiment showed that the traditional IACC could not account for the perceived changes in ASW with frequency. A similar discrepancy

between the perceived ASW and IACC predictions were reported in one of our earlier studies (Käsbach et al., 2013; see Ch. 3). The reason for the mismatch in both cases is that the IACC is not sensitive to instantaneous changes in the ear signals over time and frequency. In contrast, the $IACC_{E3}$ was found to correlate better with the ASW data which is due to the applied frequency-averaging over three octave-bands. An alternative measure based on ITD fluctuations is shown in Figure 5.2, where the right graph of each panel shows the accumulated distribution (center of gravity) of ITDs across frequencies. While this distribution is flat for the reference signal, representing a uniform distribution of ITDs, a peak with a decreasing variance becomes prominent with increasing center frequency of the compressed ITDs. The standard deviation or the difference between percentiles of such distribution was successfully used in van Dorp Schuitman et al. (2013) and Käsbach et al., 2016c (see Ch. 7), respectively, as corresponding prediction of ASW perception with a narrow distribution corresponding to a narrow ASW. The experimental results presented here could be used to formulate a weighting for the across-frequency integration of ITD fluctuations. A similar experiment could be used to determine the frequency-dependent contribution of ILD fluctuations. In addition, future investigations regarding the temporal integration of these fluctuations are required to form a single measure that reflects the perceived ASW (see Ch. 7). In Catic et al. (2013) it was reported that the compression of ITD fluctuations of a low-pass filtered speech signal at a lateral source position did not result in a loss of externalisation. However, for the here presented noise signals, listeners commented on perceiving a more internalised sound in case of a narrower ASW.

5.5 Summary and Conclusion

This study investigated the influence of ITD fluctuations on the percept of ASW by modifying these fluctuations in different frequency regions. The experimental results showed that ITD compression below 2 kHz substantially reduced the perceived ASW, whereas the modification of higher frequencies did not result in a change in ASW. This indicates that ITD fluctuations might only contribute to the perception of ASW for frequencies below 2 kHz.
Hence, for the prediction of ASW perception, ITD fluctuations are important. In contrast, the IACC was found not to be very sensitive to ITD compression,

and consequently is not applicable in general terms as a predictor of ASW.
In Käsbach et al. (2014a) it was shown that ILD fluctuations also correlate with perceptual data of ASW (see Ch. 7). Therefore, it would be important to carry out a similar study like the present one in order to clarify their contribution to the perception of ASW. The question remains on how the fluctuations of ITDs and ILDs are integrated by the auditory system across time and frequency. This important aspect will be addressed in future investigations.

Acknowledgments

This article is based on Gudrun Oskarsdottir's MSc thesis entitled "The effect of interaural-time-difference fluctuations on apparent source width".

6

The effect of hearing-aid processing on apparent source width perception in normal-hearing and hearing-impaired listeners[a]

Abstract

Objective: The first objective of this study was to compare spatial perception in hearing-impaired (HI) listeners to normal-hearing (NH) listeners by means of apparent source width (ASW). The second objective was to investigate the impact of behind-the-ear hearing aids (BTE-HAs) and wide-dynamic range compression (WDRC) on the listeners' ASW perception. Design: The ASW was measured in six NH and six HI listeners in loudspeaker-based experiments using noise, speech and music stimuli. The listeners were tested unaided, and aided with BTE-HAs using two different programs, linear and WDRC processing. Results: In the unaided condition, the results of the HI listeners showed, with the exception of a single listener, a reduced sensitivity regarding ASW compared to the NH listeners. In the aided condition, the two HA programs equally altered ASW in the NH listeners and in the sensitive HI listener. The other five HI listeners remained generally insensitive to both programs. Conclusions: Listeners who showed ASW sensitivity were affected by the BTE-HA, independent of the selected program. Thus, the WDRC did not affect ASW perception, whereas the BTE-HAs' microphone position did.

[a] This chapter has been submitted to the Int. Journal of Audiology in June 2016 (Käsbach et al., 2016c).

6.1 Introduction

The apparent source width (ASW) describes the perceived spatial extent of a sound source in a room. It represents an important subjective attribute, e.g. for the evaluation of the sound quality in concert halls (Okano et al., 1998), and can affect sound source localization and speech intelligibility (e.g., Whitmer et al., 2012). In multi-talker environments, perceptually punctate sound sources might be easier to separate from each other than sources that are perceived as spatially diffuse (Noble et al., 1997). It has been demonstrated that important auditory cues contributing to the perception of ASW are temporal fluctuations of interaural time differences (ITDs) and interaural level differences (ILDs) (Blauert and Lindemann, 1986b), which are caused by room reflections, but also by reflections from the head and torso of the listener. An increasing amount of room reflections typically causes stronger temporal fluctuations of the binaural cues. One of the consequences is that the interaural coherence (IC), reflecting the similarity between the two ear signals, decreases with increasing amount of room reverberation which, in turn, is linked to an increased ASW (Ando, 2007).

However, while such a relationship has been found in normal-hearing (NH) listeners (Blauert and Lindemann, 1986a), Whitmer et al. (2012, 2014) showed that hearing-impaired (HI) listeners indicated a roughly constant ASW in corresponding measurements, independent of the stimulus IC. It was argued that this might be caused by the HI listeners' degraded ability to detect instantaneous interaural phase differences, e.g. due to a degraded representation of temporal fine structure (TFS) information in the peripheral auditory system. Whitmer et al. (2014) suggested that the observed constant ASW values might then have been mainly driven by the ILDs of the stimuli that did not vary across the conditions in their study. Furthermore, the same authors investigated the relation between ASW perception and azimuthal localization accuracy and found only a low correlation in the data.

A major purpose of hearing aids (HAs) is to restore audibility for the HI listeners while, ideally, maintaining the natural spatial perception of the target sound and the environment. However, spatial sound perception might be affected by compression, such as wide dynamic-range compression (WDRC), which is commonly applied to provide level-dependent gain in different frequency

bands in the individual listeners. WDRC decreases the dynamic level range of an acoustic signal and is meant to compensate for loudness recruitment in HI listeners with a sensorineural hearing loss (Allen, 1996; Villchur, 1973). If no linked processing between the left-ear and right-ear HAs is provided, WDRC acts independently in the two instruments. This may lead to alterations of the interaural cues and, thus, could affect ASW perception. Besides the applied HA processing, the HA microphones' position and directivity pattern might influence spatial perception (Best et al., 2010). In a behind-the-ear (BTE) HA, the placement of the microphones results in spectral changes of the ear signals that are source-direction dependent and pronounced at high frequencies, relative to the listener's head-related transfer functions (HRTFs) in the unaided situation (Sivonen, 2011). Thus, natural pinna cues are not available to the listener when wearing BTE HAs. The locations of the microphones are typically taken into account in the HA signal processing by standardized compensation filters (e.g., Bentler and Pavlovic, 1989; Moore et al., 2010) and the use of dedicated beam forming settings to restore the directivity of the pinna (Kuk et al., 2013). However, due to the non-individualized compensation strategy, source-direction dependent spectral differences relative to the individual listener's HRTFs remain. Hence, the localization accuracy and ASW perception of the aided listener might be affected.

Several studies have investigated the impact of WDRC on ITD and ILD discrimination (Musa-Shufani et al., 2006), localization performance (Wiggins and Seeber, 2011, 2012) and ASW (Whitmer and Akeroyd, 2013). Musa-Shufani et al. (2006) considered ITD and ILD discrimination in NH and HI listeners using several WDRC settings. They found that WDRC increased the just-noticable-differences (JNDs) for ILDs in both listener groups, especially at high frequencies, whereas no effect on ITDs was observed. The ILD effect was rather small in conditions with low compression ratios (3:1), commonly used in HAs, and long attack time constants (200 ms), but increased at higher compression ratios (8:1) and faster attack time constants (2 ms), whereby the release time was kept constant at 500 ms throughout the conditions. In the HI listeners of their study, the ILD JNDs were similar to those obtained with NH listeners, whereas their ITD JNDs were about 6 times larger than in the NH listeners. Wiggins and Seeber (2011, 2012) showed that fast-acting WDRC, applied separately at both ears of the listener and only at frequencies above 2 kHz, with an attack time of 5 ms, a release time of 60 ms and a compression ratio of 3:1, affected the listeners'

lateralization performance, consistent with the results of Musa-Shufani et al. (2006). The ILDs were reduced in Wiggins and Seeber (2011, 2012) as a result of the applied compression ratio which also decreased the perceived angle of the presented sound source, mostly for signals with abrupt onsets and offsets. For signals with gradual onsets and offsets, including speech, the applied WDRC caused an increase in ASW which they referred to as "increased diffuseness". In addition, the WDRC increased the occurrences of "image splits", i.e. the perception of two sound sources instead of a single one. The lateralization effects, the increased ASW and the increased occurrences of "image splits" were found to be more pronounced in a high-pass filtered condition (with a cut-off frequency of 2 kHz) compared to the full-bandwidth condition where listeners were simultaneously presented with unprocessed low-frequency cues. In both studies (Wiggins and Seeber, 2011, 2012), only NH listeners were tested in headphone experiments using binaural simulations (based on non-individualized HRTFs) of single sources. Therefore, the influence of WDRC could only be evaluated for point-like sources (small ASW) as a reference and not for larger ASWs. In Whitmer and Akeroyd (2013), real HAs were used and ASW perception was investigated in HI listeners using a virtual sound environment. They found no effect of HA processing on ASW perception. However, the participants were wearing their own HA with individual signal-processing settings which did not allow for a more controlled stimulus presentation.
In the present study, the impact of a BTE-HA on ASW perception in NH and HI listeners for controlled stimuli was investigated using loudspeaker-based listening tests. The influence of linear amplification versus WDRC was investigated and compared to corresponding results obtained in the unaided conditions. The presented stimuli in the different conditions were analyzed in terms of their binaural cues (IC, ITDs and ILDs) to explore their contributions to ASW.

6.2 Methods

6.2.1 Listeners

6 NH (5 male and 1 female) and 6 HI (male) listeners participated in the listening experiments. The NH listeners were 27 to 32 years old and had hearing thresholds below +20 dB hearing level (HL) for the octave frequencies between

125 and 8000 Hz. The HI listeners were between 59 and 75 years old and were diagnosed with a symmetrically mild-to- moderate sensorineural hearing loss. Their audiograms are shown in Figure 6.1. The left-ear and right-ear pure-tone hearing thresholds of the individual listeners ranged between 0 and 10 dB in the frequency range from 0.25 to 2 kHz and from 0 to 15 dB at the frequencies 4 and 8 kHz. All HI listeners had been hearing-aid users for at least 4 months.

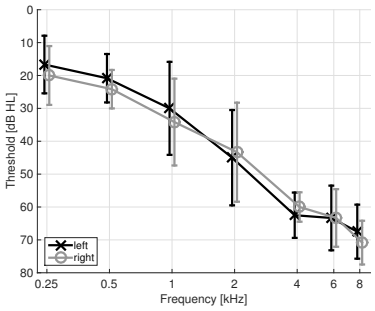


Figure 6.1: Averaged audiograms of the HI listeners for the left ear (crosses) and right ear (circles).

6.2.2 Stimuli and Procedure

Listening room and experimental setup

For the listening test, five distinct physical source widths (PSW) were reproduced via loudspeakers to generate corresponding distinctive sensations of ASW. The experimental setup is shown in Figure 6.2. Three loudspeaker pairs, with loudspeakers of type Dynaudio BM6, were installed in a line array in a standardized IEC listening room with a reverberation time of $T_{30} = 0.4$ s. The loudspeakers were positioned at opening angles of 16°, 30° and 42.5° with corresponding distances from the listening position between 1.75 and 2.30 m. In addition, a center loudspeaker (at 0° and 1.70 m distance) was used as a reference. An overview of the setup is provided in Table 6.1.

Source widening algorithm

Each loudspeaker pair produced a phantom source in the center of the line-array. The PSW was controlled by the opening angle of the loudspeakers (Frank, 2013) and by applying the "efficient source widening" algorithm described in Zotter

PSW	Parameters		
	LS ang. [°]	dist. [m]	Φ [°]
#1	0	1.70	-
#2	±16	1.75	20
#3	±30	1,95	20
#4	±30	1,95	40
#5	±42.5	2,30	30

Table 6.1: Parameters for the five distinct physical source width (PSW) values: Corresponding angular distribution of the loudspeakers, the source receiver distances and the modulation depths settings for the source widening algorithm.

and Frank (2013). The algorithm generates two filters, one for the left and one for the right loudspeaker channel. It imposes a cosine-modulated inter-channel time delay τ with a modulation period of $\Delta f = 1/T$, with T representing the modulation period. The modulated time delay corresponds to a sinusoidal modulation of the phase spectrum in the frequency domain with a modulation depth expressed as phase $\Phi = \tau/T$. The modulation of the spectrum for the left and right channel, respectively, results in:

$$H_{L,R}(\omega) = \frac{1}{\sqrt{2}} e^{\pm i\Phi \sin(\omega T)}, \tag{6.1}$$

with ω representing the angular frequency. Two finite impulse response (FIR) filters were generated, one for the left loudspeaker channel and one for the right loudspeaker channel, and convolved with the mono source signals. Each filter consisted of 5 filter coefficients, separated by $N = T \cdot f_s$ samples, with f_s denoting the sampling frequency. The free parameters of the algorithm were set to $N = 132$ samples at $f_s = 44.1$ kHz and $T = 3$ ms. The modulation depth phase Φ was set individually for each loudspeaker pair. Five distinct PSW values (#1 to #5) were generated for the experiments. The corresponding angular distribution of the loudspeakers, the modulation depths settings for the source widening algorithm and the source receiver distances are listed in Table 6.1.

Stimuli

Three different source signals were used: anechoic speech (a male talker), anechoic guitar excerpts (a classical acoustic guitar playing staccato chords) and random pink noise. For each source, five different realizations were generated, each representing one of the five PSWs. The duration of each stimulus was 6 s. The processed signals were presented to the listeners at a sound pressure level (SPL) of 70 dB (with a maximal deviation of +0.3 and -1 dB) based on the long-term root mean square (RMS) value of the individual signals.

Hearing aid conditions

The obtained stimuli were tested for three different HA conditions: A reference condition without hearing aids ("undaided") and two "aided" conditions were considered where either linear HA processing (referred to as "lin HA") or wide-dynamic range compression (referred to as "WDRC HA") was applied. In the aided conditions, behind-the-ear (BTE) HAs of type Widex Dream Fusion with in-canal receivers were bilaterally fitted to the individual listener. The receiver was inserted in a listener's ear with a closed fitting using instant double ear-tips with a 1 mm vent. In the case of linear HA processing, the insertion gain for the HI listeners was based on the average pure-tone thresholds across listeners following the NAL-R(P) rationale (Byrne et al., 1990). This was reasonable due to the homogeneous audiograms across the HI listeners. For NH listeners, no insertion gain was applied in the linear HA processing. The program with WDRC used a compression ratio of 2:1 and attack and release times of 11.5 ms and 100 ms, respectively, which were measured according to IEC 60118-2 (1983). The compression knee-point was set to 35 dB SPL in 15 frequency bands covering the frequency range from 100 Hz to 10 kHz. The insertion gain of the program with WDRC was linearly increased by 13 dB from the NAL-R(P) gain, to achieve the same overall SPL of the stimuli as in case of the linear HA processing. The HAs operated independently, i.e. no cross link was established between the left-ear and right-ear HA. No further signal processing, such as beam-forming, noise-reduction or feedback control, was applied. However, standardized filters were active that compensated for the microphone position. In the present study, one of the HI listeners, was not available for the data collection in the aided condition with the BTE HAs. For this subject, these data were collected

using a master hearing aid (MHA; Buchholz, 2013) with BTE satellites including the receiver and with the same two programs, linear amplification and WDRC, whereby it was attempted to provide a similar processing in the two HA types.

Procedure

The listeners were asked to indicate the ASW by separately identifying the left most and the right most extension of the phantom sound source (adapted from Käsbach et al., 2014a). An acoustically transparent curtain was installed in front of the entire loudspeaker setup to avoid any impact of visual cues provided by the loudspeakers on the ASW estimates. The listeners were asked to project their sensation of ASW on a visual degree scale in front of the curtain placed at the same height as that of the loudspeakers, as illustrated in Fig. 6.2. The same scale was displayed on a touchscreen that the listeners used to indicate their responses, separately for the left and the right boundary of the perceived sound. Each sound was evaluated individually. A reference sound (PSW #5 with pink noise) was available to the listener throughout the entire experimental procedure such that the listeners had the opportunity to compare the current sound with the reference. In addition, the listeners could indicate with a checkbox whether a split image was perceived for a presented stimulus, i.e. the perception of two phantom sound sources instead of a single one. In total, 45 stimuli were tested per HA condition. The HA conditions were arranged using a Latin-rectangular design and the stimuli were presented in randomized order in each HA condition. Each stimulus was presented three times.

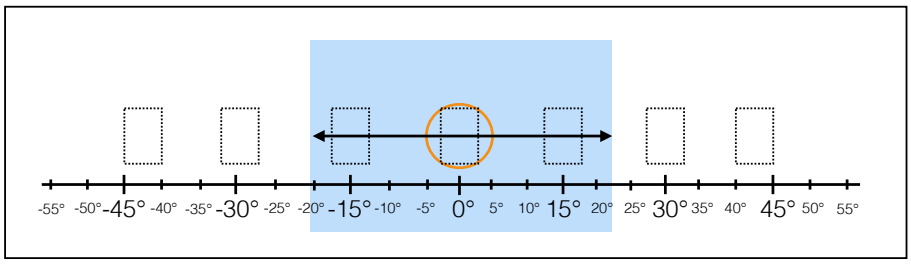


Figure 6.2: Sketch of the experimental set-up and procedure. The loudspeaker pairs (see Table 6.1) generate a phantom source at 0°. Subjects were asked to indicate the ASW in degree on the given scale, separately for the left and right boundary of the source image.

6.2.3 Statistical analysis

For the statistical analysis, a linear mixed-effects model was fitted to the data with 4 fixed factors and 2 random factors. In the model, the parameters PSW, source signal, HA condition and listener group (NH or HI listeners) were considered as fixed factors and the participants and repetitions were considered as random factors. An analysis of variance (ANOVA) was performed with the response variable ASW, which was calculated as the difference between the measured left and right boundaries. After contrasting NH and HI listeners, the two groups were analyzed separately. An ANOVA was performed on a mixed model for each group individually, with the remaining 3 fixed factors (PSW, source signal and HA condition) and the random factors participants and repetitions. A post-hoc analysis was performed using pairwise comparisons with Bonferroni corrections, with a correction factor of $c = \sum_{l=1}^{L-1} l$, where c represents the number of comparisons between the L levels of each considered fixed factor in the mixed model (which will be specified in the results section together with p_{posthoc}).

