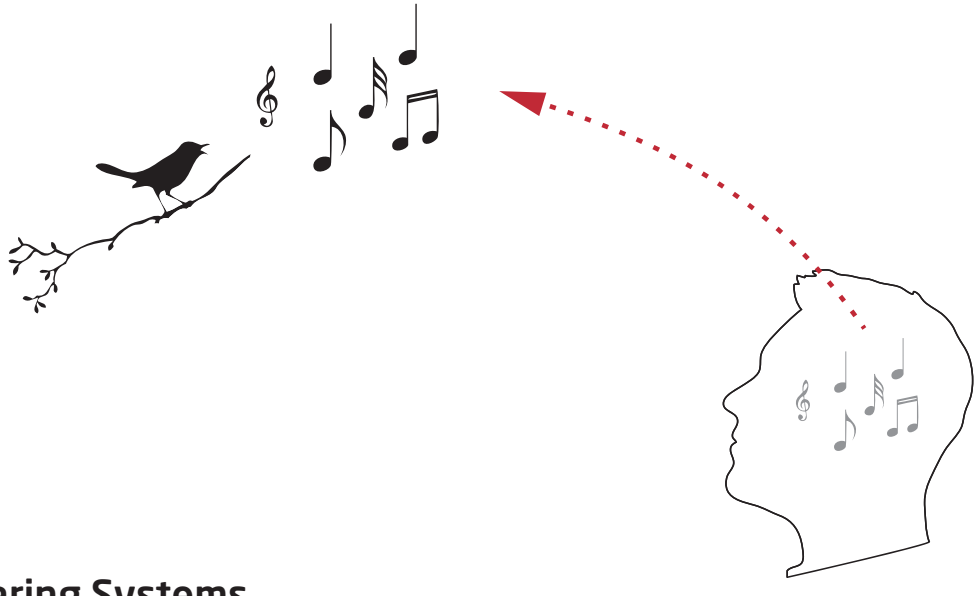
Technical University of Denmark

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 27

Henrik Gert Hassager

Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing



Hearing Systems
Department of Electrical Engineering

Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing

PhD thesis by
Henrik Gert Hassager

Preliminary version: January 30, 2017

Technical University of Denmark

2017

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 the Centre for Applied Hearing Research (CAHR) supported by Widex, Oticon, GN Resound (2/3) and by the Technical University of Denmark (1/3).

Supervisors

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

Senior research scientist Jesper Udesen, PhD
Research Group
GN ReSound A/S
Ballerup, Denmark

Abstract

One key advantage of having two ears is that it facilities localization. Besides the acoustic information in the sound, the auditory system utilizes the time and level differences between the two ears to associate the origin of a sound source with a place in our surrounding space and perceptually place the sound at this position. This process is called externalization. Unfortunately, it is very challenging to construct hearing aids that enable hearing-impaired listeners to externalize in the same way as normal-hearing listeners. This is especially true in reverberant environments, where acoustic reflections interfere with the process. The aim of this work is to understand which components of sound are necessary for externalization. This understanding will help in the development of future hearing aids with the capacity for externalization even in environments with significant reverberation. Based on several experiments with both normal-hearing and hearing-impaired listeners, an analysis of the externalization process was performed. The experiments were conducted over headphones to provide the test subjects with the same acoustic information as if the sound source was physically placed in the reverberant environment. The test subjects were asked to indicate the virtual position of the sound source. Results show that the spectral details of the direct sound plays an important role on externalization, whereas the spectral details in the reverberant sound has little influence on externalization. In addition, we investigated the effect of several hearing-aid compression strategies on externalization. It was found that compression schemes where the acoustic information is shared between the hearing instruments across ears provided the listeners with a spatially distorted sound scene even though the intrinsic level differences were preserved. The analyses suggest that the problem comes from enhanced reverberant energy and modifications of the relation between the direct and the reverberant sound should be avoided. Based on this suggestion, we devised a signal-processing scheme that is driven by the direct sound activity. Implementation and experimental tests of this method demonstrated that the natural sound scene perception is preserved.

Resumé

Fra naturens side er mennesket udstyret med to ører. Det er ikke kun fordi som man måske kunne tro, at to ører hører bedre end et, men fordi to ører hører helt anderledes end et. Ved hjælp af to ører kan man retningsbestemme og lokalisere lydkilder i rummet omkring os. Dette kan lade sig gøre da man placerer naturligt forkommende lydkilder uden for hovedet. Dette kaldes eksternalisering. Desværre viser det sig, at der er store udfordringer forbundet med at konstruere høreapparater, som sætter den hørehæmmede i stand til problemløst at lokalisere lydkilder på samme måde som normalthørende. Retningsbestemmelsen er især en udfordring i omgivelser med kraftige refleksioner, som naturligt nok kan forvirre lokaliseringen. Formålet med dette arbejde har været at bidrage til forståelsen af hvilke egenskaber i lyder der benyttes til at eksternalisere lydkilder korrekt. Denne forståelse vil kunne bidrage til udviklingen af høreapparater som netop kan hjælpe den hørehæmmede med at lokalisere lydkilder selv i rum med en kraftig efterklang. Der er gennemført en analyse af evnen til eksternalisere lydkilder på basis af eksperimenter med både normalthørende og hørehæmmede personer. I eksperimenterne udsættes forsøgspersonerne for lydsignaler afspillet over høretelefoner for at kunne skabe samme eksternaliserede lydbillede som hvis lyden var placeret i rummet. I eksperimenterne udsættes forsøgspersonerne for lydsignaler hvor der er brugt forskellige strategier til at forstærke lydsignalerne. Forsøgspersonerne bedes om at udpege den virtuelle placering af lydkilderne. Eksperimenterne peger på, at en del af problemet med at eksternalisere lydkliderne korrekt er resultatet af forstærkning af energien i den del af lydsignalet som hidrører fra de rumlige refleksioner. Det fører til den hypotese, at evnen til at lokalisere lydkilder kan bevares hvis forholdet mellem energien i den direkte lyd og den reflekterede lyd ikke ændres af forstærkningen. På basis af denne hypotese konstrueres en metode til udelukkende forstærke den direkte den af lyden for derved at kunne bibeholde lydens rummelige egenskaber.

Acknowledgments

First of all, I would like to thank my supervisors at Hearing Systems Group, Tobias May and Torsten Dau for their continued support and guidance throughout the project. It is a pleasure to thank my external supervisor Jesper Udesen, GN Resound, for motivation, initiation and many fruitful discussions. Further, I would like to thank all my colleagues at Hearing Systems. Especially, Alan Wiinberg for the great collaboration on the projects on externalization and wide dynamic range compression. Thanks also to Borys Kowalewski for the countless discussions at the blackboard on signal processing. I appreciated as well all the inspiring discussions in our "spatial hearing group" at Hearing Systems. I would also like to thank my former supervisors Frederik Gran, GN Resound, and Nicolas Le Goff for all their support. Finally, I would like to thank my girlfriend Anna Áslaug Lund, family and friends for the their never failing support during the entire project period.

Related publications

Journal papers

- Hassager, H. G., Gran, F. and Dau, T. (**2016**). "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," J. Acoust. Soc. Am. **139**, 2992–3000

- Hassager, H. G., Wiinberg, A. and Dau, T. "Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment," submitted to J. Acoust. Soc. Am.

- Hassager, H. G., May, T., Wiinberg, A. and Dau, T. "Preserving spatial perception in rooms using direct-sound driven dynamic range compression," submitted to J. Acoust. Soc. Am.

Conference papers

- Hassager, H. G., Gran, F. and Dau, T. (**2014**). "The effect of spectral details on the externalization of sounds," 7th Forum Acusticum, Krakow, Poland.

Abstracts

- Hassager, H. G., Wiinberg, A., Holtegaard P., Udesen, J., and Dau, T. (**2016**). "The effect of dynamic range compression on spatial perception in normal-hearing and hearing-impaired listeners," International Hearing Aid Research Conference, Tahoe City, California.

- Hassager, H. G., Gran, F. and Dau, T. (**2014**). "Contributions of spectral details of the direct and reverberant components to sound externalization," International Hearing Aid Research Conference, Tahoe City, California.

Contents

1

General introduction

Even in situations with multiple sound sources around us in a noisy and reverberant acoustic environment we are able to almost effortlessly segregate the incoming sound mix into individual auditory objects. We can perceive most of the sound sources as being compact in space and can correctly localize them even if no visual information was available. Segregating auditory objects into individual streams is also crucial for following a conversation with a single person in the presence of various interfering sources. The ability to form auditory objects has been shown to be supported by listening through both ears, such that binaural hearing can provide a listening advantage over monaural listening particularly in those acoustic conditions where sound localization matters (e.g., Cherry, 1953).

As a sound arrives at a listener's two ears the sound interacts with the listener's head, torso and pinna that lead to diffraction and scattering of the sound. This diffraction and scattering changes the frequency content of the sound such that when a sound arrives at the two ears it contains differences in level and time. The changes in the frequency content of the ear signals are commonly referred to as the monaural spectral cues whereas the differences across the ears are referred to as interaural level differences (ILDs) and interaural time differences (ITDs). These binaural cues typically vary as a function of frequency such that each position in space will produce a unique set of auditory cues, often referred to as the head-related transfer function (HRTF).

In everyday listening environments, the sound does not only reach the listener's ears directly from the position of the sound source but also from different directions via reflections from surrounding surfaces (e.g., walls and objects) in the reverberant space. The time-course of the reverberation can be divided into three distinct parts each having its own characteristics. The first part represents the direct sound, followed by the "early reflections" containing the first distinguishable reflections and the reverberant part where the reflections become denser and decay exponentially with time (e.g., Kuttruff, 2009).

In reverberant environments, the auditory system has been demonstrated to be able to identify the direct sound while "suppressing" the early reflections from the other spatial directions. This remarkable phenomenon has been described as the "precedence effect" or "law of the first wave front" (e.g., Wallach et al., 1949; Haas, 1951) and enables the listener to robustly identify the direction of a sound source despite the complex input from various directions, even though this process has its limitations since reverberation for certain sound sources can indeed degrade the listener's accuracy of identifying their direction (e.g., Hartmann, 1983). In addition, the relation between the direct sound and the reflections has been shown to facilitate distance perception in a reverberant environment (e.g., Mershon and King, 1975; Nielsen, 1993). While the direct sound energy decreases by 6 dB per doubling of the distance to the source, the reverberant energy remains fairly constant. This implies that the energy ratio of the direct sound and the reverberant sound (DRR) decreases for increasing distances to the sound source.

When people listen to sounds over headphones, the source of the sound is generally perceived to be located inside the listener's head, i.e. it is internalized. However, when the headphone reproduction contains the individual auditory cues provided by the HRTF and the interaction with the reverberant

environment, the same convincing externalized perception can be achieved as if the sound originated from the same location in the space. The headphone reproduction should also account for the head movements of the listener as it has been shown that head movements also provide additional dynamic cues that contribute to externalization perception (Brimijoin et al., 2013).

Noble and Gatehouse (2006) found in their questionnaire-based field study that many hearing-aid users experience difficulties with localization of sounds. In the worst case, sounds were found to be perceived as internalized instead of externalized in the space. This indicated that the hearing-aid processing distorted the auditory cues provided by HRTF and the interaction with the reverberant environment. Those findings motivated the present work. The goal was to obtain a detailed understanding of the effects of signal modifications (e.g. via hearing-aid signal processing) on the spatial cues involved in externalization. Such understanding should have major impact for the development and evaluation of new signal processing strategies, e.g. in hearing instruments and other devices were robust sound externalization is required. Ideally, the work could help design compensation strategies for hearing-impaired listeners such that their spatial perception is not compromised by the applied amplification.

1.1 Overview of the thesis

Specifically, this thesis is focused on both static and dynamic hearing-aid related modifications and their impact on externalization perception both in normal-hearing and hearing-impaired listeners. All experimental listening tests conducted in this thesis use stimuli virtualized over headphones to simulate externalized sound sources.

Chapter 2 investigates the effect of manipulating the spectral details in the

binaural transfer function on perceived externalization in a reverberant room, to understand the implications of e.g., the position of the microphones in hearing aids. The spectral details are reduced in either the direct or reverberant part of the BRIRs by smoothing the magnitude responses with band-pass filters. Listening tests are performed using broadband noise convolved with unmodified BRIRs and modified BRIRs. It is investigated how the reduction spectral details in the BRIR affects the externalization of the broadband noise. In an attempt to characterize the auditory processes involved in externalization perception, a computational model is presented and evaluated.

Chapter 3 investigates the effects of fast-acting hearing-aid compression on spatial perception in an everyday reverberant environment. It has been shown that temporal modifications on reverberant ear signals have been shown to degrade the externalization perception (Catic et al., 2013). Three compression schemes are considered: independent compression at each ear, linked compression between the two ears, and "spatially ideal" compression operating solely on the dry source signal. Linear amplification processing is used as the reference condition. The listener's task is to indicate the location and extent of their perceived sound images on the horizontal plane graphically on a touch screen. The interaural coherence of the ear signals and direct-to-reverberant energy ratio calculated from the ear signals are considered as objective outcome measures and compared to the obtained experimental data.

Chapter 4 focuses on the importance of preserving the spatial perception of listeners in a reverberant environment while applying fast-acting hearing-aid compression. It is investigated whether this can be accomplished by linearizing the hearing-aid when the signal is dominated by reverberation, while maintaining the fast-acting compression when the signals are dominated by the direct sound. The classification of the signal into direct sound and reverberation is

either done with a priori knowledge of the BRIR, or estimated blindly from the ear signals. Both linear processing and fast-acting hearing-aid compression without linearization of the reverberant part of the signal will be used as reference conditions. The listeners' task is to indicate their spatial perception of virtualized speech graphically to capture all relevant spatial attributes with respect to distance, azimuth localization, source width and the occurrence of split images.

Finally, *Chapter 5* summarizes the main findings of this thesis and discusses the implications for models of auditory signal processing and perception, particularly in relation to sound localization and externalization, as well as for future applications in hearing instruments.

2

The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment[a]

Abstract

Individual binaural room impulse responses (BRIRs) were recorded at a distance of 1.5 m for azimuth angles of 0° and 50° in a reverberant room. Spectral details were reduced in either the direct or the reverberant part of the BRIRs by averaging the magnitude responses with band-pass filters. For various filter bandwidths, the modified BRIRs were convolved with broadband noise and listeners judged the perceived position of the noise when virtualized over headphones. Only reductions in spectral details of the direct part obtained with filter bandwidths broader than one equivalent rectangular bandwidth affected externalization. Reductions in spectral details of the reverberant part had only little influence on externalization. In both conditions, externalization was not as pronounced at 0° as at 50°. To characterize the auditory processes that may be involved in the perception of externalization, a quantitative model

[a] This chapter is based on Hassager, H. G., Gran, F., Dau, T. (2016), *The Journal of the Acoustical Society of America.*

is proposed. The model includes an echo-suppression mechanism, a filterbank describing the frequency selectivity in the cochlea and a binaural stage that measures the deviations of the interaural level differences between the considered input and the unmodified input. These deviations, integrated across frequency, are then mapped to a value that corresponds to the perceived externalization.

2.1 Introduction

One fascinating aspect of the human hearing sense is its ability to capture the surrounding auditory space from the often complex acoustic input to the two ears. Even in reverberant environments and in the presence of multiple sound sources, the auditory system is able to extract the relevant acoustic information provided by the binaural room impulse responses (BRIRs), such that sound sources are perceived as externalized and as arising from the reverberant environment. The BRIRs represent the impulse responses from the sound source positions to the listener's left and right ears. The impulse responses consist of the filtering by the head, torso, and pinna, as described by the head-related transfer functions (HRTFs), and the acoustic interaction with the reflective environment in which the listener is present (Hartmann and Wittenberg, 1996; Begault et al., 2001). According to Hartmann and Wittenberg (1996), externalized sound images are perceived by the listener to be compact and correctly located in space. Convincingly externalized sound images can be obtained via headphones, if the headphone reproduction includes the filtering by the BRIRs (Begault et al., 2001). However, when sounds presented via headphones or other listening devices, such as hearing aids, lack the filtering by the BRIRs, the externalization breaks down and the sound images are most likely perceived

inside of the listener's head, i.e., internalized. In studies on distance perception, the ratio of the acoustic energy in the direct sound versus the reverberant part of the sound, i.e., the direct-to-reverberant ratio (DRR), has been demonstrated to relate to the perceived sound source distance (e.g., Zahorik et al., 2005; Kopčo and Shinn-Cunningham, 2011). However, several studies on sound externalization suggested that the spectra of the left-ear and the right-ear stimuli affect the amount of perceived externalization (e.g., Hartmann and Wittenberg, 1996; Boyd et al., 2012).

Some studies investigated the effect of modifications of the (spectral) shape of the HRTF on the spatial perception of the sound. Specifically, Kulkarni and Colburn (1998) examined how a reduction of the spectral details of individual HRTFs affects sound localization in anechoic conditions whereby no head movements were allowed but visual cues were present (Kulkarni and Colburn, 1998). In their study, a smoothing of the spectral details of the HRTF was obtained by truncating the Fourier series of the HRTF spectra. The influence of the reduced spectral details was studied for azimuth angles of a single sound source at 0°,45°, 135°, and 180°, using sounds externalized over tube phones. By comparing virtual and real loudspeaker stimuli, Kulkarni and Colburn (1998) found that the number of coefficients in the Fourier series could be reduced from 512 to 16 without affecting the perceived sound localization, i.e., all listeners reported fully externalized virtual sounds even in the conditions with the strongest spectral smoothing.

Breebaart and Kohlrausch (2001) investigated to what extent reductions in the spectral details of HRTFs were detectable. They smoothed the magnitude and phase spectra of nonindividual HRTFs with a gammatone filterbank (Patterson et al., 1988) with different filter orders to achieve different degrees of smoothing. The fourth-order gammatone filterbank had individual filter

bandwidths corresponding to the equivalent rectangular bandwidth (ERB) scale according to Glasberg and Moore (1990) which roughly reflects the frequency-selective processing in normal-hearing listeners. Listeners compared sounds that were convolved with the smoothed vs original HRTFs and were asked to rate the perceived audible difference on a three-step scale. Azimuth angles of 0°,30°, and 120° were considered in their study and it was found that a smoothing with filter orders above one did not produce any audible effect whereas smoothing with filter orders below one was detectable.

The above studies were either carried out only in an anechoic environment (Kulkarni and Colburn, 1998), i.e., in the absence of any reverberation, or were focused on pure detectability of the spectral modifications of the HRTFs (Breebaart and Kohlrausch, 2001). However, it has been demonstrated that reverberation contributes to externalization perception (e.g., Begault et al., 2001; Catic et al., 2013) via the binaural cues provided by the interaural level differences (ILDs) and interaural time differences (ITDs) in a given environment. Related to the DRR, Catic et al. (2015) found that BRIR modifications, by either BRIR truncations or making the reverberation identical in both ears, altering the binaural cues in terms of the interaural coherence (IC) are important for sound source externalization. Thus the binaural interaction of the direct part as well as early reflections and the late reverberation of the BRIR have been found to affect externalization. However, a detailed understanding of the role of the different parts of the BRIR on externalization is still missing.

The present study investigated the influence of spectral smoothing of the BRIR on externalization in a reverberant environment. Two experiments were conducted to study the effects of early reflections and late reverberation on externalization as a function of the spectral fidelity of the processing. In the first experiment, the direct part of the BRIR was spectrally smoothed, similar as

in the study of Breebaart and Kohlrausch (2001), but combined with the reverberant part of the BRIR which was left unchanged. In the second experiment, the corresponding effect of spectral smoothing of the reverberant part of the BRIR (containing the early reflections and late reverberation) was investigated whereby the direct part was left unchanged. By applying spectral smoothing only on the direct part of the BRIR, the effects of the acoustical properties of the (modified) sound source on externalization (e.g., due to spectral coloration of individual ears) were studied. When applying spectral smoothing on the reverberant part of the BRIR, the effects of the acoustical properties of delayed versions of the sound source on externalization were studied. The experiments were carried out not allowing the listeners to move their heads, since head movements have been demonstrated to affect externalization (Brimijoin et al., 2013).

Furthermore, in an attempt to characterize the auditory processes that may be involved in externalization perception, a simple quantitative model was developed. The model consisted of several stages of monaural and binaural auditory preprocessing of the input stimuli, including an echo-suppression mechanism (where the direct sound was assumed to partly suppress the lagging reverberant sound components) and a mapping between the respective internal representation of the stimuli and their corresponding externalization percept. The model was applied to the conditions of the two experiments considered in the present study.

2.2 Methods

2.2.1 Externalization measurements

Listeners

Seven listeners, six males and one female, with audiometric pure-tone thresholds below 20 dB hearing level between 125 and 8000 Hz, aged between 33 and 40 years, participated in the experiment. All listeners were familiar with psychoacoustic localization and externalization experiments. Prior to all experiments, training was conducted to make the listeners familiar with the different degrees of externalization they would encounter in the experiments.

BRIR measurements

Individual BRIRs were recorded with a source-to-listener distance of 1.5 m and for azimuth angles of 0° and 50°. The recordings were carried out in a reverberant listening room designed in accordance with the IEC 268-13 (1985) standard. Figure 2.1 shows an illustration of the experimental setup. The BRIRs were recorded with the Tucker-Davis Technologies RX8 system at 48,828 Hz, with two Etymothic Research ER-7C probe microphones placed in the ear canals of the listeners. The tip of each probe microphone was placed in an open dome used for behind-the-ear hearing aid. A maximum-length-sequence (MLS) of order 13, with 32 repetitions played from a Genelec 6010a loudspeaker, was used to obtain the speaker-to-ear impulse responses, $x(t)_{brir}$. The speaker-to-ear impulse responses are considered to be the BRIRs. The headphone-to-ear impulse responses, $x(t)_{hpir}$, from the HD650 Sennheiser headphones (placed on the listeners) to the probe microphones, were obtained with the same MLS immediately after the recordings of the headphone-to-ear impulse responses. To compensate for the headphones, the inverse headphone-to-ear impulse

response, $x(t)_{hpir,inv}$, was calculated in the time domain using the Moore-Penrose pseudoinverse. The speaker-to-ear-canal impulse responses,$x(t)_{brir}$, were convolved with the inverse headphone-to-ear-canal impulse responses, $x(t)_{hpir,inv}$, to create filters for virtual external sound source generation. Due to inaccuracies in the placement of the probe microphones in the current recording setup, the BRIRs cannot be considered to be accurate at frequencies above 6000Hz.

