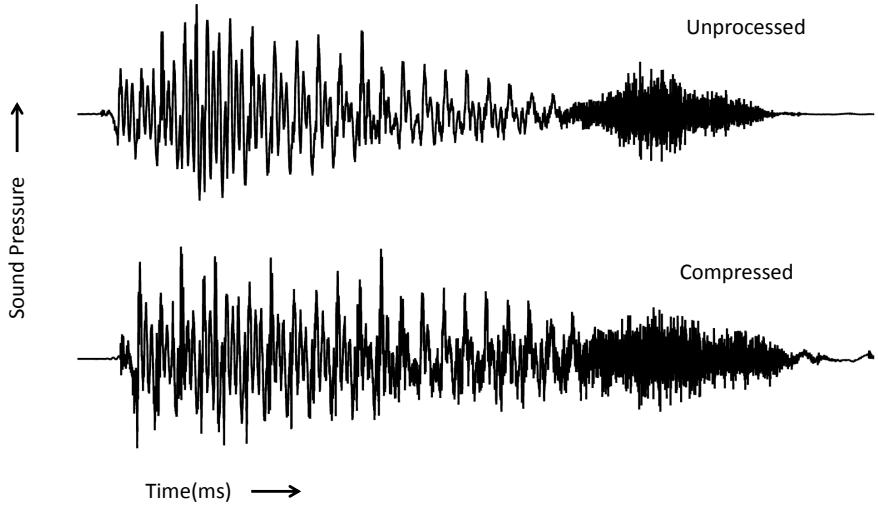
Technical University of Denmark

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 32

Alan Wiinberg

Perceptual effects of non-linear hearing aid amplification strategies



Hearing Systems
Department of Electrical Engineering

Perceptual effects of non-linear hearing aid amplification strategies

PhD thesis by
Alan Wiinberg

Preliminary version: November 8, 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**
Associate Professor **Bastian Epp**
Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Morten Løve Jepsen, PhD
Audiological Signal Processing
Widex A/S
Lynge, Denmark

Abstract

One of the most common reported complaints from people with sensorineural hearing loss is difficulties in understanding speech in complex acoustic environments. Some of these difficulties are caused by loss of cochlear compression, reflecting a reduced sensitivity to soft sounds, a steeper loudness growth function and a different temporal resolution than observed in normal-hearing people. In an attempt to compensate for the loss of cochlear compression, many modern hearing aids use a compressive amplification strategy. However, while speech perception might be improved by hearing-aid compression, the provided level-dependent amplification may lead to undesired distortions of the binaural cues which are essential for spatial perception. In the first study presented in this thesis, the influence of hearing-aid compression and hearing loss on two measures of temporal resolution was investigated. It was found that the hearing-impaired listeners showed higher modulation detection sensitivity and a lower modulation-depth discrimination sensitivity than the normal-hearing listeners. The compression processing restored the hearing-impaired listeners' modulation detection thresholds back towards the level observed in the normal-hearing listeners. However, the compression processing had no effect on the modulation-depth discrimination thresholds. Interestingly, a strong correlation between the modulation-depth discrimination sensitivity and speech intelligibility scores was observed in the normal-hearing listeners. In the second study, the influence of three speech enhancement schemes on consonant recognition was investigated. These schemes were designed to compensate for the degraded modulation-depth discrimination sensitivity observed in the hearing-impaired listeners in the first study. Two of the three speech enhancement schemes were found to improve consonant recognition scores relative to linear processing. In the third study, the effects of three hearing-aid compression schemes on spatial perception in a reverberant environment were investigated. It was found that fast-acting compression increases the diffusiveness of the perceived sound and leads to broader, sometimes internalized, sound images as well as sound-image splits. The observed spatial distortions were mainly caused by the applied compression enhanced the level of the reflected sound relative to the level of the direct sound. In the fourth study, the benefit of a direct-sound driven compression system was investigated that adaptively selects appropriate time constants in an attempt to maintain the energy ratio of the direct sound to the reflected sound. The proposed compression system was able to preserve

the listener's spatial impression. Overall, it is expected that the outcomes of this thesis facilitate the development and implementation of better hearing-aid signal processing strategies for the benefit of hearing-aid users.

Resumé

En følgevirkning af sensorineural høretab er, at det være svært at forstå tale i komplekse lydmiljøer. Det skyldes blandt andet at det indre øres evne til at forstærke lave lydniveauer op, og dermed komprimere dynamikområdet, er nedsat. Dette afspejles i en formindsket hørbarhed ved svage lydstyrker, stejlere loudness vækstfunktioner, og en ændring af hørelsens temporale opløsning. For at afhjælpe disse problemer anvender mange moderne høreapparater en niveauafhængig kompressiv forstærkningsstrategi. Strategien giver en øget hørbarhed ved svage lydstyrker, og kan normalisere loudness vækstfunktionerne. Derimod er effekten af denne type forstærkning på hørelsens temporale opløsning, og på opfattelsen af det rummelige lydbillede i lydmiljøer med rumklang, desværre stadig uklar på grund af begrænset forskning. I det første studie i denne afhandling blev effekten af høretab og kompressiv forstærkning undersøgt på to veletableret mål af hørelsens temporale opløsning. Resultaterne viste at de hørehæmmede lyttere var bedre til at detektere om et signal var amplitude moduleret, mens de derimod var dårligere til at diskriminere mellem forskellige modulationsdybder end de normalthørende lyttere. Det blev vist at kompressiv forstærkning kunne normalisere de hørehæmmedes evne til at detektere amplitude modulation til samme niveau som de normalthørende lyttere. Den kompressive forstærkning havde imidlertid ingen effekt på evnen til at diskriminere mellem forskellige modulationsdybder. Interessant nok blev der fundet en stærk korrelation mellem evnen til at diskriminere mellem forskellige modulationsdybder og taleforståelighed hos de normalthørende lytter. I det andet studie blev effekten af tre signalbehandlingsstrategier, der forventedes at forbedre taleforståeligheden, undersøgt ved hjælp af en konsonant-vokal lytte test. Disse strategier blev udarbejdet for at kompensere for den forringede modulationsdybdediskriminationsfølsomhed der blev observeret hos de hørehæmmede lyttere i det første studie. Resultaterne viste at to ud af de tre signalbehandlingsstrategier gav en forbedring af taleforståligheden i form af konsonant genkendelse i forhold til lineær processering. I det tredje studie blev virkningerne af tre kompressive forstærkningsstrategier på den rumlige lydopfattelse undersøgt i et lytte miljø med rumklang. Det blev konstateret, at hurtigvirkende kompression mindskede evnen til korrekt at lokalisere lyden og førte til bredere, undertiden internaliserede lydbilleder såvel som spaltninger af lydbilledet. De observerede rumlige forvrængninger blev hovedsageligt forårsaget af den påførte kompression, som bevirkede at niveauet af den reflekterede lyd i forhold til niveauet af den direkte lyd steg. I det fjerde studie blev fordelene ved et nyt udviklet kompressivt for-

stærkningssystem undersøgt. Systemet var designet således at tidskonstanterne blev adapteret på en sådan måde at relationen mellem den direkte lyd til den reflekterede lyd blev påvirket mindst muligt. Lytte forsøg viste at det foreslåede kompressionssystem var i stand til at bevare lytterens rumlige indtryk. Alt i alt forventes det, at resultaterne af denne afhandling kan facilitetere udvikling og implementering af bedre høreapparatsignalbehandlingsstrategier til gavn for alle der bruger høreapparater.

Acknowledgments

I would like to thank my main supervisor, Torsten Dau, for his encouragement, patience, and invaluable guidance. I would also like to express my gratitude to Bastian Epp and Morten Jepsen for their support, thoughtful questions and helpful advice. I had the pleasure to work with Johannes Käsbach, Henrik Hassager, Pernille Holtegaard and Johannes Zaar. Thank you so much for your significant contribution to this thesis.