6.2.4 Binaural cue analysis

Binaural recordings were obtained with a head-and-torso simulator (HATS) at the listener position. The recordings were used to analyze the binaural cues IC, ITD and ILD in the various listening conditions.

Interaural coherence (IC)

The interaural coherence (IC) represents the absolute maximum value of the energy-normalized interaural cross-correlation function between the left-ear and the right-ear signals for time-lags between +1 and -1 ms (Blauert and Lindemann, 1986a). The broad-band analysis of IC is termed here interaural cross-correlation coefficient ($\text{IACC}_{\text{broad}}$, Ando, 2007). In addition, a modified version of the $\text{IACC}_{\text{broad}}$, the IACC_{E3}, was considered that averages the IC over three frequency bands centered at 0.5, 1 and 2 kHz (Okano et al., 1998). $\text{IACC}_{\text{broad}}$ and IACC_{E3} were here calculated for the entire duration of the binaural recordings.

ITDs and ILDs

The statistical distributions of the ITDs and ILDs in the different stimulus conditions were considered in terms of histograms indicating the relative occurrence of these cues as a function of frequency. According to van Dorp Schuitman et al. (2013), a broader distribution of the ITDs corresponds to a wider ASW perception. The same approach was considered here for the ILDs. First, a time-frequency analysis of the left- and right-ear signals was performed, using a fourth-order gammatone filterbank with one equivalent rectangular bandwidth (ERB; Glasberg and Moore, 1990) filters in the range from 80 Hz to 18 kHz. Then, in each frequency channel, the ITD and ILD statistics were computed for time frames of 20 ms and an overlap of 50%. Cross-correlations between ITD and ILD histograms were calculated to compare different stimulus conditions to each other. The correlation per frequency band was evaluated at lag zero of the cross-correlation function and the calculations were performed for five distinct angular positions of the HATS' head, ranging from -40 to +40 degrees in 20 degree steps. This was done to provide an analysis that reflects the listening situation, where the listeners were allowed to move their head during the experiment.

6.3 Results

6.3.1 Experimental data

Figure 6.3 shows the results of the listening experiments obtained in the unaided conditions, i.e without HAs. The data for the normal-hearing (NH) listeners are shown in the left panel and the results for the hearing-impaired (HI) listeners are shown in the right panel. The measured ASW values (indicated on the abscissa) are shown for the different values of the PSW, #1 to #5, as indicated on the ordinate. The different line styles in the two panels show the results for the different stimulus types (noise, speech and guitar). For each stimulus type, two functions are shown, one representing the left boundary (negative values) and one representing the right boundary (positive values), given in degrees. The symbols represent median values obtained across listeners and repetitions. The error bars represent the 25th and 75th percentiles. For the

NH listeners (left panel), small ASWs were associated with small PSW values and an increase of PSW caused an increase of ASW. In the case of the noise source (solid black line), the ASW amounted to ±10° for PSW #1 and to ±40° for PSW #5. This stimulus produced the largest dynamic range of observed ASW values, i.e. the biggest difference between the largest and the smallest ASW. The dynamic range was reduced ($p_{\text{posthoc}} < 0.001$, $c_{\text{HA}} = 3$) for the speech (dashed grey curves) and the guitar sources (dotted light grey curve), which produced almost identical ASW values ($p_{\text{posthoc}} = 0.36$, $c_{\text{HA}} = 3$). For the speech and guitar sources, the PSW #1 led to ASW estimates corresponding to about 5° while the PSW #5 produced ASW sensations corresponding to about 30°. The main factor PSW had the largest effect size ($F(4, 20) = 110$, $p < 0.001$). The source signal ($F(2, 719) = 31.8$, $p < 0.001$) as well as an interaction term between the PSW and the source signal ($F(8, 718) = 9.5$, $p < 0.001$) were identified as significant factors. In the random effects, significant interactions were found between participants and PSW ($\chi^2 = 92.4$, $p < 0.001$) as well as between participants and repetitions ($\chi^2 = 34.5$, $p < 0.001$). The results demonstrated differences of the ASW sensations obtained for the different PSWs as was also supported by the post-hoc analysis ($p_{\text{posthoc}} = 0.001$, $c_{\text{PSW}} = 10$), except for the PSWs #1 and #2.

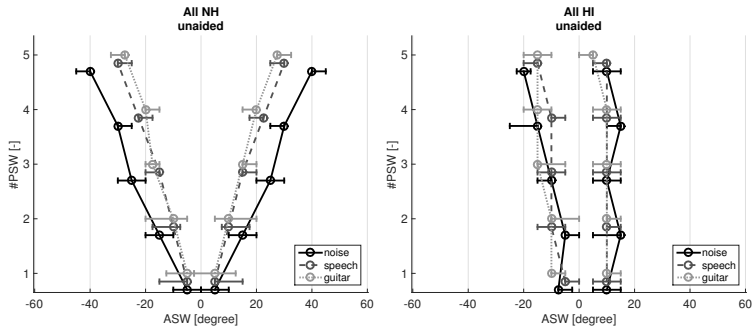


Figure 6.3: Averaged ASW results for normal-hearing (NH) (left) and hearing-impaired (HI) listeners (right). ASW is shown in degrees on the x-axis, as left and right expansion of the sound source, as a function of the physical source width (PSW), PSW #1 (narrow) to #5 (wide). Plotted are the median and respective 25th and 75th percentiles. Shown are the results for all source signals in the condition without hearing aid [unaided]. The various stimuli in each panel are represented by the different line styles and grey shades.

The results for the HI listeners (right panel of Figure 6.3) show a different pattern than those obtained for the NH listeners. In fact, the dynamic range of the ASW ratings was found to be reduced for all three source types. In contrast to the results for the NH listeners, the ASW ratings were essentially independent of the PSW. Specifically, a small PSW value (PSW #1) produced a larger ASW in

the HI listeners than in the NH listeners, whereas a large PSW value (PSW #5) resulted in smaller ASW values than in the NH listeners. This demonstrates a reduced sensitivity of the HI listeners with respect to changes in ASW compared to the NH listeners ($p_{\text{posthoc}} = 0.02$, no correction). The slight asymmetry in the HI data regarding the left and the right boundary estimates was mainly due to larger individual differences in ASW. The statistical analysis of the HI data revealed significant effects caused by the factors PSW ($F(4, 666) = 12.1$, $p < 0.001$), source signal ($F(2, 666) = 15.1$, $p < 0.001$) and an interaction between the two ($F(8, 666) = 4$, $p < 0.001$). Overall, reduced effect sizes (reduced F-values) were found for the HI data compared to the NH data. In the random effects, significant interactions were found between participants and repetitions ($\chi^2 = 204$, $p < 0.001$). A post-hoc analysis confirmed the similarity in ASW between two neighboring PSW values, but showed significant differences of ASWs for PSW values further apart ($p_{\text{posthoc}} < 0.001$, $c\,\text{PSW} = 10$) indicating that ASW perception was not entirely independent of PSW in these listeners. Significant differences were also found for these listeners when comparing the results obtained for the different source signals to each other, i.e. noise and speech ($p_{\text{posthoc}} < 0.001$, $c_{\text{HA}} = 3$), noise and guitar ($p_{\text{posthoc}} = 0.03$, $c_{\text{HA}} = 3$) as well as speech and guitar ($p_{\text{posthoc}} = 0.01$, $c_{\text{HA}} = 3$).

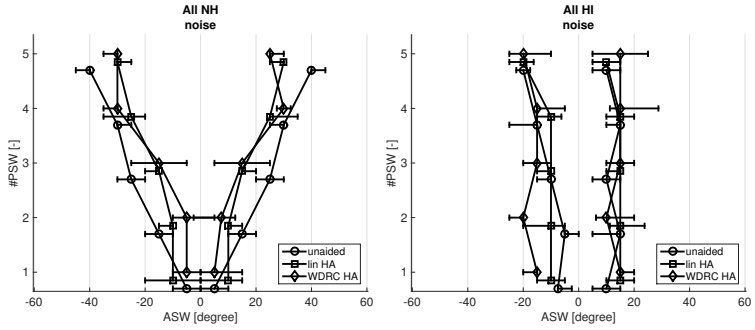


Figure 6.4: Averaged ASW results for normal-hearing (NH) (left) and hearing-impaired (HI) listeners (right). ASW is shown in degrees on the x-axis, as left and right expansion of the sound source, as a function of the physical source width (PSW), PSW #1 (narrow) to #5 (wide). Plotted are the median and respective 25th and 75th percentiles. Shown are the results for all hearing aid conditions [unaided], [lin HA] and [WDRC HA] in case of the pink noise signal. The HA conditions in each panel are represented by the different symbols.

Figure 6.4 shows the results of the listening experiment obtained in the aided conditions, i.e with HAs with the two programs "linear" or "WDRC" processing. The results are only shown for the noise source condition since the data exhibited the largest dynamic range of ASWs for this stimulus. The squares

represent the results for the linear processing and the diamonds indicate the results obtained with the WDRC processing. For comparison, the circles represent the results for the unaided condition (replot from Fig. 6.3). For the NH listeners (left panel), the HA processing caused smaller ASWs compared to the unaided condition ($p_{\text{posthoc}} < 0.01$, $c_{\text{HA}} = 3$; for the noise source), regardless of the selected program, especially for the PSWs of #3 and #5 (but not for PSW #1 and #4). Besides the HA condition as a significant main factor (F(2, 10) = 7.7, $p = 0.01$), an interaction between PSW and HA condition (F(8, 719) = 12, $p < 0.001$), source signal and HA condition (F = 5.4(4, 719), $p < 0.001$) as well as a three-way interaction between HA condition, PSW and source signal (F(16, 719) = 3.3, $p < 0.001$) were found. In the random effects, significant interactions were found between participants and HA condition ($\chi^2 = 14.2$, $p < 0.001$). Considering all source signals together, only a significant difference between linear processing [lin HA] and the unaided condition remained. For the speech and guitar sources (not shown), the reduction in ASW compared to the unaided condition was, on average, less pronounced and led to an increased ASW for PSW #1. Significant differences were found for the speech source between the unaided condition and the linear processing ($p_{\text{posthoc}} = 0.01$) and the linear and the WDRC processing ($p_{\text{posthoc}} < 0.01$), but not for the guitar source. In contrast, for the HI listeners (right panel), the HA conditions tended to produce slightly (but not significantly; $p_{\text{posthoc}} = 1$ and $p_{\text{posthoc}} = 0.32$, $c_{\text{HA}} = 3$; for the [HA lin] and [HA WDRC], respectively) larger ASW values compared to the unaided condition. Furthermore, the pattern was found to be more asymmetric for the HI listeners due to larger differences across the listeners (see Fig. 6.6). For the HI listeners, the HA condition was found to be a non-significant factor (F(2, 14) = 1.6, $p = 0.24$). However, a significant interaction between the HA condition and the source signal (F(8, 666) = 3.6, $p < 0.001$) was found, even though the posthoc analysis did not reveal significant differences between the HA conditions for any source type ($p_{\text{posthoc}} = 1$ and $p_{\text{posthoc}} = 0.32$, $c_{\text{HA}} = 3$; for the [HA lin] and [HA WDRC], respectively). In the random effects, significant interactions were found between participants and HA condition ($\chi^2 = 406.3$, $p < 0.001$) as well as repetition and HA condition ($\chi^2 = 147.8$, $p < 0.001$). Figure 5 (left) shows the individual ASW data for the six NH listeners represented in the individual panels. The results are shown for the noise source alone, contrasting the unaided condition (circles) and the aided conditions (squares for the lin HA and diamonds for the WDRC HA). The dynamic range of ASWs in the unaided

condition varied across the individual listeners, such that small PSW values were perceived in the range of ±5° to ±15° ASW and large PSW values were perceived in the range of ±30° to ±50°. However, each listener provided largely reproducible responses as indicated by the small error bars within about 5°. For the listeners shown in the top panels, KSt (top) and AP (right), the WDRC caused a strong reduction in ASW for all presented PSW values. The linear processing did also affect listener KSt in a similar fashion as the WDRC at large PSW values, whereas for listener AP the linear processing resulted in similar ASW values as in the unaided condition (except for the PSWs #1 and #5). The remaining listeners were affected less by the WDRC processing. Listener MM (left middle panel) revealed a very large dynamic range of ASW in the unaided condition which was similarly reduced in the cases of both HA programs (mainly for PSW #1 and #3). Interestingly, for listener PC (right middle panel), the linear processing led to an ASW of about 20°, independent of the PSW, whereas the WDRC program had a smaller impact on the listener's ASW. Similar effects were observed for listener CS (bottom left panel), especially for the PSWs #1 to #3. For listener KG (bottom right panel), only minor changes in ASW due to HA processing were found. The HA produced increased error bars in some of the listeners (KSt, MM, CS, KG) in several conditions (PSW #2, #2, #1 and #3, respectively), which indicates a reduced precision in determining the left and right boundaries. Listener CS was the only listener that indicated 3 image splits (one each for PSW #2, #3 and #4) in the unaided condition.

Figure 6.6 shows the individual ASW data for the six HI listeners. In the unaided condition (circles), almost constant ASW values independent of the presented PSW were obtained for most of the listeners, reflecting a low sensitivity to changes of the PSW. The obtained ASW values varied across listeners in the range from ±5° to ±20°. However, as for the NH listeners, the individual listeners were very consistent in their responses as reflected by small error bars of below 5°. Listener JG (top left panel) was the only listener who showed ASW ratings that were similar to those of the NH listeners, although with a lower precision of determining the spatial boundaries, as reflected by the larger error bars (20°). For this listener, the results were obtained with the master hearing aid (MHA). While the linear processing [lin MHA] (squares) essentially preserved the ASW performance, the WDRC program (diamonds) led to an increased ASW, especially for PSWs #1 and #2. Listener KM (top right panel) showed small dif-

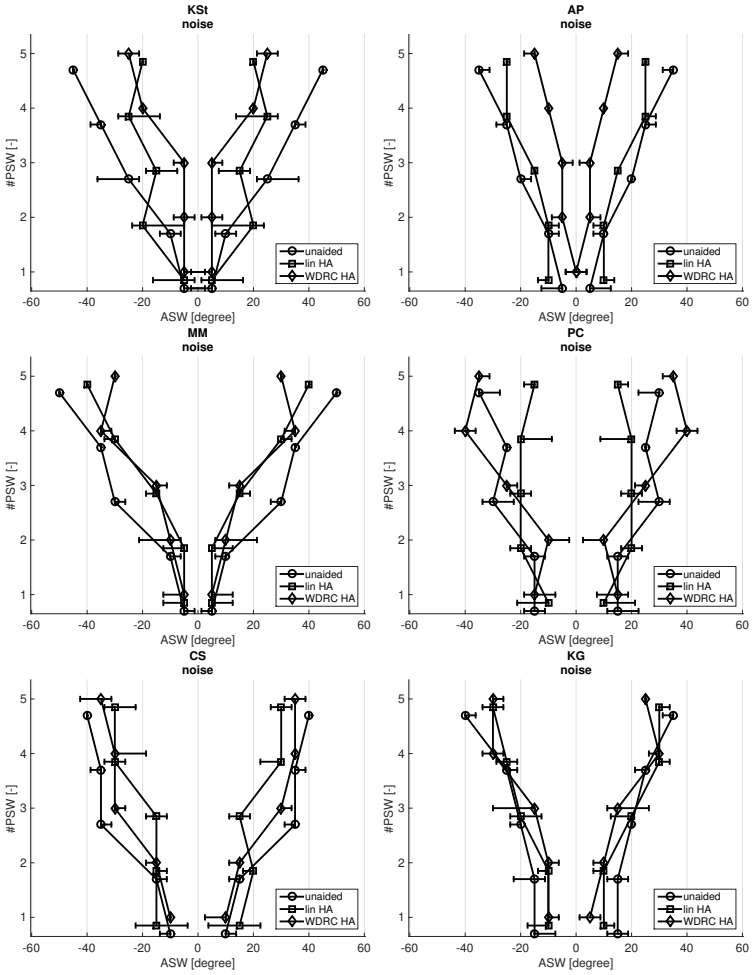


Figure 6.5: Individual ASW results for the 6 normal-hearing (NH) listeners. ASW is shown in degrees on the x-axis, as left and right expansion of the sound source, as a function of the physical source width (PSW), PSW #1 (narrow) to #5 (wide). Plotted are the median and respective 25th and 75th percentiles. Results are shown for all hearing aid conditions [unaided], [lin HA] and [WDRC HA] in case of the pink noise signal. The various HA conditions are represented by the different symbols.

ferences in ASW only for the left boundary estimate with increasing PSW which was almost unaffected by the HA processing. For the remaining listeners, a monotonic increase of ASW with PSW was absent and the ASW values observed in the unaided condition were roughly constant (e.g. listeners KL and JHN in the middle panel). Listeners KS and ET (bottom panels) indicated different values for the left and the right boundary estimates. These differences were found to be even more pronounced in the conditions with HA processing, as indicated by large shifts of the boundary estimates. These results might have

been caused by an asymmetric head-positioning during the experiments or by an asymmetric placement of the hearing aids. Listener JHN was the only one that indicated 2 image splits for the [lin HA] condition.

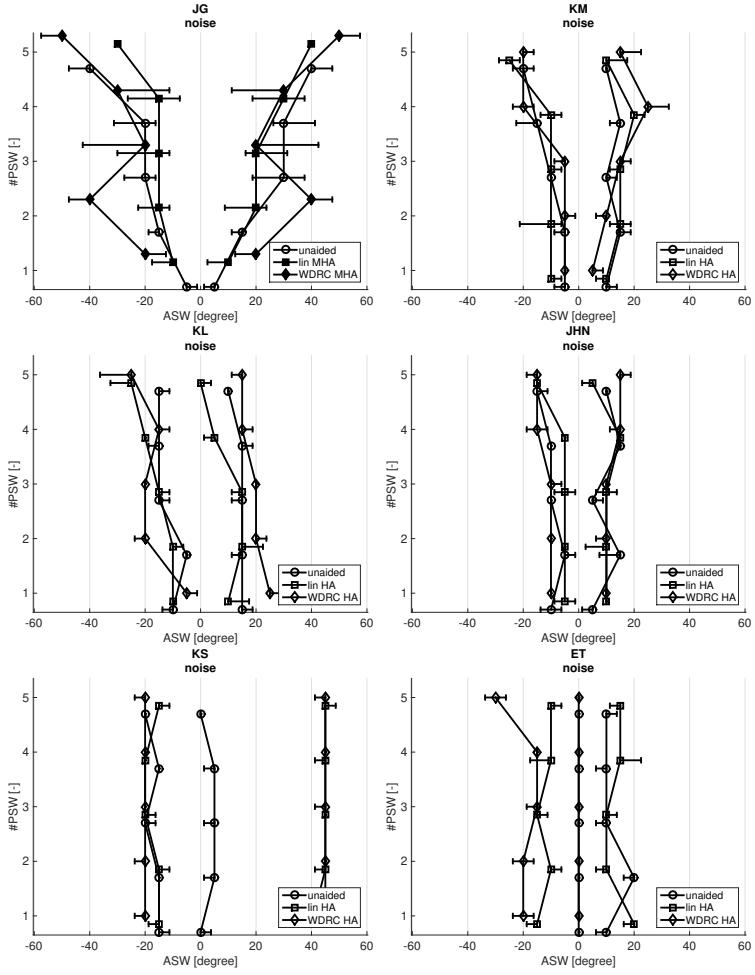


Figure 6.6: Individual ASW results for 6 hearing-impaired (HI) listeners. ASW is shown in degrees on the x-axis, as left and right expansion of the sound source, as a function of the physical source width (PSW), PSW #1 (narrow) to #5 (wide). Plotted are the median and respective 25th and 75th percentiles. Results are shown for all hearing aid conditions [unaided], [lin HA] and [WDRC HA] in case of the pink noise signal. Note that for listener JG, the HA data was not available and the MHA data is shown instead. The various HA conditions are represented by the different symbols.