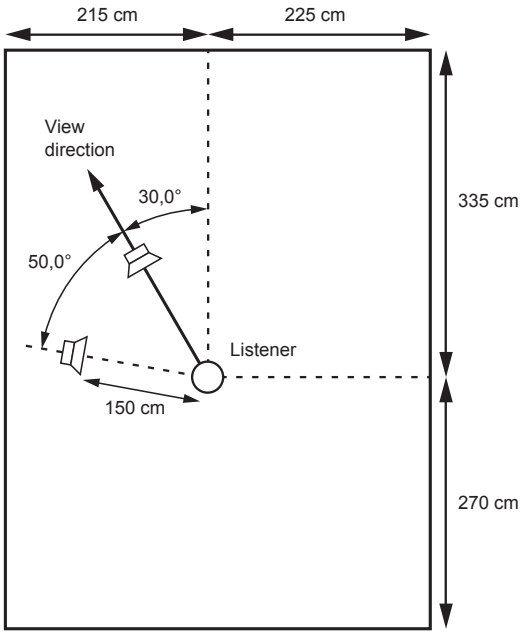


Figure 2.1: The geometry of the listening room with the placement of the listener and the loudspeakers at 0° and 50°. The 0° angle is indicated by the view direction.

BRIR modifications

To study the effect of spectral detail in the BRIRs on externalization, modifications of the direct part and the reverberant part of the BRIR were undertaken. When the listeners were seated in front of the loudspeakers, the first reflection

from the floor occurred after 3.8 ms. The direct part of the BRIRs was therefore defined as the first 3.8 ms, starting at the sample of most energy. The reverberant part of the BRIR after 3.8 ms comprised early reflections and late reverberation. A 5 ms half raised-cosine window was used to ensure a smooth transition of the BRIRs between the direct part and the reverberant part. Figure 2.2 shows the transition of the BRIR between the direct part and the reverberant part for one of the measured BRIRs. In the time domain, the sum of the direct and the

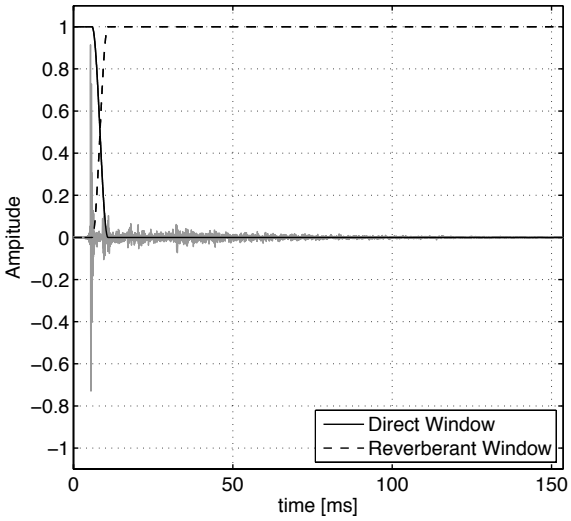


Figure 2.2: The black solid and dashed functions indicate the transition windows applied on the BRIRs (in grey) to divide the BRIRs into the direct part and the reverberant part.

reverberant part of the BRIRs is denoted as

$$x(t)_{brir} = x(t)_{dir} + x(t)_{reverb}\,. \tag{2.1}$$

The magnitude spectrum of either the direct part or the reverberant part of the BRIRs was smoothed with a gammatone filterbank. The method described in the following is similar to the procedure used by Breebaart and Kohlrausch (2001). However, in contrast to Breebaart and Kohlrausch, the different degrees of smoothing were achieved here for different bandwidths, b, of the gammatone

filters rather than different orders, n. The smoothed magnitude spectrum, $\left|Y(f)\right|$, can be found by calculating the smoothed frequency bins for each center frequency, f_c, given by

$$\left|Y(f_c)\right| = \sqrt{\frac{\int_0^\infty \left|X(f)\right|^2 \left|H(f,f_c)\right|^2 \mathrm{d}f}{\int_0^\infty \left|H(f,f_c)\right|^2 \mathrm{d}f}} \tag{2.2}$$

with $\left|X(f)\right|$ representing the original magnitude spectrum of either the direct part or reverberant part of the BRIR, $x(t)_{dir}$ or $x(t)_{reverb}$, and $\left|H(f,f_c)\right|$ denoting the magnitude spectrum of the gammatone filter at the center frequency, f_c. The approximation of the transfer function of a fourth-order gammatone filter is given as

$$\left|H(f,f_c)\right| = \left(\frac{1}{1+j(f-f_c)/b(f_c)}\right)^4. \tag{2.3}$$

To achieve magnitude smoothing of different degrees, the bandwidth $b(f_c)$ of the gammatone filters was represented as

$$b(f_c) = B \times 1.149 \times 24.7\left(\frac{4.37 \times f_c}{1000}+1\right) \tag{2.4}$$

with B denoting the bandwidth factor relative to a value of one, representing the original ERB values according to Glasberg and Moore (1990).

Since it has been shown that a minimum-phase version of the HRTF is a perceptually valid description (Kulkarni et al., 1995) for the HRTF, the magnitude smoothing of the direct part of the BRIRs was applied to minimum-phase versions of the direct parts of the BRIRs. The direct part of the BRIRs was therefore decomposed into a minimum-phase filter and an all-pass filter

$$X(f)_{dir} = \left|X(f)\right|_{dir} e^{j\varphi(f)_{dir,mp}} e^{j\varphi(f)_{dir,ap}}, \tag{2.5}$$

where $\varphi(f)_{dir,mp}$ and $\varphi(f)_{dir,ap}$ indicate the phases of the minimum-phase and the all-pass filters, respectively. The all-pass filters can be considered as pure delays. The magnitude spectra of the minimum-phase filters, $\left|X(f)\right|_{dir}$, were smoothed according to Eq. (2.2). The phase of magnitude-smoothed filters were turned into minimum-phase to ensure that the phase modifications were kept small. The magnitude smoothed minimum-phase filters were then convolved with the corresponding all-pass filters to generate the modified direct parts

$$Y(f)_{dir} = \left|Y(f)\right|_{dir} e^{j\,\psi(f)_{dir,mp}} e^{j\,\varphi(f)_{dir,ap}}, \tag{2.6}$$

where $\psi(f)_{dir,mp}$ denotes the phase for the magnitude smoothed minimum-phase filter. The corresponding unmodified reverberant part was then added to generate the BRIRs with the modified direct part to form BRIRs with magnitude smoothed direct parts

$$y(t)_{brir,dir} = y(t)_{dir} + x(t)_{reverb}. \tag{2.7}$$

The magnitude smoothing of the reverberant part of the BRIRs was applied on the magnitude response of the short-time Fourier transformation, similar to the smoothing done by Baer and Moore (1993). A short-time Fourier transformation of the reverberation part of the BRIRs was applied with an 8192-samples long Hanning window, corresponding to 160 ms, and a step size of one sample. For each window, frequency spectra were obtained,

$$X(f)_{reverb} = \left|X(f)\right|_{reverb} e^{j\,\varphi(f)_{reverb}}. \tag{2.8}$$

The magnitude spectra were smoothed, according to Eq. (2.2), while keeping

the phase spectra unmodified

$$Y(f)_{reverb} = \left|Y(f)\right|_{reverb} e^{j\,\varphi(f)_{reverb}}, \tag{2.9}$$

where $\left|X(f)\right|_{reverb}$ indicates the original magnitude spectrum, $\left|Y(f)\right|_{reverb}$ indicates the smoothed magnitude spectrum, and $\varphi(f)_{reverb}$ represents the phase. The smoothed reverberant part of the BRIRs was generated via the corresponding inverse short-time Fourier transformation. The unmodified direct parts were then added to generate the BRIRs with the modified reverberant parts

$$y(t)_{brir,reverb} = x(t)_{dir} + y(t)_{reverb}\,. \tag{2.10}$$

Figure 2.3 shows the original spectra and the smoothed spectra, for one of the listeners, for both the direct part (top panels) and the reverberant part (bottom panels) of the BRIR for the 50° conditions. The left panels show the results for the left-ear stimuli and the right panels show the corresponding results for the right-ear stimuli. The smoothed spectra were created using the same degree of smoothing as in the experiment. For better visibility, the smoothed spectra were separated by 10 dB for each degree of smoothing.

Experimental conditions

The stimuli used for the experiments were band-pass filtered noises (50-6000Hz) with a duration of 4 s and a sound pressure level (SPL) of 75 dB. The long duration of the stimuli was chosen to ensure that the listeners could provide a reliable response associated with their percept. BRIRs with either the modified direct part or the modified reverberant part were created with bandwidth factors, B, of 0.316, 0.570, 1.03, 1.85, 3.33, 6.0, 10.8, 19.5 35.0, and 63.1. The noise was convolved with either the original or the modified BRIRs and then convolved

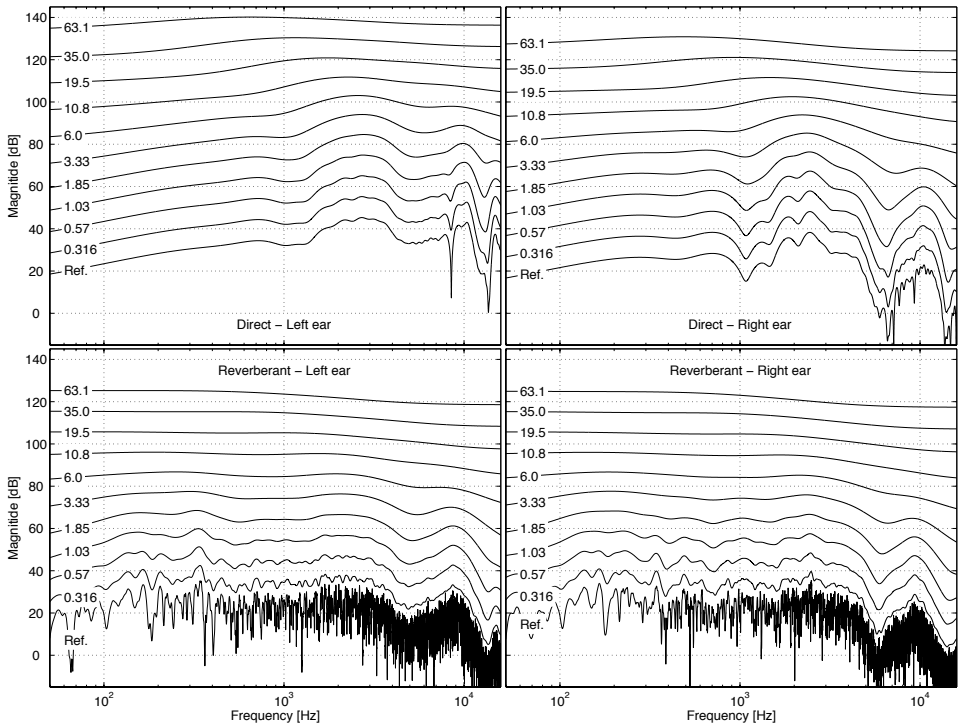


Figure 2.3: Examples of original (bottom-most curves) and smoothed spectra (separated by 10 dB for each degree of smoothing for better visibility) for both the direct part (top panels) and the reverberant part (bottom panels) of the BRIR for the 50° conditions for one of the listeners. The results for the left-ear stimuli are shown in the left panels and the corresponding results for the right-ear signals are shown in the right panels.

with the inverse headphone-to-ear filters to create signals providing different degrees of perceived externalization of the noise. These signals were presented over the same headphones as used in the BRIR measurements. The listeners were seated in front of the loudspeakers, which were placed at the measurement positions, as illustrated in Fig. 2.1. The listeners were asked to remain still and judge the perceived position of the sound source between the loudspeaker and their head. Similar to Hartmann and Wittenberg (1996), a linear scale was used to indicate the distance from the listener's head to the loudspeaker. A scale from 1 to 5 was used where 1 corresponded to a percept in the head and 5 corresponded to a percept coming from the loudspeaker. The listeners were instructed to ignore other perceptual attributes, such as frequency coloration

or apparent source width, and to only focus on the degree of perceived position of the sound source. The modified signals and the original signal were each repeated 10 times with different realizations of the noise. No response feedback was provided to the listeners. The signals were presented in random order. The experiment was first carried out at the azimuth angle of 0° and then at the azimuth angle of 50°. Prior to the actual experiment, a training session was carried out, where each of the signals was played once to get the listeners familiar with the perceptual sensation. Some listeners reported changes in coloration. However, all listeners reported the source width of the stimuli to be unchanged. Results of the experiment will be presented later in Sec.2.3.

2.2.2 Externalization model

The proposed model maps a given acoustical input signal onto a value that corresponds to the perceived externalization of the signal. The model includes several preprocessing stages, including an echo-suppression mechanism, a middle-ear filter, a gammatone filterbank describing the frequency selectivity in the cochlea, and a binaural stage that utilizes the ILD deviations of the considered stimuli from the reference ILDs of the unmodified stimuli. The deviations from the reference are then mapped to the perceived externalization. The measured BRIRs were used to calculate the reference ILDs that provide the correct externalization for the azimuth angles of 0° and 50°. It was assumed that the listener has learned the properties of the reference BRIRs, e.g., the natural ILDs given by the listener's HRTF and the statistical properties of the reverberation.

Figure 2.4 illustrates the individual processing steps of the model. First, a simple temporal weighting of the BRIRs was applied to both the left-ear and the right-ear impulse responses to simulate echo suppression, inspired by the processing proposed in Catic et al. (2015). According to Braasch et al. (2003), the

interval of the BRIR between 4.5 and 80 ms represents the summing location window for broadband sounds. The weighting function was thus assumed to take a value of one for the portion of the BRIR up to the first reflection (3.8 ms), followed by a transition to a value of 0.01 reflecting suppression, followed by a transition from 0.01 back to one using a half raised-cosine window from 10 to 160 ms. Figure 2.5 i shows the echo suppression window, represented on a dB scale, together with one of the measured BRIRs.

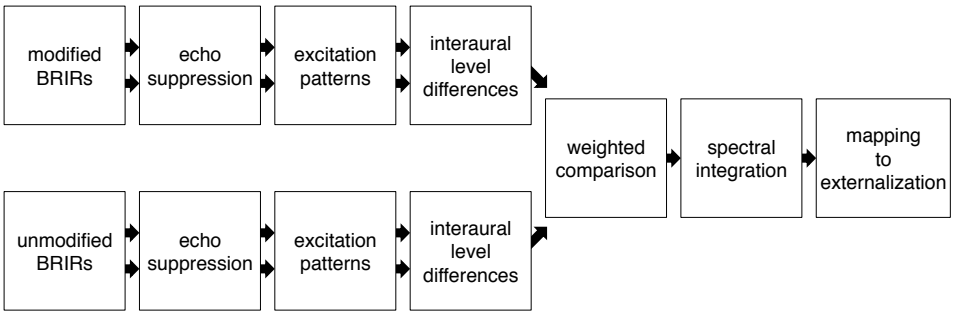


Figure 2.4: Structure of the computational sound externalization model considered in the present study. The input to the model is the noise stimulus convolved with the unmodified BRIR (lower path) and the modified BRIR (upper path). This signal is processed through several stages, including an echo-suppression mechanism, a gammatone filterbank to calculate the excitation pattern at the level of the cochlea and a binaural stage that calculates the ILDs. The deviations between the modified and unmodified ILDs are calculated at each center frequency, weighted and integrated across frequency. The deviations from the reference are then mapped to the perceived externalization.

The middle-ear filtering was simulated using a 512 tap finite impulse response filter as described by Goode et al. (1994) and Lopez-Poveda and Meddis (2001). The frequency selectivity in both ears was calculated via the excitation pattern as a function of the filter center frequency on an ERB scale, following Glasberg and Moore (1990). Next, the ILDs (in dB) at the output of the excitation patterns were calculated for center frequencies with audible content (50-6000Hz) both for the unmodified BRIRs (upper path in Fig. 2.4) and the modified BRIRs (lower path in Fig. 2.4). The deviation between the signal ILDs

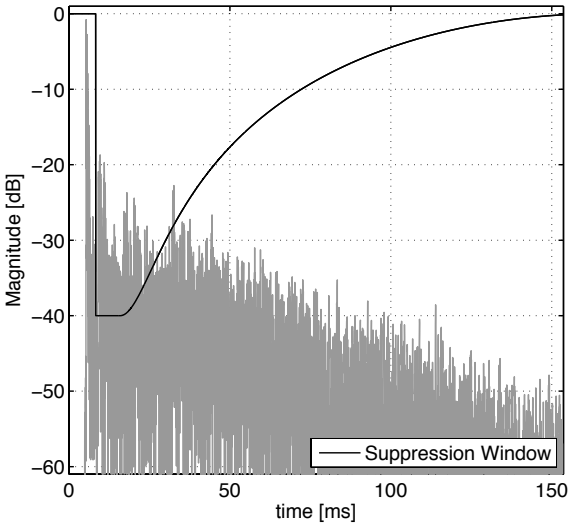


Figure 2.5: The echo suppression window used in the model is indicated by the black solid function (represented on a dB scale). The grey pattern represents one of the measured BRIRs.

and the reference ILDs was calculated at each center frequency,

$$\Delta ILD = \left| ILD_{signal} - ILD_{reference} \right| . \tag{2.11}$$

ILD discrimination thresholds have been shown to be roughly constant at about 0.5-2 dB across frequency for a reference ILD of 0 dB (e.g., Yost and Dye, 1988). In the proposed model, deviations below 1.5 dB were therefore considered to be below threshold and set to zero,

$$\Delta ILD_{trun} = \begin{cases} \Delta ILD, & \Delta ILD \geq JND \\ 0, & \Delta ILD < JND . \end{cases} \tag{2.12}$$

The ratio of the ILD deviations (from the reference ILDs) to the reference ILDs was calculated at each center frequency. This implies that deviations from small reference ILDs are weighted more strongly than deviations from large references ILDs. This is consistent with results from measurements of the minimum audible angle (e.g., Mills, 1958). An integration across the center

frequencies containing audible content was undertaken to compute the overall normalized ILD deviation Δ_{ILD},

$$\Delta_{ILD} = \frac{1}{N_{CF}} \sum_{i_{CF}=1}^{N_{CF}} \frac{\Delta ILD_{trun}}{\left|\Delta ILD_{reference}\right|}, \tag{2.13}$$

whereby it was assumed that ILDs contribute equally across frequency, as proposed by Hartmann and Wittenberg (1996). Finally, the overall normalized ILD deviation was mapped to a value corresponding to the perceived externalization of the incoming sound. The mapping is represented by a decay- ing exponential function

$$P_{ext} = c_1 e^{-z_1 \Delta_{ILD}} + c_2, \tag{2.14}$$

where $c_1 = 3.78$, $c_2 = 1$ and $z_1 = 0.99$ represent the mapping parameters. c_1 and z_1 were derived from a least-squares fit and c_2 was defined such that stimuli providing large ILD deviations correspond to an internalized percept ($P_{ext} \rightarrow 1$ for $\Delta_{ILD} \rightarrow \infty$). The parameters were fit only to the data obtained in the 0° azimuth condition where the spectral smoothing was applied to the direct sound (corresponding to a quarter of the overall data set) and were kept constant throughout the remaining conditions considered in the present study. The fitting could alternatively have been based on another subset of the experimental data without major affect on the derived parameter values.

2.3 Results

Figure 2.6 shows the mean externalization ratings (open symbols), averaged across listeners, as a function of the smoothing bandwidth factor. Simulated externalization values obtained with the proposed model are indicated by the filled symbols. The left panel shows the results obtained with the stimuli presented

from the 0° direction whereas the corresponding results for the 50° azimuth are shown in the right panel.

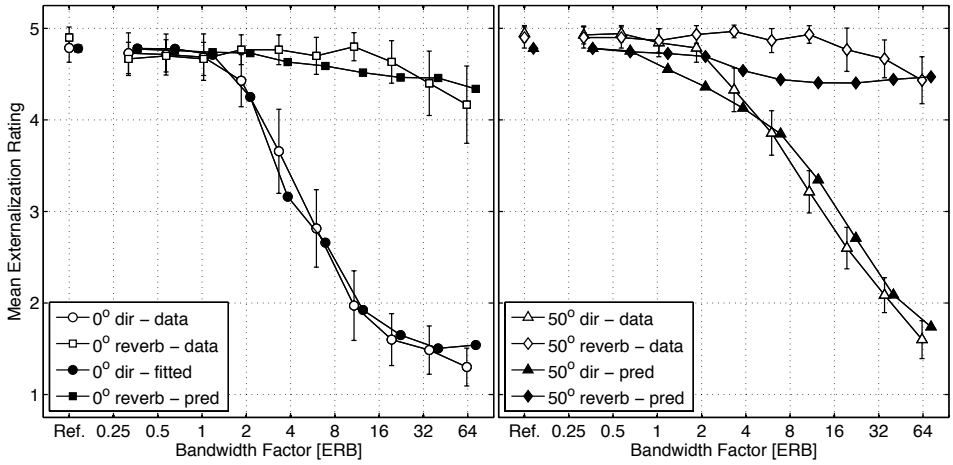


Figure 2.6: The mean of the seven listeners perceived sound source location (open symbols) as a function of the bandwidth factor and the corresponding model predictions (filled symbols). The model predictions have been shifted slightly to the right for a better visual interpretation. The error bars are one standard error of the mean.

Regarding the 0° direction (left panel), the open circles represent the data for the condition where the direct part of the BRIRs was modified while the reverberant part was kept untouched, indicated as "0° dir." It can be seen that spectral smoothing of the direct part obtained with band-width factors below about 1 ERB did not, or only marginally, affect the perceptual externalization of the virtual sound source. In contrast, spectral smoothing with a band-width factor above 1 ERB led to decreasing externalization ratings with increasing smoothing factor such that, for the largest bandwidth factors, the sound was perceived to be close to the head or almost fully internalized. In contrast, when the reverberant part of the BRIR was spectrally smoothed whereas the direct part was kept untouched, as indicated by the open squares and referred to as 0° reverb in the figure, essentially no effect of the amount of the spectral smoothing on externalization was observed, i.e., the sound was always perceived as being externalized closely to the loudspeaker.

For the 50° direction (right panel), the overall pattern of the results was similar as in the 0° conditions: The externalization ratings decreased monotonically with increasing bandwidth factor above 1 ERB when the direct part was modified and the reverberant part was left unchanged, as indicated by the open triangles and referred to as the "50° dir" condition. In the "50° reverb" condition (open diamonds), where the reverberant part of the BRIR was modified but not the direct part, the externalization was hardly affected by the amount of spectral smoothing. However, there are also differences in the results obtained for the two source directions. In particular, for bandwidth factors above 2, the ratings obtained in the condition "50° dir" were above those obtained in the condition 0° dir, i.e., the decay of externalization with increasing bandwidth factor was more gradual for the 50° than for the 0° condition and did not reach the same low values of externalization even for the largest bandwidth factors. In all considered experimental conditions, i.e., the conditions 0° dir, 0° reverb, "50° dir," and "50° reverb," the listeners reported consistently to perceive compact (non-diffuse) sound sources.

The simulations (filled symbols in Fig. 2.6) agree reasonably well with the measured data. The model describes the main effects of spectral smoothing of the BRIR on externalization both regarding the effects of modifications of the direct vs the reverberant part as well as regarding the sound source location. The simulations obtained in the (non-fitted) conditions show some deviations from the data for the intermediate bandwidth factors. However, the deviations never exceed half an externalization category.