My time at DTU was made enjoyable in large part due to the many friends from the Hearing Systems Group that became a part of my life. I am grateful for the time spent with my office mates (and good friends) Henrik Hassager and Borys Kowalewski. I will miss all the fun and the stimulating black board discussions.

Most importantly, I would like to thank my family. I thank my parents for all their love and encouragement. Thank you for all of your support and help with the children when I had to work. I thank my two sons who are filling my life with smiles, and for having patiently waited for their dad to finish his "homework". And I especially thank my beautiful wife and best friend, Louise, for all of her patience and for making my life richer than I could have ever imagined.

Related publications

Journal papers

- Wiinberg, A., Jepsen, M. L., Epp, B. and Dau, T. "Effects of hearing loss and fast-acting hearing-aid compression on amplitude modulation perception and speech intelligibility," submitted to Ears and Hearing

- Wiinberg, A., Zaar, J., Jepsen, M. L. and Dau, T. "Effects of expanding envelope fluctuations on consonant perception in hearing-impaired listeners," Submitted to Trends in Hearing

- Hassager, H. G., Wiinberg, A. and Dau, T. (**2017**). "Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment," J. Acoust. Soc. Am. **141**, 2556–2568

- Hassager, H. G., May, T., Wiinberg, A. and Dau, T. (**2017**). "Preserving spatial perception in rooms using direct-sound driven dynamic range compression," J. Acoust. Soc. Am. **141**, 4556–4566

Conference papers

- Wiinberg, A., Jepsen, M. L., Epp, B. and Dau, T. (**2015**). "Effects of dynamic-range compression on temporal acuity," Proceedings of the International Symposium on Auditory and Audiological Research 2015, Nyborg, Denmark.

- Käsbach, J., Wiinberg, A., May, T., Jepsen, M. L., Dau, T. (**2015**). "Apparent source width perception in normal-hearing, hearing-impaired and aided listeners," Forschritte der Akustik DAGA'15, Nuremberg, Germany.

Published abstracts

- Hassager, H. G., May, T., Wiinberg, A. and Dau, T. (**2017**). "Preserving spatial perception in reverberant environments using direct-sound driven dynamic range compression," The 3rd Joint Meeting of the Acoustical Society of America and the European Acoustics Association, Boston, Massachusetts

- Wiinberg, A., Jepsen M. L., Epp, B., and Dau, T. (**2016**). "Effects of dynamic-range compression on amplitude modulation processing and speech recognition," International Hearing Aid Research Conference, Tahoe City, California.

- 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.

- Wiinberg, A., Jepsen M. L., Epp, B., and Dau, T. (**2014**). "Effects of dynamic-range compression on temporal modulation transfer functions and modulation depth discrimination of normal-hearing and hearing-impaired listeners," International Hearing Aid Research Conference, Tahoe City, California.

Contents

Abstract **v**

Resumé på dansk **vii**

Acknowledgments **ix**

Related publications **xi**

Table of contents **xv**

1 Introduction **1**

2 The effect of hearing-impairment and fast-acting hearing-aid compression on temporal resolution **5**

2.1 Introduction 6

2.2 Methods 11

2.2.1 EXPERIMENT 1: Amplitude modulation detection and modulation-depth discrimination 11

2.2.2 EXPERIMENT 2: Perception of speech in noise 15

2.3 Results 18

2.3.1 Amplitude modulation detection thresholds 18

2.3.2 Modulation depth discrimination thresholds 21

2.3.3 Speech reception thresholds 22

2.3.4 Regression analyses 23

2.4 Discussion 25

2.5 Conclusions 29

3 Effects of expanding fast envelope fluctuations on consonant perception in hearing-impaired listeners **31**

3.1 Introduction 32

3.2 Methods 36

3.2.1 Listeners 36
3.2.2 Stimuli 37
3.2.3 Envelope expansion processing 38
3.2.4 Experimental design 41
3.2.5 Procedure and apparatus 41
3.2.6 Statistical analysis 42
3.3 Results 43
3.3.1 Consonant recognition scores of NH and HI listeners 43
3.3.2 Individual listener analysis and confusion matrices 46
3.4 Discussion 47
3.5 Conclusion 52

4 Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment 55
4.1 Introduction 56
4.2 Methods 59
4.2.1 Listeners 59
4.2.2 Experimental setup and procedure 60
4.2.3 Spatialization 63
4.2.4 Experimental conditions 64
4.2.5 Dynamic range compression 65
4.2.6 Statistical analysis 68
4.2.7 Analysis of spatial cues 68
4.3 Results 71
4.3.1 Analysis of spatial cues 78
4.4 Discussion 82
4.5 Conclusions 87

5 Preserving spatial perception in rooms using direct-sound driven dynamic range compression 89
5.1 Introduction 90
5.2 Compression system 94
5.2.1 Algorithm overview 94
5.2.2 Classification 97
5.2.3 Level estimation 101
5.3 Methods 102
5.3.1 Listeners 102

5.3.2 Experimental setup and procedure 102
5.3.3 Spatialization 104
5.3.4 Stimuli and processing conditions 105
5.3.5 Statistical analysis 106
5.3.6 Analysis of spatial cues 107
5.4 Results 108
5.4.1 Experimental data 108
5.4.2 Analysis of spatial cues 111
5.5 Discussion 113
5.6 Conclusion 116

6 General discussion **119**
6.1 Summary of main findings 119
6.2 Implications and perspectives 122

Bibliography **125**

Collection volumes **141**

1

General introduction

Hearing is an important sense for humans. The sense of hearing enables the listener to identify where sounds in the world are coming from, and to analyze what the meaning of those sounds is. Survival in urban environments can in some situations depend on accurate localization and categorization of sound sources. For example, we need to jump the right way when a rapidly approaching car is heard. The sense of hearing also enables the listener to understand speech. Speech is one of the main means by which we communicate with one another. Because the sense of hearing allows listeners to detect, process, and interpret sounds continuously and without conscious effort, this sense is usually taken for granted by normal-hearing (NH) listeners. However, some people do not take it for granted as they face the reality of a degraded sense of hearing. In the United States, the National Institute on Deafness and Other Communication Disorders (NIDCD) estimates that hearing loss is the reality for one in eight above the age of 12. For adults, the most common form of hearing loss is sensorineural, i.e. resulting from damages to the inner ear (cochlea) or to the nerve pathways from the inner ear to the brain. One of the typical symptoms of a sensorineural hearing loss is major challenge to follow a conversation in a noisy and reverberation environment (e.g., Festen and Plomp (1990)). Hence, for listeners with hearing loss, large groups and noisy venues can make it difficult to understand what other people are saying and thereby make it difficult to be engaged in a conversation. As a direct consequence, it can be a challenge for

hearing-impaired (HI) people to maintain an active social life and hearing loss may thereby lead to feelings of loneliness and isolation.

Hearing aids attempt to compensate for the deficits caused by a sensorineural hearing loss. The use of hearing aids, particularly in a quiet environment, has been shown to effectively improve speech intelligibility relative to the unaided listening situation (e.g., Metselaar et al., 2008). However, although hearing aids usually allow for better audibility of speech sounds, HI listeners continue to have particular difficulty in background noise (e.g., Gnewikow et al., 2009; Metselaar et al., 2008). Hence, hearing aids are far from being able to restore speech intelligibility back to the level observed in NH listeners. Also, hearing aids unfortunately face challenges with preserving the natural spatial perception of the surrounding sound environment. For example, Van den Bogaert et al. (2006) demonstrated that hearing aids do not preserve localization cues. Noble and Gatehouse (2006) found in their questionnaire-based field study that hearing-aid users perceive external sound sources as originating from inside the head, rather than from out in the world. In other words, there is still room for considerable improvement in terms of hearing-aid benefit.