6.3.2 IC Analysis

Figure 6.7 shows $IACC_{broad}$ (top panels) and $IACC_{E3}$ (bottom panels) predictions for the same stimuli as used for the experiments for the unaided condition

(left panels). As for the experimental data, two boundaries are shown in the simulations. The model produced only a single output value such that the two boundaries represent mirrored versions of each other. The $IACC_{broad}$ estimation captures the trend in the measured data for the speech signal (dashed dark grey line) and the guitar (dotted light grey line), except for PSW #5, where the $IACC_{broad}$ estimation does not further increase in contrast to the measured data. However, for the noise source, the model is insensitive with respect to changes of PSW, which is inconsistent with the data. The model correctly predicts a generally larger ASW for the noise source compared to the other two sources. The right panel of Fig. 6.7 shows the predictions obtained for the aided conditions. Here, the model remains insensitive and provides the same values as in the unaided condition, independent of the selected HA processing. The left bottom panel of Fig. 7 shows the $IACC_{E3}$-based prediction for the different source signals in the unaided condition. In contrast to the $IACC_{broad}$-based prediction, the $IACC_{E3}$ correctly captures the trend of the data for all three source signals. However, $IACC_{E3}$ is not consistent with the measured data in some conditions. For example, the model shows a decrease between PSW #4 and PSW #5 in the cases of the noise and the guitar sources and predicts equal values for PSW #1 and PSW #2 in the case of the guitar source. Also, the perceptual differences between the noise source and the other two sources are not captured by this model. Similarly, $IACC_{E3}$ (right bottom panel) cannot reveal any differences between the HA conditions. Comparisons of the performance between the $IACC_{E3}$-based (average of three octave bands) and $IACC_{broad}$-based (broadband analysis) ASW predictions suggest that a frequency-selective analysis seems to be important in the calculations.

6.3.3 ITD and ILD Analysis

The analysis of the $IACC_{broad}$ or $IACC_{E3}$ did not show any differences between the HA conditions. Therefore, other cues must have contributed to the measured differences in ASW caused by the HA processing. Figure 6.8 shows the ITD histograms for the noise source in the unaided condition for PSW #1 (left panel) and PSW #5 (right panel). The ITDs are shown as a function of frequency. A dark color indicates a high probability of occurrence of ITDs. When comparing PSW #1 and PSW #5, it can be seen that the spread of ITDs increases as well, especially at frequencies below 1 kHz. For frequencies above 1 kHz, ambiguous

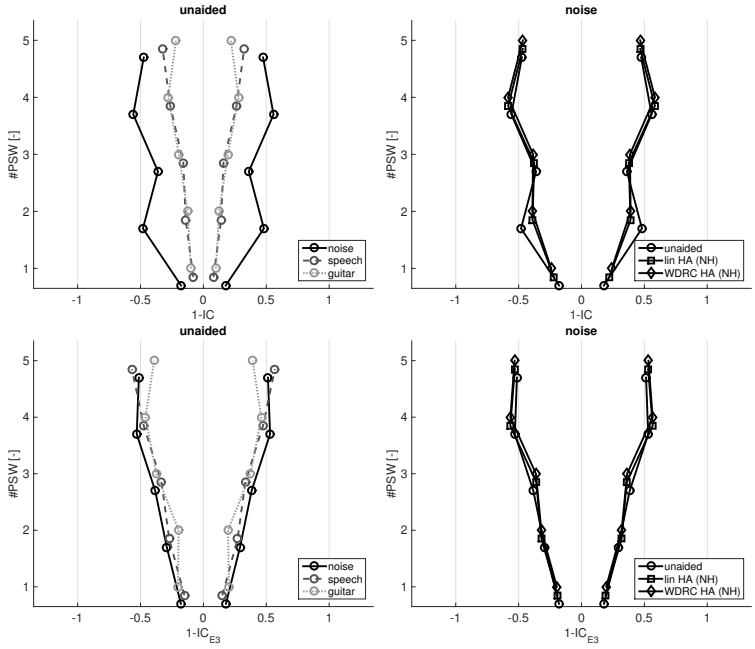


Figure 6.7: Evaluation of $IACC_{broad}$ (top panels) and $IACC_{E3}$ (bottom panels) as a function of the physical source width (PSW). Left: For the presented source signals (represented by the different line styles and grey shades) in the unaided condition. Right: For the various HA conditions for NH listeners (represented by the different symbols) in case of the pink noise source. Left and right boundaries are identical and are shown to be consistent with the psychoacoustic data (see Figure 6.3).

ITDs occur which decrease with increasing frequency. They are caused by the periodicity in the interaural cross-correlation function used to extract the ITDs and are equal to the reciprocal of the corresponding center frequency in each frequency channel.

Figure 6.9 shows the histograms for the ILDs for PSW #1 (left panel) and PSW #5 (right panel). The spread of ILDs is larger for PSW #5, especially at frequencies between 500 and 2000 Hz, even though less pronounced as for the ITDs. Thus, both cues, ITDs and ILDs, show an increase in their fluctuations over time with increasing the PSW.

To further quantify the changes in the physical properties of the stimulus between the aided and the unaided condition, cross-correlations between their ITD and ILD histograms were calculated. Figure 6.10 displays the correlation across frequency bands between ITD histograms (top panel) and ILD histograms (bottom panel). The figure compares the WDRC HA and the unaided condition

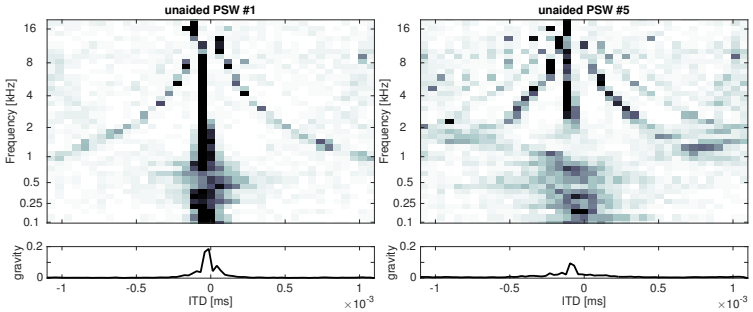


Figure 6.8: Normalized occurrences of ITDs per frequency channel shown in form of histograms. Results are here shown for the pink noise source for the extrema of the physical source width (PSW), narrow (PSW #1) in the left panel and wide (PSW #5) in the right panel. A dark color represents a high relative occurrence of ITDs. The accumulated distribution across all frequency channels (center of gravity) is shown at the bottom of each panel.

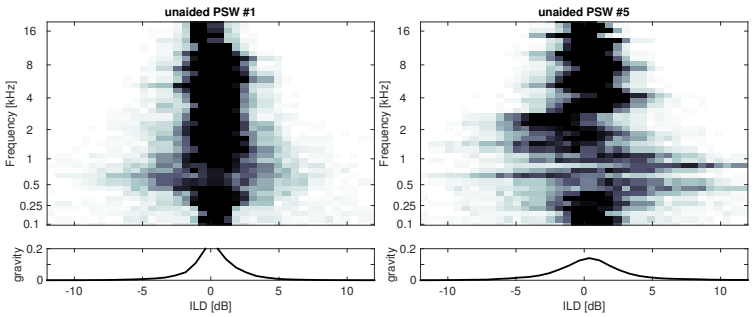


Figure 6.9: Normalized occurrences of ILDs per frequency channel shown in form of histograms. Results are here shown for the pink noise source for the extrema of the physical source width (PSW), narrow (PSW #1) in the left panel and wide (PSW #5) in the right panel. A dark color represents a high relative occurrence of ILDs. The accumulated distribution across all frequency channels (center of gravity) is shown at the bottom of each panel.

in the case of the noise source and PSW #5, where significant differences in ASW for NH listeners were measured between both HA conditions. The correlation for each angular position of the HATS' head is indicated by the different grey-shades. The mean correlation across all head orientations is shown in the same plot as a thick red line. Its average and the corresponding rms-error across the frequency bands, indicated by vertical dashed lines, are displayed on top of each panel. For the ITDs, the average correlation between the WDRC HA and the unaided conditions is high and drops at frequencies above about 1.5 kHz. This is due to small wavelengths at high frequencies, where small differences between the measurement positions in the aided and the unaided conditions resulted in large differences in ITDs, and hence a low correlation. Therefore, above 1.5 kHz, a comparison between unaided and aided conditions based on

their ITD histograms becomes inaccurate. The average correlation of the ITDs up to this frequency limit was 0.85. Thus, the ITDs - as well as the IC - were mainly maintained by the HA. The agreement between both measures, ITDs and IC, might be explained by the fact that both are extracted from the interaural cross-correlation function. For the ILDs a lower average correlation of 0.75 was obtained. Here, only frequencies between 0.5 and 10 kHz were considered, since ILDs were maintained below 0.5 kHz and the HA's cut-off frequency was at 10 kHz. ILDs were mostly modified at 1, 2 and between 4 and 8 kHz. The lower correlation of ILDs as opposed to the correlation of ITDs shows that ILDs have been altered more strongly than ITDs. The same analysis for the the unaided and the linear HA condition revealed similar results as shown for the WDRC processing. Therefore, based on the analysis of histogram correlations, modifications of ILDs might have constituted the lower ASW in the aided compared to the unaided condition, independent of the HA program.

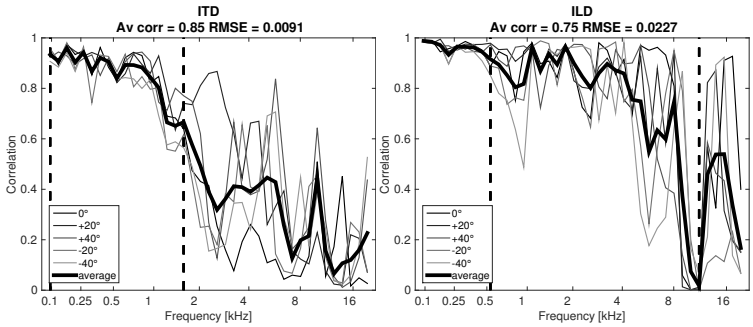


Figure 6.10: Correlation plots of ITD histograms (left) and ILD histograms (right) for different rotation angles of the HATS' head (indicated by different grey-shades) in the case of the noise source for PSW #5. The mean correlation and RMS-error (displayed on top of each panel) are averaged across all five angular positions for the indicated frequency range by the dashed vertical lines.

6.4 Discussion

6.4.1 ASW perception in "unaided" NH and HI listeners

The results showed that the NH listeners could well distinguish the ASWs of the presented PSW values for a given source signal. The pink noise source was generally associated with a larger ASW than the speech and guitar sources. This might be due to the spectral differences between the three sources. The pink

noise signal had more energy at low frequencies than the other two signals and low frequencies generally cause a larger ASW than high frequencies in sounds with constant IC (e.g.; Blauert and Lindemann, 1986a and Käsbach et al., 2014a. In contrast to the NH listeners, the HI listeners generally showed a reduced dynamic range of ASW values for all three source signals. In acoustically dry environments (such as represented by PSW #1 and corresponding to a high IC), the HI listeners assigned larger ASW values to the sounds than the NH listeners, whereas in reverberant spaces (represented by PSW #5 and corresponding to a low IC), the HI listeners' ASW was found to be smaller than for the NH listeners. However, the within-listener variability of ASW was similar in both listener groups. Overall, consistent with earlier results reported in Whitmer et al. (2012, 2014), the results of the present study demonstrated a reduced sensitivity to changes in ASW across conditions in most HI listeners, which might be detrimental for sound source separation, e.g. in multi-talker conditions.

6.4.2 ASW perception in aided conditions

Wearing the BTE-HAs created a smaller ASW in the NH listeners in the case of the noise source, especially for PSW #1 and #3. In addition, some listeners showed a reduced precision in terms of determining the boundaries of the sound images for several PSWs. Both effects were independent of the selected HA program. Thus, the change in ASW perception could not be associated with WDRC. In the HI listeners, the HA did neither restore nor negatively affect the ASW compared to the unaided condition, with the exception of one listener (JG), whose performance was similar to that of the NH listeners. This suggests that, in comparison to the NH listeners, the HI listeners were generally less sensitive to listening through the HA, for both HA programs. Neither the HI listeners (besides only two incidences) nor the NH listeners reported occurrences of image splits for either HA program. However, the smaller ASW ratings in the NH listeners obtained in the two aided conditions were probably related to changes in the ILD distributions relative to the unaided conditions. These occurred primarily at high frequencies presumably due to the microphone placement behind the ear of the listener. For frontal source directions, resonance frequencies measured behind the ear of a listener are different, especially for frequencies above 2 kHz, from those measured at the entrance of the ear canal capturing pinna reflections (Sivonen, 2011). This suggests that the applied compensation

filters for the microphone position were insufficient to maintain the natural ASW perception in the individual listener. While changes in ILDs at high frequencies might have affected the ASW of the NH listeners, such cues probably remained inaudible for most of the tested HI listeners. One reason might be the generally higher just noticeable differences for ILDs in HI listeners (Musa-Shufani et al., 2006). Another reason might be that the applied insertion gain in the BTE-HA amplified resonance frequencies that deviated from individual pinna resonances which might be crucial for a correct ASW perception. To further investigate whether high-frequency ILDs change the ASW in a HI listener, an unaided condition with high-frequency amplification of the stimuli should be tested, e.g. using the experimental setup of the present study. If verified, it would be recommendable in the clinical profiling of the HI listeners to include measurements of ILD thresholds, besides measurements of ITD thresholds (or instantaneous interaural phase difference detection, e.g.; Whitmer et al., 2014), which might influence the choice of the HA type. Especially, it would be important to identify individual HI listeners that remain sensitive regarding ASW perception (as has been seen for listener JG in this study), since a BTE-HA might be detrimental for their ASW perception. Instead, such listeners could be offered in-the-ear (ITE) or completely-in-canal (CIC) HAs which preserve the listener's pinna cues.

6.4.3 Effect of WDRC vs linear processing on ASW

Both HA programs led, on average, to the same ASW results. The setup was symmetrical which implied smaller natural ILDs as compared to a laterally presented sound source. The 2:1 compression of these small ILDs by the WDRC processing was probably below the listeners' just noticeable difference for ILDs, such that differences between the two programs were, on average, not noticeable. Even sound sources with larger ASW (at higher PSW values), revealing larger ILDs compared to the compact source (see Fig.6.10), remained unaffected by the WDRC. A larger impact of WDRC on ASWs might be expected for laterally presented sound sources, which imply larger ILDs, and applying higher compression ratios than used in the current study. However, the compact speech source, presented through a single loudspeaker at 0° (PSW #1), was the only signal where significant differences between the linear and the WDRC processing were observed in the NH listeners. The ASW was found to be increased

here which is consistent with the results in Wiggins and Seeber (2011, 2012), where also an increased ASW for compact speech sources presented at ±60° was observed. Therefore, it seems that compact speech sources, as occurring in an acoustically dry environment (corresponding to a high IC), are affected by the WDRC processing for lateral as well as central source positions.

6.4.4 Perspectives

In this study, a reduced ASW sensitivity was found in the HI relative to the NH listeners. It is not clear to what extent limitations in ASW perception affect source segregation in a multi-talker environment. Such effects could, for example, be investigated by measuring speech intelligibility of a target-talker in the presence of maskers using compact as well as spatially diffuse sound sources. Furthermore, the findings of this study revealed strong similarities to results obtained in studies on externalization (Boyd et al., 2012). Their results showed that, in contrast to NH listeners, HI listeners revealed a reduced dynamic range of externalization, i.e. sounds were neither fully externalized nor internalized. Furthermore, a BTE-HA simulation with linear processing reduced the degree of externalization for NH listeners, but not in HI listeners. Thus, for NH listeners, the dynamic range of both, externalization and ASW, was reduced by the linear BTE-HA condition. Likewise, for HI listeners, neither externalization nor ASW were affected by such condition. It would be valuable to examine the relation between externalization and ASW perception in more detail. Finally, in the current study, the binaural cues (IC, ITDs and ILDs) were analyzed separately. Considering these cues together in an auditory modelling framework of ASW perception would help to further evaluate HA signal processing strategies and could provide guidelines of how to restore spatial perception in individual listeners. For example, the model suggested by van Dorp Schuitman et al. (2013) could be considered in such an investigation. However, their model is based on the evaluation of ITD fluctuations whereas the results from the current study suggest to also include ILDs in the analysis. Such a combined analysis of ITDs and ILDs was proposed in the functional auditory model of Mason et al., 2005b.

6.5 Summary and Conclusion

In the present study, ASW perception was investigated in NH and HI listeners. While the NH listeners were able to distinguish the ASW as a function of the physical source width, it was found that HI listeners were less sensitive in this task, which is consistent with the results in previous studies (Whitmer et al., 2012, 2014). On the one end of the scale, sounds that NH listeners assigned with a small ASW (as occurring in dry listening conditions) were typically associated with a larger ASW by the HI listeners. On the other end of the scale, sounds that the NH listeners assigned with a large ASW (as occurring in reverberant listening conditions) were rated smaller by the HI listeners. The overall dynamic range of ASW ratings was, therefore, clearly reduced in the HI group in comparison to the NH group. The influence of HA processing on ASW perception was also evaluated considering either linear or WDRC processing. The results showed a similar effect of the two programs: While for the NH listeners, the dynamic range of the ASW ratings was found to be reduced, HI listeners rated the ASW to be the same as in the unaided condition. An analysis of the IC and the statistics of ITDs and ILDs was performed for the stimuli used in the experiments. All three binaural cues were sensitive to changes in PSW. While IC and ITDs remained largely unaffected by the HA processing, the statistics of ILDs showed differences between the aided and unaided conditions. The microphone placement in the HAs was found to affect ASW more than the HA's signal processing itself, suggesting that the listener's natural HRTFs are important to maintain for an authentic spatial perception.