Figure 2.7 shows the relationship between the measured externalization ratings (replotted from Fig. 2.6)and the model output after the preprocessing stages [Eq. (2.13)] and before the final mapping to perceived externalization [Eq. (2.14)]. Thus, the measured externalization ratings obtained in all conditions

are now represented as a function of the overall normalized ILD deviation from the reference ILD (as defined in Sec. 2.2.2) for the respective stimulus conditions 0° dir (circles), 0° reverb (squares), 50° dir (triangles), and 50° reverb (diamonds). It can be seen that stimuli from different experimental conditions that produced a similar externalization percept in the measurements exhibit similar values of the integrated ILD deviation. The solid function in Fig. 2.7 shows the least squares fit of the externalization ratings to the corresponding overall normalized ILD deviation and represents the mapping function [Eq. (2.14)] used in the final model step.

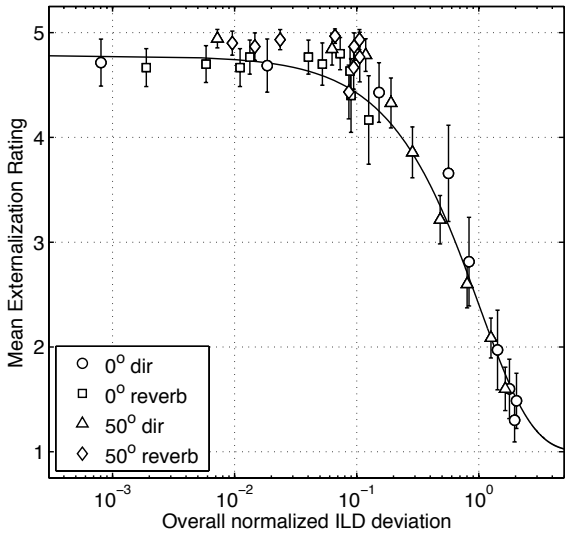


Figure 2.7: The perceived sound source location as function of the overall normalized ILD deviation together with the exponential fit representing the mapping function. The overall normalized ILD deviation (x axis) is show on a logarithm scale to display the data point linearly. The error bars are one standard error of the mean.

2.4 Discussion

The results from the present study showed that, when visual cues are provided and no head movements are allowed, the spectral details of the direct part of the BRIR contribute to sound externalization in a reverberant environment.

These observations seem not consistent with the informal observations in the anechoic condition by Kulkarni and Colburn (1998), where all listeners reported fully externalized virtual sounds. However, the finding that spectral alterations affect the perceived externalization is in agreement with the results obtained in Hartmann and Wittenberg (1996) and Boyd et al. (2012). In the presence of early reflections and late reverberation, spectral modifications in the direct part of the BRIR (corresponding to the HRTF in an anechoic environment) are important for sound externalization in reverberant listening environments.

In studies on distance perception in reverberant environments, it has been demonstrated that the DRR of the stimuli represents an informative indicator, such that the distance of a sound source is perceived to be further away with decreasing amount of DRR (e.g., Zahorik et al., 2005; Kopčo and Shinn-Cunningham, 2011). In order to test whether this metric also successfully accounts for the externalization data obtained in the present study, DRRs were calculated by convolving the direct and reverberant parts of the individual BRIRs with the noise stimuli and by computing the total power contained in the two parts, similar to Kopčo and Shinn-Cunningham (2011). Figure 2.8 shows the mean DRR for the modified BRIRs[a]. The mean DRRs are shown for the listeners' left-ear stimuli for the two frontal conditions, 0° dir and 0° reverb. Similar characteristics would be observed in the 50° conditions (not shown). In the condition 0° reverb (squares), the DRR increases with increasing amount of smoothing, suggesting a decrease of the perceived distance. This would be qualitatively consistent with the (slightly) decreasing amount of externalization in the data (open squares in Fig. 2.6). In the condition 0° dir (circles), the DRR

[a] According to Eq. (2.2), the smoothing process ensures a constant power across frequency. Since the frequency range of the stimulus (50 to 6000 Hz) and that of the smoothing process (0 to 24 414 Hz) are different, the DRRs changed as a function of the bandwidth factor used in the smoothing process.

decreases with increasing amount of smoothing, suggesting an increase of the perceived distance according to Kopčo and Shinn-Cunningham (2011). In this condition, however, the data from the present study (open circle in Fig. 2.6) indeed demonstrated a less externalized sound image with increasing amount of smoothing. Thus, the DRR metric does not seem to account for the conditions where the smoothing was applied to the direct part of the BRIR. Spectral modifications of the HRTF (i.e., the direct part of the BRIR) introduce deviations of the natural ILDs. The model proposed in the present study showed that the deviations of the natural ILDs could account for the internalization of the sound stimuli, suggesting that binaural cues are crucial for correct sound externalization. However, binaural cues may not be important for robust distance perception, which may be mainly driven by monaural cues, as represented in the DRR.

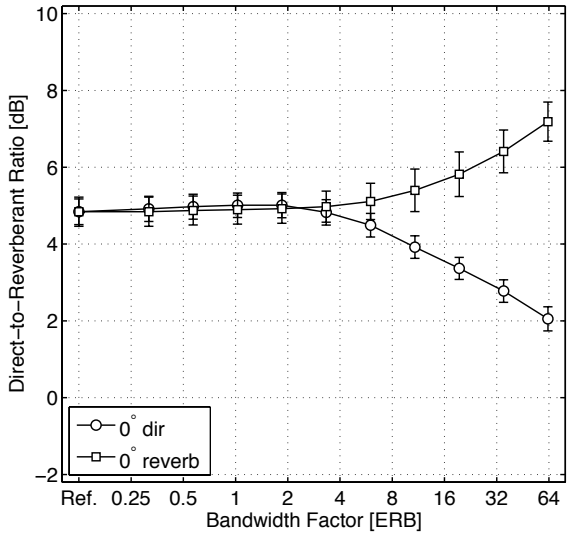


Figure 2.8: The mean DRR of the left ear of the listeners for frontal conditions plotted as a function of the bandwidth factor. The error bars are one standard error of the mean.

To illustrate the effect of some of the processing stages of the model on simulated externalization, Fig. 2.9 shows the results obtained with two modifications

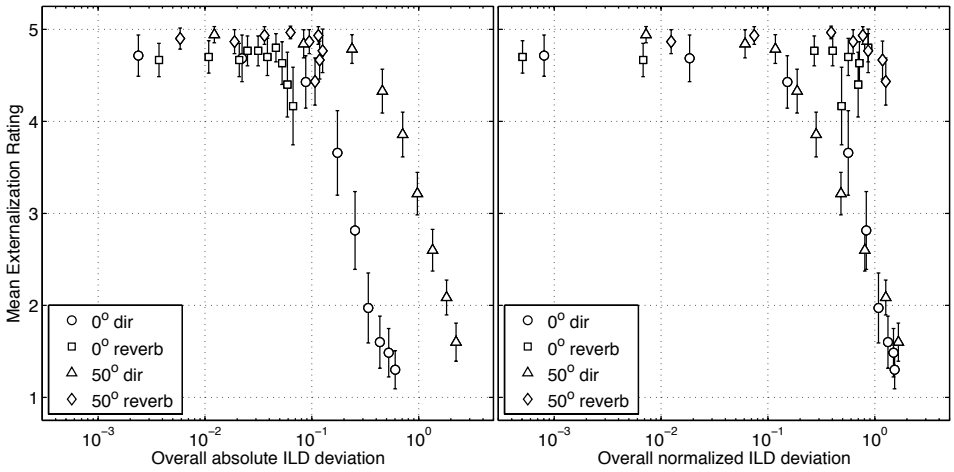


Figure 2.9: The perceived sound source location as function of the overall normalized ILD deviation for the two model modifications. In the left panel the absolute deviations from the reference ILDs were considered and in the right panel shows the situation when echo suppression was omitted. The overall normalized ILD deviation (x axis) is show on a logarithm scale to display the data point linearly. The error bars are one standard error of the mean.

in the model. The left panel shows simulations where the absolute deviations from the reference ILDs were considered, instead of the relative deviations as defined in Eq. (2.13). For the stimulus conditions 0° dir (circles) and 0° reverb (squares), the model no longer produces simulated externalization values along a single function (as in Fig. 2.7 obtained with the original model). Particularly for the spectral smoothing conditions that led to medium and small externalization ratings in the data, the simulated values for the 0° dir and the 50° dir conditions deviated strongly from each other, in contrast to the findings in the data and the original model. Thus, with this modified model version, the reduced externalization for the conditions at 0°, compared to the conditions at 50°, cannot be accounted for. The right panel of Fig. 2.9 illustrates simulations when no echo suppression was considered in the model. Also for this modification, the model fails to account for the data. The simulations for the different conditions no longer follow a single function. Without echo suppression, the spectral details in the early reflections and late reverberation are given too much weight and,

thus, the model underestimates the amount of externalization in the conditions 0° reverb and 50° reverb, i.e., predicts more internalized values than observed in the data (and in the original model).

The presented model can only be considered as a first step towards a computational model of perceptual externalization. The assumptions in the model regarding echo suppression are qualitative and conceptual and also the processes describing frequency selectivity, excitation pattern calculation, ILD comparison, and integration across frequency have been pragmatic choices. While the individual steps have been inspired by existing auditory modeling work and psychoacoustic data, alternative implementations might be more powerful. The concept of matching the BRIR to existing templates was used to estimate the amount of externalization (vs internalization). However, the template- matching method only considered ILDs and, thus, did neither take effects of ITDs nor stimulus coloration into account, which have been argued to also contribute to externalization perception (e.g., Hartmann and Wittenberg, 1996). Furthermore, dynamic changes of the convolution of BRIRs and stimuli, e.g., resulting from hearing-aid compression, are not accounted for in the model. Nevertheless, despite its simplicity, the proposed model might provide a valuable basis for further investigations of the auditory processes underlying externalization perception. Such investigations could include experimental conditions with stimuli different from those considered here, presented from different source positions and in various acoustic environments.

Changes of the BRIRs, e.g., resulting from static hearing-aid processing, will affect the pattern of ILDs (and ITDs) and might thus have an effect on the listeners' ability to externalize sound sources. Spectral modifications applied on the direct part of the BRIR will mainly affect static ILDs. For example, the position of the microphones in a hearing aid has an effect on the spectral details

of the HRTF and thereby influences the static ILDs (e.g., in the case of behind-the-ear hearing aids). Thus, if the modified static ILDs due to a change of the micro- phone positions are not compensated for, this should affect the perceived externalization of the sounds. Less critical should be situations were fluctuating ILDs are affected, e.g., due to spectral modifications applied on the reverberant part of the BRIR (but not the direct part).

2.5 Summary and conclusion

In the present study, the effect of manipulating the spectral details in the binaural transfer function on perceived externalization in a reverberant room was investigated. The spectral details were reduced in either the direct or reverberant part of the BRIRs by smoothing the magnitude responses with band-pass filters. For various filter bandwidths, the modified BRIRs were convolved with broadband noise and listeners were asked, while keeping their head still, to judge the perceived position of the noise when virtualized over headphones. The data showed that reductions of the spectral details in the direct part of the BRIR in most experimental conditions had an effect on externalization. This is different from the findings obtained with corresponding spectral manipulations in an anechoic environment where no effect of spectral smearing on externalization was found. Reductions in the spectral details of the reverberant part of the BRIR did hardly affect externalization. A simple computational model was presented in an attempt to account for the data from the present study, obtained for the two source positions (0° and 50° azimuth) and for all tested modifications of the BRIR. The simulations suggested that perceived externalization can be estimated based on the deviations of the ILDs for the given modified signal from those for the (unmodified) reference signal, after

some stages of auditory preprocessing (including an echo-suppression mechanism). The results from the present study might be valuable for the further investigation of the auditory processes involved in externalization perception in complex acoustic environments.

Acknowledgments

This project has been carried out in connection to the Centre for Applied Hearing Research (CAHR) supported by Widex, Oticon, GN Resound and the Technical University of Denmark. Thanks to Jesper Udesen (GN Resound A/S) who performed the BRIR recordings. We also thank the Associate Editor, Frederick Gallun, and the two reviewers for their helpful and constructive feedback on an earlier version of this paper.

3

Effects of hearing-aid dynamic range compression on spatial perception in an everyday reverberant environment[a]

Abstract This study investigated the effects of fast-acting hearing-aid compression on normal-hearing and hearing-impaired listeners' spatial perception in a reverberant environment. Three compression schemes - independent compression at each ear, linked compression between the two ears, and "spatially ideal" compression operating solely on the dry source signal - were considered using virtualized speech and noise bursts. Listeners indicated the location and extent of their perceived sound images on the horizontal plane graphically on a touch screen. A linear amplification scheme was considered as the reference condition. The results showed that both independent and linked compression resulted in more diffuse and broader sound images as well as internalization and image splits, whereby more image splits were reported for the noise bursts than for speech. Only the spatially ideal compression provided the listeners with a spatial percept similar to that obtained with linear processing. The same general pattern was observed for

[a] This chapter represents a manuscript submitted to The Journal of the Acoustical Society of America.

both listener groups. An analysis of the interaural coherence and direct-to-reverberant ratio suggested that the spatial distortions associated with independent and linked compression resulted from enhanced reverberant energy. Thus, modifications of the relation between the direct and the reverberant sound should be avoided in amplification strategies that attempt to preserve the natural sound scene while restoring loudness cues.

3.1 Introduction

Loudness recruitment is a typical consequence of sensorineural hearing loss (Fowler, 1936; Moore, 2004; Steinberg and Gardner, 1937). To compensate for recruitment and thereby restore the normal dynamic range of audibility, multi-band fast-acting dynamic range compression (DRC) algorithms for hearing aids have been developed (Allen, 1996; Villchur, 1973). DRC algorithms amplify soft sounds and provide progressively less amplification to sounds whose level exceeds a defined compression threshold. In anechoic acoustic conditions, it has been shown that DRC systems that operate independently in the left and the right ear can lead to a distorted spatial perception of sounds, as reflected by an impaired lateralization performance, an increased sensation of diffuseness as well as the perception of split sound images (Wiggins and Seeber, 2011; Wiggins and Seeber, 2012). However, other studies conducted in anechoic acoustic conditions found only a minor effect of independent compression on sound localization (Keidser et al., 2006; Musa-Shufani et al., 2006). In the case of independent compression of the two ear signals, less amplification is typically provided to the ear that is closer to the sound source than to the ear that is farther away from the sound source, such that the intrinsic interaural

level differences (ILDs) given by the acoustic shadow of the listener's head are reduced. Wiggins and Seeber (2011) and Wiggins and Seeber (2012) ascribed the detrimental effects of independent compression on spatial perception to the mismatch between the reduced intrinsic ILDs and the unprocessed interaural time differences (ITDs) coming from a given sound source Brown et al. (see also 2016)(see also Brown et al., 2016).

With the aim of preserving the naturally occurring ILDs, state-of-the-art bilaterally fitted hearing aids share the measured sound intensity information in one hearing aid with that in the other hearing aid via a wireless link. The ear signal with the higher sound intensity in a given acoustic sound source scenario is typically chosen as the one providing the input to the level-dependent gain function in both (left-ear and right-ear) DRC systems (Korhonen et al., 2015). For hearing-impaired listeners with a symmetrical hearing loss, this shared processing, often referred to as "synchronization" or "link", implies that the amplification provided by the two DRC systems is the same such that the intrinsic ILDs are preserved. For hearing-impaired listeners with an asymmetrical hearing loss with different prescribed DRC gain settings (i.e., gain levels in the linear region, compression thresholds and compression ratios) for the left and right ear, the synchronization of the provided input level to the gain functions does not necessarily lead to a preservation of the intrinsic ILDs.

It has been demonstrated that linked fast-acting DRC systems, as compared to independent DRC systems, can improve speech intelligibility in the presence of a spatially separated stationary noise interferer for normal-hearing listeners in anechoic conditions (Wiggins and Seeber, 2013). In reverberant conditions, linked fast-acting DRC systems have been shown to improve the ability of normal-hearing listeners to attend to a desired target in an auditory scene with spatially separated maskers as compared to independent compres-

sion (Schwartz and Shinn-Cunningham, 2013). However, the effects of both independent and linked compression on more fundamental measures of spatial perception (such as distance, localization and source width) in reverberant conditions have only received little attention. In particular, the effects of compression on the direct part of a sound as well as its early reflections and late reverberation in a given environment have not yet been examined.

Catic et al. (2013) demonstrated that modifications of the interaural cues provided by the reverberation inside an enclosed space degrade the listeners' ability to perceive natural sounds as "externalized", i.e. as compact and properly localized both in direction and distance (Hartmann and Wittenberg, 1996). In a given reverberant environment, correct localization of an acoustic source is, among other factors, based on the interaural coherence between the listeners' ear signals (Catic et al., 2015), which is determined by the interaction between the direct sound and the reverberant part of the sound.

The hypothesis of the present study was that both independent as well as linked compression schemes affect the interaural cues provided by the reverberation, e.g., the interaural coherence and, thus, impair the spatial perception of the sound scene in a reverberant environment. In contrast, a compression scheme where the DRC operates on the "dry" source before its interaction with the reverberant environment, i.e. a "spatially ideal" DRC, should preserve the relation between the direct sound and the interaural cues provided by the reverberation and thus lead to robust spatial perception. To test this hypothesis, the effects of (fast-acting) independent, linked and spatially ideal compression schemes on the spatial auditory perception in a reverberant environment were examined in a group of normal-hearing listeners and a group of sensorineural hearing-impaired listeners with a symmetrical hearing loss. Linear processing, i.e. level-independent amplification, was considered as a reference condition.

The sounds in the different conditions were virtualized over headphones in a standard listening room using individual binaural room impulse responses (BRIRs). Listeners indicated their spatial perception graphically to capture all relevant spatial attributes with respect to distance, azimuth localization, source width and the occurrence of split images. The deviations of the listeners' ratings in the different compression conditions from those in the reference condition were considered to reflect the amount of spatial distortion. Transient sounds as well as speech were used as test stimuli to investigate the effects of the compression schemes on both the direct sound and the reverberant part of the sound. To quantify the distortion of the spatial cues in the different conditions, the interaural coherence (IC) and the direct-to-reverberant energy ratio (DRR) of the ear signals were considered as objective metrics.

3.2 Methods

3.2.1 Listeners

Two groups of listeners participated in the present study. The normal-hearing group consisted of twelve listeners (eight males and four females) aged between 25 and 58 years. All had audiometric pure-tone thresholds below 20 dB hearing level at frequencies between 125 Hz and 8 kHz. The hearing-impaired group consisted of fourteen listeners (eleven males and three females), aged between 62 and 80 years. All had symmetrical sloping mild-to-moderately-severe high-frequency sensorineural hearing loss, with a maximum difference of 15 dB between their left and right ear. Figure 3.1 shows the average pure-tone thresholds for the hearing-impaired listeners. Only three of the fourteen hearing-impaired listeners used hearing-aids on a regular, daily basis. Two of the hearing-impaired listeners were excluded from further analysis since they

perceived sounds that were presented diotically via headphones to be externalized, i.e. the sound was perceived as originating from outside of the head. Diotic signals are known to be internalized, i.e. perceived to be inside the head, by normal-hearing listeners (e.g., Boyd et al., 2012; Catic et al., 2013). It was considered important in the present study, in terms of the reliability of the spatial perception data, that the recruited listeners consistently could differentiate between internalized and externalized sound images. All listeners signed an informed consent document and were reimbursed for their efforts.

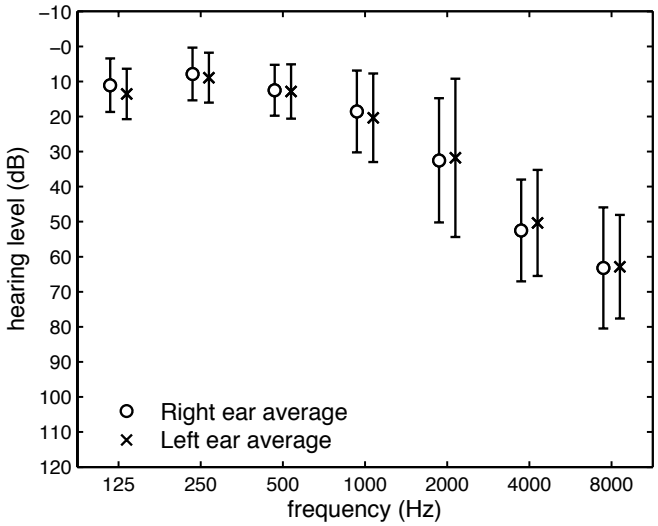


Figure 3.1: Audiometric pure-tone threshold averages for the right and left ear of the hearing-impaired listeners. The error bars represent one standard deviation of the thresholds.

3.2.2 Experimental setup and procedure

The experiments took place in a reverberant listening room designed in accordance with the IEC 268-13 (1985) standard. The room had a reverberation time T_{30} of approximately 500 ms, corresponding to a typical living room environment. Figure 3.2 shows the top view of the listening room and the experimental setup as placed in the room. The dimensions of the room were 752 cm × 474 cm

× 276 cm (L × W × H). Twelve Dynaudio BM6 loudspeakers were placed in a circular arrangement with a radius of 150 cm, distributed with equal spacing of 30 degrees on the circle. A chair with a headrest and a Dell s2240t touch screen in front of it were placed in the center of the loudspeaker ring. The listeners were seated on the chair with view direction on the loudspeaker placed at the azimuth angle of 0 degree. The chair was positioned at a distance of 400 cm from the wall on the left and 230 cm from the wall behind.

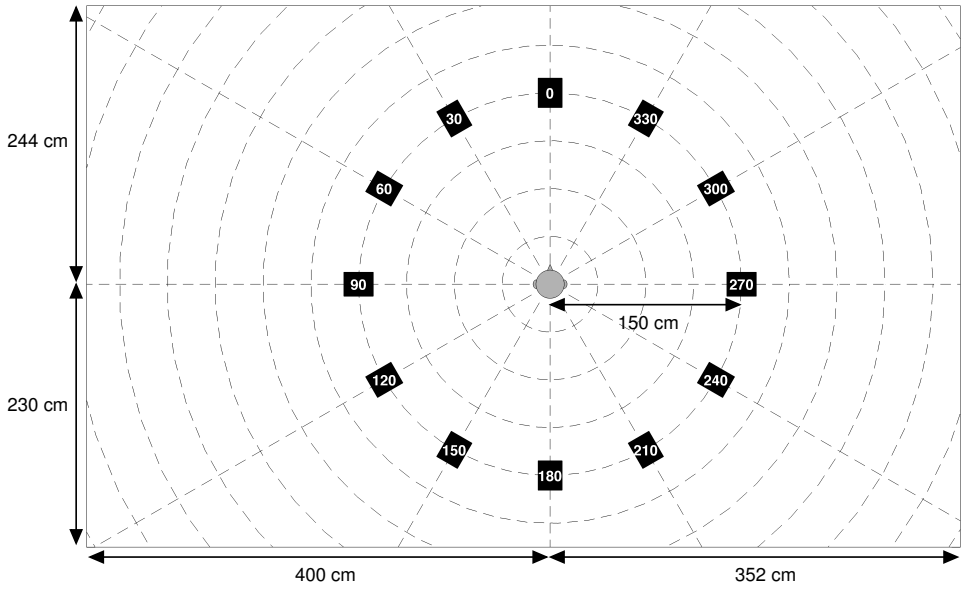


Figure 3.2: The top view of the experimental setup. The loudspeaker positions are indicated by the black squares. The grey circle in the center indicates the position of the chair, where the listener was seated. The listeners had a view direction on the loudspeaker placed at the 0° degree azimuth. The graphical representation was also shown on the touch screen, without the room dimensions shown in the figure.