In the past, hearing aids typically used linear amplification to compensate for the perceptual consequences of a hearing loss. However, the problem with linear amplification is that, if sufficient amplification is provided to make low-intensity sounds audible, high-intensity sounds are perceived as louder than normal and can even exceed the threshold of pain. This is because the growth of loudness as a function of sound intensity is usually far steeper (and hence abnormal) in HI listeners than in NH listeners (Fowler, 1936; Steinberg and Gardner, 1937). This abnormal loudness coding in HI listeners is commonly referred to as loudness recruitment and appears to be a consequence of damaged outer hair-cells (OHCs) and, thereby, a loss of the level-dependent

cochlear amplification the OHCs provide. In an attempt to mimic and thereby compensate for the loss of compressive amplification provided by the OHCs, modern hearing aids use compression amplification (Allen, 1996; Villchur, 1973) whereby level-dependent amplification is provided in various frequency bands. By amplifying the low-intensity sounds more strongly than the higher-intensity sounds, hearing-aid compression can provide audibility of the low-intensity portions of the sound while improving loudness comfort at high intensities. It has been shown that fast-acting hearing-aid compression can restore normal loudness summation and differential loudness growth in HI listeners (Strelcyk et al., 2012). It has also been demonstrated that measures of auditory temporal resolution that are assumed to be affected by recruitment, such as gap detection and forward masking, can at least partly be restored in HI listeners by fast-acting compression (e.g., Moore et al. (2001), Brennan et al. (2015), and Kowalewski et al. (2015)).These findings suggest that fast-acting hearing-aid compression can improve the ability of HI listeners to process the time-varying and dynamic aspects of speech in noise.

In this thesis, behavioral methods are used to study the effects of non-linear hearing-aid amplification strategies on HI listeners' spatial perception, temporal resolution and speech perception. The objectives are to better understand the consequences of conventional hearing-aid compression on the different behavioral outcome measures and to analyze the effects of various modifications of this amplification strategy.

In **Chapter 2**, the influence of fast-acting hearing-aid compression on two measures of temporal resolution (modulation detection and modulation depth-discrimination thresholds) and speech intelligibility is examined in NH and HI listeners. The results are expected to provide insights into the interplay between loss of cochlear compression, fast-acting hearing-aid compression,

speech intelligibility and temporal resolution.

Based on these findings, **Chapter 3** investigates whether artificially increasing the modulation depth of the speech envelope facilitates better extraction of speech cues and thereby improves speech intelligibility in HI listeners. The perceptual consequences of three speech enhancement schemes based on multi-band non-linear expansion of temporal envelope fluctuations between 10 and 20 Hz are examined in NH and and HI listeners. The speech enhancement schemes are expected to improve speech recognition in background noise, particularly the intelligibility of stop consonants.

In **Chapter 4**, the effects of three fast-acting hearing-aid compression schemes on spatial perception in a reverberant environment are investigated. The results are expected to improve our understanding of how fast-acting hearing aid compression distorts the listener's natural spatial perception. Based on these findings, a novel hearing-aid compression strategy is proposed in **Chapter 5** that adaptively selects appropriate time constants in the attempt to preserve the listener's spatial perception. It is expected that the proposed compression system is able to provide the listeners with a spatial perception similar to that obtained with linear processing.

Finally, **Chapter 6** summarizes the main findings of this thesis and discusses possible implications and perspectives.

2

The effect of hearing-impairment and fast-acting hearing-aid compression on temporal resolution[a]

Abstract This study investigated the effects of fast-acting hearing-aid compression on normal-hearing and hearing-impaired listeners' temporal envelope sensitivity and speech intelligibility. Two measures of temporal envelope sensitivity were obtained: Amplitude modulation detection thresholds and modulation-depth discrimination thresholds. Speech intelligibility was assessed by obtaining reception thresholds for speech in steady and fluctuating background noise. To estimate the effect of the modulation depth reduction by hearing-aid compression, all thresholds were obtained with and without compression while controlling for the effect of audibility. The results showed that properly set fast-acting hearing-aid compression can restore the hearing-impaired listeners' modulation detection thresholds to the level observed in normal-hearing listeners. Despite this "normalization" of the modulation detection thresholds, hearing-aid compression does not seem to provide a benefit for speech intelligibility. Furthermore, hearing-aid compres-

[a] This chapter is based on Wiinberg, A, Jepsen, M. L., Epp, B., Dau, T. (in review), *Ears and Hearing*.

sion was not able to restore the modulation depth discrimination thresholds to the level of "normal" processing suggesting that the two measures of sensitivity to amplitude modulations represent different limits of resolution. The results obtained in the normal-hearing listeners show that the ability to discriminate modulation depths is correlated with speech intelligibility in stationary noise.

2.1 Introduction

Loudness recruitment is a typical consequence of sensorineural hearing loss (Fowler, 1936; Steinberg and Gardner, 1937; Moore, 2004). In an attempt to restore normal loudness perception, many hearing aids use multi-band dynamic range compression (DRC, Allen, 1996; Villchur, 1973) which provides level-dependent gain such that low-level sounds are amplified more strongly than higher-level sounds. DRC thus provides audibility of the low-level portions of the sound while improving loudness comfort at high levels. It has been shown that multi-band DRC with compression attack and release times constants of 10 ms can restore normal loudness summation and differential loudness growth in hearing-impaired (HI) listeners (Strelcyk et al., 2012). It has also been demonstrated that measures of auditory temporal resolution that are assumed to be affected by recruitment, such as gap detection and forward masking, can at least partly be restored by fast-acting DRC with attack and release times less than 60 ms in HI listeners (e.g., Moore et al., 2001; Brennan et al., 2015; Kowalewski et al., 2015).

Another measure of temporal resolution is amplitude modulation detection. Recruitment has been shown to enhance the internal representation of the signal envelope in the auditory system, mainly as a result of reduced amplitude

compression of the signal at the level of processing on the basilar membrane (BM) in the impaired auditory system, relative to the processing in the healthy auditory system (Moore et al., 1996). As a perceptual consequence of reduced BM compression, a larger perceived modulation strength of modulated tones has been reported in listeners with a unilateral hearing loss in the conditions where the stimuli were presented to the impaired ear relative to the results in the conditions when the stimuli were presented to the healthy ear (Moore et al., 1996). Consistent with this result, HI listeners have shown lower modulation detection thresholds (MDTs), i.e. a higher sensitivity, than normal-hearing (NH) listeners when using tonal carriers presented at the same sensation level (Moore and Glasberg, 2001; Bianchi et al., 2016). Since fast-acting DRC reduces the depth of modulations in the temporal envelope, this type of processing might be able to restore the "normal" internal representation of the envelope in the HI listeners. Brennan et al. (2013) obtained MDTs with noise carriers in conditions with and without DRC in HI listeners and found higher MDTs in the conditions with DRC, consistent with the idea that DRC may compensate for recruitment. However, the study did not consider to what extent the DRC processing restored the performance to the level found in NH listeners. Furthermore, since noise carriers contain intrinsic fluctuations that can mask the imposed signal modulation, in contrast to determinstic tonal carriers without any intrinsic fluctuations, MDTs obtained with noise carriers may be dominated by modulation masking effects whereas MDTs obtained with tonal carriers can only be limited by internal noise (e.g., Dau et al., 1997).