6.6 Acknowledgment

This article is based on the study "Control of spatial impression by means of apparent source width" under the Danish Sound Innovation Network consortium (http://www.danishsound.org). The author would like to thank all participating members from DTU, Widex A/S and Bang & Olufson, for their contribution and support.

7

Modelling apparent source width perception[a]

Abstract

A primary function of the human auditory system is the spatial perception of sound sources in a local environment. This requires that the sound sources are externalized, referring to a correct localization in space (position and distance) and a compact sound image (Hartmann and Wittenberg, 1996). The present study focused on the compactness of sound sources, which is commonly referred to as the apparent source width (ASW). To elicit this perception, the auditory system has access to fluctuations of binaural cues, the interaural time differences (ITDs), interaural level differences (ILDs) and the interaural coherence (IC). To quantify their contribution to ASW, a functional model of ASW perception was developed. The model determines the left- and right-most boundary of a sound source using a statistical representation of ITDs and ILDs based on percentiles integrated over time and frequency. The model's performance was evaluated against psychoacoustic data obtained with noise, speech and music signals in loudspeaker-based experiments. A robust model prediction of ASW was achieved based on an analysis of IC or ITDs, in contrast to an analysis based on a combination of ITDs and ILDs where the performance slightly decreased.

[a] This chapter is an extended version of Käsbach et al. (2016b) (peer-reviewed article at the Proc. of ICA, Buenos Aires, Argentina, 2016).

7.1 Introduction

Spatial perception implies decoding the auditory scene surrounding a listener. Each sound source in such a scene has a certain location and distance with respect to the listener. This spatial separation helps the listener to distinguish concurrent sources from each other, e.g. a target speaker from interfering noise sources. The perceived horizontal extent of sound sources is typically described by the apparent source width (ASW). A reduced sensitivity to ASW as, e.g., found in hearing-impaired listeners (Whitmer et al., 2012; Whitmer et al., 2014; Käsbach et al., 2016c; see Ch. 6) may have consequences for the ability to spatially separate sound sources. Therefore, it is important to understand the contributing cues to ASW perception. Three binaural cues are commonly considered to contribute to the ASW of a sound source: The interaural time differences (ITDs), the interaural level differences (ILDs) and the interaural coherence (IC). Due to reflections in rooms and from the head and torso of the listener, all three cues fluctuate over time. With increasing amount of room reflections, the IC decreases and larger variations in ITDs and ILDs occur, leading to an increased ASW. The relation between the binaural cues and ASW can be exploited by binaural auditory models.

Traditional models of ASW have been used to evaluate the quality of concert halls by analyzing the interaural cross-correlation (IACC) function (Ando, 2007). Based on the IACC, the interaural coherence (IC) can be extracted as the absolute maximum value normalized by the root-mean-square (RMS) value of the left- and right-ear signals. Hereby, an inverse relation between IC and ASW exists. Okano et al. (1998) proposed a frequency-specific weighting of the IC, termed $\mathrm{IACC}_{\mathrm{E3}}$, that averages the IC in three octave bands centered at 0.5, 1 and 2 kHz. The $\mathrm{IACC}_{\mathrm{E3}}$ is calculated for the first 80 ms of the binaural room impulse responses (BRIRs) since early reflections are known to contribute mostly to ASW (Bradley, 2011). Zotter and Frank (2013) observed a high correlation of $r = 0.97$ between the $\mathrm{IACC}_{\mathrm{E3}}$ and the perceptual data obtained in a stereo loudspeaker measurement setup. Similar ideas as the ones provided by Okano et al. (1998) were implemented in a complex binaural auditory model by van Dorp Schuitman et al. (2013) which divides the input signal in a direct and a reverberant stream. From the direct stream, the model derives ITDs up to 2 kHz and estimates the ASW by averaging their standard deviation. In contrast to the

traditional IC-based measures, this model is applied on binaural recordings. The model showed higher correlations with perceptual data obtained using room simulations than the $IACC_{E3}$. Blauert and Lindemann (1986b) proposed that both ITD and ILD fluctuations contribute to ASW. They combined the standard deviation of the two cues with equal weights and reported a higher correlation with the perceptual data ($r = 0.75$) than an IC-based model ($r = 0.61$). Later, Mason et al. (2005b) developed an ASW model that combined both ITDs and ILDs according to the duplex theory, by using ITDs at low frequencies and ILDs at high frequencies (Macpherson, 2002). Furthermore, the loudness of the stimuli was estimated and integrated in their model. Also, Okano et al. (1998) and van Dorp Schuitman et al. (2013) considered the sound pressure level (SPL) as an additional cue for ASW. Thus, several models of ASW have been presented, each validated on different perceptual datasets.
The present study investigated the generalizability of such models by evaluating their performance across three experimental datasets that were obtained for band limited and broadband noise, as well as for speech and music signals (Käsbach et al., 2014a; Käsbach et al., 2016c; Grosse et al., 2015). Here, it was investigated whether correlation-based approaches, i.e. using ICs or ITDs (as suggested by Okano et al. (1998) and van Dorp Schuitman et al. (2013), respectively) are sufficient for the estimation of ASW. Furthermore, it was tested whether the analysis in the three octave bands at 0.5, 1 and 2 kHz or below 2 kHz, is most appropriate or whether high-frequency ICs or ITDs also contribute to ASW. Finally, it was evaluated if a model combining ITDs and ILDs (as suggested by Blauert and Lindemann (1986b) and Mason et al. (2005b)) is feasible.

7.2 ASW models

All considered models operated on binaural signals for the same source signals as used in the experiments. Binaural recordings were obtained with a head and torso simulator (HATS) that was placed at the listener's position. Two binaural models were used to predict ASW: a functional model developed in this study and a more complex auditory model as described in van Dorp Schuitman et al. (2013).

7.2.1 Functional ASW model

Figure 7.1 shows a schematic diagram of the functional ASW model. The auditory processing is based on the auditory-front-end (AFE) developed by the TWO!EARS consortium (May et al., 2015; TWO!EARS, 2013-2016). It consists of various processing stages, including gammatone filtering (Hohmann, 2002), inner haircell transduction (IHC) and absolute threshold of hearing (ATH). The signals are processed in each monaural channel by a gammatone filterbank (Hohmann, 2002) to mimic the frequency selectivity of the basilar membrane. The 35 filters have a bandwidth of one equivalent rectangular bandwidth (ERB) in the frequency range between 80 to 11891 Hz. In the next stage, the IHC transduction is simulated, i.e. the loss of phase locking to the stimulus' fine structure at high frequencies. The IHC processing is performed according to Bernstein et al. (1999), suggesting a cut-off frequency of 425 Hz and simulating basilar-membrane compression, as in the auditory model of localization perception by Faller and Merimaa (2004). In the following stage, frequency bands are selected for further processing according to their estimated level. The signals are calibrated to a RMS value corresponding to the 70 dB SPL of the experimental stimuli. Frequency bands with an output level below the ATH as defined in Terhardt (1979) are not considered further in the processing. From the two monaural (left and right) channels, the model extracts ITDs, ILDs and IC per time-frequency units for the ASW prediction. The signals in both channels are analyzed in short-time hanning windows of 20 ms duration, with an overlap of 50%, resulting in a time-frequency representation in each channel. The IC and ITD are extracted from the normalized interaural cross-correlation function per time-frame. The IC is equal to the maximal coherence (see Eq. 2.3) and the ITD corresponds to the timelag at this value. Timelags were limited to a range of ±1.1 ms. The ILDs were defined as the energy difference in dB between the two ear signals. The ASW estimation was based on the statistical distribution of the binaural cues. The width of this distribution was represented by percentiles and resembled the ASW. Hereby, the left- and right-most boundary of the sound source corresponded to the lower and upper percentile from the distribution's median.

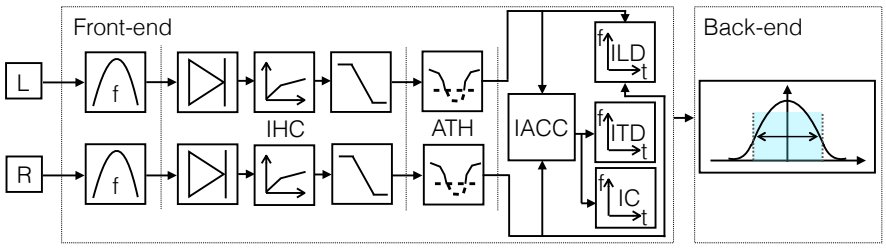


Figure 7.1: Schematic diagram of the functional ASW model. The front-end consists of a gammatone filterbank, inner haircell transduction (IHC) and an absolute threshold of hearing (ATH). The binaural cues (ITDs, ILDs and ICs) are calculated in time-frequency units. In the back-end, ASW is calculated based on percentiles of the binaural cues.

7.2.2 Complex auditory-based ASW model

The complex model was developed by van Dorp Schuitman et al. (2013) and is available in the Room Acoustic Analyzer (RAA) toolbox. It was designed for the analysis of room acoustical parameters including ASW, but also reverberance, clarity, and listener envelopment, with the aim to accurately reflect human perception. The model considers temporal and spectral masking effects, in contrast to the functional model, and is capable to follow dynamic changes in the sound pressure level by capturing non-linearities of the human auditory system. To simulate the binaural interaction, the complex model uses Breebaart's equalization-cancellation (EC) approach with excitation-inhibition (EI)-type elements (Breebaart et al., 2001). Further, the model has the ability in its central processor to split the incoming sound into two streams, a direct stream, capturing source-related perception, and a reverberant stream that is associated with the perception of the acoustical environment. The ASW is estimated using the direct stream only. Figure 7.2 shows a schematic diagram of the model. The fourth-order gammatone filter bank consists of 41 frequency bands with center frequencies from 27 to 20577Hz. The IHC device applies a fifth-order lowpass filter with a cutoff frequency of 770 Hz after half-wave rectification of the signal. The ATH is considered in the model according to Terhardt (1979). In addition, the model comprises five adaptation loops (AL1-5) to simulate the non-linearities of the auditory system regarding temporal masking and intensity coding (Dau et al., 1996). The ALs act effectively as a logarithmic compressor with an input-output relation of approximately $y \approx x^{1/32}$ for stationary signals. The ASW prediction, p_{RAA}, is calculated in the central processor as a linear com-

bination of the standard deviation of ITDs, σ_{ITD}, in the frequency bands from 387 Hz to 1.84 kHz and the monaural output level, L_{LOW}, at the low-frequency bands from 168 to 387 Hz according to:

$$p_{\mathrm{RAA}} = \alpha L_{\mathrm{LOW}} + \log_{10}(1 + \beta \sigma_{\mathrm{ITD}} 10^3), \tag{7.1}$$

with $\alpha = 2 \cdot 10^{-2}$ and $\beta = 5.63 \cdot 10^2$ being weighting factors that were obtained by a parameter optimization using a genetic algorithm. Since the model requires signals with a duration of at least 10 s, the binaural recordings were repeated and truncated accordingly.

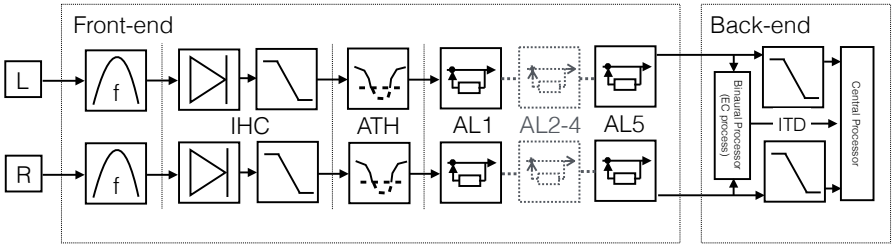


Figure 7.2: Simplified schematic diagram of the complex binaural ASW model. The front-end comprises a gammatone filterbank, inner haircell transduction (IHC) and an absolute threshold of hearing (ATH). In addition, five adaptation loops (ALs) allow for following dynamic changes in the input signal's SPL. In the back-end the binaural interaction is simulated using the equalisation-cancellation (EC) principle. For further details, see van Dorp Schuitman et al. (2013).

7.2.3 Calibration of the models

A calibration stage was required to map the output of each model to ASW in degrees. Using a linear fitting approach, the calibrated model output was $y_{\mathrm{cal}} = a\,y + b$, where a is a sensitivity parameter, b an offset and y the uncalibrated model output.

7.3 Methods

7.3.1 Selected reference data of ASW perception

A successful ASW model should be generalizable across various conditions. Here, three experimental ASW studies were selected to evaluate the models' performance. The data comprised conditions with source signals of varying frequency content, sound pressure level and duration, as well as data obtained when using hearing aids. Specifically, the data from the studies of Käsbach et al. (2014a) (see Ch. 4), Käsbach et al. (2016c) (see Ch. 6) and Grosse et al. (2015) (external dataset) were considered, in the following referred to as Exps. A, B and C, respectively. In these studies, distinct sensations of ASW were generated by using loudspeaker setups.

The ASW was measured as a function of the physical source width (PSW) which was controlled by two experiment-specific settings, the loudspeaker layout and the applied signal processing. In Exps. A and B, stereo loudspeaker setups were used, where the listeners perceived a phantom sound image in the center of the two loudspeakers due to summing localization. In the measurement procedure, the listeners indicated the perceived ASW on a degree scale, as illustrated in Figure 7.3. In Exp. B, the listeners indicated the left- and right-most boundary of the sound source separately, whereas in Exp. A the response was given symmetrically. In Exp. A, the stereo setup at an angle of ±30 degrees was used, as indicated by the red dashed rectangles in Figure 7.3. Five distinct PSW values, denoted by PSW #1 to PSW #5, were generated by varying the coherence between the two loudspeaker channels accordingly to $\mathrm{IC}_{\mathrm{LS}} = 1, 0.8, 0.6, 0.3$ and 0. The source signal was either Gaussian white noise, bandpass (BP) filtered with a bandwidth of 2 octaves at a center frequency of 0.25 kHz, 1 kHz or 4 kHz, or highpass (HP) filtered at 8 kHz. The stimuli had a duration of 2 s and were presented at 70 dB SPL. Two additional SPL conditions were tested at 50 and 60 dB. For further details on this experiment, the reader is referred to Ch. 4. In Exp. B, the PSW was controlled by varying the angle between the stereo speakers. In addition, a source widening algorithm was applied as described in Zotter and Frank (2013). Specifically, a line-array of 3 stereo loudspeaker pairs and an additional loudspeaker in the center of the array were used, as indicated by the gray rectangles in Figure 7.3. In total, five distinct PSW values

were generated. The source signals were pink noise, male speech and a guitar sample. The stimuli had a duration of 6 s and were presented at 70 dB SPL. In addition, three hearing-aid (HA) conditions were tested: A reference condition without hearing aids ("undaided") and two "aided" conditions were considered where either linear HA processing (referred to as "lin HA") or wide-dynamic range compression (referred to as "WDRC HA") was applied. For further details on this experiment and the HA settings, the reader is referred to Ch. 6.
In contrast to Exps. A and B, Exp. C (Grosse et al., 2015) used a half-ring of 31 loudspeakers in front of the listener, ranging from −90° to +90° and separated by 5.625°. The PSW was controlled by the number of loudspeakers used for the playback. The listeners provided their ASW ratings in degrees, by matching a visual marker to their auditory perception. The visual feedback was given by LEDs that were installed on top of the loudspeaker ring and could be adjusted accordingly via a slider, as illustrated in Figure 7.4. The response of the left- and right-most boundary had to be given symmetrically. The number of loudspeakers allowed for the playback of 16 PSW values, denoted by PSW #1 to PSW #16 PSW, using decorrelated white noise. Four different frequency bands were used, including the conditions broadband (0.1 to 10 kHz), lowpass (LP) filtered (0.1 to 1 kHz), BP filtered (1 to 3 kHz) and HP filtered (3 to 10 kHz), presented with a duration of 2000 ms at 65 dB SPL. In addition, shorter signal durations of 50, 150 and 1000 ms were tested. For further details, the reader is referred to Grosse et al. (2015).

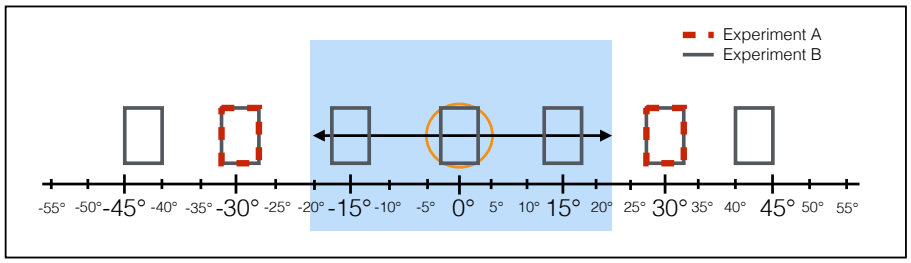


Figure 7.3: Sketch of the experimental set-up for Exp. A and B. The loudspeaker pairs generate a phantom source at 0 degree. Listeners were asked to indicate the ASW in degree, for both boundaries of the source image. For further details, see Ch. 4 (Käsbach et al., 2014a) and Ch. 6 (Käsbach et al., 2016c).

Figure 7.5 shows the measured ASW responses for Exp. A (left panel; 70 dB SPL), Exp. B (middle panel; undaided condition) and Exp. C (right panel, 2000 ms signal duration). Each panel shows the ASW ratings, in degrees, as a function of

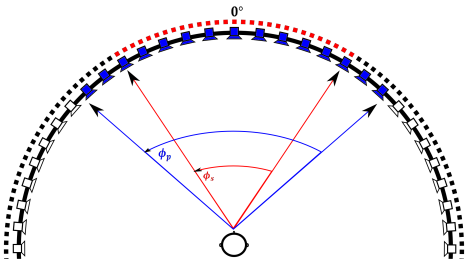


Figure 7.4: Sketch of the experimental set-up for Exp. C (taken from Grosse et al., 2015). The PSW was controlled via the number of loudspeakers that played back decorrelated noise signals. Listeners were asked to indicate the ASW in degree, for both boundaries of the source image via LEDs controlled with a slider. For further details, see Grosse et al. (2015).

PSW, averaged across listeners. The error bars represent the standard deviation across listeners. It can be seen that the ASW increases with increasing PSW for all three experiments. Experiment A (left panel) shows that the BP filtered signal at 0.25 kHz (blue circles) and the white noise signal (black squares) were perceived with larger ASW than the HP filtered signal at 8 kHz (green diamonds). The BP filtered source signals at 1 kHz and 4 kHz (not shown here) show a similar ASW as the white noise source (see Ch. 4).