The graphical representation of the room and the setup, as illustrated in Fig. 3.2 were also shown on the touch screen, without the information regarding the room dimensions. Besides the loudspeakers, a Fireface UCX sound card operating at a sampling frequency of 48000 Hz, two DPA high sensitivity microphones and a pair of HD850 Sennheiser headphones were used to record the individual BRIRs for the listeners (see 3.2.3). The BRIRs were measured

from the loudspeakers placed at the azimuth angles of 0, 30, 150 180, 240 and 300 degrees. The listeners were instructed to support the back of their head on the headrest while remaining still and to fixate on a marking located straight ahead (0°) both during the BRIR measurements and during the sound presentations. On the touch screen, the listeners were asked to place circles on the graphical representation as an indication of the perceived position and width of the sound image in the horizontal plane. By placing a finger on the touch screen, a small circle appeared on the screen with its center at the position of the finger. When moving the finger while still touching the screen, the circumference of the circle would follow the finger. When the desired size of the circle was reached, the finger was released from the screen. By touching the center of the circle and moving the finger while touching the screen, the position of the circle would follow along. By touching the circumference of the circle and moving the finger closer to or farther away from the center of the circle while touching the screen, the circle would decrease or increase in size, respectively. A double tap on the center of the circle would delete the circle. If the listeners perceived a split of any parts of the sound image, they were asked to place multiple circles reflecting the positions and widths of the split images. The listeners were instructed to ignore other perceptual attributes, such as sound coloration and loudness. Each stimulus was presented three times from each of the six loudspeaker positions. This was done for each of the test conditions: Linear processing, independent compression, linked compression and spatially ideal compression. No response feedback was provided to the listeners. The test conditions and active loudspeaker position were presented in random order within each run.

3.2.3 Spatialization

Individual BRIRs were measured to simulate the different conditions virtually over headphones. Individual BRIRs were used since it has been shown that the use of individual head-related transfer functions (HRTFs), the Fourier transformed head-related impulse responses, improve sound localization performance compared to non-individual HRTFs (e.g., Majdak et al., 2014), as a result of substantial cross-frequency differences between the individual listeners' HRTFs (Middlebrooks, 1999). Individual BRIRs were measured from the loudspeakers placed at the azimuth angles of 0, 30, 150 180, 240 and 300 degrees. The BRIR measurements were performed as described in Hassager et al. (2016). The microphones were placed at the ear-canal entrances and were securely attached with strips of medical tape. A maximum-length-sequence (MLS) of order 13, with 32 repetitions played individually from each of the loudspeakers, was used to obtain the impulse response, h_{brir}, representing the BRIR for the given loudspeaker. The headphones were placed on the listeners and corresponding headphone impulse responses, h_{hpir}, were obtained by playing the same MLS from the headphones. To compensate for the headphone coloration, the inverse impulse response, h_{hpir}^{inv}, was calculated in the time domain using the Moore-Penrose pseudoinverse. By convolving the room impulse responses, h_{brir}, with the inverse headphone impulse responses, h_{hpir}^{inv}, virtualization filters with the impulse responses, h_{virt}, were created. Stimuli convolved with h_{virt} and presented over the headphones produced the same auditory sensation in the ear-canal entrance as the stimuli presented by the loudspeaker from which the filter, h_{brir}, had been recorded. Hence, a compressor operating on an acoustic signal convolved with h_{brir} behaves as if it was implemented in a completely-in-canal hearing aid.

To validate the BRIRs, the stimuli were played in random order first from the

loudspeakers and then via the headphones filtered by the virtual filters h_{virt}. In this way, it could be tested if the same percept was obtained when using loudspeakers or headphones. By visual inspection, the graphical responses obtained with the headphone presentations were compared to the graphical responses obtained with the corresponding loudspeaker presentations. Apart from several front-back confusions (representing cone-of-confusion errors) in some of the listeners in the case of the headphone presentations, the graphical responses confirmed that all listeners had a very similar spatial perception in the two conditions. Generally, the response variability was found to be higher in the validation than in the actual experiment, especially for the elderly hearing-impaired listeners, which most likely is caused by the validation also served as training in evaluating the auditory perception on the graphical user interface.

3.2.4 Experimental conditions

Two types of stimuli were considered to investigate the effect of the different compression schemes on spatial perception. A 1.6-s long clean speech sentence from the Danish hearing in noise test corpus (Danish HINT; Nielsen and Dau, 2011), and 4 s of 20 noise bursts (transients), whereby each of the transients had a duration of 50 ms. Four conditions were tested: Independent compression, linked compression, spatially ideal compression as well as linear processing that served as a reference. The technical details of the DRC system will be described in 3.2.5. Figure 3.3 shows the block diagrams of the different conditions illustrating how the DRC systems were combined with the binaural impulse response that is represented by its left part, $h_{brir,l}$, and its right part, $h_{brir,r}$. In the independent compression scheme (top panel), the input signal, s_{in}, was first convolved with $h_{brir,l}$ and $h_{brir,r}$ and then passed through two DRC systems operating independently in each ear. In the linked compression scheme

(middle panel), after convolving with $h_{brir,l}$ and $h_{brir,r}$ as in the condition with the independent DRC systems, the signals were passed through a synchronized pair of DRC systems that, on a sample-by-sample basis in each of the seven frequency channels (Sect. II.E.), applied the lowest gain of the two level-dependent gain functions to both ears. In the spatially ideal compression scheme (bottom panel), the input signal, s_{in}, was first passed through a single DRC system and the output was then convolved with $h_{brir,l}$ and $h_{brir,r}$. The spatially ideal compression scheme thus consisted of a compression of the dry signal before the interaction with the room (i.e. the convolution with $h_{brir,l}$ and $h_{brir,r}$). In practice, since the dry signal is typically not available, such a system would require a deconvolution of $h_{brir,l}$ and $h_{brir,r}$ before compression, followed by a convolution with $h_{brir,l}$ and $h_{brir,r}$ to provide the listener with the spatial cues.

To create the signals for the condition with linear processing, the stimuli were convolved with $h_{brir,l}$ and $h_{brir,r}$. To compensate for the effect of the headphones, the outputs $s_{out,l}$ and $s_{out,r}$ in all conditions were convolved with $h^{inv}_{hpir,l}$ and $h^{i}_{hpir,r}nv$, respectively, i.e. the left and right parts of h^{inv}_{hpir}. For the normal-hearing listeners, the sound pressure level (SPL) at the ear closest to the sound source was 65 dB in all conditions. For the hearing-impaired listeners, the headphone outputs were amplified with the NAL-R(P) linear gain prescription (Byrne et al., 1990) according to the listener's individual audiometric pure-tone thresholds to ensure audible high-frequency content.

3.2.5 Dynamic range compression

To represent a modern multi-band hearing aid compressor, an octave-spaced seven-band DRC system was implemented. The incoming signal was windowed in time using a 512-sample long Hanning window (corresponding to a 10.7 ms time window at the sampling frequency of 48000 Hz) with a frame-to-frame step

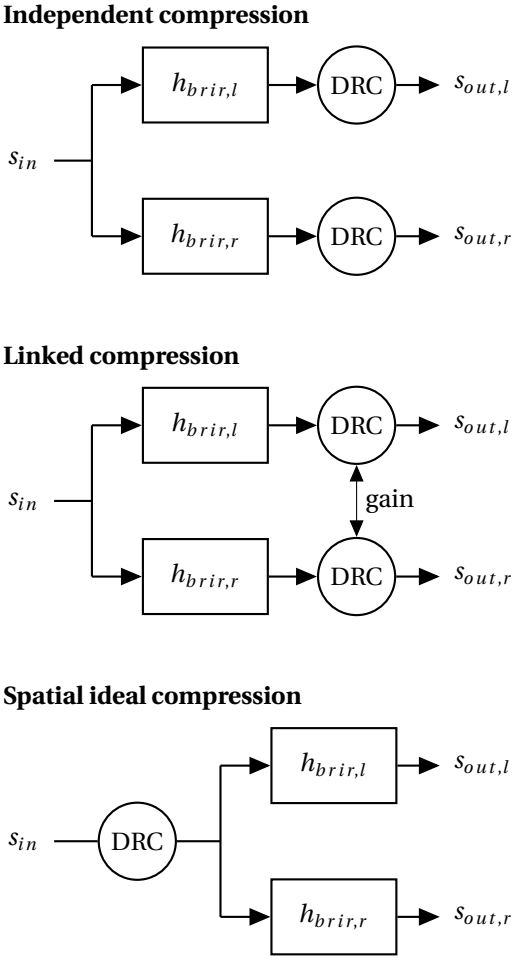


Figure 3.3: Block diagrams of the three compression conditions: Independent compression (top), linked compression (middle) and spatially ideal compression (bottom). For the independent and linked compression schemes, the dry signal, s_{in}, is convolved with the left and right BRIR, $h_{brir,l}$ and $h_{brir,r}$, respectively, and then processed by the DRC system. In the case of linked compression, the arrow between the two DRC systems indicates that the DRC gain is synchronized between the left and the right ear. In the case of spatially ideal compression, the dry signal is processed by DRC and then convolved with the left and right BRIR. The output in the left-ear and right-ear channels in the different schemes are denoted as, $s_{out,l}$ and $s_{out,r}$, respectively.

size of 128 samples. Each of the windowed segments was padded with 256 zeros in the beginning and with 256 zeros at the end and transformed to the spectral domain using a 1024-sample fast Fourier transform (FFT). The power values of the resulting frequency bins were combined to seven octave-wide frequency bands with center frequencies ranging from 125 Hz to 8 kHz. The power in

each band was smoothed using a peak detector (Eq. 8.1 in Kates (2008)). The attack and release time constants, measured according to IEC 60118-2 (1983), were 10 ms and 60 ms, respectively. The smoothed envelopes were converted to dB sound pressure level (SPL). A broken-stick gain function (with a linear gain below the compression threshold and a constant compression ratio above the threshold) was applied to the processed power envelopes. The resulting band-wise gains were then smoothed in the frequency domain using a piecewise cubic interpolation to avoid aliasing artifacts. The frequency smoothed gains were applied to the bins of the short-time Fourier transformed input stimulus, and an inverse FFT was applied to produce time segments of the compressed stimuli. These time segments were subsequently windowed with a tapered cosine window to avoid aliasing artifacts, and combined using an overlap-add method to provide the processed temporal waveform. The compression thresholds and compression ratios were calculated from NAL-NL2 prescription targets (Keidser et al., 2011) for audiometric pure-tone thresholds corresponding to the average audiometric pure-tone thresholds of the hearing-impaired listeners. The compression thresholds (CT) and compression ratios (CR), as derived from the NAL-NL2 prescription, are summarized in Table3.1 for the seven respective frequency bands. The simulated input level to the compressor operating closest to the sound source was 75 dB SPL.

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
CT (dB SPL)	31	36	40	32	34	31	9
CR	2.2:1	2.2:1	1.8:1	1.9:1	2.2:1	2.9:1	2.6:1

Table 3.1: The compression thresholds (CT) and compression ratios (CR) in the seven octave frequency bands.

3.2.6 Statistical analysis

The graphical responses provided a representation of the perceived sound image in the different conditions. To quantify deviations in the localization from the loudspeaker position across the different conditions, the root-mean-square (RMS) error of the Euclidean distance from the center of the circles to the loudspeakers was calculated. To reduce the confounding influence of front-back confusions as a result of the virtualization method, the responses placed in the opposite hemisphere (front versus rear) of the virtually playing loudspeaker were reflected across the interaural axis to the mirror symmetric position.

An analysis of variance (ANOVA) was run on four-factor mixed-effect models to assess the effects of hearing impairment, compression condition, stimulus and loudspeaker position on both the RMS error and the radius of the placed circles. The hearing status (normal hearing versus impaired hearing) was treated as a between-listener factor, and the compression condition, stimulus type (speech versus transients) and loudspeaker position were treated as within-listener factors. The radius data were square-root transformed to correct for heterogeneity of variance. Tukey's HSD corrected post hoc tests were conducted to test for main effects and interactions. A confidence level of 1 % was considered to be statistically significant.

3.2.7 Analysis of spatial cues

In order to quantify the effect of the different compression schemes on the spatial cues, ICs and DRRs were calculated. To visualize the effect of compression on the relation between the direct and reverberant energy, "temporal energy patterns" were calculated, i.e. the energy of the processed signal as a function of time.

Interaural coherence

The interaural coherence (IC) can be defined as the absolute maximum value of the normalized cross-correlation between the left and right ear output signals $s_{out,l}$ and $s_{out,r}$ occurring over an interval of $|\tau| \leq 1$ ms (e.g., Blauert and Lindemann, 1986; Hartmann et al., 2005):

$$\mathrm{IC} = \max_{\tau} \left| \frac{\sum_t s_{out,l}(t+\tau) s_{out,r}(t)}{\sqrt{\sum_t s_{out,l}^2(t) \sum_t s_{out,r}^2(t)}} \right|. \tag{3.1}$$

For each individual listener, the left- and right-ear output signals were filtered with an auditory inspired "peripheral" filterbank consisting of complex fourth-order gammatone filters with equivalent rectangular bandwidth spacing (Glasberg and Moore, 1990). The ICs were subsequently computed from the filtered output signals. The just-noticeable difference (JND) in IC is about 0.04 for an IC equal to 1 and increases to 0.4 for an IC equal to 0 (Gabriel and Colburn, 1981; Pollack and Trittipoe, 1959). The IC distribution was estimated by applying a Gaussian kernel-smoothing window with a width of 0.02 (half of the smallest JND) on the IC histograms.

Temporal energy patterns

Temporal energy patterns were obtained from the bandpass filtered output signals. The temporal envelope was calculated by convolving the absolute value of the complex outputs with a 20 ms rectangular window. The power of the windowed segments was calculated for the left- and right-ear segments and converted to dB SPL.

Direct-to-reverberant energy ratio

The direct part of the BRIRs, $h_{brir;dir}$, was defined as the first 2.5 ms of the impulse response, and the reverberant part, $h_{brir;reverb}$, was defined as the remaining subsequent samples of the BRIRs. The 2.5 ms transition point was chosen since the first reflection occurred immediately after this point in time. The reverberant part contained both the early reflections and the late reverberation. The gain values provided by the DRC systems in the processing of the left- and right-ear stimuli were extracted for each of the compression conditions. The impulse responses $h_{brir,l}$ and $h_{brir,r}$ (in Fig. 3.3) were replaced by their direct parts $h_{brir;dir,l}$ and $h_{brir;dir,r}$ and the extracted gain values were applied such that the outputs $s_{out;dir,l}$ and $s_{out;dir,r}$ only contained the effect of the compression on the direct part of the signal. Correspondingly, the outputs $s_{out;reverb,l}$ and $s_{out;reverb,r}$, representing the outputs that contained the effect of the compression on the reverberant part of the signal, were obtained by replacing the impulse responses $h_{brir,l}$ and $h_{brir,r}$ with their reverberant parts $h_{brir;reverb,l}$ and $h_{brir;reverb,r}$. Besides the effect of the compression on the direct and reverberant part of the signal, the extracted gain values were applied on the time aligned dry signal such that the outputs $s_{out;dry,l}$ and $s_{out;dry,r}$ only contained the effect of the compression on the dry signal.

To estimate the effect of the different compression schemes on the reverberant content of the processed stimuli, the direct-to-reverberant energy ratio (DRR) was calculated for the left- and right-ear signals for the four conditions. For the compression conditions, the DRR was calculated in the frequency domain:

$$DRR_k = 10 \cdot \log_{10} \left(\frac{\sum_f \frac{\left|S_{out;dir,k}(f)\right|^2}{\left|S_{out;dry,k}(f)\right|^2}}{\sum_f \frac{\left|S_{out;reverb,k}(f)\right|^2}{\left|S_{out;dry,k}(f)\right|^2}} \right), \tag{3.2}$$

where $S_{out;dir,k}(f)$, $S_{out;reverb,k}(f)$, and $S_{out;dry,k}(f)$ indicate the frequency-domain versions of the time signals $s_{out;dir,k}$, $s_{out;reverb,k}$, and $s_{out;dry,k}$ with respect to frequency f for $k \in [l;r]$ (left- and right-ear signal). For the linear processing condition, the DRR was calculated directly from the direct part ($h_{brir;dir,l}$ and $h_{brir;dir,r}$) and the reverberant part ($h_{brir;reverb,l}$ and $h_{brir;reverb,r}$) of the BRIR, respectively. DRRs were calculated for the frequency range from 100 Hz to 10 kHz.

3.3 Results

Experimental data

Figure 3.4 shows a graphical representation of all normal-hearing listeners' responses, including repetitions, obtained for speech virtualized from the loudspeaker positioned at 300° azimuth. The upper left panel represents the responses for the linear processing (the reference condition), whereas the responses obtained with independent compression, linked compression and spatially ideal compression are shown in the upper right, lower left and lower right panels, respectively. The responses of each individual listener in a given condition are indicated as transparent filled (colored and gray) circles with a center and size corresponding to the associated perceived sound image in the top-view perspective of the listening room (including the loudspeaker ring and the listening position in the center of the loudspeakers). Overlapping areas of circles obtained from different listeners are reflected by the increased cumulative intensity of the respective color code. To illustrate when a listener experienced a split in the sound image and, therefore, indicated more than one circle on the touch screen, only the circle the listener placed nearest to the loudspeaker (including positions obtained by front-back confusions) was

indicated in color whereas the remaining locations were indicated in gray.

In the reference condition (upper left panel in Fig. 3.4), apart from some front-back confusions (i.e. errors on the cone of confusion), the sound was perceived as coming from the loudspeaker position at 300° azimuth. In contrast, in the independent compression condition (upper right panel), the sound was generally perceived as being wider and, in some cases, as occurring closer to the listener than the loudspeaker or between the loudspeakers at 240° and 300° azimuth. One of the listeners even internalized the speech stimulus. In some of the listeners, the independent compression also led to split images as indicated by the gray circles. In the linked compression condition (lower left panel), the sound images were reported to be scattered around and located between the loudspeakers at 240° and 300° azimuth, similar as in the condition with independent compression. Likewise, the sound images were indicated to be of larger width and were commonly perceived to be closer to the listener and not at the position of the loudspeaker. As in the condition with independent compression, the linked compression led to image splits and internalization in some of the listeners. Most of the listeners reported verbally that the sound image was more diffuse in the conditions with independent and linked compression than in the reference condition. Furthermore, in the independent and linked compression conditions, some of the listeners reported that they perceived part of the reverberation as enhanced and being located at a different place than the "main sound" leading to split images. In the spatially ideal compression condition (lower right panel), the listeners perceived the sound image as being compact and located mainly at the loudspeakers at 240° and 300° azimuth. None of the listeners experienced image splits in this condition.

In summary, in the normal-hearing listeners, independent and linked compression provided similar results. In both conditions, the results differed sub-

stantially from the results obtained in the condition with linear processing. In contrast, in the condition with the spatially ideal compression, similar results were observed as in the condition with linear processing.

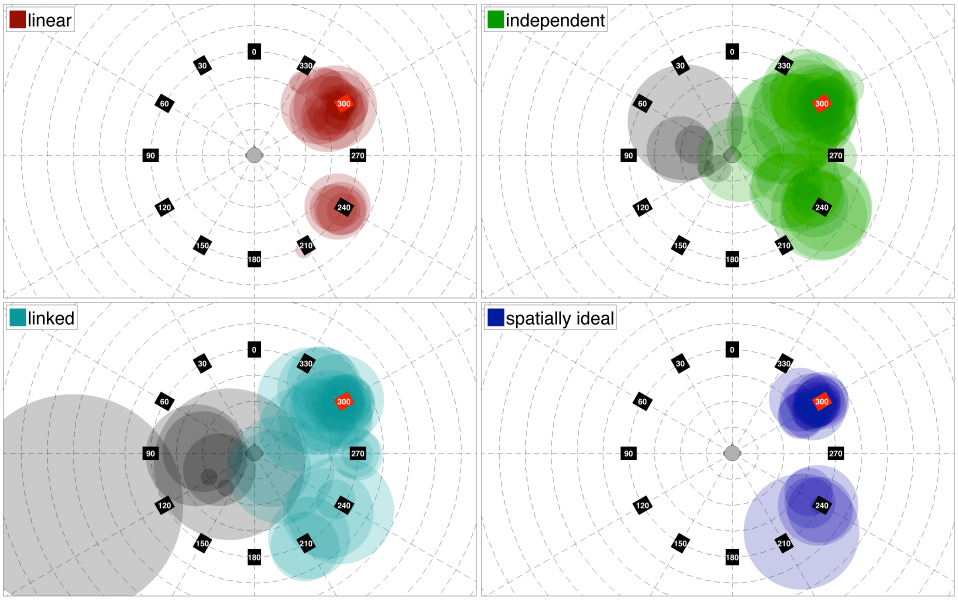


Figure 3.4: Graphical representations of the normal-hearing listeners' responses obtained with the speech stimulus virtually presented from the 300 degrees position in the listening room. The upper left panel shows the results for linear processing (reference condition). The results for independent, linked and ideal spatial compression are shown in the upper right, lower left, and lower right panels, respectively. The response of each individual listener is indicated as a transparent filled circle with a center and width corresponding to the associated perceived sound image. The main sound images are indicated by the different colors in the different conditions whereas split images are indicated in gray.

Figure 3.5 shows the corresponding results for the hearing-impaired listeners. The general pattern of results across conditions was similar to that found for the normal-hearing listeners (from Fig. 3.4). However, the hearing-impaired listeners typically perceived the sound images to be less compact than the normal-hearing listeners and the responses were characterized by a larger variability across listeners. For example, in the reference condition (upper left panel), the hearing-impaired listeners perceived the sound to be positioned at and around the loudspeakers at 240°, 270° and 300° azimuth. Some of the

listeners perceived the sound to occur between themselves and the loudspeakers while other listeners perceived the sound to be coming from beyond the loudspeakers. Both independent and linked compression (upper right and lower left panels of Fig. 3.5) caused wider and more spatially distributed sound images than in the reference condition whereas, in the case of ideally spatial compression (lower right panel), the sound was perceived to be more compact and similar to the sound presented in the reference condition. As observed for the normal-hearing listeners, some of the hearing-impaired listener also experienced split images in the independent and linked compression conditions.