Much evidence has been provided that speech intelligibility in a given environment is affected by the processing of the amplitude modulations contained in the speech (and the interferers) by the auditory system. Speech intelligibility models, such as the speech transmission index (STI, Houtgast and Steeneken,

1985), the spectro-temporal modulation index (STMI, Elhilali et al., 2003) or the speech-based envelope power spectrum model (sEPSM, Jørgensen and Dau, 2011) have accounted for effects of different types of processing channels (such as a room or a nonlinear processor) on the envelope representations of the signals and their relation to speech intelligibility. However, the link between the sensitivity to amplitude modulations and speech intelligibility has been controversial. It has been argued that magnified masker fluctuations in the internal representation could be a distraction that would reduce speech intelligibility (Kale and Heinz, 2010; Schlittenlacher and Moore, 2016). Nonetheless, MDTs obtained with noise carriers seem to only poorly correlate with speech intelligibility obtained with modulated maskers in HI listeners (e.g., Takahashi and Bacon, 1992; Feng et al., 2010). Regarding effects of fast-acting DRC on speech intelligibility, some studies reported a degradation of speech intelligibility (Drullman and Smoorenburg, 1997; Reinhart et al., 2016; Noordhoek and Drullman, 1997). For example, Noordhoek and Drullman (1997) found that fast-acting DRC increased the speech reception threshold (SRT) in stationary noise. The authors argued that the detrimental effect of fast-acting DRC on speech intelligibility was caused by the processing distorts the temporal envelope of the speech signal and thereby distorts important speech cues. However, other studies found either no effect of fast-acting DRC on speech intelligibility (e.g., Boothroyd et al., 1988; Souza and Turner, 1996; Drullman and Smoorenburg, 1997) or even an improved speech intelligibility (Souza and Turner, 1998; Yund, 1995; Gatehouse et al., 2006). For example, Souza and Turner (1998) found increased speech recognition scores in HI listeners using multi-band fast-acting DRC at low speech levels when compared to the scores obtained with linear amplification. The authors ascribed this improvement to differences in amplification; the DRC scheme applied more amplification and thereby raised the

low-level speech stimuli above the audibility threshold by a larger amount than the linear scheme, whereas no improvement was observed at higher speech levels.

Since most modulations inherent in speech are well above threshold, i.e. above MDT (Edwards, 2004; Schlittenlacher and Moore, 2016), supra-threshold measures of modulation processing, such as modulation-depth discrimination (MDD) thresholds, might provide stronger links to speech intelligibility performance than MDTs. Schlittenlacher and Moore (2016) observed higher MDD thresholds (with tonal carriers) in HI listeners than in NH listeners. It was argued that recruitment increases the perceived amount of amplitude modulation, called fluctuation strength, such that this sensation "saturates" at lower modulation depths in HI listeners than in NH listeners. Hence, differences between modulation depths may become less noticeable in HI listeners than in NH listeners when the fluctuation strength is at ceiling level in the HI listeners and below ceiling level in the NH listeners. Following this argumentation, fast-acting DRC might be able to compensate for the enhanced fluctuation strength and thereby restore "normal" MDD thresholds. Alternatively, it is possible that MDT versus MDD thresholds represent two different resolution limits. In fact, Ewert and Dau (2004) demonstrated that MDD thresholds obtained at different reference modulation depths follow Weber's law, i.e. thresholds increase roughly proportional with increasing reference modulation depth. Two different "internal noise" processes can then be assumed to limit temporal resolution: an "absolute" internal noise to represent MDT and an internal noise after a logarithmic compression to represent the MDD threshold. If cochlear recruitment causes an enhancement of the internal representation of the envelope, relative to normal hearing, this may affect the MDT but not necessarily the MDD threshold. Likewise, DRC may compensate for the effect of recruitment and

thus readjust MDT back to “normal” but this may not affect the MDD threshold.

In the framework of the sEPSM (Jørgensen and Dau, 2011), speech intelligibility is related to the speech-to-noise power ratio at the output of an audio-frequency and modulation-frequency-selective process. An earlier version of the model, the envelope power spectrum model (EPSM, Ewert and Dau 2000), was developed to account for modulation detection and masking data and can also account for the MDD data of Ewert and Dau (2004) in NH listeners. It is possible that aspects of a hearing loss that are beyond loudness recruitment, such as supra-threshold deficits due to, e.g., inner hair cell loss, are reflected by increased MDD thresholds as well as degraded speech intelligibility in some conditions.

The current study addressed the relations between temporal envelope sensitivity (in terms of MDT and MDD thresholds), speech intelligibility, hearing-impairment and fast-acting DRC. Two measures of temporal envelope sensitivity were obtained: (i) temporal modulation transfer functions (TMTFs) with tonal carriers for which MDTs were obtained as a function of modulation frequency (e.g., Kohlrausch et al., 2000; Moore and Glasberg, 2001) and (ii) MDD thresholds, for which the just-noticeable increase in modulation depth from a (supra-threshold) standard modulation depth was measured as a function of modulation frequency (e.g., Ewert and Dau, 2004). Since previous work suggested that both slow envelope fluctuations (<16 Hz) as well as fast fluctuations (16-300 Hz) contribute to speech intelligibility in competing-talker conditions (Stone et al., 2012; Stone et al., 2008; Christiansen et al., 2013), MDD thresholds and MDTs were obtained for modulation frequencies in the range from 8 to 256 Hz. Tonal carriers were considered, instead of noise carriers, since tonal carriers do not contain intrinsic envelope fluctuations which may mask, and thereby limit, the detectability of the imposed modulation (e.g., Dau et al., 1997; Dau

et al., 1999). In the case of narrowband (i.e. also tonal) carriers, the imposed modulation introduces spectral sidebands which may be resolved if they are sufficiently far from the carrier frequency (e.g., Kohlrausch et al. 2000). However, as long as the modulation frequency is within the range where spectral resolution does not play a major role, results obtained with tonal carriers may provide a better measure of the temporal resolution of the auditory system than results obtained with noise carriers. In the first experiment, MDTs and MDD thresholds were obtained with and without DRC for listeners with normal and impaired hearing. In the second experiment, SRTs for speech in both steady and fluctuating noise were obtained with and without DRC for the same group of listeners.

2.2 Methods

2.2.1 EXPERIMENT 1: Amplitude modulation detection and modulation-depth discrimination

Listeners

Two groups of listeners participated in the experiments, a NH group and a HI group. The NH group consisted of 12 adults (6 males and 6 females). The mean age was 29 years and the range was 21 to 60 years. All had absolute thresholds better than 20 dB HL for the octave frequencies between 0.125 and 8 kHz. The HI group consisted of 16 adults (10 males and 6 females) with symmetrical mild-to-moderately-severe sensorineural hearing losses. One listener dropped out of the study without completing the TMTF measure with the 5 kHz carrier. The mean age was 68 years and the range was 50 to 80 years. The absolute thresholds for the test ear, measured using conventional audiometry, are shown in Fig. 2.1. All listeners (except the first author, who served as one of the NH

listeners) signed an informed consent document and were reimbursed for their efforts. Approval for the study was granted by the Science Ethics Committee of the Capital Region in Denmark ("De Videnskabsetiske Komitéer for Region Hovedstaden").

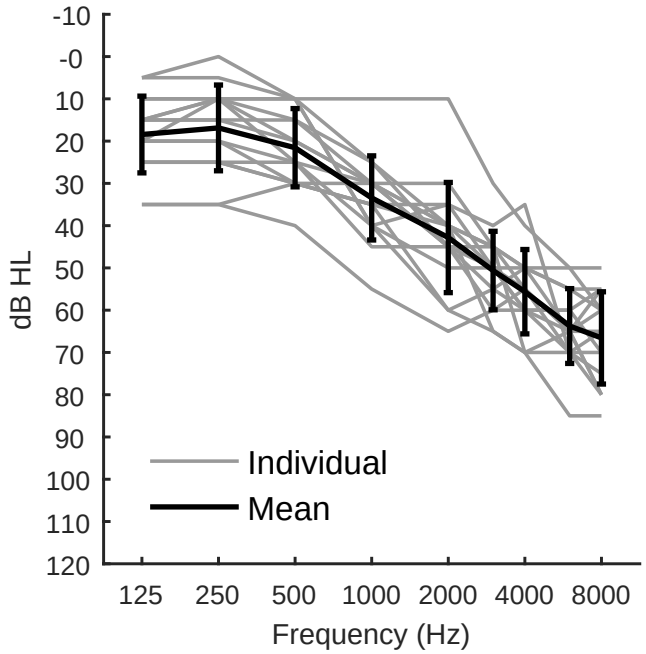


Figure 2.1: Individual and mean absolute thresholds for the tested ear of the HI listeners, measured using conventional manual audiometry, and expressed in dB HL. Error bars represent ±1 standard deviation.