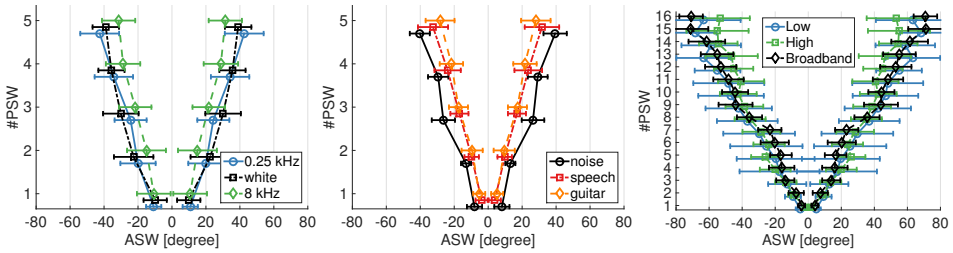


Figure 7.5: Perceptual results of ASW for Exp. A (left panel; 70 dB SPL condition), Exp. B (middle panel, unaided condition) and Exp. C (right panel; condition with 2000 ms signal duration). ASW is shown in degree as a function of the physical source width (PSW). Plotted are the mean and standard deviation. The different symbols and line styles represent the different source signals.

In Exp. B (7.5, middle panel), it can be seen that the ASW increases with PSW in a similar manner as in Exp. A (left panel). Small differences are noticeable between the source signals with the noise source having a larger assigned ASW than the speech and guitar signals. In Exp. C (7.5, right panel), the ASW increases with PSW also for larger PSW values. However, the ASW ratings are underestimated at large PSW values. For instance, at PSW #16, corresponding to loudspeakers at ±90°, the ASW was only ±70°. The source signals confirm the observed trend in Exp. A, i.e. low frequencies (blue circles) and the broad

band signal (black squares) show a larger ASW than high frequencies (green diamonds). The BP filtered source between 1 to 3 kHz (not shown here) produces a similar ASW as the HP filtered source (see Grosse et al., 2015).
Figure 7.6 shows the data for the sound pressure level (Exp. A; left panel), the hearing aid (Exp. B; middle panel) and the signal duration conditions (Exp. C; right panel). The left panel displays the ASW ratings as a function of the SPL in Exp. A in the case of the white noise source. With increasing SPL also the ASW increases. The middle panel shows the HA conditions of Exp. B in the case of the noise source. The ASW is smaller in the aided conditions compared to the unaided condition, whereas no differences were found between the two programs, [lin HA] and [WDRC HA]. The right panel displays the results obtained for the four different signal durations in Exp. C in the case of the HP filtered source (Grosse et al., 2015). For a very short signal duration of 50 ms, the ASW is smaller than for longer signal durations, indicating a build-up effect of ASW over time. This effect seems to saturate though for signal durations of about 250 ms (Grosse et al., 2015).

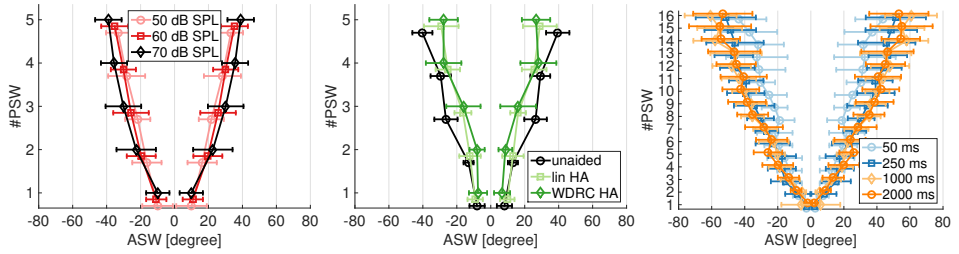


Figure 7.6: Perceptual results (further experimental conditions) of ASW for Exp. A (left panel; white noise source signal), Exp. B (middle panel; noise source signal) and Exp. C (right panel; HP filtered source signal). ASW is shown in degree as a function of the physical source width (PSW). Plotted are the mean and standard deviation. The different symbols and colors represent the different experimental conditions.

7.3.2 Back-end settings of the functional ASW model

The functional model was tested with different back-ends. The first back-end, termed DUPLEX, combined the percentiles of the ITDs and ILDs according to the duplex theory (Macpherson, 2002) which was motivated by Blauert and Lindemann (1986b) and Mason et al. (2005b). The combination of both binaural cues required the normalization of each cue. ITDs were normalized with 1.1 ms and ILDs with 12 dB SPL, which corresponded to the observed maxima, respec-

tively, in the percentiles across stimuli. According to the duplex theory, ITDs contribute up to 1.5 kHz and ILDs contribute above this frequency value. The final prediction of the left and right boundaries was then obtained by calculating the mean value across all frequency channels of the lower and upper percentile, respectively. In a second back-end, termed ITD_{low}, only the ITD-percentiles were analyzed with an upper frequency limit of 1.93 kHz comparable to the implementation in van Dorp Schuitman et al., 2013. The third back-end used the IC for the ASW prediction, termed IC_{E3}, resembling a short-term analysis of the $IACC_{E3}$. In total, 16 gammatone filters of the front-end were selected corresponding to the frequency range between 0.35 to 2.83 kHz, defined by the octave-wide filters in $IACC_{E3}$ at 0.5, 1 and 2 kHz. The frame-based values of IC were averaged with equal weights across all frames and frequency channels. The $IACC_{E3}$ according to Okano et al. (1998), served as a reference.

7.3.3 Evaluation of the ASW models

For the calibration of all models, two data points were used, PSW #1 and PSW #5 of the white noise stimulus in Exp. A. (see Section 7.3.1). The overall performance of the models was assessed across all experimental conditions, comprising all source signals in Exp. A at 70 dB SPL, all source signals and HA conditions in Exp. B and all source signals and signal durations (excluding 50 ms duration) in Exp. C. The individual model performance was accessed by calculating Pearson's correlation coefficient r^2 (and r) and the RMS error between the left and right boundaries of the calibrated model outputs and the experimental data. The presented ASW models, $IACC_{E3}$, IC_{E3}, ITD_{low}, DUPLEX and p_{RAA} were also compared in a statistical analysis considering Exp. A and Exp. B together. A 3-way analysis of variance (ANOVA) was performed using the model type, PSW and source signal as factors. In contrast to the correlation coefficient r^2, this allowed for a more detailed model analysis across both factors PSW and source signal. The evaluation was based on the Akaike information criterion (AIC) (using 13 degrees of freedom, dof = 13) which is a relative criterion, whereby a lower AIC indicates a better model performance.

7.4 Results and discussion

7.4.1 Effect of percentiles in the functional ASW model

Figure 7.7 shows the estimated percentiles per frequency channel for ITDs (upper panels) and ILDs (bottom panels) estimated in the functional ASW model in the case of the noise source in Exp. B. The percentiles increase from PSW #1 (narrow distribution in gray) to PSW #5 (wide distribution in red), especially for the ITDs (upper panels). The left panels show an example of the percentiles [30 70]% (left and right pointing triangles, respectively). Considering the percentiles that are further away from the median, here illustrated for percentiles [10 90]% (squares and circles, respectively), it can be seen that the values of the ITDs (upper right panel) and ILDs (lower right panel) increase, but their dynamic range, i.e. the difference between PSW #1 and PSW #5, is similar. For the following analysis, the [30 70]% percentiles were chosen to obtain a higher outlier rejection.

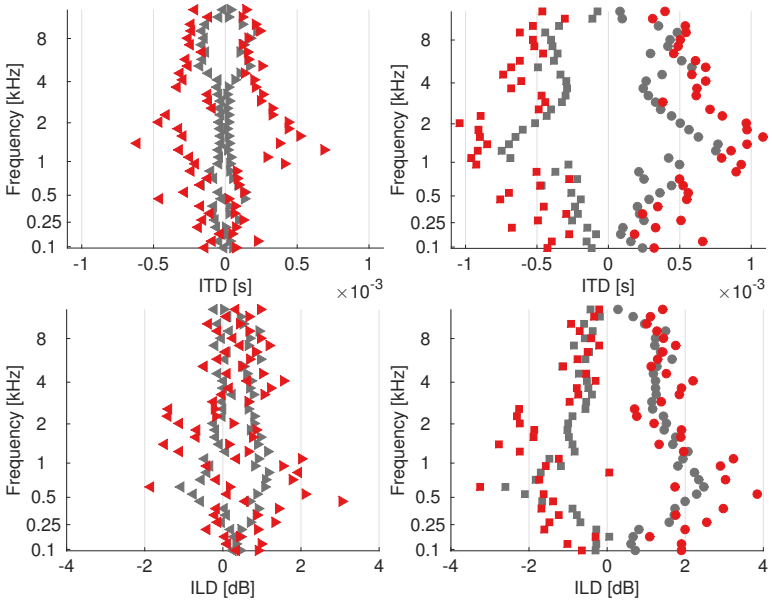


Figure 7.7: ITD-percentiles (upper panels) and ILD-percentiles (bottom panels) as a function of frequency in case of the pink noise source in Exp. B. Shown are the [30 70]% percentiles (left and right pointing triangles, respectively) and the [10 90]% percentiles (squares and circles, respectively) for PSW #1 (gray) and PSW #5 (red).

7.4.2 Overall pattern of results

Table 7.1: Model performances in terms of correlation coefficient r^2, r, RMS error and the Akaike information criterion (AIC).

Model	r^2	r	RMS error [°]	AIC (dof = 13; only Exps. A + B)
$IACC_{E3}$	0.94	0.97	8.9	159
IC_{E3}	0.88	0.95	12.2	128
ITD_{low}	0.94	0.96	9.5	142
DUPLEX	0.93	0.95	12.1	143
p_{RAA}	0.71	0.84	19.9	195
$IACC_{broad}$	0.84	0.92	15.3	-
IC_{broad}	0.86	0.93	13.5	-
ITD_{broad}	0.91	0.95	11.5	-
$DUPLEX_{short}$	0.92	0.96	11.7	-
ILD	0.73	0.85	20	-

The performance indicators of the models are displayed in Table 7.1 for the reference model $IACC_{E3}$, the functional models, IC_{E3}, ITD_{low} and DUPLEX and the complex model, p_{RAA} (all highlighted in gray). In general, all three types of the functional model and $IACC_{E3}$ provided a high correlation with the perceptual data, ranging from $r^2 = 0.88$ to $r^2 = 0.97$. This is due to the fact that PSW is the dominating factor compared to the source stimulus which is captured correctly by all models. The $IACC_{E3}$ model and the ITD_{low} model achieve the highest correlation with $r^2 = 0.94$ ($r = 0.97$), which corresponds to the findings regarding $IACC_{E3}$ in Zotter and Frank (2013). The ITD_{low} model reveals, however, a slightly larger RMS error of 9.4° compared to 8.7° for the $IACC_{E3}$ model. Since both, $IACC_{E3}$ and the ITD_{low} model, are derived from the IACC, their similar performance is plausible. The DUPLEX model has a slightly lower correlation ($r^2 = 0.93$) than the ITD_{low} model. Therefore, adding ILDs in the analysis does not provide a further benefit. Considering the model denoted by IC_{E3}, the performance decreased to $r^2 = 0.88$. This suggests that a short-term analysis of the IC (including the IHC and ATH model stages) and a higher frequency resolution (16 gammatone filters as opposed to 3 octave-wide filters in the $IACC_{E3}$) are not required to account for the perceptual data. Compared to the functional models, the complex model, p_{RAA}, achieved a lower correlation of $r^2 = 0.71$. This might indicate that the additional stages in this model are not required for the ASW analysis. Alternatively, the two components in Eq. 7.1, the monaural loudness (L_{LOW}) and the ITD statistics (σ_{ITD}), could be underestimated by the

model.
The results from the statistical analysis of the presented models, considering the AIC, is also listed in Table 7.1. The IC_{E3} model performed best (AIC = 128), the ITD_{low} and the DUPLEX provided similar performance (AIC = 142 and 143, respectively), the $IACC_{E3}$ (AIC = 159) model performed less well and the p_{RAA} model provided the lowest performance (AIC = 195). In the post-hoc analysis with Bonferroni correction (correction factor of 10), only a significant difference could be found for the p_{RAA} compared to the other models ($p_{posthoc} = 0.01$) that did not reveal significant differences ($p_{posthoc} = 1$) to each other.
The individual model outputs regarding the different test conditions are presented in the following. An analysis of some modified model versions (highlighted with white background in the Table 7.1) will be presented further below (section 7.5).

7.4.3 Influence of the source signal

Figure 7.8 shows the outputs of the models $IACC_{E3}$, IC_{E3}, ITD_{low}, DUPLEX and p_{RAA} for Exp. A (left panels), Exp. B (middle panels) and Exp. C (right panels). Note that the $IACC_{E3}$ and IC_{E3} models are inversely proportional to ASW and therefore shown as $1-IACC_{E3}$ and $1-IC_{E3}$, respectively. The $IACC_{E3}$, IC_{E3} and the p_{RAA} model produce a single output value and are therefore shown as symmetric functions. The ASW data (first row) are replotted from Fig. 7.5 but without standard deviations to facilitate comparisons to the model results. It can be seen that all models are able to predict the general trends in the data, i.e. that the perceived ASW increases with PSW. Differences occur with respect to the slopes of the predicted boundaries of the ASW and between source signals in all three experiments. The $IACC_{E3}$ (second row) correctly captures the dynamic range of the ASW data in Exps. A and B, i.e. the difference between the smallest and the largest ASW. This results in the high model performance ($r^2 = 0.94$). However, the model prediction does not account for all observed perceptual effects. For example, in contrast to the data in Exp. C, the model saturates for PSW values larger than PSW #10 (corresponding to ±56.25°). Further, this model only reveals minor differences between the source signals. In comparison to the $IACC_{E3}$, the IC_{E3} predictor (third row) shows a reduced sensitivity as a function of the PSW in all three experiments, i.e. a more shallow slope of the boundaries, which explains the reduced model performance. Also the IC_{E3} saturates for PSW

values larger than PSW #10 in Exp. C. The saturation of both models, the $IACC_{E3}$ and the IC_{E3}, indicates that either other cues cause the further increase of ASW in the data or the considered back-end is not sensitive enough. In contrast to the $IACC_{E3}$, the IC_{E3} partially captures source signal differences in Exp. A, e.g. larger ASWs for low frequencies (0.25 kHz; blue circles) compared to high frequencies (8 kHz; green diamonds), but contradicts the data for the white noise source (black rectangles). This indicates that a short-term analysis of the IC allows the prediction for narrowband signals, as in the case of the 2 octave-wide BP filtered signal at 0.25 kHz, but not for broadband signals, as in the case of the white noise. This is further supported by the fact that no differences are captured for the source signals used in Exps. B and C which also comprise a wider bandwidth. The output of the ITD_{low} model (fourth row) shows a dynamic range similar to that in the data for all three experiments. Specifically, in Exp. C this model does not show a saturation for larger PSW values, as opposed to the $IACC_{E3}$ and IC_{E3} models. Thus, the ITDs might be the dominating cues for ASW at lateral source positions. However, the ITD_{low} prediction is also more asymmetric, especially in the case of Exp. A and Exp. B, which results in the increased RMS error compared to the $IACC_{E3}$. This is due to the fact that the boundaries are estimated separately by the corresponding percentiles, such that a potentially asymmetric HATS positioning becomes more crucial. Source signal differences are not captured by the ITD_{low} model. The addition of ILDs in the DUPLEX model (fifth row) provides a similar output to the ITD_{low} model. Therefore, there are no significant contributions of the ILDs to ASW as a function of the PSW and the source signal.

The output of the complex model, p_{RAA} (bottom row), shows a smaller dynamic range in Exp. A and Exp. B than the data and the functional models, indicating a reduced sensitivity to PSW. For the HP filtered noise source in Exp. A, even a non-monotonic increase of ASW with PSW is seen. In Exp. C, the model reveals a larger sensitivity at low PSW values, but for larger PSW values it also provides a small increase of ASW with PSW. Thus, it does not resemble the output of the ITD_{low} model in the three experiments, even though both models are based on an ITD analysis. In contrast to the ITD_{low} model, the p_{RAA} computes the standard deviation instead of the percentiles of the ITDs which provides a smaller dynamic range of ITD fluctuations. Further, the p_{RAA} model comprises additional modelling stages, such as the segregation in a direct and a reverberant

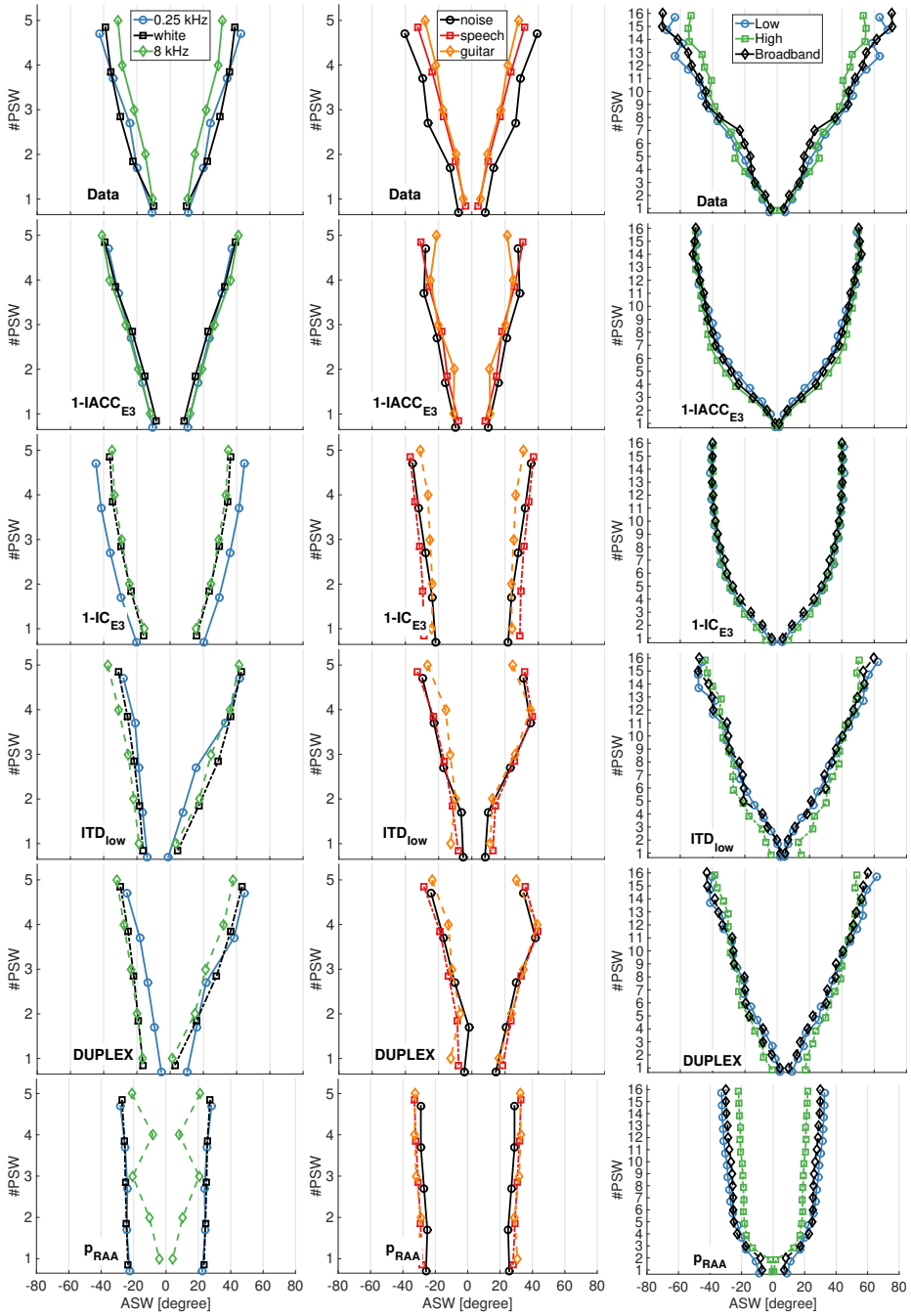


Figure 7.8: Perceptual data (top row, replotted from Fig. 7.5 w/o std deviations) and modelling results of ASW for Exp. A (left panels), Exp. B (middle panels) and Exp. C (right panels) in degrees. From top to bottom: Perceptual data, $1-\mathrm{IACC}_{\mathrm{E3}}$, $1-\mathrm{IC}_{\mathrm{E3}}$, $\mathrm{ITD}_{\mathrm{low}}$, DUPLEX and $\mathrm{p}_{\mathrm{RAA}}$. ASW is shown as a function of the physical source width (PSW), denoted by PSW #1 (narrow) to #5 (wide; Exps. A and B) or #16 (wide, Exp. C). The different symbols and line styles represent the different source signals.

stream and non-linear components which might cause further discrepancies. However, the p_{RAA} model captures the frequency-dependency better than the functional models. At least in Exp. A and Exp. C, the high-frequency sources are assigned correctly with a lower ASW than the low-frequency and broadband sources.