Thus, overall, the hearing-impaired listeners typically showed a degraded spatial sensation relative to the normal-hearing listeners, i.e. they experienced more diffuse and spatially distributed sound images. However, the hearing-impaired listeners showed similar effects of independent, linked and spatially ideal compression on spatial perception as in the normal-hearing listeners.

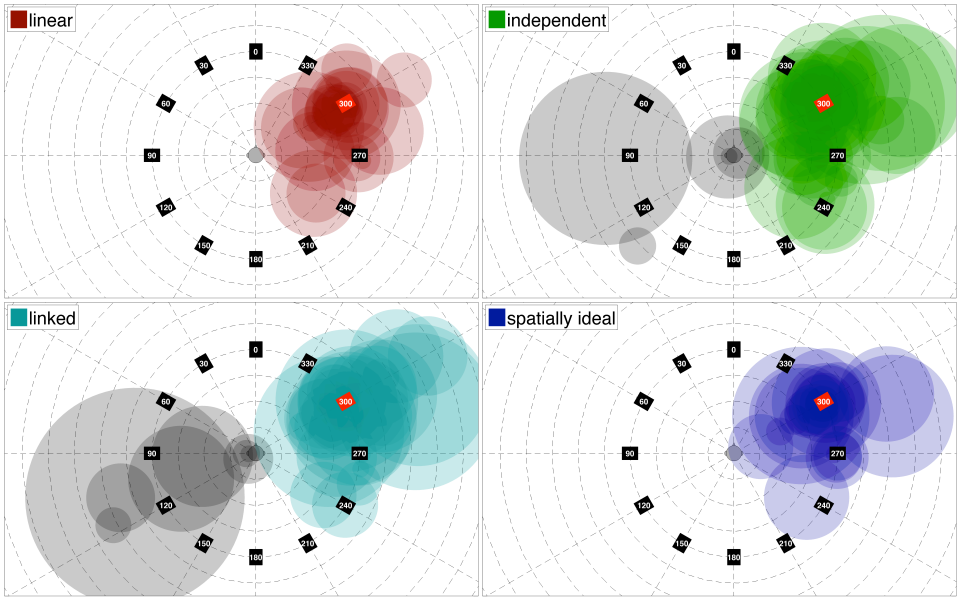


Figure 3.5: Same as Fig. 3.4, but for the hearing-impaired listeners.

The results obtained with the transients are shown in Fig. 3.6 for the normal-

hearing listeners and Fig. 3.7 for the hearing-impaired listeners. The general pattern of results across conditions was similar to that observed for the speech stimulus, i.e. (i) the listeners' spatial perception was largely affected by both independent and linked compression whereas spatially ideal compression provided similar results as in the reference conditions, and (ii) the hearing-impaired listeners indicated wider and more spatially distributed sound images than the normal-hearing listeners. However, in both listeners groups, the transients were generally perceived as more compact than speech, as indicated by the smaller circles in Fig. 3.6 and 3.7 compared to those in Figs. 3.4 and 3.5. Furthermore, more image splits were documented for the transients than for speech in the independent and linked compression conditions.

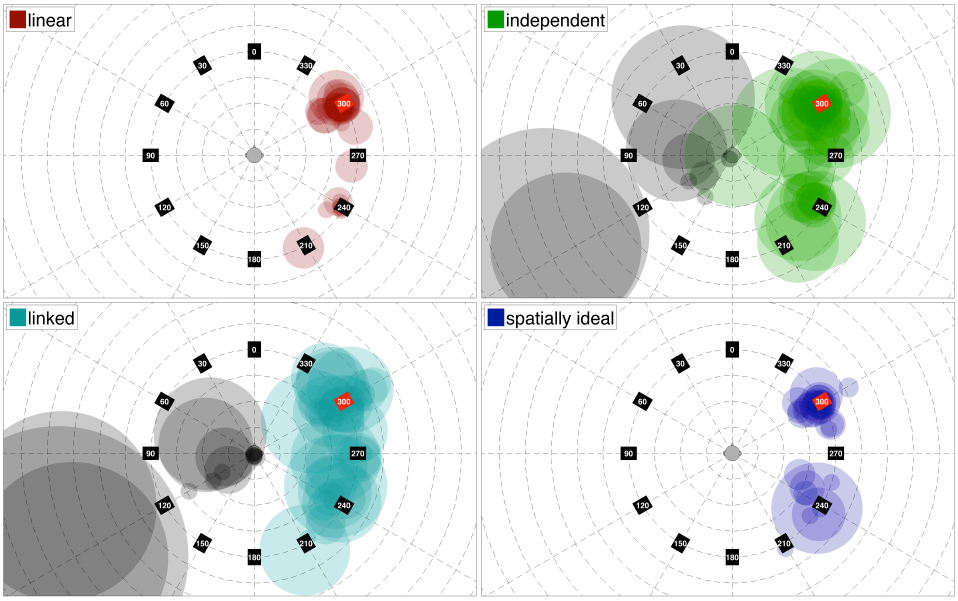


Figure 3.6: Same as Fig. 3.4, but for the normal-hearing listeners and transients.

The overall pattern of results obtained in the other five loudspeaker positions (0°, 30°, 150°, 180° and 240° azimuth) was similar to that observed for the loudspeaker positioned at 300° azimuth (Figs. 3.4-3.7). For the radius of the placed circles, indicating the perceived width of the sound image, the ANOVA

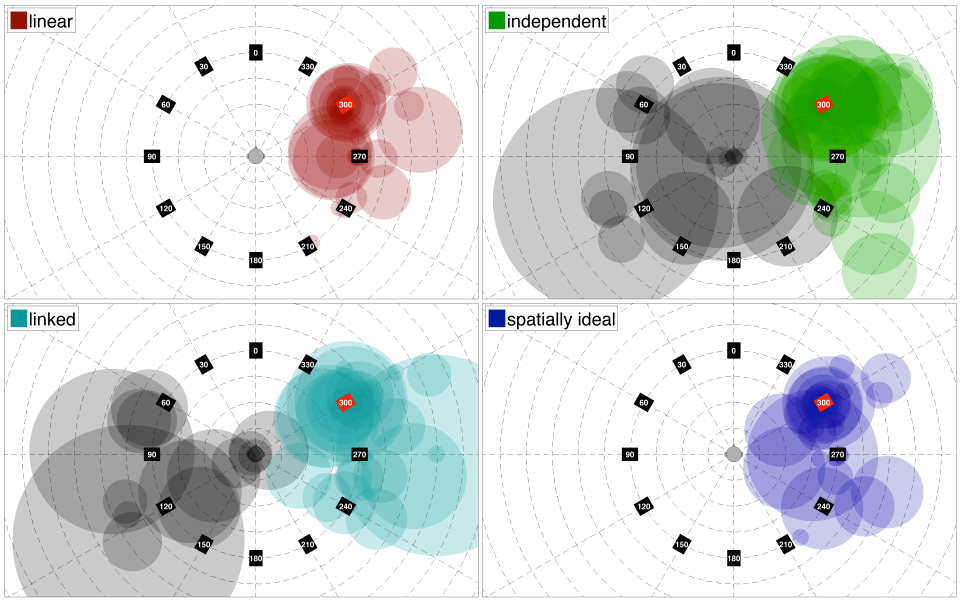


Figure 3.7: Same as Fig. 3.4, but for the hearing-impaired listeners and transients.

revealed an effect of compression condition [$F(3,66) = 61.54$, $p \ll 0.001$] and stimulus [$F(1,22) = 13.48$, $p = 0.001$] and loudspeaker position [$F(5,110) = 3.97$, $p \ll 0.001$]. Post-hoc comparisons confirmed that the listeners reported wider sound widths in the independent and the linked compression conditions than in the linear processing and spatially ideal compression conditions [$p \ll 0.001$]. No differences between the independent and the linked compression conditions [$p = 0.88$], and between the linear processing and spatially ideal compression conditions [$p = 0.11$] were found. Furthermore, post-hoc comparisons revealed that the indicated perceived sound width was similar for all combinations of loudspeaker positions, except between the loudspeakers positioned at 180° azimuth and 300° azimuth [$p = 0.004$]. The post-hoc estimated radius was higher for the speech than for the transients. For the RMS error, the ANOVA showed an effect of hearing status [$F(1,22) = 7.07$, $p = 0.01$], compression condition [$F(3,69) = 7.52$, $p \ll 0.001$] and loudspeaker position [$F(5,115) = 3.92$, $p = 0.003$]. Post-hoc comparisons confirmed that the RMS error was higher in the independent

compression and linked compression conditions than in the linear processing and spatially ideal compression conditions [$p \ll 0.001$]. No differences between the independent and the linked compression conditions [$p = 0.86$], and between the linear processing and spatially ideal compression conditions [$p = 0.99$] were found. The post-hoc estimated RMS error was higher for the hearing-impaired listeners than for the normal-hearing listeners. Furthermore, post-hoc comparisons revealed that the estimated RMS error was higher for the lateral loudspeaker positions than for the loudspeaker positioned at 0° azimuth. For the reported image splits, no differences between the independent and the linked compression conditions [$p = 0.91$] was found in a mixed-effects logistic regression analysis. However, the regression analysis confirmed that there was a higher proportion of reported image splits in the trials with the transients than in the trials with the speech [$p = 0.001$]. A significantly lower proportion of front-back confusions was obtained in the linear processing and spatially ideal compression conditions than in the independent and linked compression conditions [$p < 0.05$] according to a mixed-effects logistic regression analysis. The proportion of front-back confusions in the different conditions was 23.6 % in the case of linear processing, 23.9 % for the spatially ideal compression, 30.3 % for independent compression and 28.6 % for linked compression, respectively.

3.3.1 Analysis of spatial cues

Figure 3.8 shows the IC distributions for linear processing and the three compression conditions for the speech (upper panel) and the transients (lower panel) virtualized from the frontal loudspeaker. For simplicity, only the results at the output of the gammatone filter tuned to 2000 Hz are shown, but many other frequency channels show similar characteristics. The red, green, light blue and dark blue curves represent the IC distributions for linear processing,

independent compression, linked compression and spatially ideal compression, respectively. For both stimuli, the IC distributions for linear processing and spatially ideal compression are similar to each other, and the distributions for independent and linked compression are similar to each other. The distributions obtained with linear processing and spatially ideal compression show their maxima at interaural correlations of about 0.92, both for the speech and the transients. In contrast, the maxima of the distributions for the independent and linked compression conditions are shifted towards lower values of about 0.87 in the case of speech stimulation and between 0.66 and 0.77 for the transients. The computation of the IC based on the temporal envelope instead of the temporal waveform revealed the same pattern of results across the four processing conditions. Thus, in the conditions with independent and linked compression, the interaural correlation of the stimuli was substantially decreased due to the compression-induced changes to the temporal envelope on each ear.

Figure 3.9 shows temporal energy patterns for the linear processing and the three compression conditions for the speech stimulus (upper panel) and the transient stimulus (lower panel) virtualized from the frontal loudspeaker. The energy patterns were computed from the stimulus presented to the right ear of one of the listeners. Again, for illustration, only the output of the gammatone filter tuned to 2000 Hz is shown. The red, green, light blue and dark blue functions represent the results for linear processing, independent compression, linked compression and spatially ideal compression, respectively. For dry stimuli, the effect of compression is reflected by the difference between the patterns obtained with spatially ideal compression versus linear processing. For the transient stimulus (bottom panel), the effect of compression is small due to the short duration of the transients relative to the time constants of the DRC system, while for the speech stimulus (upper panel) the effect of

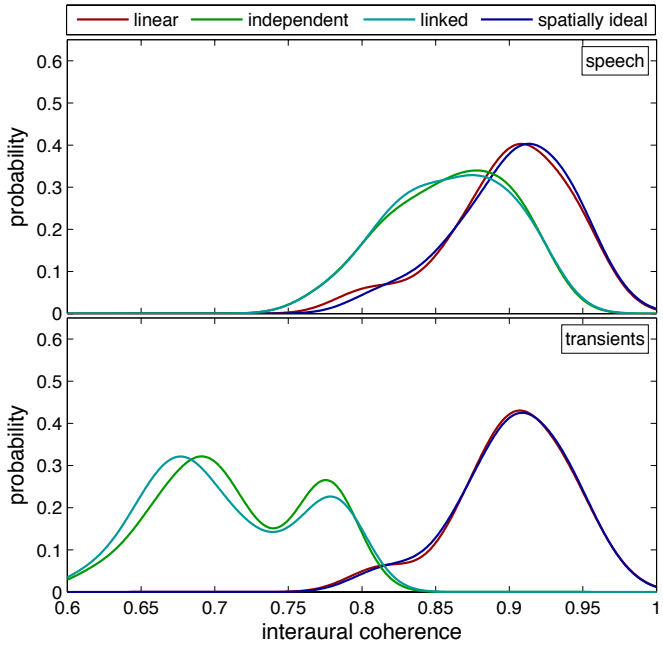


Figure 3.8: IC distributions of the ears signals, pooled across all listeners, at the output of the gammatone filter tuned to 2000 Hz. Results are shown for the speech (upper panel) and the transients (lower panel) virtualized from the frontal loudspeaker position. The red, green, light blue and dark blue functions represent the IC distribution for linear processing, independent compression, linked compression and spatially ideal compression, respectively.

compression is more prominent as revealed by the reduced modulation depth in the temporal pattern. For reverberant stimuli, the effect of compression is reflected by the difference between the patterns obtained with independent and linked compression versus the pattern obtained with linear processing. For the transients (bottom panel), the reverberant decay rate is clearly reduced in the independent and linked compression conditions relative to the linear processing condition. The same can be observed for the speech (upper panel) at time instances where reverberation is dominating, e.g. at 0.38 s, 0.55 s and 1.7 s. This indicates that these compression schemes increase the amount of reverberant energy relative to the direct sound energy. This is also reflected in the direct-to-reverberant ratios which amounts to 6.1 dB in the case of linear processing as well as spatially ideal compression (for this loudspeaker position).

In contrast, the direct-to-reverberant ratio reduces to 4.2 dB for the speech stimulus and 0.2 dB for the transients both in the condition with independent and linked compression. This behavior is consistent with the different amounts of IC reduction observed in Fig. 3.8 for the two stimulus types. The reduced decay rate in the case of independent/linked compression is more prominent for the transients than for the speech stimulus since the effect of reverberation is partly "masked" by the ongoing speech stimulus.

Thus, both objective metrics (IC distributions and temporal energy patterns) show similar results for independent and linked compression. Furthermore, both metrics also show similar results for linear processing and ideal spatial compression. These patterns are consistent with the main observations in the behavioral data from Figs. 3.4-3.7.

3.4 Discussion

The spatial cue analysis showed that both independent and linked compression increased the energy of the reverberant sound relative to the direct sound. The reason for this is that the segments of the stimuli that are dominated by reverberation often exhibit a lower signal level and are therefore amplified more strongly than the stimulus segments that are dominated by the direct sound. Compared to the speech stimulus, the transients contained more segments that were dominated by reverberation. The enhanced reverberant energy was reflected by a similar decrease of the DRR as well as a similar change of the IC statistics for independent and linked compression relative to linear processing, particularly for the transient stimulus. Thus, in the reverberant environment considered in the present study, compression modifies the relation between the direct and reverberant sound energy which, in turn, affects the IC that underlie

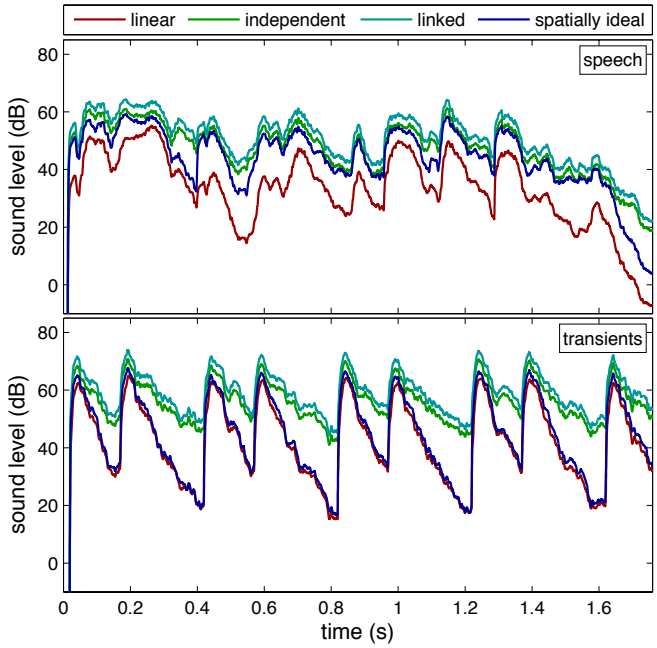


Figure 3.9: Temporal energy patterns of the speech stimulus (upper panel) and the transient stimulus (lower panel) virtualized from the frontal loudspeaker position. Only the output of the signals processed by the gammatone filter at 2000 Hz is shown. The different colors represent the different processing conditions (red: linear processing; green: independent compression; light blue: linked compression; dark blue: spatially ideal compression. For better visualization of the trends, the functions have been displaced by 3 dB (spatially ideal compression), 6 dB (independent compression) and 9 dB (linked compression).

spatial perception. The decreased IC of the processed stimuli in the case of independent/linked compression was consistent with the higher proportion of image splits reported for the transients than for the speech stimulus and the perception of broader, more diffuse sound images as compared to linear processing. It has been demonstrated that listeners localize sound sources in reverberant environments by responding to the spatial cues carried by the direct sound and suppressing the spatial cues carried by the early reflections. This perceptual phenomena is termed “the precedence effect” (see Brown et al., 2015, , for a review). In the present study, the early reflections were most likely not enhanced sufficiently by the independent and linked compression to overcome the precedence effect and thereby affect the listeners’ perceived localization

of the stimuli, i.e., cause the image splits. Instead, the perceived split images might result from the enhancement of the late reverberation carrying spatial cues unrelated to the sound source. Thus, the results suggest that the energy ratio between the direct and the reverberation sound should ideally be preserved to provide the listener with undistorted cues for spatial perception.

The results are consistent with Blauert and Lindemann (1986) who demonstrated that a reduction in the IC results in both image splitting as well as a broadening of the sound image for normal-hearing listeners. However, in contrast to the findings of the present study, earlier studies (Whitmer et al., 2012; Whitmer et al., 2014) found that hearing-impaired listeners were relatively insensitive to changes in IC, as measured by perceived width when using stationary noise stimuli. The different results might have been caused by the differences in the stimuli used in the present study and the ones of Whitmer et al. (2012) and Whitmer et al. (2014). In the present study, the reduction of the IC by compression was caused by changes to the binaural temporal envelope whereas in Whitmer et al. (2012) and Whitmer et al. (2014) the change in IC was driven by changes in the binaural temporal fine structure, which is also the reason why the reported insensitivity was correlated with the ability to detect interaural phase differences (Whitmer et al., 2014). It has previously been shown that, in contrast to temporal fine structure sensitivity, the sensitivity to temporal envelope cues is similar in hearing-impaired listeners and normal-hearing listeners (e.g., Moore and Glasberg, 2001).

The increased amount of front-back confusions in the independent and linked compression conditions suggests that these compression schemes distorted the monaural spectral cues (e.g., Middlebrooks and Green, 1991) that listeners in combination with head movement cues (Brimijoin et al., 2013) normally use to resolve forward from rearward sources. Thus, both independent

and linked compression seem to make it more difficult for the listeners to distinguish between frontal and rearward sources.

In contrast to independent compression, linked compression is expected to restore the listener's natural spatial perception in anechoic environments due to the preservation of ILDs (Wiggins and Seeber, 2011; Wiggins and Seeber, 2012). However, no effect of preserving the intrinsic ILDs by linked compression, as compared to independent compression, was found in the reverberant condition considered in the present study. Thus, the beneficial effect of preserving the ILDs is not apparent in reverberation, which most likely is a result of the dominating effect of fast-acting compression reducing the rate of the reverberant decay and, thereby, reducing the IC. Nonetheless, linked fast-acting compression has, in reverberant conditions, been shown to partly restore the ability to attend to a desired target in an auditory scene with spatially separated maskers, in contrast to independent compression (Schwartz and Shinn-Cunningham, 2013). However, the performance obtained with linked compression did not reach the level obtained with linear processing, potentially as a result of the reduced IC due to this compression scheme. It is possible that, based on the results of the present study, spatially ideal compression would produce similar results as linear processing since the spatial cues would be preserved.

It has been demonstrated that listeners can adapt to artificially produced changes of the spatial cues responsible for correct sound source location (for a review, see Mendonça, 2014). This plasticity in spatial hearing has been demonstrated both in the horizontal and vertical plane for various manipulations of the localization cues. For example, by modifying the direction-dependent spectral shaping of the outer ear by inserting ear molds in both of the listener's ears (Hofman et al., 1998) or only in one of the ears (Van Wanrooij and Van Opstal, 2005), listeners can reacquire accurate sound localization performance within

a few weeks. It might be argued that such "remapping" processes also occur for other modifications of the acoustic cues, such as the ones considered in the present study. However, the signal-driven changes of the binaural cues considered here might be difficult to learn, since they affect the sound location, sound width and give rise to image splits. Although the performance of sound localization can be reacquired, the increased sound width and image splits originating from the altered reverberation will most likely be difficult to remap as these are signal dependent and dynamic due to the characteristics of the fast-acting compression schemes. Consistent with this reasoning, it has been shown that not all modifications can be remapped. An example of this is ear swapping (Hofman et al., 2002; Young, 1928), where adaptation to switched binaural stimuli was not found for periods as long as 30 weeks.

Only the spatially ideal compression scheme, operating on the dry signal, provided the listeners with a similar spatial percept as the linear processing scheme. The processing did not distort the listeners' spatial perception in terms of source localization, at least not in the conditions considered in the present study. However, spatially ideal compression requires a priori knowledge of the BRIRs which is not a feasible solution in realistic applications where the BRIR is unknown. Instead, a feasible approach could be to estimate the amount of reverberation in the stimulus, e.g. via an estimation of the DRR as a function of time, such that compression is only applied in moments where the DRR is above a certain criterion and otherwise switched off or reduced. Such a system might be particularly useful for hearing-instrument amplification strategies where the goal is to preserve the natural sound scene around the listener while still providing sufficient dynamic range compression restoring proper loudness cues.