Stimuli

The stimulus was a sinusoidally amplitude modulated tone:

$$s(t) = (1 + m\cos(2\pi f_{\mathrm{m}} t))\sin(2\pi f_{\mathrm{c}} t), \tag{2.1}$$

where t is time, m is the modulation depth taking values between 0 and 1, and f_{c} and f_{m} represent the frequency of the carrier and modulator, respectively.

The frequency of the carrier was either 1 or 5 kHz for the measurement of the

MDTs, and 1 kHz for the measurement of the MDD thresholds. The modulation frequencies were 8, 16, 32, 64, 128, and 256 Hz in the unprocessed condition and 8, 16, and 32 Hz in the DRC condition. The stimulus duration was 600 ms, including 50 ms raised-cosine onset and offset ramps. Intervals were separated by 500-ms pauses. The overall level of the stimuli in each interval was kept the same, regardless of the modulation depth and DRC processing. For the NH listeners, the SPL of the stimuli was 65 dB. For the HI listeners, the 65 dB stimuli were linearly amplified according to the NAL-R(P) frequency-dependent prescription rule by amounts depending on the individual audiometric thresholds (Byrne et al., 1990). The frequency-dependent amplification was provided using a bank of seven octave-wide bandpass linear-phase, finite-impulse-response filters with center frequencies between 0.125 to 8 kHz.

All signals were generated digitally on a PC equipped with a RME UCX Fireface sound card at a sampling rate of 44.1 kHz. The stimuli were presented in a sound-attenuating booth via DT 770 PRO Beyerdynamic headphones to the better ear of the listeners, as derived from the average of the absolute threshold at 500 Hz, 1000 Hz and 2000 Hz. The transfer function of each earpiece of the headphones was digitally equalized (1001 point FIR filter) to produce a flat frequency response, measured with an ear simulator (B&K 4153) and a flat plate adaptor as specified in IEC 60318-1 (2009).

Experimental procedure

MDTs were measured using an adaptive three-interval, three-alternative forced-choice paradigm. A three-down one-up procedure was used to track the 79.4% point on the psychometric function (Levitt, 1971). The carrier was unmodulated (the reference intervals) in two of the intervals and amplitude modulated in the other (the target interval). In each trial, the intervals were presented in random

order and the listeners had to select the interval containing the modulated carrier. For each carrier frequency, the different modulation frequencies were tested in random order. The thresholds and step sizes were represented in terms of the modulation depth, in decibels: $20\log_{10}(m)$. A run started with a modulation depth of -5 dB. The step size was initially 5 dB, and the step size was then reduced to 2 dB following the first incorrect response. Each run was terminated after seven reversals, and the threshold estimate for that run was computed as the mean value of the modulation depth at the last six reversals. Reported thresholds represent the mean over three runs.

For the measurement of MDD thresholds, the procedure was the same as for the amplitude modulation detection procedure, except that the carrier was modulated with a constant standard modulation depth (m_s) in the reference intervals and modulated with a higher modulation depth in the target interval. The standard modulation depth was -15 dB, which was 10-15 dB above the amplitude modulation detection thresholds typically found for a 1-kHz carrier for the range of the modulation frequencies studied here (Moore and Glasberg, 2001; Kohlrausch et al., 2000). A threshold run started with the target at a modulation depth of -3 dB. The step size of the modulation depth of the target was initially 2 dB, and the step size was then reduced to 1 dB following the first incorrect response.

The order of the measurement of MDTs and MDD thresholds with and without DRC was randomized for each individual listener.

Single channel dynamic range compression system

The single-channel DRC system was implemented in MATLAB version 2013b (The MathWorks, Inc., Natick, Massachusetts, United States). The envelope of the signal was extracted using the Hilbert transform and 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 and 60 ms, respectively. The smoothed envelope was converted to dB SPL. A broken-stick gain function (with linear gain below the compression threshold and constant compression ratio above threshold) was applied to the processed envelope. The resulting sample-wise gain was applied to the input stimulus. The static compression ratio was set to 2:1 and the compression threshold was set to 20 dB SPL.

2.2.2 EXPERIMENT 2: Perception of speech in noise

Stimuli

The Danish version of the hearing in noise test (HINT) was used to measure SRTs (Nielsen and Dau, 2011). Two noise maskers provided by the International Collegium for Rehabilitative Audiology (ICRA) were used, ICRA-1 steady speech-shaped noise and ICRA-6 speech-shaped noise with the modulation characteristics of two-talker babble (Dreschler et al., 2001).

Listeners

The same listeners as in experiment 1 participated in experiment 2. However, the HI listener who dropped out during experiment 1 did not participate in experiment 2. The SRT data from another HI listener were excluded from further analysis because of the listener's familiarity with the HINT speech corpus from his work as a clinical audiologist.

Experimental procedure

The SPL of the speech was kept constant and the level of the noise masker was varied using an adaptive one-up one-down procedure. The listeners were instructed to verbally repeat the sentences as accurately as possible and to

guess if they were uncertain. All of the words in a sentence need to be correct for the sentence to be scored as correct (Nielsen and Dau 2011). The SRT was measured for each of the noise maskers with and without DRC applied to the mixture of speech and noise. The output level of the DRC was equalized to the input signal's root-mean-square (RMS) level. The sentence lists were randomly selected for each of the four conditions (2 noise conditions x 2 DRC conditions). In order to familiarize the listeners with the task and to reduce possible learning effects, training was conducted before the data collection using two practice lists. For the NH listeners, the speech had a level of 65 dB SPL in the unprocessed condition. For the HI listeners, the mixture of speech and noise was linearly amplified according to the NAL-R(P) frequency dependent prescription (Byrne et al. 1990). The SPL for the mixture of speech and noise was the same with and without DRC. Signal generation and playback were the same as in experiment 1.

Multi-band dynamic range compression system

The multi-band DRC system used in the second experiment contained seven frequency bands, in contrast to the wideband DRC system used in experiment 1. The input signal was Hanning windowed in time frames of 256 samples, approximately 6 ms in duration, with 75% overlap between frames. Each of the windowed segments was padded with 128 zeros at the beginning and with 128 zeros at the end and transformed to the spectral domain using a 512-point fast Fourier transform (FFT). The power of the resulting frequency bins was combined into seven octave-wide frequency bands with center frequencies between 0.125 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 and 60 ms, respectively. The smoothed envelopes were converted to dB 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 static compression ratio was set to 2:1 and the compression threshold was set to 20 dB SPL. The resulting band-wise gains were then smoothed in the frequency domain using a shape-preserving 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.

Statistical analysis

Three analyses of variance (ANOVA) assessed the effects of hearing impairment and modulation frequency on the MDTs for the 1-kHz, the 5-kHz carrier, and the MDD thresholds, respectively. Hearing status (NH versus HI) was treated as a between-listener factor and modulation frequency was treated as a within-listener factor. Two additional ANOVAs also assessed the effect of DRC for modulation frequencies between 8 and 32 Hz. The additional factors, DRC (on versus off) and carrier frequency (1 versus 5 kHz) were treated as within-listener factors. The SRT data were analyzed separately for each noise type using a two-factor repeated-measure ANOVA that included the factors of hearing status and DRC condition. Tukey's Honestly Significant Differences (HSD) post-hoc tests were conducted to test for main effects and interactions. A confidence level of 5% was considered to be statistically significant. In the case that even full modulation was not detectable, the threshold was set to a modulation depth corresponding to full modulation.