7.4.4 Impact of the sound pressure level

The complex model, p_{RAA}, comprising non-linear components, was tested in terms of variations of the SPL in Exp. A. The functional models were not considered here since they do not contain level sensitive components. Figure 7.9 shows the output of the p_{RAA} model (left panel) together with the perceptual data (right panel; replotted from Fig. 7.6 w/o std deviation) for Exp. A for all levels in the case of the white noise source. The output of the model only slightly increases with the level. The nonlinear components of the model allow for the right trends in the ASW data, but with a decreased sensitivity. Thus, both PSW and the effect of level are underestimated by the model. This indicates that both components in Eq. 7.1, the ITD statistics, σ_{ITD}, and the monaural sound pressure level component, L_{LOW}, require a higher sensitivity factor to match the dynamic range of the ASW data.

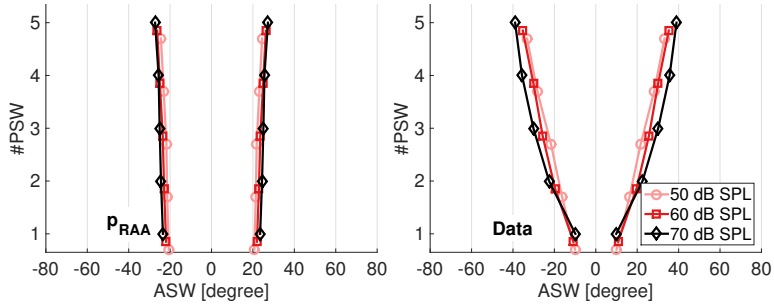


Figure 7.9: Output of the p_{RAA} model (left) and perceptual data (right; replotted from Fig. 7.6 w/o std deviation) for Exp. A in case of the white noise source for all SPLs. ASW is shown in degrees as a function of the physical source width (PSW). The different symbols and colors represent the different SPL conditions.

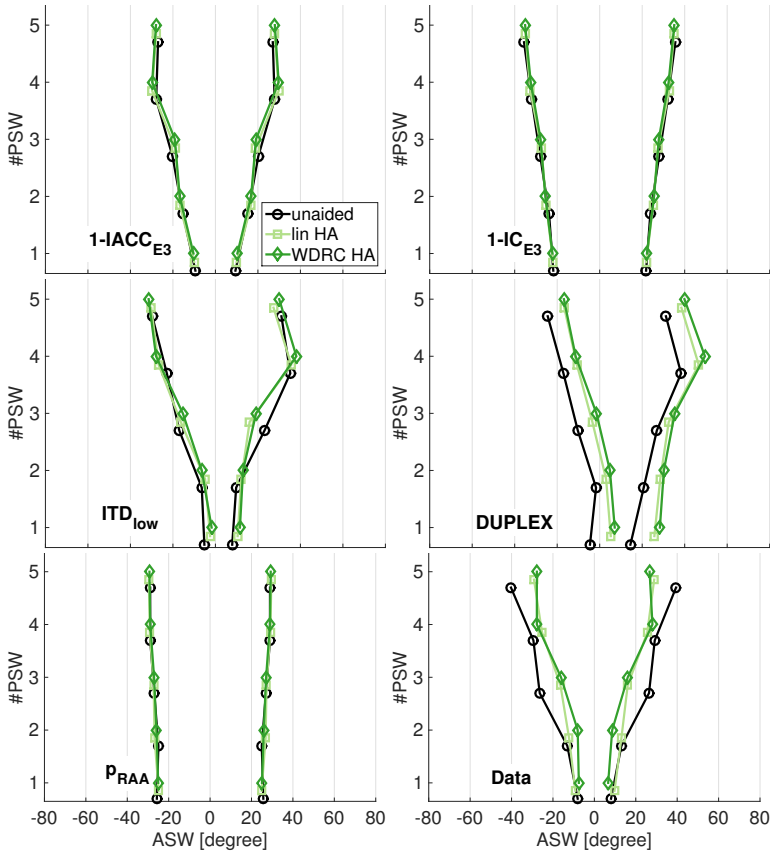


Figure 7.10: Perceptual data (bottom right panel, replotted from Fig. 7.6 w/o std deviation) and modelling results (remaining panels) of ASW for the different HA conditions in Exp. B in the case of the noise signal. From top left to bottom right panel: $1-\mathrm{IACC}_{\mathrm{E3}}$, $1-\mathrm{IC}_{\mathrm{E3}}$, $\mathrm{ITD}_{\mathrm{low}}$, DUPLEX, $\mathrm{P}_{\mathrm{RAA}}$ and perceptual data. ASW is shown in degree as a function of the physical source width (PSW), denoted by PSW #1 (narrow) to #5 (wide). The different symbols and colors represent the different HA conditions.

7.4.5 Effects of hearing-aid processing on ASW

All models were further tested for the HA conditions in Exp. B. Figure 7.10 shows the outputs of the models $\mathrm{IACC}_{\mathrm{E3}}$, $\mathrm{IC}_{\mathrm{E3}}$, $\mathrm{ITD}_{\mathrm{low}}$, DUPLEX, and $\mathrm{p}_{\mathrm{RAA}}$ for the different HA conditions in Exp. B in the case of the noise source. The perceptual data (bottom right panel, replotted from Fig. 7.6 w/o std deviation) show a smaller ASW for the unaided compared to the two aided conditions, which cannot be captured by any of the models. Only the DUPLEX model differentiates between the unaided and aided conditions (besides some minor sensitivity in the $\mathrm{ITD}_{\mathrm{low}}$ model) as represented by a shift of both estimation boundaries to the right. This results in a correct prediction for the left but not the right boundary.

Even though the prediction of the DUPLEX model is not fully correct, the results indicate that the ILDs become important in the analysis of ASW perception in aided conditions. This is in line with the findings presented in Ch. 6.

7.4.6 Impact of the signal duration

The various signal durations in Exp. C were used to test the performance of the functional models especially for short durational signals. These models do not have a dedicated building block to analyze the signal duration with an appropriate integration time constant.

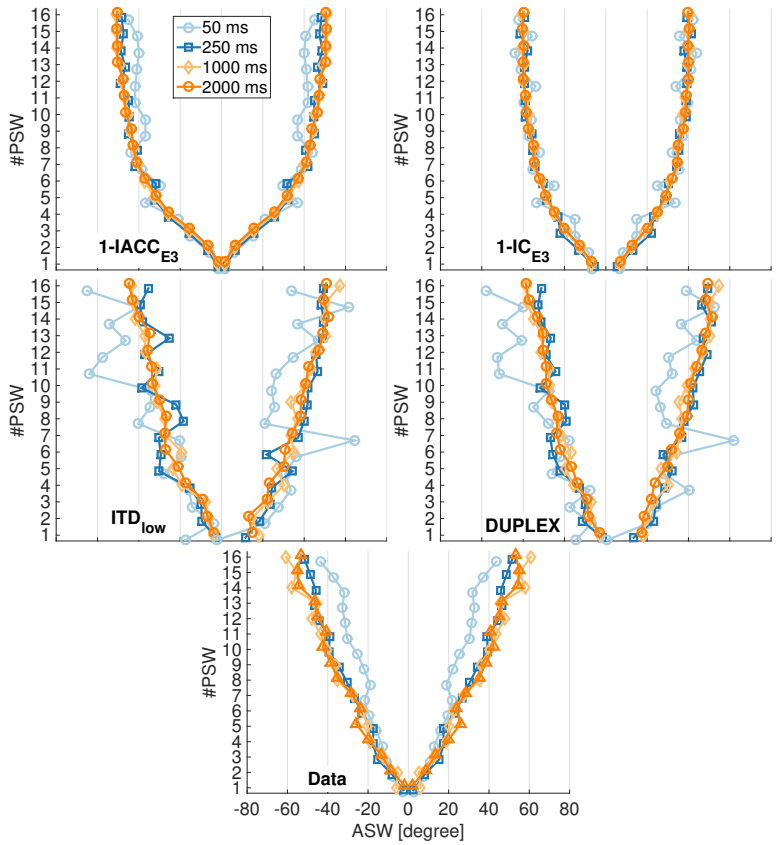


Figure 7.11: Perceptual data (bottom) and modelling results (remaining panels) of ASW for the duration conditions in Exp. C in the case of the HP filtered signal. From top left to bottom: $1-\mathrm{IACC}_{E3}$, $1-\mathrm{IC}_{E3}$, ITD_{low}, DUPLEX and perceptual data. ASW is shown in degrees as a function of the physical source width (PSW), denoted by PSW #1 (narrow) to #16 (wide). The different symbols and colors represent the different signal durations.

However, the ASW prediction is based on time averages which depend on the

signal duration. The p_{RAA} model was excluded for this condition since it requires signals with a duration of at least 10 s. Figure 7.11 shows the results of the functional models and the reference model ($IACC_{E3}$) for the different signal durations in Exp. C in the case of the HP filtered source. The perceptual data (bottom panel, replotted from Fig. 7.6 w/o std deviation) show a smaller ASW for the 50-ms long signal compared to the other durations. This is only reflected by the $IACC_{E3}$ model. For the ITD_{low} and DUPLEX models, the prediction is inaccurate in the case of the 50-ms long signal because only 4 data frames of 20 ms each are used in the analysis. Thus, an evaluation based on percentiles is not meaningful here. In contrast, the low number of analysis frames does not affect the IC_{E3} model which, however, remains insensitive to the signal duration as opposed to the $IACC_{E3}$ model. This indicates that for an accurate ASW prediction of short signals (< 250 ms) based on the IC, the analysis window duration needs to match the signal duration.

7.5 General discussion

7.5.1 Broadband analysis in the functional model

In Table 7.1, the performances of the $IACC_{E3}$, IC_{E3} and ITD_{low} models are also shown for the case when including the entire bandwidth for the analysis (denoted with the subscript 'broad'). The corresponding performance was found to be decreased compared to their low frequency estimates, resulting in $r^2 = 0.84$ for the $IACC_{broad}$ (compared to $r^2 = 0.94$ for $IACC_{E3}$), $r^2 = 0.86$ for the IC_{broad} (compared to $r^2 = 0.88$ for IC_{E3}) and $r^2 = 0.91$ for the ITD_{broad} (compared to $r^2 = 0.94$ for ITD_{low}). These results suggest that high-frequency components in IACC-based measures do not provide useful information for ASW. It can be noted though that in the case of a broadband analysis, the short-term analysis of IC (IC_{broad} with $r^2 = 0.86$) becomes favorable compared to its long-term analysis ($IACC_{broad}$ with $r^2 = 0.84$).

7.5.2 The contribution of ILDs in the functional model

It was shown that including ILDs in the model predictions did not improve the overall model performance compared to the ITD_{low} model (besides a sensitivity in the HA conditions). In Figure 7.7, an analysis of the [30 70]% percentiles (marked by the opposite pointing triangles) for the ITDs (left panel) and ILDs (third panel) across frequency is presented for PSW #1 and PSW #5 in the case of the noise source in Exp. B. While the percentiles of the ITDs increase substantially (roughly by 400 ms) from PSW #1 to #5, the percentiles of the ILDs only increase by less than 1 dB for frequencies below 2 kHz and, thereby, exploit a small dynamic range of ILD fluctuations.

Despite the small dynamic range of ILDs, this information is sufficient for a pure ILD model resulting in a correlation of $r^2 = 0.73$ (see Table 7.1). The consideration of ILDs is interesting because ILD fluctuations alone can generate the perception of an increased ASW. Considering a dichotic headphone presentation, using a sinusoidal or broadband signal which is sinusoidally modulated in antiphase between the left and right channel will generate ILDs corresponding to the modulation frequency (binaural modulation). The ASW of the perceived internalized stationary sound image is increased when the ILDs exceed the threshold of binaural sluggishness around 80 Hz (Blauert, 1972; Thompson, 2009). Interestingly, a larger ASW can be perceived in this example, even though the ear signals are correlated (IC=1), which should elicit a compact source (small ASW), and no ITDs are present. This example highlights that ILDs should be considered in ASW models.

The analysis window duration and shape also play a role in the analysis of the ILDs. With a short analysis window, the resulting ILDs might be larger as compared to longer windows. The reason is that a shorter average time of the signal's energy allows to capture instantaneous level differences between the ear signals. This might improve the dynamic range of the ILDs in the percentiles. This was tested for a 1 ms window duration with a 0.5 ms overlap for the ILD analysis while maintaining the 20 ms window and the 0.5 ms overlap for the ITD analysis. While for this approach ($DUPLEX_{short}$), the dynamic range of ILDs was doubled to 2 dB, the correlation further decreased to $r^2 = 0.92$ compared to the original DUPLEX model with $r^2 = 0.93$ (see Table 7.1). Even though the contribution of ILDs to ASW cannot be supported in the current study with the considered stationary stimuli (even for the speech and music signals, the variations across

PSW were small), ILDs might become more relevant for the ASW estimation of sound sources in real rooms. At least in the HA conditions of Exp. B, the DUPLEX model revealed differences between the unaided and aided condition by including ILDs whereas the other models remained insensitive.

7.5.3 Perspectives

The suggested models should be further investigated in real rooms and in virtual sound environments (VSE) where the binaural statistics might be different from the one observed here for a stereo loudspeaker setup. Furthermore, the ASW model might be helpful in the evaluation and validation process of VSEs in auditory research. It could be investigated whether an acoustic scene in a real acoustical space generates the same ASW prediction as the same scene reproduced in a VSE. Such analysis should also consider the individual binaural cue statistics across frequency bands, as they might be helpful in identifying deviances of the reproduced from the real acoustic scene.
The model estimated ASW based on binaural cue statistics under the assumption that the time course of the binaural cues is irrelevant. Hence, time-sensitive effects, such as summing localization (delay < 1 ms) and the precedence effect (1 ms < delay < 80 ms) (Blauert and Braasch, 2005) as well as signal onsets (van Dorp Schuitman et al., 2013), were not considered. These effects could be taken into account by differentiating perceptual streams as it was realized in the complex non-linear model. Furthermore, David et al. (1959) showed that a mismatch between the localization cues ITDs and ILDs also causes a broadened sound image. It might be possible that an increased ASW in rooms is enhanced by reflections that lead to "antagonistic" binaural cues at the listeners ears. This effect could be considered by a model stage that compares ITDs and ILDs on a time frame basis. Finally, a blurred auditory image with increased ASW is perceived when the fluctuations of ITDs and ILDs are fast enough due to binaural sluggishness (Siveke et al., 2007; see Ch. 2 and 8). This threshold is only 3 Hz for ITDs (Blauert, 1997) and about 80 Hz for ILDs (Stellmack et al., 2005). A model could consider only contributing ITDs and ILDs, for example, by implementing a second filterbank that estimates the frequency of the fluctuations in each peripheral frequency channel.

7.6 Summary and Conclusion

In this study, three experiments were considered where the ASW was measured as a function of the PSW. The stimuli were analyzed by a binaural functional model and a complex model to predict ASW. The functional model was tested with different back-ends. One back-end combined ITDs and ILDs according to the duplex theory (DUPLEX) and was compared to the back-ends IC_{E3}, and ITD_{low}. The $IACC_{E3}$ model that represents a standard predictor for ASW perception in room acoustics dealt as a reference model. The models based on the interaural cross-correlation function (either extracting IC or ITD) produced similar results for the estimation of ASW. The best performance was obtained by the $IACC_{E3}$ and ITD_{low} models. Apparently, the signals were stationary enough such that a long-term analysis with the $IACC_{E3}$ was sufficient. In contrast, the short-term performance with the IC_{E3} model showed a poorer performance. The best model performance was obtained when considering previously suggested frequency regions for the analysis with cross-correlation based models. For the $IACC_{E3}$ and IC_{E3} models, this comprised averaging across three octave bands at 0.5, 1 and 2 kHz and for the ITD_{low} model, considering frequencies only below 2 kHz. Adding higher frequency components deteriorated the ASW estimation in all models. The DUPLEX model that also included ILDs could not provide any further benefit in the ASW estimation, possibly due to the stationary character of the chosen stimuli. A complex non-linear model showed a poorer performance than the ITD_{low} model, reflected by a reduced sensitivity to changes in PSW. Apart from a linear combination of low-frequency ITDs and monaural signal power, the model also used a segregation stage to focus on the direct sound, which could be the reason for the poorer performance. However, the model could account for frequency-dependent effects and moderate sound pressure level effects observed in the data. The non-linear processing as applied in the complex model, could be further investigated within the framework of the here suggested functional model by replacing the gammtone filterbank with the dual-resonance non-linear (DRNL) filter bank as implemented in Jepsen et al. (2008).

Acknowledgement

This work was supported by EU FET grant TWO!EARS (ICT-618075) and by the Centre for Applied Hearing Research which is a research consortium with Oticon, Widex and GNResound. The binaural functional models were implemented together with Manuel Hahmann and are based on the auditory front-end (AFE) of the TWO!EARS consortium (May et al., 2015; TWO!EARS, 2013-2016). The author would also like to thank Julian Grosse for making his data (Exp. C), presented in Grosse et al., 2015, available for this study. Further thanks to Jesper van Dorp Schuitman who supplied us with the complex model of the Room Acoustic Analyzer (RAA) toolbox.