In the present study, no ambient noise in the listening room was added to

the input of any of the processing conditions. Typical everyday environments are likely to include some level of background noise which could influence the results since background noise will reduce the valleys of the temporal envelope of the sound. Thus, in such a condition, less amplification would be provided by the compression in the segments of the stimuli that exhibit a lower signal level than in the corresponding quite situation, such that the reverberant portions of the stimulus would be enhanced less. Furthermore, the added background noise may perceptually mask some of the reverberation, decreasing the detrimental impact of compression on spatial perception. Hence, in everyday listening environments with ambient noise, the impact of compression on spatial perception might be less prominent than the effects reported in the present study.

3.5 Conclusions

This study investigated the effect of dynamic range compression in reverberant environments on spatial perception in normal-hearing and hearing-impaired listeners. The following was found:

(i) Both independent and linked fast-acting compression resulted in more diffuse and broader sound images, internalization and image splits relative to linear processing.

(ii) No differences in terms of the amount of spatial distortions were observed between the linked and independent compression conditions.

(iii) Spatially ideal compression provided the listeners with a spatial percept similar to that obtained with linear processing.

(iv) More image splits were reported for the noise bursts than for speech both

for independent and linked compression.

(v) The spatial resolution of the hearing-impaired listeners was generally lower than that of the normal-hearing listeners. However, the effects of the compression schemes on the listeners' spatial perception were similar for both groups.

(vi) The stimulus-dependent distortion due to the linked and independent compression was shown to be a result of a reduced interaural-cross correlation of the ear signals as a result of enhanced reverberant energy.

Overall, the results suggest that preserving the ILDs by linking the left- and right-ear compression is not sufficient to restore the listener's natural spatial perception in reverberant environments relative to linear processing. Since spatial distortions were introduced via an enhancement of reverberant energy, it would be beneficial to develop compressor schemes that minimize the distortion of the energy ratio between the direct and the reverberant sound.

Acknowledgments

This project was carried out in connection to the Centre for Applied Hearing Research (CAHR) supported by Widex, Oticon, GN ReSound and the Technical University of Denmark. We thank Ruksana Giurda and Pernille Holtegaard for their assistance with recruiting the listeners and collecting the data, and Jesper Udesen from GN ReSound for helpful comments and stimulating discussions. We also wish to thank two anonymous reviewers who helped us improve an earlier version of this manuscript.

4

Preserving spatial perception in rooms using direct-sound driven dynamic range compression[a]

Abstract Fast-acting hearing-aid compression systems typically distort the auditory cues involved in the spatial perception of sounds in rooms by enhancing low-level reverberant energy portions of the sound relative to the direct sound. The present study investigated the benefit of a direct-sound driven compression system that adaptively selects appropriate time constants to preserve the listener's spatial impression. Specifically, fast-acting compression was maintained for time-frequency units dominated by the direct sound while the processing of the compressor was linearized for time-frequency units dominated by reverberation. This compression scheme was evaluated with normal-hearing listeners who indicated their perceived location and distribution of sound images in the horizontal plane for virtualized speech. The experimental results confirmed that both independent compression at each ear and linked compression across ears resulted in broader, sometimes internalized, sound images as well as image split. In contrast, the

[a] This chapter represents a manuscript submitted to The Journal of the Acoustical Society of America.

linked direct-sound driven compression system provided the listeners with a spatial perception similar to that obtained with linear processing that served as the reference condition. The independent direct-sound driven compressor created a sense of movement of the sound between the two ears, suggesting that preserving the interaural level differences via linked compression is advantageous with the proposed direct-sound driven compression scheme.

4.1 Introduction

In everyday acoustic environments, the sound that reaches a listener's ears contains the direct sound stemming from the different sound sources as well as reflections from obstacles in the surroundings. Despite the mixture of direct sound, early and late reflections that are typically present in rooms, normal-hearing listeners commonly perceive sound sources as being compact and correctly localized in the space. It has been shown that both monaural cues, such as the sound pressure level at the ear drums and the direct-to-reverberant energy ratio (DRR; Zahorik, 2002), as well as binaural cues, such as interaural time and level differences (Catic et al., 2013; Hartmann and Wittenberg, 1996) contribute to reliable sound source localization in reverberant environments. Specifically, robust distance perception has been shown to be based on estimations of the DRR (Zahorik et al., 2005) whereas the sensation of externalized sound images, their azimuthal orientation in the space and their apparent source width, have been argued to be driven by binaural cues (e.g., Whitmer et al., 2012; Catic et al., 2015).

People with a sensorineural hearing impairment typically suffer from loudness recruitment, such that low-level sounds are not detectable while high-level

sounds produce a close-to-normal loudness perception (e.g., Fowler, 1936; Steinberg and Gardner, 1937). To compensate for this reduced dynamic range of levels in the hearing-impaired listeners, level-dependent amplification is commonly applied in hearing aids, such that low-level sounds are amplified more than higher-level sounds (Allen, 1996). This corresponds to a compressive processing of the input level range to the smaller dynamic range of levels that can be perceived by the listener. If such dynamic range compression in hearing aids operates independently in the left-ear and right-ear channels, less amplification is typically provided to the ear signal that is closer to a given sound source than to the ear signal that is farther away from the sound source, such that the intrinsic interaural level differences (ILDs) in the sound are reduced. To avoid this, state-of-the-art bilaterally fitted hearing aids share the measured sound intensity information across both devices via a wireless link (Korhonen et al., 2015). This shared processing is commonly referred to as "linked" compression, such that in the case of a symmetrical hearing loss the amplification provided by the two compressors is the same in both ears and, as a consequence, the intrinsic ILDs are preserved. This has been shown to improve the ability of normal-hearing listeners to attend to a desired target in an auditory scene with spatially separated maskers as compared to independent compression in reverberant conditions (Schwartz and Shinn-Cunningham, 2013).

However, as demonstrated in Hassager et al. (2017), both independent and linked fast-acting compression can strongly distort the spatial perception of sounds in reverberant acoustic environments. Both compression strategies can lead to an increased diffusiveness of the perceived sound and broader, sometimes internalized ("in the head"), sound images as well as sound image splits. Such spatial distortions were observed both in normal-hearing and hearing-impaired listeners when either linked or independent compression was applied

to the signals. It was demonstrated that the observed spatial distortions mainly resulted from the applied compression enhancing the level of the reflected sound relative to the level of the direct sound. It was concluded that compressive hearing-aid processing needs to maintain the energy ratio of the direct sound to the reflected sound in order to preserve the natural spatial cues in the acoustic scene.

Ideally, a dereverberation of the binaural room impulse responses (BRIRs) for each of the sound sources would be required to apply compression to the individual "dry" sound sources, followed by a convolution of the individual sound sources with the respective BRIRs to reintroduce and preserve the spatial characteristics of a given scene. It was shown by Hassager et al. (2017) that this approach provided the listener with an undistorted spatial perception. However, such idealized processing requires *a priori* knowledge of the dry source signals and the respective BRIRs. In addition, blind dereverberation strategies will inevitably introduce artifacts and distortions, which may limit the applicability in hearing-aid applications.

An alternative approach of preserving the natural spatial properties of a sound scene would be to effectively "linearize" the compressive processing by using time constants that are longer than the reverberation time. However, such processing would compromize the restoration of loudness perception obtainable by fast-acting compression (Strelcyk et al., 2012). In the present study, it was investigated whether fast-acting compression that preserves the listener's spatial impression could be achieved by adaptively adjusting the time constant of the compressor depending on a binary decision reflecting direct-sound activity. The idea was to maintain fast-acting compression in time-frequeny (T-F) units dominated by the direct sound while linearizing the processing via longer time constants of the compressor in T-F units dominated by reverberation.

If BRIR information was avaliable, the short-term estimate of the signal-to-reverberant energy ratio (SRR) could be used to identify T-F units that are dominated by the direct sound. Specificially, the BRIR could be split into its direct and reverberant parts (Zahorik, 2002). Then, the energy ratio of the direct sound (the source signal convolved with the direct part of the BRIR) to the reverberant sound (the source signal convolved with the reverberant part of the BRIR) could be used as a decision metric. For a given criterion (e.g., SRR $>$ 0 dB), an *a priori* classification could be performed to identify those T-F units that are dominated by the direct sound. However, this technique is not feasible in practical applications since the BRIRs are typically not available. Therefore, several "blind" algorithms have been developed to estimate the presence of reverberation in signals without *a priori* knowledge of the BRIRs. For example, the interaural coherence (IC) can be used to estimate the amount of reverberation in a signal since reverberation reduces the IC (e.g., Thiergart et al., 2012; Westermann et al., 2013; Zheng et al., 2015). Hazrati et al. (2013) developed an algorithm operating on monaural signals to identify direct-sound dominated T-F units by extracting a variance-based feature from the reverberant signal and comparing it to an adaptive threshold. The algorithm generates a binary T-F classification which was applied on the signal to suppress reverberation. The authors reported significant speech intelligibility improvements in cochlear implant users.

The present study focused on the spatial perception of speech presented in an everyday reverberant environment. The speech signals were processed by fast-acting hearing-aid compression with and without a binary classification stage to linearize the compressive processing of T-F units dominated by reverberation. Besides the classification using the short-term SRR based on *a priori* knowledge of the BRIRs, the blind classification method by Hazrati et al.

(2013) was tested both in independent and linked compression settings of the simulated hearing aid. The compression without the binary classification stage corresponded to conventional compression schemes described in the literature (e.g., Kates, 2008), whereas the compression with the binary classification stage represented the proposed direct-sound driven compression system. Linear processing, i.e. level-independent amplification, was considered as the reference condition. Only normal-hearing listeners were used in the present study. The main goal was to evaluate the feasibility of the approach motivated by the results from Hassager et al. (2017). To quantify the distortion of the spatial cues in the different conditions, the IC of the ear signals was considered as an objective metric.

4.2 Compression system

4.2.1 Algorithm overview

Figure 4.1 shows the block diagram of the algorithm. Both the independent and linked hearing-aid compression systems were based on short-time Fourier transformations (STFTs) and operated in seven octave-spaced frequency channels. In the STFT block, the left- and right-ear signals, sampled at a rate of 48000 Hz, were divided into overlapping frames of 10.7 ms duration (512 samples) with a shift of 2.6 ms (128 samples). Each frame was Hanning-windowed and zero padded to a length of 1024 samples and transformed into the frequency domain by applying a 1024-point discrete Fourier transform (DFT). In the left and right filterbank (FB), the power of the DFT bins was integrated into seven octave-wide frequency bands with center frequencies ranging from 128 Hz to 8 kHz. Similarly, the direct-sound classification stages (see Sect. 4.2.2) consisted of seven octave-wide frequency bands. The power and the corresponding binary

classification of the seven frequency bands were used to estimate the gain level (see Sec. 4.2.3). The estimated levels for the individual T-F units were converted to sound pressure level (SPL) in dB, and a broken-stick gain function (with a linear gain below the compression threshold and a constant compression ratio above the threshold) was applied. The compression thresholds and compression ratios were calculated from NAL-NL2 prescription targets (Keidser et al., 2011) for the N_3 audiogram corresponding to a flat and moderately sloping hearing-loss as defined in Bisgaard et al. (2010). The compression thresholds (CT) and compression ratios (CR) for the seven respective frequency bands are summarized in Table 4.1.

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
CT (dB SPL)	45	50	49	40	48	44	32
CR	3.4:1	3.2:1	2.3:1	2.7:1	3.6:1	3.8:1	4.0:1

Table 4.1: The compression thresholds (CT) and compression ratios (CR) in the seven octave frequency bands.

The simulated input level to the compressor operating closest to the sound source was 75 dB SPL. In the case of independent processing, the gain values for the individual T-F units were kept untouched. In the case of linked processing, the minima of the left and right gain values were taken as the gain values in both ears. In the inverse filterbank (IFB), the resulting gains were then interpolated in the frequency domain using a piecewise cubic interpolation to avoid aliasing artifacts and applied to the STFT bins of the input stimulus. Finally, an inverse DFT of the STFT coefficients was computed to produce time segments of the compressed stimuli. These time segments were subsequently windowed with a tapered cosine window to avoid aliasing artifacts, and combined using an overlap-add method to provide the processed temporal waveform presented to the left and right ear.

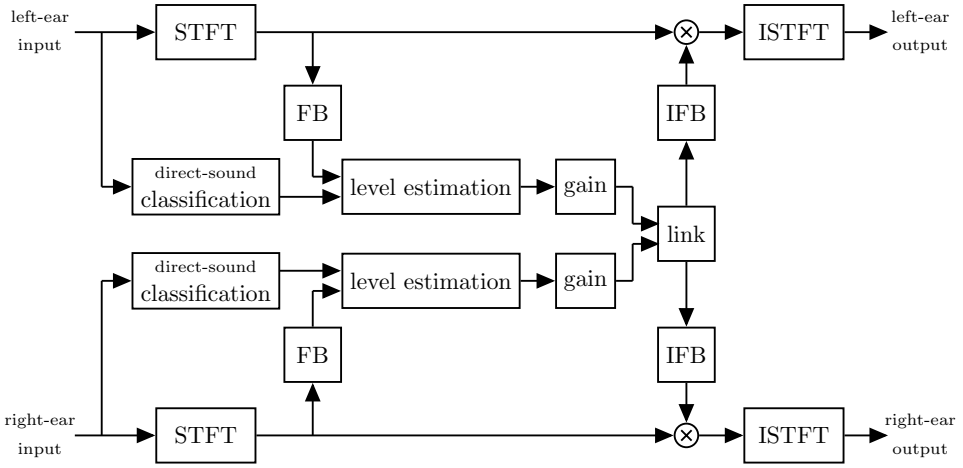


Figure 4.1: Block diagram of the proposed direct-sound driven compressor. First the left- and right-ear signals are windowed in time segments and transformed into the frequency domain by a short-time Fourier transforms (STFT). The frequency bins in each time window are combined into seven octave spaced frequency bands by the filterbank (FB), thereby creating T-F units. In the direct sound classification block a binary classification is performed whether T-F units are dominated by the direct sound. In the level estimation and gain blocks, the T-F units are smoothed across time with time constants determined by the classification and the gain values for T-F units are found. In the link block, the gain values are either kept untouched or the minima of the left and right gain values are used as the gain values in both ears. In the inverse filterbank (IFB), the gains were then interpolated in the frequency domain and applied to the STFT bins of the input stimulus. Finally, an inverse STFT (ISTFT) was computed and the resulting temporal waveform was presented to the left and right ear.

Figure 4.2 illustrates the different processing stages of the proposed system in relation to a conventional compression system. Panel (a) shows anechoic speech at the output of an octave-wide bandpass filter tuned to 1000 Hz. Panel (b) shows the corresponding output for reverberant speech, illustrating the impact of reverberation on the dry source signal. The blind classification of direct-sound signal components (blue) is shown in panel (c) together with a conventional compressor (light green) using a fixed compression mode with short time constants (fast-acting). The gain functions of the proposed direct-sound driven compressor (blue) and the conventional compressor (light green) are shown in panel (d). Panel (e) shows the waveform of the compressed reverber-

ant speech using the proposed direct-sound driven compressor, and panel (f) shows the waveform of the compressed reverberant speech processed with the conventional compressor. It is apparent that the conventional compressor amplifies the low-level portions of the sound and thereby enhances the reverberant components. In contrast, the proposed direct-sound driven compressor only acts on the direct signal components and applies a linear gain to the reverberant sound components.

4.2.2 Classification

The proposed direct-sound driven compressor requires a binary classification of individual T-F units into direct-sound and reverberant signal components. This classification was either based on the short-term SSR using *a priori* knowledge of the BRIRs or on the blind classification method described by Hazrati et al. (2013). The details of the two approaches are described below.

Signal-to-reverberant ratio classification

Assuming *a priori* knowledge about the BRIR, the short-term SRR was used as a decision metric to identify T-F units that are dominated by the direct sound. Specifically, the BRIRs were split into their direct and reverberant parts (Zahorik, 2002). The direct part was defined as the first 2.5 ms of the impulse response and the reverberant part was defined as the remaining subsequent samples of the BRIRs. The 2.5 ms transition point was chosen here since the first reflection occurred immediately after this point in time. The reverberant part contained both the early reflections and the late reverberation. The direct signal and the reverberant signal were obtained by convolving the dry speech (source signal) with the direct part and the reverberant part of the BRIR, respectively. The direct signal, D, and the reverberant signal, R, were segmented into overlapping

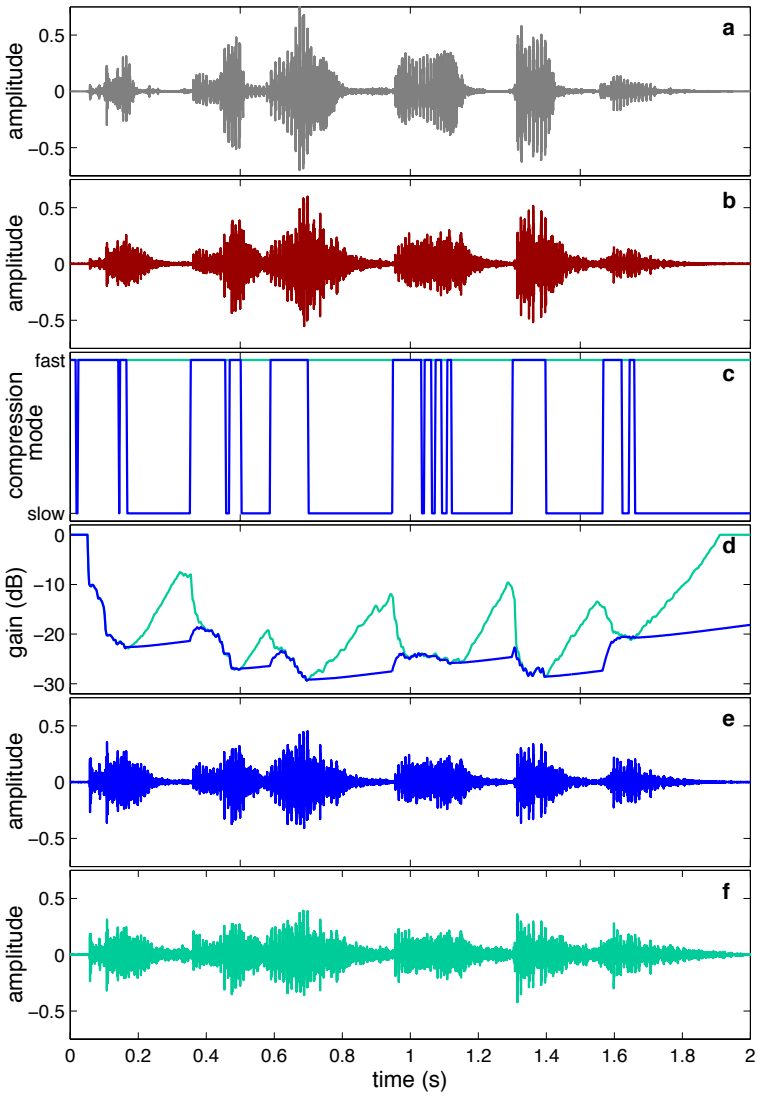


Figure 4.2: Example illustrating a bandpass filtered HINT sentence extracted at the center frequency of 1000 Hz. (a) Anechoic sentence, (b) reverberant sentence, (c) the blind binary classification (blue) where a value of one indicates direct-sound activity, (d) the corresponding gain function for conventional compression (light green) and the direct-sound driven compression (blue), (e) the reverberant sentence processed by the proposed direct-sound driven compression, and (f) the reverberant sentence processed by conventional compression.

frames and decomposed into seven octave-wide frequency channels using the same parameters as the compressor. The power was thereafter smoothed in

time (t) by recursive averaging as follows:

$$D_s(t,f) = \lambda \cdot D_s(t-1,f) + (1-\lambda) \cdot |D(t,f)|^2 \tag{4.1}$$

and

$$R_s(t,f) = \lambda \cdot R_s(t-1,f) + (1-\lambda) \cdot |R(t,f)|^2, \tag{4.2}$$

where D_s and R_s represent the smoothed versions, and ? represents the smoothing constant which was determined by $\lambda = exp(-T_{step}/\tau)$ for a time constant, τ, of 10 ms and a step time T_{step} of 2.6 ms. The SRR was calculated, as:

$$SRR(t,f) = 10 \cdot \log_{10}\left(\frac{D_s(t,f)}{R_s(t,f)}\right) \tag{4.3}$$

The classification of T-F units was performed by applying a local criterion to the short-term SSR, such that units greater than 0 dB were assigned a value of one and zero otherwise, creating a binary SRR classification:

$$C^{SRR}(t,f) = \begin{cases} 1, & SRR(t,f) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

Blind classification

The blind detection of direct-sound components without prior knowledge was performed using the method described by Hazrati et al. (2013). The reverberant signal was band-pass filtered by seven octave-spaced filters to match the frequency resolution of the compressor. The band-pass filtered signals were then segmented into overlapping frames denoted by S, and a variance-based feature labeled as F was calculated. Specifically, the variance of the signal raised to a power, α, was divided by the variance of the absolute value of the signal

converted to dB:

$$F(t,f) = 10 \cdot \log_{10}\left(\frac{\sigma^2(|S(t,f)|^{\alpha})}{\sigma^2(|S(t,f)|)}\right) \tag{4.5}$$

where the exponent, α, was set to 1.75. This variance-based feature was smoothed across time using a 3-point median filter. To obtain the binary classification, C^{Blind}, the variance-based feature, F, were compared to an adaptive threshold T (Otsu, 1979), producing a binary blind classification:

$$C^{Blind}(t,f) = \begin{cases} 1, & F(t,f) > T(t,f), \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

The adaptive threshold was based on a temporal context of 80 ms and was calculated for each frequency channel separately. The parameters used in the blind classification method were experimentally optimized to achieve the highest performance using the short-term SRR classification described in Section 4.2.2 as a reference condition.

Classification parameters

To quantify the performance of the blind classification, the hit rate minus the false-alarm rate (H-FA) was computed by comparing the detection of direct-sound components to the short-term SRR classification in the seven frequency channels. Clean speech sentences from the Danish hearing in noise test corpus (Danish HINT; Nielsen and Dau, 2011) and BRIRs corresponding to room A and B of the Surrey database (Hummersone et al., 2010) were used in the analysis. The results are shown in Table 4.2. The hit rate (H) was defined as the percentage of correctly classified direct-sound dominant T-F units, while the false-alarm rate (FA) was defined as the percentage of wrongly classified T-F units dominated by reverberation. The parameters of the blind classification were adjusted to

account for an SRR threshold criterion of 0 dB, as opposed to a local criterion of -8 dB which was used in the study by Hazrati et al. (2013). Apart from the lowest frequency band (at 125 Hz), where the FA is higher than at all other frequencies, the blind classification produces reasonably high performance values in terms of the H-FA metric.

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
H	94.0 %	94.2 %	96.0 %	93.0 %	79.6 %	86.2 %	79.2 %
FA	54.9%	28.6%	33.1%	12.7%	14.7%	24.3%	26.3%
H-FA	39.2 %	65.6%	62.9%	80.3%	64.9%	61.9%	52.9%

Table 4.2: The blind classification performance in terms of the H, HA, and H-FA for the seven octave frequency channels.