2.3 Results

2.3.1 Amplitude modulation detection thresholds

Figure 2.2 shows the MDTs as a function of the modulation frequency for the carrier frequency of 1 kHz for the NH listeners (upper panels) and the HI listeners (bottom panels). Panels A and C show the results obtained without DRC, and panels B and D show the results obtained with DRC. The across-listener variation was larger for the HI group (with a standard deviation of about 5.4 dB) than for the NH group (with a standard deviation of about 3.5 dB). The MDT decreased from 8 Hz to 16 Hz and remained constant from 16 Hz to 64 Hz. The increased MDT at 8 Hz may reflect a temporal interference between the carrier onset and the onset of modulation, as shown by Sheft and Yost (1990). The MDT decreased with increasing modulation frequency above 64 Hz. The decrease in MDT at high modulation frequencies reflects the detection of the spectral sidebands (Kohlrausch et al. 2000). On an individual level, eleven of the sixteen HI listeners showed a monotonically decreasing MDT above 64 Hz, similar to the NH listeners, whereas the MDT remained roughly constant above 64 Hz for the remaining five HI listeners. This is most likely a consequence of reduced frequency selectivity for these five HI listeners in the frequency region around 1 kHz. To avoid confounding the measure of temporal modulation sensitivity with the ability to detect spectral sidebands, MDTs for the 1-kHz carrier above 64 Hz were excluded from the statistical analysis. There was no effect of hearing status [$F(1,26)=0.46$, $p=0.503$] and no interaction between hearing status and modulation frequency [$F(3,78)=0.81$, $p=0.503$]. However, there was an effect of modulation frequency [$F(3,78)= 21.40$, $p<0.001$].

Figure 2.3 shows the MDTs as a function of the modulation frequency for the carrier frequency of 5 kHz for the NH listeners (upper panels) and the HI lis-

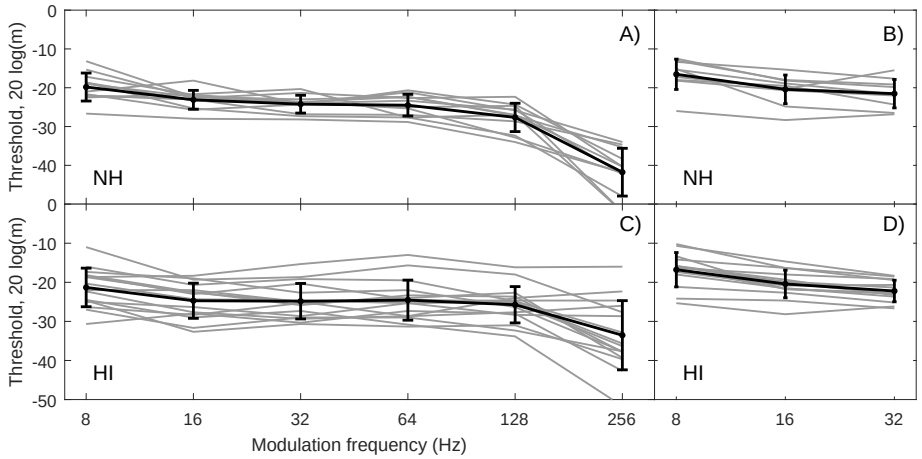


Figure 2.2: Individual (gray) and mean (black) MDTs for a carrier frequency of 1 kHz. The MDT ($20\log m$) is plotted as a function of the modulation frequency. Panels A and B show the results for the NH listeners, and panels C and D show the results for the HI listeners. Panels A and C show the results obtained without DRC, and panels B and D show the results obtained with DRC. The error bars represents ±1 standard deviation.

teners (bottom panels). There was an effect of modulation frequency [$F(5,125)=44.14$, $p<0.001$], but no effect of hearing status [$(F(1,25) = 0.93$, $p=0.343$]. The interaction between hearing status and modulation frequency was significant [$F(5, 125) = 10.43$, $p<0.001$]. Post-hoc comparisons revealed that the threshold decreased from 8 to 16 Hz, remained constant from 16 to 128 Hz, and increased between 128 and 256 Hz. Below 64 Hz, the MDT thresholds were about 5.5 dB (with a standard error of 3 dB) lower for the HI listeners than for the NH listeners, while the MDTs were similar in the two groups for modulation frequencies above 64 Hz. Two of the HI listeners showed much higher MDTs above 32 Hz than the other HI listeners. The increase of the MDT at high modulation frequencies for the 5-kHz carrier reflects a limitation in the temporal processing of fast amplitude modulation in the auditory system (e.g., Ewert and Dau, 2000; Kohlrausch et al., 2000).

Panels B and D in Figs. 2.2 and 2.3 represent the MDTs for the NH listeners and HI, respectively, when DRC was applied. There were significant effects

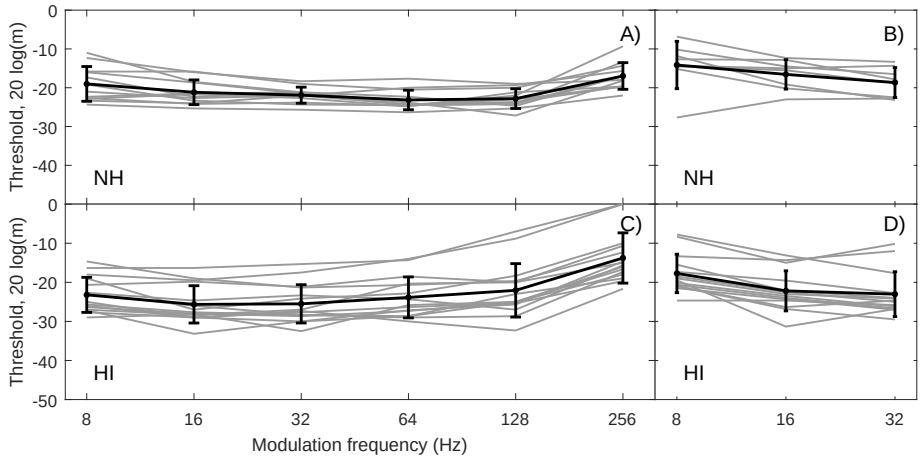


Figure 2.3: As Fig.2.2, but for a carrier frequency of 5 kHz.

of DRC [F(1,19)=73.28, p<0.001], hearing status [F(1,45)=4.23, p=0.045], modulation frequency [F(2,40)=60.42, p<0.001] and an interaction between DRC and modulation frequency [F(2,145)=6.89, p=0.001]. No other effects were significant. The change in the MDT due to DRC as a function of the modulation frequency is shown in Fig. 2.4. The change in the MDT due to DRC (solid curve) is roughly consistent with the physical reduction of the modulation depth (dashed curve). The physical reduction of the modulation depth was numerically derived from the Hilbert envelope before and after compression for an input modulation depth of -15 dB (Stone and Moore, 1992).

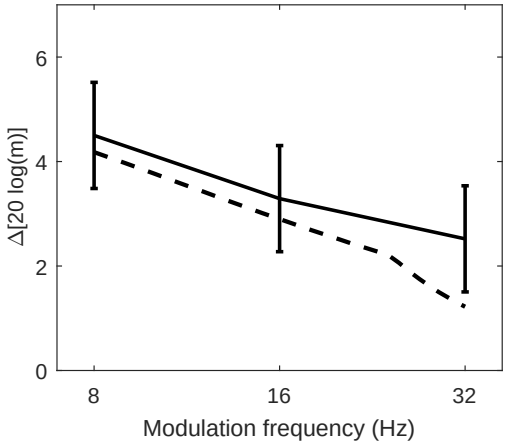


Figure 2.4: The effect of DRC on the MDTs (solid curve) averaged across listeners and carrier frequencies. The error bars represent ±2 standard errors of the mean. For comparison, the dashed curve shows the physical reduction of the modulation depth.