8

Overall discussion

8.1 Summary of main results

In this thesis, measurements were conducted in stereo loudspeaker arrangements to characterize ASW perception. The presented ASW measurements with stereo loudspeaker setups confirmed the findings from the literature that were based on headphone experiments and data obtained in concert halls. Therefore, some of the "gaps" in ASW perception between headphone-based, loudspeaker-based and room acoustical studies have been closed. It was found that ASW increases with decreasing the stimulus IC for broadband and bandpass filtered Gaussian noise signals consistent with related findings in the literature. The main difference of stereo loudspeaker presentations compared to headphone presentations is the presence of cross-talk and head- and torso reflections from the listener (neglecting room reflections). For a constant stimulus IC, it was shown that the cross-talk diminished the assigned ASW compared to a condition without cross-talk, independent of the stimulus center frequency (Ch. 3). In a next step, ASW measurements were performed where IC was measured at the listener's ears. Absolute measurements of ASW on a degree-scale made it possible to study frequency- and level-dependent effects (Ch. 4). For a constant stimulus IC at the listener's ears, low frequencies were assigned with a larger ASW than high frequency stimuli. In addition, ASW and also the apparent source height (ASH; indicating a vertical expansion of the sound source) were proportional to the stimulus level, such that an increase in level led to an increase of both, ASW and ASH, i.e. the apparent source area. These findings are consistent with literature studies with headphones (direct control of the ear signals; Blauert and Lindemann, 1986a) and in concert halls (also considering room reflections; Okano et al., 1998). In Blauert et al. (1986c), it was shown that the frequency region around 1 kHz is particularly sensitive for the per-

ception of spaciousness. This was confirmed in the measurements where the ITD statistics, generated by a stereo listening setup, were artificially modified (Ch. 5). Also here, the ITDs of about 1 kHz contributed mostly to a change in ASW perception, whereas frequencies above 2 kHz did not contribute. It was further shown that the $IACC_{E3}$ predicted these changes best whereas the frequency-independent version, the IACC, remained insensitive to the applied ITD modifications. Since similar results were obtained for the different presentation methods (headphones, loudspeakers and in real rooms), it can be concluded that ASW perception seems consistent across the different presentation methods due to similar cues that elicit the perception.

In a follow-up study (Ch. 6), differences in ASW perception between NH and HI listeners were found. In contrast to the NH listeners, the HI listeners showed a clearly reduced dynamic range in ASW perception. Sounds that were assigned with a large ASW by the NH listeners were rated smaller by the HI listeners. In contrast, sounds that were assigned with a small ASW by the NH listeners were rated with a larger ASW by the HI listeners. Furthermore, the influence of a BTE-HA was tested on ASW perception using two programs, linear and WDRC processing. The listeners who could distinguish small from large ASWs in the unaided condition perceived a change in ASW by listening through the HA. The WDRC had generally no influence on the ASW perception but the HA's microphone position caused, on average, a smaller ASW perception. This was due to the modification of the ILD cues compared to the unaided listening condition demonstrating the importance of maintaining the listener's natural pinna cues for a natural spatial sound reproduction. The measurement setup, consisting of 3 stereo loudspeaker pairs, showed accurate measurement results with a low within-listener variability for the NH as well as the HI listeners. Such setup might be interesting for clinical studies of ASW perception since it requires a smaller amount of loudspeakers and also less space than a virtual sound environment.

In a functional binaural model of ASW perception, the contributions of the binaural cues, ITDs, ILDs and IC, were further investigated (Ch. 7). It was shown that cross-correlation based metrics, i.e. IC and ITDs, most successfully predicted the ASW data, whereby a long-term analysis using the $IACC_{E3}$ was sufficient. Frequencies up to 2 kHz were most important for the ASW analysis whereas including higher frequencies deteriorated the ASW prediction. This was in line with the findings in Ch. 5 where a broadband analysis of the IACC did

not successfully capture the perceptual differences, in contrast to the $IACC_{E3}$. Including high-frequency ILDs in addition to low-frequency ITDs according to the duplex theory, could not provide a benefit in the prediction compared to solely using ITDs. However, the duplex model was the only sensitive approach that could discriminate unaided from aided conditions in the previous experiment (Ch. 6). This indicates that including ILDs in the prediction of spatial perception using personal hearing devices, such as HAs, can be important. In contrast, a complex nonlinear model showed a reduced correlation with the perceptual data compared to the functional model and the $IACC_{E3}$. The model prediction was based on a linear combination of the monaural level and ITD fluctuations, but only showed a small dynamic range regarding the two components.

8.2 Interpretation of apparent source width perception and its relation to other spatial sensations

ASW perception is elicited by fluctuations of ITDs and ILDs at the ears of the listener, including spectral modifications. The source of these fluctuations can be caused by modifications of the original signal due to a transmission line or by the source itself. The transmission line might be a room or appropriate signal processing in loudspeaker or headphone presentations. Large sound sources also elicit ASW. For example, a large ASW is perceived when a choir is singing unisono, i.e. the same melody is sung by all members of the choir. Small synchronization and amplitude differences between the otherwise "coherent" signals of the singers lead to binaural cue fluctuations and, hence, to an increased ASW (the same principle is exploited in a chorus effect where the original signal is superimposed with copies of the original signal using several delay lines). Another example of a large sound source emitting coherent signals might be a string section in an orchestra or even a grand piano. In other cases of large sound sources, the emitted signals from the smaller composing units are not coherent due to noise-like sounds as, for example, a tree with rustling leaves, a waterfall or a motor-highway, all eliciting a large ASW. This highlights that ASW is elicited by binaural cue fluctuations irrespective of its source.

In this thesis, ASW perception was studied while other spatial cues were attempted to remain constant. However, in real-life listening scenarios, many perceptual effects happen simultaneously and might be correlated with each other. An important step would be to link ASW perception to related spatial hearing phenomena. Since ASW perception is based on the fluctuation of the important localization cues ITDs and ILDs, the impact of ASW on the perceived location of a sound source could be investigated. In a room, the enlargement of ASW (influenced by early reflections) might be accompanied by an increased distance perception (related to the direct-to-reverberant ratio). Also, ASW perception has been considered to be an important component of externalization perception (Hartmann and Wittenberg, 1996). It was reported that changes in externalization can also produce changes in ASW perception (Catic et al., 2013). Future studies could identify these correlations and link them to the related physical cues. This would provide further insights into the individual building blocks of spatial hearing, especially in the larger context of communication in a cocktail-party scenario. Further, it would help to improve the design of VSEs or signal processing strategies in hearing devices for HI listeners.

Localizability, localization blur and auditory spatial resolution are terms often used in the literature to describe uncertainties in human sound localization. Here, an attempt is made to link ASW perception to these attributes. Figure 8.1 shows a diagram on sound source localization (left side), discrimination (middle) and segregation (right side) and the potential impact of ASW perception on these three features.

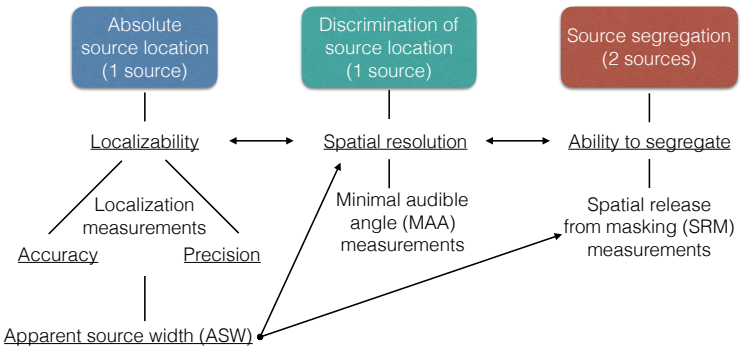


Figure 8.1: Scheme on localization, discrimination and segregation of sound sources. The apparent source width (ASW) might affect the spatial resolution and the listener's ability to segregate concurrent sound sources from each other.

In auditory localization tasks, the listener is asked to estimate the location of a single sound source (left part of Fig. 8.1). The *localizability* of the sound source (Hartmann, 1983, Lindau et al., 2014) is determined by the localization *accuracy*, also known as *unsigned bias*, and the localization *precision* of the listener. While the localization accuracy refers to the difference between the listener's (mean) estimation and the physical location of the sound source, the localization precision refers to the variance in the listener's responses (Whitmer et al., 2014). The listener's localization accuracy and precision decrease for lateral sound sources compared to frontal sound sources. This is due to increasing temporal ITD and ILD fluctuations for lateral sound sources caused by stronger head and torso reflections (Goupell and Hartmann, 2007).

The uncertainties in determining the location of a sound source, especially for lateral directions, might be caused by temporal limitations of the auditory system in following the fluctuations of the binaural cues. This effect is called *binaural sluggishness* (Siveke et al., 2007). The perception of ASW might as well be related to the binaural sluggishness of the auditory system. Room reflections, for instance, create binaural fluctuations which vary fast enough to be affected by the binaural sluggishness. Since early reflections arrive in the time window of the precedence effect (< 80 ms), the arriving reflections are perceptually bound to the direct sound (direct stream). Hence, the listener perceives a fused sound image which is stable in its location instead of perceiving multiple sound sources associated with each reflection. However, the sound source is not perceived as a point source. It is perceived as locally blurred with a spatial expansion in the horizontal dimension, which is the perception of ASW. Therefore, it makes sense to predict ASW perception in auditory models with the variability of the binaural cues using the standard deviation or percentiles (see Ch. 7).

In addition to determining the sound source location, the discrimination between sound source locations is important to consider (middle part of Fig. 8.1). The displacement that is needed to recognize a change of the sound source's location is limited by the *spatial resolution* of the auditory system. Commonly, the spatial resolution is reflected by measurements of the minimal audible angle (MAA), also referred to as *localization blur* by Blauert (1997). The MAA is smallest for frontal sound sources and increases for lateral sources. Therefore, the limitations of the auditory system in discriminating the sound source's location might as well be related to the increasing fluctuations of the binaural cues towards lateral sources. Therefore, both localizability and spatial resolution might

be limited by the same principles of binaural sluggishness. However, while the perception of ASW can be considered as a resulting effect of localization uncertainties, it remains unclear whether ASW might be considered as a cause for the limited spatial resolution. In such case, ASW measurements could reflect the limited spatial resolution.

The considerations so far have been limited to a single sound source. The situation becomes more complex when adding a second concurrent sound source (right part of Fig. 8.1). In a listening situation where one source is target speech and the second source is a masker, speech intelligibility improves when the two sources are spatially separated. This effect is known as *spatial release from masking (SRM)*. Exploiting binaural mechanisms, the auditory system uses time glimpses across frequency bands to localize the sound sources which helps segregate them and hence makes the speech intelligible (e.g. Faller and Merimaa, 2004, Cooke, 2006). Also monaural cues might help as the ear with the better signal-to-noise ratio might provide a benefit (*better-ear listening*). The SRM is typically measured with the speech recognition threshold (SRT) for several target and masker configurations.

ASW perception might influence the segregation abilities of sound sources, i.e. the SRM: Large appearing sources might be more difficult to segregate than point-like sources. Culling et al. (2001) mentioned that a change of a sound source's ASW might be a cue for listeners in listening tasks of binaural masking level differences (BMLD) or even in binaural intelligibility level differences (BILD). However, this has not been further explored yet and should be tested in future studies. ASW could represent an important spatial sensation in relation to sound source segregation.

8.3 Perspectives

8.3.1 Effects of apparent source width perception on source segregation abilities

It was hypothesized in the previous section and throughout this thesis (Ch. 1, Ch. 2 and Ch. 6) that ASW perception might influence or reflect the ability of a listener to segregate sound sources in multi-talker environments. Sound

sources with a large ASW, as they occur in reverberant environments, might be more difficult to segregate than sound sources with small ASW, as occurring in acoustically dry environments. This might affect, e.g., the intelligibility of a target talker in a noisy environment. To test such hypothesis, SRTs could be measured in a spatial-release-from-masking experiment by varying the ASW of both target and masker. The experiment could be conducted in a VSE employing room simulations whereby the ASW of the sound sources could be controlled by using only the direct sound and varying the amount of early reflections. However, the energy of the early reflections should remain constant since it can contribute to an improvement of the SRTs (Arweiler and Buchholz, 2011). The late reverberation part should be excluded since it can elicit additional spatial sensations, such as reverberance and listeners envelopment, and can be detrimental on speech intelligibility (Arweiler and Buchholz, 2011). The obtained SRT data could be correlated with individual ASW measurements. In the case that ASW does affect sound source segregation, ASW measurements as presented in this thesis could aid the profiling process of individual HI listeners in clinical studies.

8.3.2 Further evaluation of the functional ASW model

In Chapter 7, a functional ASW model was suggested and validated with perceptual data primarily obtained in stereo-loudspeaker setups. The contribution of ILDs, combining ITDs and ILDs according to the duplex theory, could not be proven in such setup due to the small dynamic range of the ILD fluctuations. Therefore, the model should be also tested in real acoustical spaces or in simulated acoustical environments using VSEs. Both cases probably exploit a larger dynamic range in ILD fluctuations, such that the ILD analysis component in the model could be justified.

8.3.3 Evaluation of virtual sound environments

Virtual sound environments have recently become more extensively used for auditory research. However, the hearing research community seeks for robust measures that can quantify the reproduction quality of such systems. In Cu-

bick and Dau (2016) the performance of a VSE was perceptually evaluated by comparing speech intelligibility measurements in the simulated environment versus the real room. An objective comparison was carried out regarding room acoustic parameters, including e.g. the IACC coefficient. The functional ASW model developed in this thesis could provide further insights in the perceptual validation process. This includes investigations of the individual binaural cue statistics (e.g. in terms of histograms) and the sensitivity of listeners to differences in these statistics between real and simulated environments. The perceptual differences could be addressed by measurements of ASW perception, localization performance as well as distance and externalization perception.

8.3.4 Representation of apparent source width in the auditory system

The contribution of binaural cues on ASW were studied in this work by employing a binaural functional model. One could further investigate the auditory features involved in ASW perception using a synthesis approach. Specifically, the contribution of individual binaural cue statistics on ASW perception would be interesting to consider in this context. McDermott et al. (2011) developed a monaural framework to evaluate the contributions of individual statistics on the perception of sound textures, comprising sounds such as bee humming, waterfalls and fire cracking. Such framework could be extended to a binaural framework for the investigation of ASW perception. The synthesized sounds could be perceptually validated by comparing the sound source's location and ASW of the synthesized and the original sounds. For the testing, binaural recordings of stationary sound sources, including sound textures, in anechoic and reverberant spaces would be suited.

8.3.5 Binaural cue coding and binaural cue synthesis

The artificial generation of spatial sensations like ASW in sound reproduction has been used for artistic and experimental purposes, using e.g. source widening algorithms (Zotter and Frank, 2013) as well as chorus effects. The underlying principle is based on the artificial generation of binaural cue fluctuations across the audible frequency range. For example, Pessentheiner (2011) and Zotter and

Frank (2013) used digital filters that modulate the phase or amplitude spectrum to produce the binaural differences. However, the resulting binaural cues do not resemble those of real acoustic spaces, leading to perceptual differences between the real and reproduced acoustic space. To obtain a natural spatial perception in sound reproduction, an attempt could be made to imitate binaural fluctuations of real spaces in a binaural cue synthesis (BCS). In Baumgarte and Faller (2003) and Faller and Baumgarte (2003), the concept of binaural cue coding (BCC) was introduced. The audio image was spatialized by imposing (encoded) binaural information on a monaural audio signal. A similar principle could be employed for a natural spatial reproduction in a binaural cue synthesis. One option could be to use adaptive filters that encode the binaural information from binaural room impulse response (BRIR) recordings. Another way to produce the desired binaural cues would be to adapt digital filter designs (Pessentheiner, 2011, Zotter and Frank, 2013) to be frequency-band specific. This includes the synthesis from binaural histograms with a known probability distribution. In general, a BCS algorithm would provide a powerful research tool that allows to systematically test the perceptual effect of ITD and ILD fluctuations on ASW but also on the externalization of sound sources. It might further help to compensate for the distorted spatial perception in HI listeners using hearing devices. Individual HI listeners might have different sensitivities regarding ITDs and ILDs, such that the algorithm could differentiate the enhancement of these cues adapted to the individual sensitivity.

Bibliography

Allen, J. B. (1996). *DeRecruitment by multiband compression in hearing aids. Psychoacoustics, speech, and hearing aids*. World Scientific Press, Editor B. Kollmeier, Singapore, pp. 141–152.

Ando, Y. (2007). *Concert hall acoustics based on subjective preference theory.* pp. 351-386. The Springer Handbook of Acoustics (Springer Science + Business Media, New York).

Arweiler, I. and J. Buchholz (2011). "The influence of spectral characteristics of early reflections on speech intelligibility". In: *J. Acoust. Soc. Am.* 130.2, pp. 996–1005.

Baumgarte, F. and C. Faller (2003). "Binaural Cue Coding - Part I: Psychoacoustic Fundamentals and Design Principles". In: *IEEE Trans. Speech Audio Process.* 11.6, pp. 509–519.

Bentler, R. A. and C. V. Pavlovic (1989). "Transfer Functions and Correction Factors Used in Hearing Aid Evaluation and Research". In: *Ear Hear.* 10.1, pp. 58–63.

Bernstein, L. R., S. van de Par, and C. Trahiotis (1999). "The normalized interaural correlation: accounting for NoS_π thresholds obtained with Gaussian and "low-noise" masking noise". In: *J. Acoust. Soc. Am.* 106.2, pp. 870–876.

Blauert, J. (1972). "On the lag of lateralization caused by interaural time and intensity differences". In: 11.5, pp. 265–270.

Blauert, J. (1978). "Some aspects of three-dimensional hearing in rooms". In: 3.3, pp. 65–68.

Blauert, J. (1997). *Spatial Hearing: The psychophysics of human sound localization.* Revised edition. MIT Press.

Blauert, J. and J. Braasch (2005). "Acoustic Communication: The Precedence Effect". In: *Forum Acusticum.* Budapest, Hungary.

Blauert, J. and W. Lindemann (1986a). “Spatial mapping of intracranial auditory events for various degrees of interaural coherence”. In: *J. Acoust. Soc. Am.* 79.3, pp. 806–813.

Blauert, J. and W. Lindemann (1986b). “Auditory spaciousness: Some further psychoacoustic analyses.” In: *J. Acoust. Soc. Am.* 80.2, pp. 533–542.

Blauert, J., U. Möbius, and W. Lindemann (1986c). “Supplementary Psychoacoustical Results on Auditory Perception”. In: *Acustica* 59. Letter to the Editors, pp. 292–293.

Boyd, A. W., W. M. Whitmer, J. J. Soraghan, and M. Akeroyd (2012). “Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers”. In: *J. Acoust. Soc. Am.* 131.3, pp. 268–274.