4.2.3 Level estimation

The levels of the T-F units were estimated by smoothing the power of the T-F units across time using recursive averaging:

$$X_s(t,f) = c \cdot X_s(t-1,f) + (1-c) \cdot |X(t,f)|^2, \tag{4.7}$$

where $|X|^2$ represents the power of the individual T-F units, X_s the smoothed power and c the smoothing constant. The smoothing constant, c, was updated according to the following criteria:

$$c = \begin{cases} c_{\text{attack}}^{\text{fast}} & \text{when} \quad |X(t,f)|^2 \geq X_s(t-1,f) \quad \text{and} \quad C(t,f)=1 \\ c_{\text{release}}^{\text{fast}} & \text{when} \quad |X(t,f)|^2 < X_s(t-1,f) \quad \text{and} \quad C(t,f)=1 \\ c_{\text{attack}}^{\text{slow}} & \text{when} \quad |X(t,f)|^2 \geq X_s(t-1,f) \quad \text{and} \quad C(t,f)=0 \\ c_{\text{release}}^{\text{slow}} & \text{when} \quad |X(t,f)|^2 < X_s(t-1,f) \quad \text{and} \quad C(t,f)=0 \end{cases} \tag{4.8}$$

with C being either C^{SRR} or C^{Blind} and the smoothing constants, $c_{\text{attack}}^{\text{fast}}$, $c_{\text{release}}^{\text{fast}}$, $c_{\text{attack}}^{\text{slow}}$, and $c_{\text{release}}^{\text{slow}}$, found according to IEC 60118-2 (1983), to be 10 ms, 60 ms,

2000 ms and 2000 ms, respectively. When C is being one the compression mode is fast-acting and when C is zero the compression mode is slow-acting.

4.3 Methods

4.3.1 Listeners

Eighteen normal-hearing listeners (10 males and 8 females), aged between 19 and 35 years, participated in the experiment. All had audiometric pure-tone thresholds below 20 dB hearing level at frequencies between 125 Hz and 8 kHz. All listeners signed an informed consent document and were reimbursed for their efforts.

4.3.2 Experimental setup and procedure

The experimental setup and procedure were similar to the ones described in Hassager et al. (2017). The experiments took place in a reverberant listening room designed in accordance with the IEC 268-13 (1985) standard. The room had a reverberation time T_30 of approximately 500 ms, corresponding to a typical living room environment. Figure 4.3 shows the top view of the listening room and the experimental setup as placed in the room. The dimensions of the room were 752 cm × 474 cm × 276 cm (L × W × H). Twelve Dynaudio BM6 loudspeakers were placed in a circular arrangement with a radius of 150 cm, distributed with equal spacing of 30 degrees on the circle. A chair with a headrest and a Dell s2240t touch screen in front of it were placed in the center of the loudspeaker ring. The listeners were seated on the chair with view direction to the loudspeaker placed at 0 degree azimuth. The chair was positioned at a distance of 400 cm from the wall on the left and 230 cm from the wall behind. The graphical representation of the room and the setup, as illustrated in Fig. 4.3

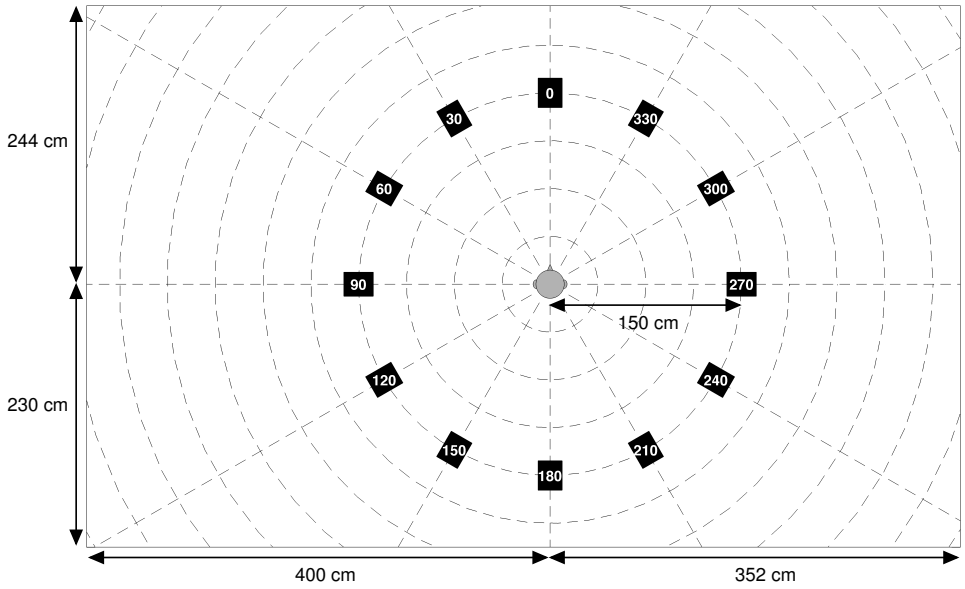


Figure 4.3: The top view of the experimental setup. The loudspeaker positions are indicated by the black squares. The grey circle in the center indicates the position of the chair, where the listener was seated. The listeners had a view direction on the loudspeaker placed at the 0°degree azimuth. The graphical representation was also shown on the touch screen, without the room dimensions shown in the figure.

were also shown on the touch screen, without the information regarding the room dimensions. Besides the loudspeakers, a Fireface UCX sound card operating at a sampling frequency of 48000 Hz, two DPA high sensitivity microphones and a pair of HD850 Sennheiser headphones were used to record the individual BRIRs for the listeners (see Section 4.3.3). The BRIRs were measured from the loudspeakers placed at the azimuth angles of 0 and 300 degrees. The listeners were instructed to support the back of their head on the headrest while remaining still and to fixate on a marking located straight ahead (0°) both during the BRIR measurements and during the sound presentations. On the touch screen, the listeners were asked to place circles on the graphical representation as an indication of the perceived position and width of the sound image in the horizontal plane. By placing a finger on the touch screen, a small circle appeared on the screen with its center at the position of the finger. When moving the finger

while still touching the screen, the circumference of the circle would follow the finger. When the desired size of the circle was reached, the finger was released from the screen. By touching the center of the circle and moving the finger while touching the screen, the position of the circle would follow along. By touching the circumference of the circle and moving the finger closer to or farther away from the center of the circle while touching the screen, the circle would decrease or increase in size, respectively. A double tap on the center of the circle would delete the circle. If the listeners perceived a split of any parts of the sound image, they were asked to place multiple circles reflecting the positions and widths of the split images. The listeners were instructed to ignore other perceptual attributes, such as sound coloration and loudness. Each stimulus was presented three times from each of the two loudspeaker positions. No response feedback was provided to the listeners. The test conditions, stimuli and loudspeaker position were presented in random order within each run.

4.3.3 Spatialization

Individual BRIRs were measured to simulate the different conditions virtually over headphones. Individual BRIRs were used since it has been shown that the use of individual head-related transfer functions (HRTFs), the Fourier transformed head-related impulse responses, improve sound localization performance compared to non-individual HRTFs (e.g., Majdak et al., 2014), as a result of substantial cross-frequency differences between the individual listeners' HRTFs (Middlebrooks, 1999). Individual BRIRs were measured from the loudspeakers placed at the azimuth angles of 0 and 300 degrees. The BRIR measurements were performed as described in Hassager et al. (2017). The microphones were placed at the ear-canal entrances and were securely attached with strips of medical tape. A maximum-length-sequence (MLS) of order 13,

with 32 repetitions played individually from each of the loudspeakers, was used to obtain the impulse response, h_{brir}, representing the BRIR for the given loudspeaker. The headphones were placed on the listeners and corresponding headphone impulse responses, h_{hpir}, were obtained by playing the same MLS from the headphones. To compensate for the headphone coloration, the inverse impulse response, h_{hpir}^{inv}, was calculated in the time domain using the Moore-Penrose pseudoinverse. By convolving the room impulse responses, h_{brir}, with the inverse headphone impulse responses, h_{hpir}^{inv}, virtualization filters with the impulse responses, h_{virt}, were created. Stimuli convolved with h_{virt} and presented over the headphones produced the same auditory sensation in the ear-canal entrance as the stimuli presented by the loudspeaker from which the filter, h_{brir}, had been recorded. Hence, a compressor operating on an acoustic signal convolved with h_{brir} behaves as if it was implemented in a completely-in-canal hearing aid.

To validate the BRIRs, the stimuli were played first from the loudspeakers and then via the headphones filtered by the virtual filters h_virt. In this way, it could be tested if the same percept was obtained when using loudspeakers or headphones. By visual inspection, the graphical responses obtained with the headphone presentations were compared to the graphical responses obtained with the corresponding loudspeaker presentations. This comparison confirmed that all listeners had a very similar spatial perception in the two conditions (see also Hassager et al., 2017).

4.3.4 Stimuli and processing conditions

Speech sentences from the Danish hearing in noise test corpus (Danish HINT; Nielsen and Dau, 2011) were used as stimuli. The clean speech signals were convolved with the listener's BRIRs, h_{brir}, and then processed by the fast-acting

compression conditions. As listed in Table 4.3, a set of six different compressor systems were tested: 1) Conventional independent compression that processed the binaural signals independently, 2) conventional linked compression that synchronizes the processing of the binaural signals, 3) independent compression with an SSR classification stage, 4) independent compression with a blind classification stage, 5) linked compression with an SSR classification stage, 6) linked compression with a blind classification stage. Linear processing was used as a reference condition. To compensate for the effect of the headphones, the left- and right-ear signals were afterwards convolved with the left and right parts of h_{hpir}^{inv}, respectively. The SPL of the stimulus at the ear closest to the sound source was 65 dB in all conditions.

method	binaural link	compression mode	estimator
independent	off	conventional	-
linked	on	conventional	-
independent SRR	off	direct-sound driven	short-term SRR
independent blind	off	direct-sound driven	blind
linked SRR	on	direct-sound driven	short-term SRR
linked blind	on	direct-sound driven	blind

Table 4.3: Overview of the different processing conditions involving compression.

4.3.5 Statistical analysis

The graphical responses provided a representation of the perceived sound image in the different conditions. To quantify deviations in the localization from the loudspeaker position across the different conditions, the root-mean-square (RMS) error of the Euclidean distance from the center of the circles to the loudspeakers was calculated. To reduce the confounding influence of front-back confusions as a result of the virtualization method, the responses placed in the opposite hemisphere (front versus rear) of the virtually playing loudspeaker

were reflected across the interaural axis to the mirror symmetric position.

An analysis of variance (ANOVA) was conducted on two mixed-effect models to evaluate whether the processing condition and loudspeaker position had an effect on the dependent variable, which was either the RMS error or the radius of the placed circles. In the mixed-effect models, listeners were treated as a random block effect nested within the repeated within-listener measures of repetition, processing condition and loudspeaker position. Repetitions were treated as a random effect, while the processing condition and loudspeaker position were treated as fixed effects. The radius data were square-root transformed and the RMS error was log transformed to correct for heterogeneity of variance. Tukey's HSD corrected post hoc tests were conducted to test for main effects and interactions. A confidence level of 5 % was considered to be statistically significant.

4.3.6 Analysis of spatial cues

In order to quantify the effect of the different compression schemes on the spatial cues, interaural coherences (ICs) were calculated. The IC can be defined as the absolute maximum value of the normalized cross-correlation between the left and right ear output signals $s_{out,l}$ and $s_{out,r}$ occurring over an interval of $|\tau| \leq 1$ ms (e.g., Blauert and Lindemann, 1986; Hartmann et al., 2005):

$$\mathrm{IC} = \max_{\tau} \left| \frac{\sum_t s_{out,l}(t+\tau) s_{out,r}(t)}{\sqrt{\sum_t s_{out,l}^2(t) \sum_t s_{out,r}^2(t)}} \right|. \tag{4.9}$$

For each individual listener, the left- and right-ear output signals were filtered with an auditory inspired "peripheral" filterbank consisting of complex fourth-order gammatone filters with equivalent rectangular bandwidth spacing (Glasberg and Moore, 1990). The ICs were subsequently computed from the filtered

output signals. The just-noticeable difference (JND) in IC is about 0.04 for an IC equal to 1 and increases to 0.4 for an IC equal to 0 (Gabriel and Colburn, 1981; Pollack and Trittipoe, 1959). The IC distribution was estimated by applying a Gaussian kernel-smoothing window with a width of 0.02 (half of the smallest JND) to the IC histograms.

4.4 Results

4.4.1 Experimental data

Figures 4.4 and 4.5 show graphical representations of the listeners' responses, including repetitions, virtualized from the loudspeaker positioned at 300°azimuth. The pattern of results obtained at the loudspeaker positioned at 0°azimuth was similar to that observed for the loudspeaker positioned at 300°, and are therefore not shown. In Fig. 4.4, the upper left panel represents the responses for the linear processing (reference) condition, whereas the responses obtained with conventional linked compression, direct-sound driven linked compression based on SRR classification and direct-sound driven linked compression based on blind classification are shown in the upper right, lower left and lower right panel, respectively. The responses of each individual listener in a given condition are indicated as transparent filled (colored and gray) circles with a center and size corresponding to the associated perceived sound image in the top-view perspective of the listening room (including the loudspeaker ring and the listening position in the center of the loudspeakers). Overlapping areas of circles obtained from different listeners are reflected by the increased cumulative intensity of the respective color code. To illustrate when a listener experienced a split in the sound image and, therefore, indicated more than one circle on the touch screen, only the circle the listener placed nearest to

the loudspeaker (including positions obtained by front-back confusions) was indicated in color whereas the remaining locations were indicated in gray.

In the reference condition (upper left panel in Fig. 4.4), the sound was perceived as coming from the loudspeaker position at 300°azimuth. In contrast, in the conventional linked compression condition (upper right panel), the sound was generally perceived as being wider and, in some cases, as occurring closer to the listener than the loudspeaker or between the loudspeakers at 240°and 300°azimuth. For some of the listeners, the conventional linked compression also led to split images as indicated by the gray circles. These results are consistent with the results obtained in Hassager et al. (2017). In the direct-sound driven linked compression conditions based on SRR classification (lower left panel) and blind classification (lower right panel), the listeners perceived the sound image as being compact and located mainly at the loudspeaker at 300°azimuth. None of the listeners experienced image splits with the direct-sound driven compression based on the SRR classification, while some image splits were experienced with the direct-sound driven compression using the blind classification. Nonetheless, in contrast to the conventional linked compression the experienced image splits were, concentrated mainly in the region around the loudspeaker that the sound was virtualized from.

Figure 4.5 shows the corresponding results for independent compression. The general pattern of results was similar to that found for linked compression (from Fig. 4.4). However, the responses for direct-sound driven independent compression based on the SRR classification (lower left panel) and the blind classification (lower right panel) contained considerably more image splits than the corresponding responses for conventional linked compression (upper right panel of Fig. 4.4). The reported image splits were in both direct-sound driven compression conditions placed around the position of the head. The

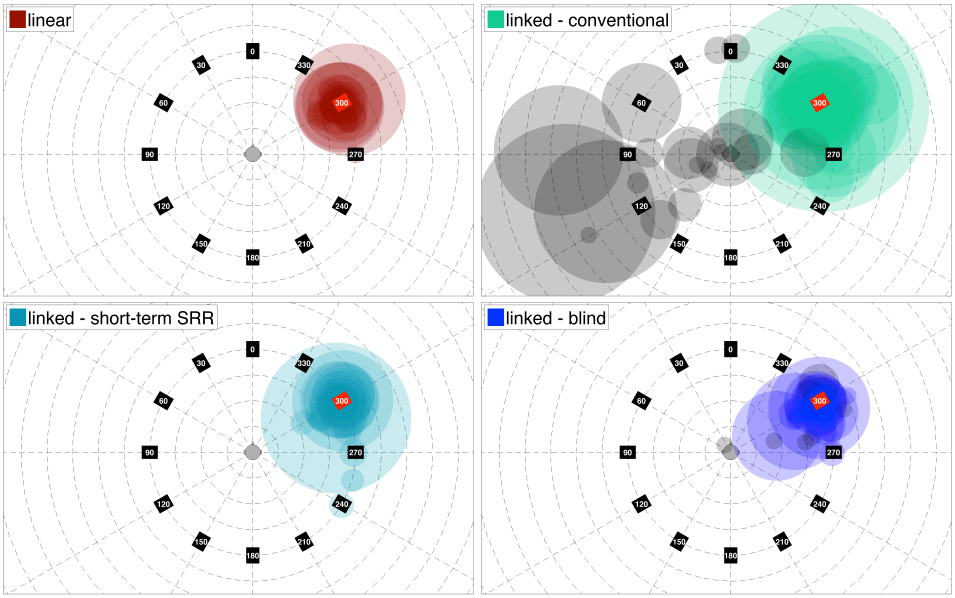


Figure 4.4: Graphical representations of the listeners' responses obtained with the speech virtually presented from the 300 degrees position in the listening room. The upper left panel shows the results for linear processing (reference condition). The results for conventional linked compression, direct-sound driven linked compression based on SRR classification and direct-sound driven linked compression based on blind classification are shown in the upper right, lower left, and lower right panel, respectively. The response of each individual listener is indicated as a transparent filled circle with a center and width corresponding to the associated perceived sound image. The main sound images are indicated by the different colors in the different conditions whereas split images are indicated in gray.

listeners who indicated image splits reported verbally that they perceived a sense of movement of the sound between the two ears. Nonetheless, the listeners generally perceived the main sound as being compact and located mainly at the loudspeaker at 300°azimuth in the two classification conditions.

For the radius of the placed circles, indicating the perceived width of the sound image, the ANOVA revealed an effect of processing condition [$F(6,42) = 65.62$, $p \ll 0.001$]. Post-hoc comparisons revealed that the listeners placed significantly [$p \ll 0.001$] larger radii in the two conventional compression conditions than in the linear processing and the direct-sound driven compression conditions. The radii in the conventional compression conditions were 34.6 cm and 37.0 cm for the linked and the independent compression condition,

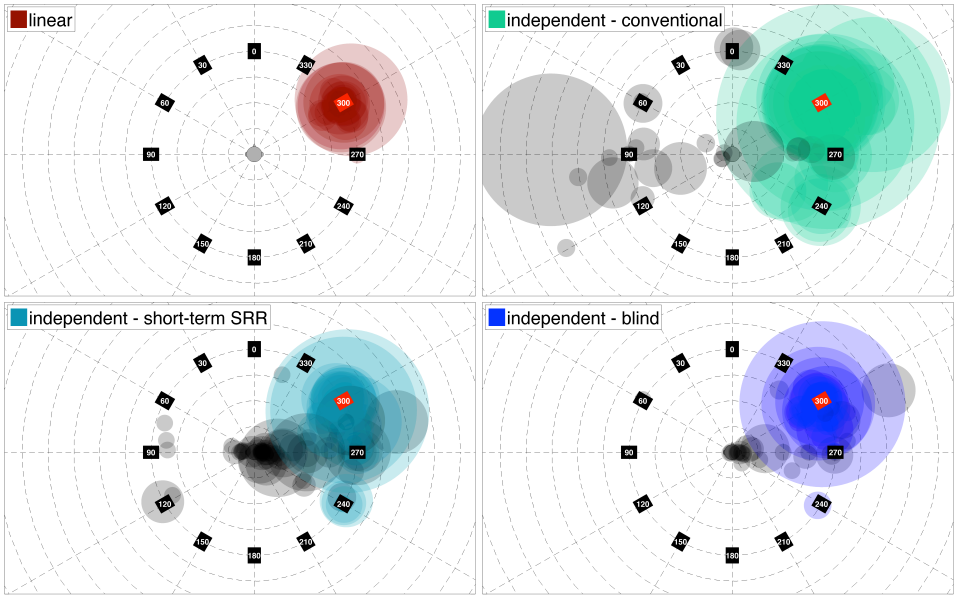


Figure 4.5: Same as Fig. 4.4, but for the independent compression conditions.

respectively, while the radii in the other conditions were between 3.3 cm and 9.1 cm. For the RMS error, the ANOVA showed an effect of the loudspeaker position [$F(1,17) = 6.82$, $p = 0.02$]. Post-hoc comparisons showed that the RMS error was slightly higher at the 300°azimuth loudspeaker position than at the frontal loudspeaker position. This is consistent with previous studies (e.g., Mills, 1958) demonstrating a higher localization acuity for frontal than for lateral positioned sound sources.

4.4.2 Analysis of spatial cues

Figure 4.6 shows the IC distributions for linear processing and the linked compression conditions (conventional, direct-sound driven with either SRR or blind classification) for the speech virtualized from the frontal loudspeaker. For simplicity, only the results at the output of the gammatone filter tuned to 1000 Hz are shown. The solid red, dashed light green, dashed light blue and dashed blue curves represent the IC distributions for linear processing, conventional linked

compression, direct-sound driven linked compression using SRR classification and direct-sound driven linked compression with blind classification, respectively. The IC distributions for the linear processing and the direct-sound driven linked compression with either short-term SRR or blind classification are similar to each other whereas the distribution for the conventional linked compression has its maximum at a much lower value. The distribution obtained with the linear processing shows a maximum at an IC of about 0.85. In contrast, the maxima of the distributions for the conventional linked compression condition are shifted towards a lower value of about 0.79. The same trends were observed for the independent compression conditions (not shown explicitly).

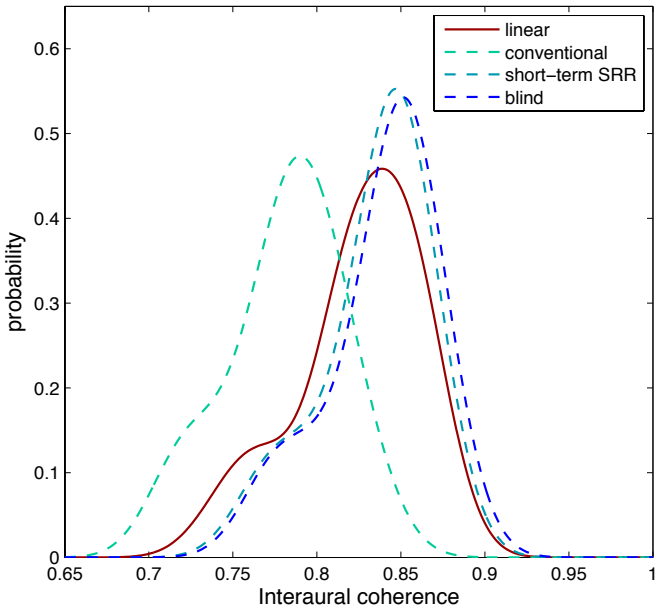


Figure 4.6: IC distributions of the ears signals, pooled across all listeners, at the output of the gammatone filter tuned to 1000 Hz. Results are shown for the speech virtualized from the frontal loudspeaker position. The solid red, dashed light green, dashed light blue and dashed blue curves represent the IC distributions for linear processing, conventional linked compression, direct-sound driven linked compression with SRR classification and direct-sound driven linked compression with blind classification, respectively.