2.3.2 Modulation depth discrimination thresholds

Figure 2.5 shows the MDD thresholds as a function of the modulation frequency for the carrier frequency of 1 kHz. The results for the NH listeners are shown in panels A and B and the results for the HI listeners are shown in panels C and D. The MDD thresholds are expressed as the ratio of the modulation depth of the target to the modulation depth of the reference $[20\log(m/m_s)]$. Panels A and C show the results obtained without DRC, and panels B and D show the results obtained with DRC. The across-listener variation in the MDD thresholds was larger for the HI group with a standard deviation of about 2.1 dB than for the NH group with a standard deviation of about 1.3 dB. An ANOVA on the data for no-DRC revealed an effect of hearing status [$F(1,26)=5.26$, $p=0.029$], an effect of modulation frequency [$F(3,78)= 7.36$, $p<0.001$] and no interaction between hearing status and modulation frequency [$F(3,78)=0.45$, $p=0.717$]. The MDD thresholds were 1.6 dB (with a standard error of 0.6 dB) lower for the NH listeners than for the HI listeners. Post-hoc comparisons revealed that the MDD threshold decreased from 8 to 16 Hz for the NH listeners, and remained

constant from 16 to 64 Hz. As for the modulation detection experiment, the increased MDD threshold at low modulation frequencies was probably caused by the gating of the carrier (Lee and Bacon, 1997).

Panels B and D in Fig. 2.5 show the MDD thresholds for the NH listeners and HI for DRC. There was no effect of DRC [F(1,50)=0.79, p=0.38] and no interaction between DRC and modulation frequency [F(2,53)=0.24, p=0.78].

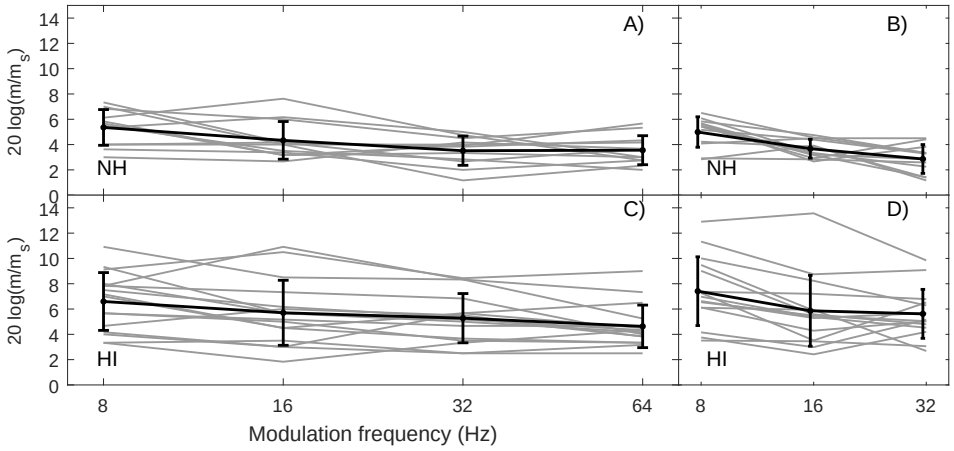


Figure 2.5: Individual and mean MDD thresholds for the 1-kHz carrier. The modulation discrimination threshold ($20\log(m/m_s)$) is plotted as a function of the modulation frequency. Panels A and B show the results for the NH listeners, and panels C and D show the results for the HI listeners. Panels A and C show the results obtained without DRC, and panels C and D show the results obtained with DRC. The error bars represent ±1 standard deviation.

2.3.3 Speech reception thresholds

Figure 2.6 shows the average SRTs for the NH listeners (open symbols) and the HI listeners (closed symbols) obtained in the two different noise conditions with and without multi-band DRC processing. The NH listeners showed, on average, lower SRTs than the HI listeners in all conditions. The SRT difference between the steady noise (ICRA-1) and the speech modulated noise (ICRA-6), the masking release, can be regarded as a measure of the benefit from the dips in the modulated noise. The masking release was smaller for the HI listeners

(0.9 dB) than for the NH listeners (3.3 dB); five of the HI listeners showed a negative masking release. The variation of SRTs across listeners was larger for the HI group than for the NH group, especially for the speech-modulated noise (ICRA-6).

For the steady noise masker (ICRA-1), there was an effect of DRC [$F(1,25) = 5.76,\ p = 0.02$], an effect of hearing status [$F(1,25) = 48.11,\ p < 0.001$] but no interaction [$F(1,25) = 1.31,\ p = 0.26$]. Post-hoc analysis showed that the SRT was 0.5 dB higher (with a standard error of 0.2 dB) for the DRC-on condition than for the DRC-off condition. For the speech-modulated noise masker (ICRA-6), there was an effect of hearing status [$F(1,25) = 36.64,\ p < 0.001$], but no effect of DRC [$F(1,25) = 0.04,\ p = 0.83$] and no interaction [$F(1,25) = 0.43,\ p = 0.51$].

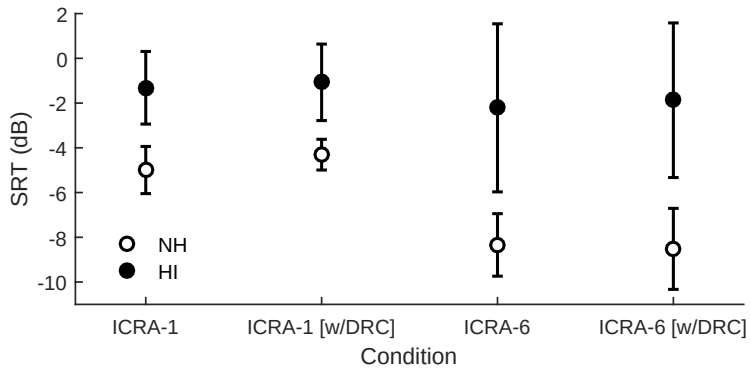


Figure 2.6: Mean SRTs for the NH listeners and HI for the different noise conditions with and without DRC. ICRA-1 is steady speech-shaped noise and ICRA-6 is speech-shaped noise with the modulation characteristics of two-talker babble. The error bars represent ±1 standard deviation.

2.3.4 Regression analyses

Correlations between the various measures were calculated separately for the NH and the HI groups and indicated in Table 2.1. The MDTs variables and the MDD variable in the table represent the mean value of the MDTs and MDD thresholds across modulation frequencies, respectively. For the MDTs at 1 kHz,

only the thresholds below 128 Hz were used to compute the mean values.

For the NH listeners, there was only a significant correlation between the SRT obtained with the stationary noise masker and the mean MDD threshold. For the HI listeners, no significant correlation between the SRTs and the mean MDD threshold was observed. Figure 2.7 shows a scatterplot of the SRT for the steady noise masker against the mean MDD threshold. For the NH listeners, the SRT increases with increasing MDD threshold. For both listener groups, there was no significant correlation between the mean MDD threshold and the mean MDT for the 1-kHz carrier.

NH listeners	$N = 12$				
SRT_{fluc}	−0.04	−			
MDD1k	−0.85***	−0.19	−		
MDT1k	0.01	0.30	−0.07	−	
MDT5k	−0.14	0.20	−0.04	0.10	−
	SRT_{stat}	SRT_{fluc}	MDD1k	MDT1k	MDT5k
HI listeners	$N = 14$				
SRT_{fluc}	0.91***	−			
MDD1k	−0.01	−0.19	−		
MDT1k	0.18	0.34	0.37	−	
MDT5k	−0.08	0.08	0.03	0.28	−
	SRT_{stat}	SRT_{fluc}	MDD1k	MDT1k	MDT5k

Table 2.1: Correlation tables for the NH and the HI groups. SRTs for fluctuating noise and stationary noise are denoted by SRTfluc and SRTstat, respectively. The mean MDD across modulation frequency for the 1-kHz carrier is denoted by MDD1k. The mean MDTs across modulation frequency for the 1-kHz and 5-kHz carrier are denoted by MDT1k and MDT5k, respectively. Triple asterisk, ***, denote a significant correlation at $p < 0.001$. Double asterisk, **, denote a significant correlation at $0.01 > p \geq 0.001$. A single asterisk, *, denote a significant correlation at $0.05 > p \geq 0.01$. N denotes the number of listeners used in correlations.