Bradley, J. S. (2011). “Review of objective room acoustics measures and future needs”. In: *Appl. Acoust.* 72.10, pp. 713–720.

Breebaart, J., S. van de Par, and A. Kohlrausch (2001). “Binaural processing model based on contralateral inhibition. I. Model structure”. In: *J. Acoust. Soc. Am.* 110.2, pp. 1074–1088.

Bronkhorst, A. W. and R. Plomp (1992). “Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing”. In: *J. Acoust. Soc. Am.* 92.6, pp. 3132–3139.

Brüggen, M. (2001). “Sound coloration due to reflections and its auditory and instrumental compensation”. PhD thesis. Ruhr-Universität Bochum, Germany.

Buchholz J., M. (2013). “A real-time hearing-aid research platform (HARP): Realization, calibration, and evaluation”. In: *Acta Acust. United Acust.* 99, pp. 477–492.

Byrne, D., A. Parkinson, and P. Newall (1990). “Hearing-aid gain and frequency-response requirements for the severely profoundly hearing-impaired”. In: *Ear Hear.* 11.1, pp. 40–49.

CIPIC (2004). *Reference to the CIPIC HRTF database.*

Catic, J., S. Santurette, J. M Buchholz, F. Gran, and T. Dau (2013). “The effect of interaural-level-difference fluctuations on the externalization of sound.” In: *J. Acoust. Soc. Am.* 134.2, pp. 1232–1241.

Chernyak, R. I. and N. A. Dubrovsky (1968). “Pattern of noise images and binaural summation of loudness for different interaural correlation of noise”. In: *Int. Congress on Acoustics*. Tokyo, Japan, pp. 53–56.

Cooke, M. (2006). “A glimpsing model of speech perception in noise.” In: *J. Acoust. Soc. Am.* 119.3, pp. 1562–1573.

Cubick, J. and T. Dau (2016). “Validation of a Virtual Sound Environment System for Testing Hearing Aids”. In: *Acta Acust. united Ac.* 102.3, pp. 547–557.

Cubick, J., S. Santurette, S. Laugesen, and T. Dau (2014). “Influence of high-frequency audibility on the perceived distance of sounds”. In: *Forum Acusticum*. Krakow, Poland.

Culling, J. F., H. S. Colburn, and M. Spurchise (2001). “Interaural correlation sensitivity”. In: *J. Acoust. Soc. Am.* 110.2, pp. 1020–1029.

Dau, T., D. Püschel, and A. Kohlrausch (1996). “A quantitative model of the ’effective’ signal processing in the auditory system. I. Model structure.” In: *J. Acoust. Soc. Am.* 99.6, pp. 3615–3622.

David, E. E., N. Guttman, and van Bergeijk W. A. (1959). “Binaural Interaction of High-Frequency Complex Stimuli”. In: *J. Acoust. Soc. Am.* 31.6, pp. 774–782.

Dillon, H. (2001). *Hearing aids*. Boomerang Press.

Faller, C. and F. Baumgarte (2003). “Binaural Cue Coding - Part II: Schemes and Applications”. In: *IEEE Trans. Speech Audio Process.* 11.6, pp. 520–531.

Faller, C. and J. Merimaa (2004). “Source localization in complex listening situations: selection of binaural cues based on interaural coherence.” In: *J. Acoust. Soc. Am.* 116.5, pp. 3075–3089.

Favrot, S. E. and J. Buchholz (2010). “LoRA: A Loudspeaker-Based Room Auralization System”. In: *Acustica United With Acta Acustica* 96.2, pp. 364–375.

Frank, M. (2013). “Phantom Sources using Multiple Loudspeakers in the Horizontal Plane”. PhD thesis. University of Music and Performing Arts Graz, Austria.

Gabriel, K. J. and H. S. Colburn (1981). “Interaural Correlation Discrimination: I. Bandwidth and level dependence”. In: *J. Acoust. Soc. Am.* 69.5, pp. 1394–1401.

Glasberg, B. R. and B. C. J. Moore (1990). “Derivation of auditory filter shapes from notched-noise data”. In: *Hearing Research* 47.1–2, pp. 103–138.

Goupell, M. J. and W. M. Hartmann (2007). “Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models.” In: *J. Acoust. Soc. Am.* 122.2, pp. 1029–1045.

Griesinger, D. (1999). “Objective measures of spaciousness and envelopment”. In: *Audio Eng. Soc.* Helsinki, Finland, pp. 27–41.

Grosse, J., F. Hungar, S. Klockgether, and S. van de Par (2015). "Wahrgenommene Quellbreite einer Lautsprecheranordnung in Abhängigkeit der physikalischen Quellbreite". In: *German Acoust. Association (DAGA)*. Nürnberg, Germany, pp. 898–901.

Hartmann, W. M. (1983). "Localization of Sound in Rooms". In: *J. Acoust. Soc. Am.* 74.5, pp. 1380–1391.

Hartmann, W. M. and Y. J. Cho (2011). "Generating partially correlated noise–a comparison of methods". In: *J. Acoust. Soc. Am.* 130.1, pp. 292–301.

Hartmann, W. M. and A. Wittenberg (1996). "On the externalization of sound images". In: *J. Acoust. Soc. Am.* 99.6, pp. 3678–3688.

Haustein, B. G. and W. Schirmer (1970). "A measuring apparatus for the investigation of the faculty of directional localization". In: *Hochfrequenztechnik u. Elektroakustik* 79, pp. 96–101.

Hohmann, V. (2002). "Frequency analysis and synthesis using a Gammatone filterbank". In: *Acta Acoust. united with Acust.* 88, pp. 433–442.

Hummersone, C. (2011). *Binaural Impulse Response Measurements*. Tech. rep. Institute of Sound Recording, University of Surrey, pp. 1–7.

ITU-R BS.1534-1 (2003). *BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems*. Tech. rep., pp. 1–18.

Jepsen, M. L., S. D. Ewert, and T. Dau (2008). "A computational model of human auditory signal processing and perception". In: *J. Acoust. Soc. Am.* 124.1, pp. 422–438.

Käsbach, J., M. Marschall, B. Epp, and T. Dau (2013). "The relation between perceived apparent source width and interaural cross-correlation in sound reproduction spaces with low reverberation". In: *German Acoust. Association (DAGA)*. Merano, Italy, pp. 1613–1616.

Käsbach, J., T. May, N. Le Goff, and T. Dau (2014a). "The importance of binaural cues for the perception of apparent source width at different sound pressure levels Model predictions of ASW". In: *German Acoust. Association (DAGA)*. Oldenburg, Germany, pp. 668–669.

Käsbach, J., T. May, G. Oskarsdottir, C. H. Jeong, and J. Chang (2014b). "The effect of interaural-time-difference fluctuations on apparent source width". In: *Forum Acusticum*. Krakow, Poland.

Käsbach, J., M. Hahmann, T. May, and T. Dau (2016b). "Assessing the contribution of binaural cues for apparent source width perception via a functional model". In: *Int. Congress on Acoustics*. Buenos Aires, Argentina.

Käsbach, J., A. Wiinberg, T. May, M. L. Jepsen, and T. Dau (2016c). "The effect of hearing-aid processing on apparent source width perception in normal-hearing and hearing-impaired listeners". In: *Int. J. Audiol.*

Kuhl, W. (1978). "Spatiousness as a component of the auditory impression". In: *Acustica* 40, pp. 167–181.

Kuk, F., P. Korhonen, C. Lau, D. Keenan, and M. Norgaard (2013). "Evaluation of a Pinna Compensation Algorithm for Sound Localization and Speech Perception in Noise". In: *Amer. J. Audiol.* 22.1, pp. 84–93.

Leakey, D. M. (1959). "Some measurements on the effects of interchannel intensity and time differences in two channel sound systems". In: *J. Acoust. Soc. Am.* 31.7, pp. 977–986.

Lindau, A., V. Erbes, S. Lepa, H. J. Maempel, F. Brinkman, and S. Weinzierl (2014). "A Spatial Audio Quality Inventory (SAQI)". In: *Acta Acust. united with Acust.* 100.5, pp. 984–994.

Lindevald, I. M. and A. H. Benade (1986). "Two-ear correlation in the statistical sound fields of rooms". In: *J. Acoust. Soc. Am.* 80.2, pp. 661–664.

Lipshitz, S. P. (1986). "Stereo microphone techniques: Are the purists wrong?" In: *J. Audio Eng. Soc.* 34.9, pp. 716–744.

Macpherson E. A .and Middlebrooks, J. C. (2002). "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited." In: *J. Acoust. Soc. Am.* 111.5, pp. 2219–2236.

Mason, R., T. Brookes, and F. Rumsey (2005a). "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli". In: *J. Acoust. Soc. Am.* 117.3, pp. 1337–1350.

Mason, R., T. Brookes, F. Rumsey, and T. Neher (2005b). *Perceptually motivated measurement of spatial sound attributes for audio-based information systems*. EPSRC Project Reference: GR/R55528/01. URL: http://iosr.uk/projects/PMMP/index.php.

May, T., R. Decorsière, C. Kim, and A. Kohlrausch (2015). *The auditory front-end framework (TWO!EARS consortium, WP2, D2.2 and D2.3)*. EU-Project, no. 618075, Coordinator: Prof. Dr. Alexander Raake, TU Berlin. URL: http://twoears.aipa.tu-berlin.de/wp-content/uploads/deliverables/.

Mills, A. W. (1958). "On the minimum audible angle". In: *J. Acoust. Soc. Am.* 30.4, pp. 237–246.

Moore, B. C. J., M. Wojtczak, and D. A. Vickers (1996). “Effect of loudness recruitment on the perception of amplitude modulation”. In: *J. Acoust. Soc. Am.* 100.1, pp. 481–489.

Moore, B. C. J., B. R. Glasberg, and M. A. Stone (2010). “Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF”. In: *Int. J. Audiol.* 49.3, pp. 216–227.

Morimoto, M., K. Ueda, and M. Kiyama (1995). “Effects of Frequency Characteristics of the Degree of Interaural Cross-Correlation and Sound Pressure Level on the Auditory Source Width”. In: *Acustica* 81, pp. 20–25.

Morimoto, M., M. Jinya, and K. Nakagawa (2007). “Effects of frequency characteristics of reverberation time on listener envelopment”. In: *J. Acoust. Soc. Am.* 122.3, pp. 1611–1615.

Musa-Shufani, S., M. Walger, H. von Wedel, and H. Meister (2006). “Influence of dynamic compression on directional hearing in the horizontal plane.” In: *Ear Hear.* 27.3, pp. 279–285.

Noble, W., D. Byrne, and K. Ter-Horst (1997). “Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners”. In: *J. Acoust. Soc. Am.* 102.4, pp. 2343–2352.

Okano, T., L. Beranek, and T. Hidaka (1998). “Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls”. In: *J. Acoust. Soc. Am.* 104.1, pp. 255–265.

Peissig, J. and B. Kollmeier (1997). “Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners”. In: *J. Acoust. Soc. Am.* 101.3, pp. 1660–1670.

Perrott, D. R. and T. N. Buell (1982). “Judgements of sound volume: Effects of signal duration, level, and interaual characteristics on the perceived extensity of broadband noise”. In: *J. Acoust. Soc. Am.* 72.5, pp. 1413–1417.

Pessentheiner, H. (2011). *Audio Signal Decorrelation for Phantom Source Widening*. Tech. rep. University of Music and Performing Arts Graz, Austria.

Plack, C. J. (2005). *The Sense of Hearing*. Lawrence Erlbaum Associates.

Plenge, G. (1972). “Über das Problem der Im-Kopf-Lokalisation”. In: *Acustica* 26.5, pp. 241–252.

Pollack, I. and W. J. Trittipoe (1959). “Binaural listening and interaural noise cross correlation”. In: *J. Acoust. Soc. Am.* 31.9, pp. 1250–1252.

Pulkki, V. (2007). “Spatial sound reproduction with directional audio coding”. In: *J. Audio Eng. Soc.* 55.6, pp. 503–516.

Rumsey, F. (2002). “Spatial Quality Evaluation For Reproduced Sound: Terminology, Meaning and a Scene-Based Paradigm”. In: *J. Audio Eng. Soc.* 50.8, pp. 651–666.

Rumsey, F., S. Zieliński, R. Kassier, and S. Bech (2005). “On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality”. In: *J. Acoust. Soc. Am.* 118.2, pp. 968–976.

Santala, O. and V. Pulkki (2011). “Directional perception of distributed sound sources”. In: *J. Acoust. Soc. Am.* 129.3, pp. 1522–1530.

Schroeder, M. R., D. Gottlob, and K. F. Siebrasse (1974). “Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters”. In: *J. Acoust. Soc. Am.* 56, pp. 1195–1201.

Siveke, I., S. D. Ewert, and L. Wiegrebe (2007). “Perceptual and Physiological Characteristics of Binaural Sluggishness”. In: *Hearing - From Sensory Processing to Perception*. Ed. by B. Kollmeier et al. Springer-Verlag Berlin Heidelberg, pp. 467–474.

Sivonen, V. P. (2011). “Binaural Directivity Patterns for Normal and Aided Human Hearing”. In: *Ear Hear.* 32.5, pp. 674–677.

Spors, S., H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter (2013). “Spatial sound with loudspeakers and its perception: A review of the current state”. In: *Proc. IEEE* 101.9, pp. 1920–1938.

Steinberg, J. C. (1937). “The Dependence of Hearing Impairment on Sound Intensity”. In: *J. Acoust. Soc. Am.* 8.3, p. 207.

Stellmack, M., N. F. Viemeister, and A. J. Byrne (2005). “Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers”. In: *J. Acoust. Soc. Am.* 118.4, pp. 2507–2518.

Strutt, J. W. (1907). “On our perception of sound direction”. In: *Philos. Mag.* 13, pp. 214–232.

TWO!EARS, c. (2013-2016). *A computational framework for modeling active exploratory listening that assigns meaning to auditory scenes*. EU-Project, no. 618075, Coordinator: Prof. Dr. Alexander Raake, TU Berlin. URL: http://twoears.aipa.tu-berlin.de.

Terhardt, E. (1979). “Calculating Virtual Pitch”. In: *Hear. Res.* 1.2, pp. 155–182.

Tervo, S., J. Pnen, A. Kuusinen, and T. Lokki (2013). “Spatial decomposition method for room impulse responses”. In: *J. Audio Eng. Soc.* 61.1-2, pp. 16–27.

Thompson, E. R. (2009). “Characterizing binaural processing of amplitude modulated sounds”. PhD thesis. Technical University of Denmark.

Van den Bogaert, T., T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters (2006). “Horizontal localization with bilateral hearing aids: without is better than with”. In: *J. Acoust. Soc. Am.* 119.1, pp. 515–526.

Villchur, E. (1973). “Signal processing to improve speech intelligibility in perceptive deafness”. In: *J. Acoust. Soc. Am.* 53.6, pp. 1646–1657.

Whitmer, W. M. and M. A. Akeroyd (2013). “The effects of simulated room acoustics on the perception of auditory source width by hearing impaired adults with and without amplification”. In: *Int. J. Audiol.* 52.4, pp. 242–302.

Whitmer, W. M., B. U. Seeber, and M. A. Akeroyd (2012). “Apparent auditory source width insensitivity in older hearing-impaired individuals”. In: *J. Acoust. Soc. Am.* 132.1, pp. 369–379.

Whitmer, W. M., B. U. Seeber, and M. A. Akeroyd (2014). “The perception of apparent auditory source width in hearing-impaired adults”. In: *J. Acoust. Soc. Am.* 135.6, pp. 3548–3559.

Wiggins, I. M. and B. U. Seeber (2011). “Dynamic-range compression affects the lateral position of sounds”. In: *J. Acoust. Soc. Am.* 130.6, pp. 3939–3953.

Wiggins, I. M. and B. U. Seeber (2012). “Effects of Dynamic-Range Compression on the Spatial Attributes of Sounds in Normal-Hearing Listeners”. In: *Ear Hear.* 33.3, pp. 399–410.

Zahorik, P., D. S. Brunsgart, and A. Bronkhorst (2005). “Auditory Distance Perception in Humans: A summary of past and present research”. In: *Acta Acust. united Ac.* 91, pp. 409–420.

Zotter, F. and M. Frank (2013). “Efficient Phantom Source Widening”. In: *Arch. Acoust.* 38.1, pp. 27–37.

de Villiers Keet, W. (1968). “The influence of early lateral reflections on spatial impression”. In: *Int. Congress on Acoustics*. Tokyo, Japan, pp. 53–57.

de Vries, D., E. M. Hulsebos, and J. Baan (2001). “Spatial fluctuations in measures for spaciousness”. In: *J. Acoust. Soc. Am.* 110.2, pp. 947–954.

van Dorp Schuitman, J. (2011). “Auditory Modelling for Assessing Room Acoustics”. PhD thesis. Delft University of Technology, Netherlands.

van Dorp Schuitman, J., D. de Vries, and A. Lindau (2013). “Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model”. In: *J. Acoust. Soc. Am.* 133.3, pp. 1572–1585.

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.

Vol. 2: *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.

Vol. 3: *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.

Vol. 4: *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.

Vol. 5: *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.

Vol. 6: *Helen Connor*, Hearing aid amplification at soft input levels, 2010.

Vol. 7: *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.

Vol. 8: *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.

Vol. 9: *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.

Vol. 10: *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.

Vol. 11: *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.

Vol. 12: *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.

Vol. 17: *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.

Vol. 18: *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.

Vol. 19: *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.

Vol. 20: *Michał Fereczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.

Vol. 21: *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.

Vol. 22: *Federica Bianchi*, Complex-tone pitch representations in the human auditory system, 2016.

Vol. 23: *Johannes Zaar*, Measures and computational models of microscopic speech perception, 2016.

Vol. 24: *Gusztáv Lőcsei*, Lateralized speech perception with normal and impaired hearing, 2016.

Vol. 25: *Johannes Käsbach*, Characterizing apparent source width perception, 2016.

The end.

To be continued...

The perception of apparent source width (ASW) characterizes the expansion of sound sources in rooms due to room reflections. Another way of interpreting this perception is that it describes how focussed a listener perceives sound sources. This might have consequences on the listener's ability to discriminate the conversation partner's speech from other interfering background noises. This thesis presents fundamental measurements of ASW in respect to its contributing cues. Consequences for hearing-impaired listeners and the effect of hearing-aid processing on ASW are further investigated. Finally, a binaural model is presented to understand the underlying auditory processes in ASW perception. Besides the fundamental aspects, the results of this thesis might contribute to the design procedure and the evaluation of acoustical spaces, loudspeaker systems and personal hearing devices.

DTU Electrical Engineering
Department of Electrical Engineering

Ørsteds Plads
Building 348
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 38 00
Fax: (+45) 45 93 16 34
www.elektro.dtu.dk