4.5 Discussion

The present study compared conventional independent and linked fast-acting compression with direct-sound driven independent and linked compression. The classification stage in the direct-sound driven compressor was either based on the short-term SRR using *a priori* knowledge of the BRIRs or on the blind classification method by Hazrati et al. (2013). A spatial cue analysis showed that, in an everyday reverberant environment, conventional compression markedly reduced the IC of the stimulus between the ears relative to linear processing. The reason for this is that the segments of the stimuli dominated by reverberation often exhibit a lower signal level and are therefore amplified stronger by the compression scheme than the stimulus segments that are dominated by the direct sound (see also Hassager et al. (2017)). In contrast, the IC was largely maintained in the case of the direct-sound driven compression schemes relative to linear processing, implying that the energy ratio of the direct-sound to reverberation was preserved by linearizing the processing of the T-F units that are dominated by reverberation.

Consistent with the IC analysis, the direct-sound driven linked compression provided the listeners with a similar spatial percept as the linear processing scheme, while the conventional linked compression resulted in more diffuse and broader sound images as well as image splits. In the independent compression conditions, the general pattern of results was similar to that found for linked compression, except that the direct-sound driven compressor in the independent configuration led to the perception of a sound image that is moving between the two ears. Previous studies have demonstrated that, in anechoic conditions, independent compression can lead to such perceived lateral movements of the sound image (Wiggins and Seeber, 2011; Wiggins and Seeber, 2012), probably due to slow ILDs changes over time. Interestingly, according

to the verbal reports of most of the listeners in the present study, the sense of movement was not experienced in the case of the conventional independent compression condition, potentially because in this condition the increased amount of reverberation masks the occurrence of the ILD distortions stemming from the direct sound. Further investigation is needed to fully understand this phenomenon.

Instead of reconstructing the anechoic source signal, which would allow for the application of a "spatially ideal" compressor (Hassager et al., 2017), the proposed compression scheme utilizes short-term estimates of direct-sound components as a control signal to adaptively select the appropriate time constants, thus avoiding artifacts and signal distortions inevitably introduced by dereverberation. The results indicated that the proposed processing scheme does not introduce artifacts other than the enhanced reverberation due to miss classification of the reverberation. The performance analysis of the blind classification revealed that fast-acting compression, in fact, is applied to T-F units dominated by the direct sound, as reflected in the observed large hit rates, whereas the T-F units dominated by reverberation are classified less accurately as reverberation, as represented by the false alarm rates (see Table 4.2). Nevertheless, the behavioral results did not show significant spatial distortions in the two linked direct-sound driven compression schemes, indicating that the binary classification performance was reasonably high.

The experiments were conducted on normal-hearing listeners who have normal loudness perception and thus do not need level-dependent amplification, i.e., hearing-aid compression. Normal-hearing listeners were considered here since Hassager et al. (2017) demonstrated that hearing-aid compression affected hearing-impaired and normal-hearing listeners to a similar degree. While the hearing-impaired listeners showed generally less accurate localization ratings

than the normal-hearing listeners, the distortions resulting from conventional compression dominated the results and were similar in both listener groups. However, it will of course be crucial to perform corresponding experiments with the proposed direct-sound driven compression system with hearing-impaired listeners to further evaluate its significance and effectiveness. Furthermore, in the experiments considered in the present study, only a single sound source was used. In fact, localization accuracy plays an even greater role when dealing with multiple competing sources. With several sound sources, the impact of distorted spatial cues by conventional compression may limit the benefit that users are able to gain from current hearing aids. Thus, studying the influence of the direct-sound driven compression in multi-source scenario will be highly relevant. The blind estimation might be able to provide a robust estimation of direct-sound activity in multi-source scenarios because it does not require knowledge about the number or the spatial distribution of the sound sources.

There are certainly various ways to improve the detection of direct-sound components, e.g. by combining the monaural cues employed by Hazrati's method with binaural cues, such as the interaural coherence. Moreover, the adaptive threshold could be replaced by supervised learning approaches which were shown to enable accurate sound source localization in multi-source environments (May et al., 2011; May et al., 2015). The present study was not focused on providing an optimized "solution" and parameter set of a compression system. Instead, the main goal was to demonstrate the principal effect of a compression system that is controlled via the surrounding reverberation statistics, such that the spatial perception of the acoustic scene becomes less distorted by the effects of compression on the reverberant portions of the ears' input signals.

4.6 Conclusion

This study presented a direct-sound driven compression scheme that applied fast-acting compression in T-F units dominated by the direct sound while linearizing the processing via longer time constants in T-F units dominated by reverberation. It was demonstrated that such a direct-sound driven compression scheme can strongly reduce spatial distortions that are introduced by conventional compressors due to the enhancement of reverberant energy. It was found that linked direct-sound driven compression provided the listeners with a spatial percept similar to that obtained with linear processing. This was confirmed by the interaural coherence of the ear signals that was similar to that in the case of linear processing. A blind classification method was shown to provide accurate classification of direct-sound dominated T-F units. Its performance was similar to that obtained with a classification based on the short-term SRR using *a priori* knowledge of the BRIRs. In general, such a classification stage was found to be necessary and ensured that fast-acting compression was only applied to the speech signal. The T-F units dominated by reverberation were classified less accurately which, however, did not produce a detrimental effect on the spatial percetion ratings. In addtion, it was found that, in the conditions with independent direct-sound driven compression, a sense of movement of the sound between the two ears was observed. Thus, linking the left- and right-ear compression in combination with the proposed direct-sound driven compression scheme might be a successful strategy to provide a natural spatial perception while restoring normal loudness.

Acknowledgments

This project was carried out in connection to the Centre for Applied Hearing Research (CAHR) supported by Widex, Oticon, GN ReSound and the Technical University of Denmark.

5

Overall summary

In this thesis, psycho-acoustic measurements were conducted over headphones to characterize normal-hearing and hearing-impaired listener's externalization perception.

Chapter 2 investigated the effect of reducing the spectral details in the binaural transfer function on perceived externalization in an everyday reverberant listening room. The magnitude responses of either the direct or reverberant part of the BRIR were smoothed with band-pass filters. Broadband noise was convolved with the modified BRIRs for various filter bandwidths. Listeners with normal hearing indicated the perceived position of the noise virtualized over headphones while keeping their head still. The results showed that reductions of spectral details in the direct part of the BRIR had an effect on the perceived externalization, whereas reductions in the spectral details of the reverberant part of the BRIR did hardly affect externalization. In contrast, no effect of reductions of spectral details on externalization found in corresponding experiments carried out in anechoic conditions. A simple model based on the deviations of the ILDs for the modified stimulus from those for the (unmodified) reference stimulus, was able to account for the experimental data. Apart from some stages of auditory preprocessing, the model consisted of a conceptual echo-suppression mechanism that was found to be necessary in order to account for the experimental data.

Chapter 3 focused on the effect of dynamic range compression in a similar

everyday reverberant environment. These experiments were carried on both normal-hearing and hearing-impaired listeners. Linked and independent fast-acting compression were investigated and compared. It was found that both the linked and independent compression results in more diffuse and broad images in comparison with linear processing. These conventional compression schemes gave rise to both internalization and image splits. There was not observed any difference in the amount of spatial distortion and image splits resulting from the two compression schemes. As a further result it was demonstrated that compression on the dry source signal before the convolution of the binaural room impulse response provides the listeners with a spatial percept similar to that obtained with linear processing. It was found the spatial distortions were introduced via an enhancement of reverbernt energy, which indicated that preserving the ILDs by linking the left- and right-ear compression is not sufficient to provide the listener with a natural spatial perception in reverberant environments similar to the perception created by linear processing.

These findings from chapter 3 therefore lead naturally to the study presented in *Chapter 4*. Here a direct-sound driven compression scheme was proposed. The direct-sound driven compression scheme only applied fast-acting compression on components of the stimulus dominated by the direct sound while linearizing the processing via longer time constants for components dominated by reverberation. Effectively this scheme preserves the energy ratio of the direct sound to the reverberant sound. Both a classification method using a priori knowledge of the binaural room impulse response and a blind classification method by Hazrati et al. (2013) were used to determine the component that were dominated by direct sound and those dominated by reverberation. It was shown that the proposed direct-sound driven compression scheme could avoid spatial distortion introduced via enhancement of reverberant energy by conventional

compression while providing the listener with a compressed stimulus. It was found that the direct-sound driven compression in a linked configuration where the compression in the two ears were synchronized provided the listeners with a spatial perception similar to that obtained with linear processing. Whereas, the direct-sound driven compression in a independent configuration where the compression in the two ear operated independently of each other created a sense of movement of the sound between the two ears. This indicates that preserving the ILDs by linking the left- and right-ear compression is still required even though the direct-sound driven compression scheme is preserving the energy ratio of the direct sound to the reverberant sound.

The work presented in chapter 4 has concentrated on the influence of the direct sound driven compression on spatial perception. As Schwartz and Shinn-Cunningham (2013) have demonstrated linked fast-acting compression improves the ability of normal-hearing listeners to attend to a desired target with spatially separated maskers in reverberant environment as compared to independent compression. Therefore in future studies it would be interesting to investigate if the proposed direct-sound driven compression scheme in a linked configuration will improve speech intelligibility in the presence of a spatially separated noise masker compared to the conventional linked compression.

Bibliography

Allen, J. B. (1996). “Derecruitment by multiband compression in hearing aids”. In: *Psychoacoustics, Speech and Hearing Aids*. World Scientific. Chap. III, p. 372.

Baer, T. and B. C. Moore (1993). “Effects of spectral smearing on the intelligibility of sentences in noise”. In: *The Journal of the Acoustical Society of America* 94, pp. 1229–1241.

Begault, D. R., E. M. Wenzel, and M. R. Anderson (2001). “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source”. In: *Journal of the Audio Engineering Society* 49, pp. 904–916.

Bisgaard, N., M. S. Vlaming, and M. Dahlquist (2010). “Standard audiograms for the IEC 60118-15 measurement procedure”. In: *Trends in amplification* 14.2, pp. 113–120.

Blauert, J and W Lindemann (1986). “Spatial mapping of intracranial auditory events for various degrees of interaural coherence”. In: *The Journal of the Acoustical Society of America* 79.3, pp. 806–813.

Boyd, A. W., W. M. Whitmer, J. J. Soraghan, and M. A. Akeroyd (2012). “Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers”. In: *The Journal of the Acoustical Society of America* 131, EL268–EL274.

Braasch, J., J. Blauert, and T. Djelani (2003). "The precedence effect for noise bursts of different bandwidths. I. Psychoacoustical data". In: *Acoustical Science and Technology* 24, pp. 233–241.

Breebaart, J. and A. Kohlrausch (2001). "The perceptual (ir)relevance of HRTF magnitude and phase spectra". In: *110th AES Convention*. Preprint No. 5406, Amsterdam, The Netherlands.

Brimijoin, W. O., A. W. Boyd, and M. a. Akeroyd (2013). "The contribution of head movement to the externalization and internalization of sounds." In: *PloS one* 8.12, e83068.

Brown, A. D., G. C. Stecker, and D. J. Tollin (2015). "The precedence effect in sound localization". In: *Journal of the Association for Research in Otolaryngology* 16.1, pp. 1–28.

Brown, A. D., F. A. Rodriguez, C. D. Portnuff, M. J. Goupell, and D. J. Tollin (2016). "Time-varying distortions of binaural information by bilateral hearing aids: effects of nonlinear frequency compression". In: *Trends in Hearing* 20, pp. 1–15.

Byrne, D., A. Parkinson, and P. Newall (1990). "Hearing Aid Gain and Frequency Response Requirements for the Severely/Profoundly Hearing Impaired". In: *Ear and Hearing* 11.1, pp. 40–49.

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: *The Journal of the Acoustical Society of America* 134, pp. 1232–1241.

Catic, J., S. Santurette, and T. Dau (2015). "The role of reverberation-related binaural cues in the externalization of speech." In: *The Journal of the Acoustical Society of America* 138, p. 1154.

Cherry, E. C. (1953). “Some experiments on the recognition of speech, with one and with two ears”. In: *The Journal of the acoustical society of America* 25.5, pp. 975–979.

Fowler, E. P. (1936). “A method for the early detection of otosclerosis: A study of sounds well above threshold”. In: *Archives of Otolaryngology* 24.6, pp. 731–741.

Gabriel, K. J. and S. H. Colburn (1981). “Interaural correlation discrimination: I. Bandwidth and level dependence”. In: *The Journal of the Acoustical Society of America* 69.5, pp. 1394–1401.

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

Goode, R. L. R., M. Killion, K. Nakamura, and S. Nishihara (1994). “New knowledge about the function of the human middle ear: development of an improved analog model”. In: *Otology & Neurotology* 15, pp. 145–154.

Haas, H. (1951). “Über den Einfluβ eines Einfachechos auf die Hörsamkeit von Sprache”. In: *Acta Acustica united with Acustica* 1.2, pp. 49–58.

Hartmann, W. and A. Wittenberg (1996). “On the externalization of sound images”. In: *The Journal of the Acoustical Society of America* 99.6, pp. 3678–3688.

Hartmann, W. M. (1983). “Localization of sound in rooms”. In: *The Journal of the Acoustical Society of America* 74.5, pp. 1380–1391.

Hartmann, W. M., B. Rakerd, and A. Koller (2005). “Binaural coherence in rooms”. In: *Acta acustica united with acustica* 91.3, pp. 451–462.

Hassager, H. G., F. Gran, and T. Dau (2016). “The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment”. In: *The Journal of the Acoustical Society of America* 139.5, pp. 2992–3000.

Hassager, H. G., A. Wiinberg, and T. Dau (2017). "Effects of hearing-aid dynamic range compression on spatial perception in an everyday reverberant environment". In: *Submitted to The Journal of the Acoustical Society of America*.

Hazrati, O., J. Lee, and P. C. Loizou (2013). "Blind binary masking for reverberation suppression in cochlear implants". In: *The Journal of the Acoustical Society of America* 133.3, pp. 1607–1614.

Hofman, P. M., J. G. Van Riswick, and A. J. Van Opstal (1998). "Relearning sound localization with new ears." In: *Nature neuroscience* 1.5, pp. 417–21.

Hofman, P. M., M. S.M. G. Vlaming, P. J. J. Termeer, and A. J. Van Opstal (2002). "A method to induce swapped binaural hearing". In: *Journal of Neuroscience Methods* 113.2, pp. 167–179.

Hummersone, C., R. Mason, and T. Brookes (2010). "Dynamic precedence effect modeling for source separation in reverberant environments". In: *IEEE transactions on audio, speech, and language processing* 18.7, pp. 1867–1871.

IEC 268-13 (1985). *Sound System Equipment. Part 13: Listening Tests on Loudspeaker*. International Electrotechnical Commission, Geneva, Switzerland.

IEC 60118-2 (1983). *Hearing aids. Part 2: Hearing aids with automatic gain control circuits*. International Electrotechnical Commission, Geneva, Switzerland.

Kates, J. M. (2008). *Digital hearing aids*. Plural, p. 449.

Keidser, G, H. Dillon, M Flax, T Ching, and S Brewer (2011). "The NAL-NL2 prescription procedure". In: *Audiology Research* 1.1S, e24.

Kopčo, N. and B. G. Shinn-Cunningham (2011). "Effect of stimulus spectrum on distance perception for nearby sources." In: *The Journal of the Acoustical Society of America* 130, pp. 1530–41.

Korhonen, P., C. Lau, F. Kuk, D. Keenan, and J. Schumacher (2015). "Effects of coordinated compression and pinna compensation features on horizontal

localization performance in hearing aid users." In: *Journal of the American Academy of Audiology* 26.1, pp. 80–92.

Kulkarni, A, S. Isabelle, and H. Colburn (1995). "On the minimum-phase approximation of head-related transfer functions". In: *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*. IEEE, pp. 84–87.

Kulkarni, A. and H. S. Colburn (1998). "Role of spectral detail in sound-source localization". In: *Nature* 396, pp. 747–749.

Kuttruff, H. (2009). *Room acoustics*. Crc Press.

Lopez-Poveda, E. A. and R. Meddis (2001). "A human nonlinear cochlear filterbank". In: *The Journal of the Acoustical Society of America* 110, pp. 3107–3118.

Majdak, P., R. Baumgartner, and B. Laback (2014). "Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization". In: *Frontiers in Psychology* 5.APR, pp. 1–10.

May, T., S. van de Par, and A. Kohlrausch (2011). "A probabilistic model for robust localization based on a binaural auditory front-end". In: *IEEE Transactions on Audio, Speech, and Language Processing* 19.1, pp. 1–13.

May, T., N. Ma, and G. J. Brown (2015). "Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues". In: *40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2679–2683.

Mendonça, C. (2014). "A review on auditory space adaptations to altered head-related cues". In: *Frontiers in neuroscience* 8, p. 219.

Mershon, D. H. and L. E. King (1975). "Intensity and reverberation as factors in the auditory perception of egocentric distance". In: *Perception & Psychophysics* 18.6, pp. 409–415.

Middlebrooks, J. C. (1999). “Individual differences in external-ear transfer functions reduced by scaling in frequency”. In: *The Journal of the Acoustical Society of America* 106.3, pp. 1480–1492.

Middlebrooks, J. C. and D. M. Green (1991). “Sound localization by human listeners”. In: *Annual review of psychology* 42.1, pp. 135–159.

Mills, A. W. (1958). “On the minimum audible angle”. In: *The Journal of the Acoustical Society of America* 30, pp. 237–246.

Moore, B. C. and B. R. Glasberg (2001). “Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners.” In: *The Journal of the Acoustical Society of America* 110.2, pp. 1067–1073.

Moore, B. C. J. (2004). “Testing the concept of softness imperception: Loudness near threshold for hearing-impaired ears”. In: *The Journal of the Acoustical Society of America* 115.6, pp. 3103–3111.

Nielsen, J. B. and T. Dau (2011). “The Danish hearing in noise test.” In: *International Journal of Audiology* 50.3, pp. 202–208.

Nielsen, S. H. (1993). “Auditory distance perception in different rooms”. In: *Journal of the Audio Engineering Society* 41.10, pp. 755–770.

Noble, W. and S. Gatehouse (2006). “Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing scale (SSQ)”. In: *International Journal of Audiology* 45.3, pp. 172–181.

Otsu, N. (1979). “An automatic threshold selection method based on discriminate and least squares criteria”. In: *Denshi Tsushin Gakkai Ronbunshi* 63, pp. 349–356.

Patterson, R., I. Nimmo-Smith, J. Holdsworth, and P. Rice (1988). *SVOS final report (Part A): The Auditory Filterbank*. Tech. rep. August.

Pollack, I. and W. Trittipoe (1959). “Interaural Noise Correlations: Examination of Variables”. In: *The Journal of the Acoustical Society of America* 31.12, pp. 1616–1618.

Schwartz, A. H. and B. G. Shinn-Cunningham (2013). “Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners.” In: *The Journal of the Acoustical Society of America* 133.4, pp. 2329–2339.

Steinberg, J. C. and M. B. Gardner (1937). “The dependence of hearing impairment on sound intensity”. In: *The Journal of the Acoustical Society of America* 9, pp. 11–23.

Strelcyk, O., N. Nooraei, S. Kalluri, and B. Edwards (2012). “Restoration of loudness summation and differential loudness growth in hearing-impaired listeners”. In: *The Journal of the Acoustical Society of America* 132.4, pp. 2557–2568.

Thiergart, O., G. Del Galdo, and E. A. Habets (2012). “Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones.” In: *ICASSP 2012*, pp. 309–312.

Van Wanrooij, M. M. and a. J. Van Opstal (2005). “Relearning sound localization with a new ear.” In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25.22, pp. 5413–5424.

Villchur, E. (1973). In: *The Journal of the Acoustical Society of America* 53.6, p. 1646.

Wallach, H., E. B. Newman, and M. R. Rosenzweig (1949). “The Precedence Effect in Sound Localization”. In: *American Journal of Psychology* 62.3, p. 315.

Westermann, A., J. M. Buchholz, and T. Dau (2013). “Binaural dereverberation based on interaural coherence histogramsa)”. In: *The Journal of the Acoustical Society of America* 133.5, pp. 2767–2777.

Whitmer, W., B. U. Seeber, and M. a. Akeroyd (2014). “The perception of apparent auditory source width in hearing-impaired adults.” In: *The Journal of the Acoustical Society of America* 135.6, pp. 3548–3559.

Whitmer, W. M., B. U. Seeber, and M. A. Akeroyd (2012). “Apparent auditory source width insensitivity in older hearing-impaired individuals”. In: *The Journal of the Acoustical Society of America* 132.1, pp. 369–379.

Wiggins, I. M. and B. U. Seeber (2011). “Dynamic-range compression affects the lateral position of sounds”. In: *The Journal of the Acoustical Society of America* 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 and Hearing* 33.3, pp. 399–410.

Wiggins, I. M. and B. U. Seeber (2013). “Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise”. In: *The Journal of the Acoustical Society of America* 133.2, p. 1004.

Yost, W. A. and R. H. Dye (1988). “Discrimination of interaural differences of level as a function of frequency”. In: *The Journal of the Acoustical Society of America* 83, pp. 1846–1851.

Young, P. T. (1928). “Auditory localization with acoustical transposition of the ears.” In: *Journal of Experimental Psychology* 11.6, pp. 399–429.

Zahorik, P., D. Brungart, and A. Bronkhorst (2005). “Auditory distance perception in humans: A summary of past and present research”. In: *Acta Acustica united with Acustica* 91, pp. 409–420.

Zahorik, P. (2002). “Direct-to-reverberant energy ratio sensitivity”. In: *The Journal of the Acoustical Society of America* 112.5, pp. 2110–2117.

Zheng, C., A. Schwarz, W. Kellermann, and X. Li (2015). “Binaural coherent-to-diffuse-ratio estimation for dereverberation using an ITD model”. In: *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pp. 1048–1052.

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

Vol. 25: *Gusztáv Lőcsei*, Lateralized speech perception with normal and impaired hearing, 2016.

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation with cochlear implants, 2016.

The end.

To be continued...

Externalization refers to the percept of an auditory object as appearing at the same location and with the same apparent size as the actual acoustic source within the surrounding space. This ability relies on the listener using two ears rather than just a single ear. For listeners lacking this ability sounds sources are internalized, i.e. perceived inside the listener's head. Many hearing-aid users experience sound sources to be internalized even with state-of-the-art hearing aids. This has consequences for the hearing-impaired listener's spatial abilities and most likely contributes to difficulties when following a conversation in a noisy listening environment. In this thesis the influence of hearing-aid signal processing on the auditory cues responsible for externalization is investigated. It is expected that the results will stimulate and guide the development of new signal processing algorithms that can provide hearing-aid users with normal externalization perception.

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