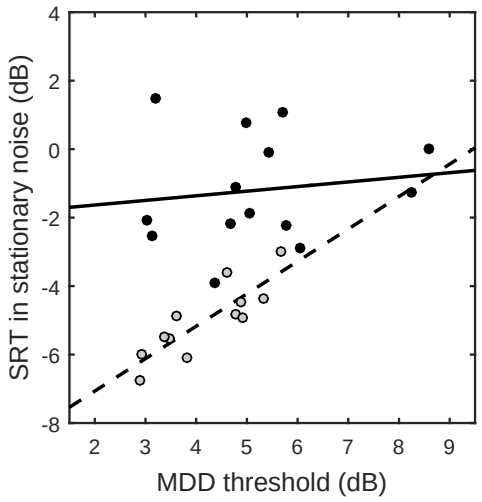


Figure 2.7: Scatterplot of the SRT in steady noise against the MDD threshold. Each symbol represents one listener. The gray and black symbols show results for the NH and HI listeners, respectively. The dashed and solid lines represent the regression lines for the two groups obtained from the correlation models.

2.4 Discussion

This study investigated modulation detection, modulation depth discrimination and speech intelligibility in noise in NH and HI listeners, both in conditions with and without fast-acting hearing-aid DRC. In the conditions without DRC, the HI listeners showed lower MDTs (for the 5-kHz carrier) than the NH listeners, consistent with the idea that the internal representation of the signal envelope in the impaired auditory system may be enhanced due to a loss of compression at the level of BM processing (e.g., Moore et al., 1996; Kale and Heinz, 2010). Hearing-aid DRC was found to increase the MDTs in the HI listeners, such that they were close to the corresponding MDTs observed in the NH listener group. Thus, proper DRC processing may be able to "restore" MDTs of HI listeners to the level observed in NH listeners.

In contrast, MDD thresholds were found to be higher for the HI listeners than for the NH listeners and there was no effect of DRC on the MDD thresh-

olds. Hence, the modulation depth reduction introduced by DRC is not able to restore the MDD thresholds of the HI listeners to the level of the NH listeners. This result is not consistent with the hypothesis that the differences between modulation depths are less noticeable for HI listeners than for NH listeners since the (enhanced) envelope fluctuations are processed at or near ceiling level (Schlittenlacher and Moore, 2016). DRC should have moved the "operation point" away from ceiling level to the "normal" range of processing levels, such that lower MDD thresholds for the HI listeners should have been expected when DRC was applied, compared to the condition without DRC. This was clearly not the case. Instead, the results seem to support the idea that two different internal noise sources can be assumed that limit different aspects of temporal resolution: An absolute internal noise term that limits MDTs (in the case of deterministic carriers, like the tonal carriers considered here) and another internal noise term that limits supra-threshold modulation processing sensitivity. Modeling has not been the focus of the present study, such that no specific model framework and quantitative predictions are presented here. However, a potential process could be a constant-amplitude internal noise after a logarithmic compression of the modulation amplitude, as represented in the framework of the EPSM (Ewert and Dau, 2000). In such a framework, the MDD threshold would be represented by the ratio of the signal modulation power and the internal noise power. This ratio may be considered as an "efficiency factor" of supra-threshold processing, conceptually similar to the efficiency factor represented in the classical power spectrum model of masking in the audio-frequency domain (Moore and Patterson, 1986). Thus, an increased MDD threshold would be accounted for by an increased signal-to-noise ratio in such a model, without affecting the MDT. Likewise, an enhancement of the internal representation of the envelope in the HI listeners due to a loss of BM compression, relative to normal hearing,

may affect the MDT but not necessarily the MDD threshold. This approach is conceptually similar to the concept of simulating the hearing threshold (or absolute detection threshold) vs the intensity discrimination threshold via different internal noise processes (e.g., Dau et al., 1996a; Dau et al., 1996b; Glasberg et al., 2001).

It is not clear which physiological processes may underlie a decreased "efficiency" (or increased signal-to-noise ratio required for signal detection) in terms of supra-threshold amplitude modulation processing. Increased MDD thresholds could be a consequence of a reduced population of auditory-nerve fibers and synaptic elements in the IHCs. Lopez-Poveda and Barrios (2013) proposed that the conversion of the envelope waveform to electrical discharges in the auditory nerve by the inner hair cells resembles a stochastic digitalization of the original continuous waveform. The ability to detect the envelope modulation may depend on the fidelity of the "discretization process" and thereby on whether the population of intact auditory-nerve fibers and synaptic elements in the IHCs is sufficient to capture differences in modulation-depths in terms of aggregated discharges. A reduced population of auditory-nerve fibers and synaptic elements in the IHCs may lead to higher thresholds in the MDD task. For the MDTs, this effect might be balanced by the loss of compression at the level of BM processing.

MDTs of NH listeners (using tonal carriers) have been shown to decrease with increasing SPL (Kohlrausch et al., 2000) and, in fact, MDTs for NH and HI listeners have been found to be similar when measured at the same SPL (Grose et al., 2016; Moore and Glasberg, 2001). Hence, the lower MDTs obtained for the HI listeners in the present study could be, at least to some extent, a consequence of the stimulation at a higher SPL and might thus not be an effect of hearing-impairment *per se*. In contrast, MDD thresholds were not found to be

affected by the presentation level, consistent with the results of Schlittenlacher and Moore (2016). Regarding the potential confounding influence of age, no correlation between age and the measures of temporal envelope sensitivity were found in any of the listener groups. This is also consistent with Schlittenlacher and Moore (2016) who did not find any clear effect of age on MDD thresholds.

The DRC processing considered in the present study was found to adversely affect speech intelligibility in stationary noise when the effect of audibility was controlled for, even though the effect was small. This suggests that the effect of restoring MDTs back to "normal" does not lead to improved speech perception performance which is consistent with earlier findings showing that MDTs only poorly correlate with speech intelligibility in HI listeners (e.g., Takahashi and Bacon, 1992; Feng et al., 2010). In contrast, interestingly, the observed strong correlation between speech intelligibility in stationary noise and MDD thresholds for the NH listeners suggests that MDD thresholds are linked to speech intelligibility performance. However, no correlation was observed for the HI listener group of the present study suggesting that increased MDD threshold may only reflect a single impairment factor related to speech-in-noise intelligibility. A supra-threshold hearing loss might be characterized by additional auditory processing deficits (such as, e.g., reduced frequency selectivity and/or deficits of temporal fine structure coding) as well as cognitive factors in the group of HI listeners that can be considered to be more heterogeneous (despite similar audiograms across the HI listeners) than the normal-hearing listener group. Such individual differences across measures of auditory and cognitive functions have not been assessed in the present study. Further investigations, including auditory modeling, may help to further clarify the relation between MDD processing, other auditory components characterizing supra-threshold listening, and speech intelligibility in noise both in normal-hearing, hearing-impaired

and aided hearing-impaired listeners.

2.5 Conclusions

This study investigated the effects of fast-acting DRC on NH and HI listeners' speech intelligibility and two measures of temporal envelope sensitivity (MDTs and MDD thresholds). It was found that the HI listeners showed lower MDTs (for the 5-kHz carrier) than the NH listeners, and fast-acting DRC was able to readjust the HI listeners' MDTs back toward the level observed in the NH listeners. This was consistent with the idea that the internal representation of the signal envelope in the impaired auditory system may be enhanced due to a loss of compression at the level of BM processing. In contrast, higher MDD thresholds were observed for the HI than the NH listeners, and fast-acting DRC had no effect on the MDD thresholds. This result was not consistent with the hypothesis that the differences between modulation depths are less noticeable for HI listeners than for NH listeners since the (enhanced) envelope fluctuations are processed at or near ceiling level. These results suggest that MDTs and MDD thresholds represent two largely independent aspects of temporal resolution. It was also found that there was no beneficial effect of restoring MDTs to the level observed in the NH listeners on speech perception, and that MDD thresholds appear to provide a stronger link to speech intelligibility performance than MDTs.

Acknowledgements

We thank Nicoline Thorup and Pernille Holtegaard for their assistance in recruiting the listeners with hearing loss. We further thank the Audiological department at Bispebjerg Hospital for providing support through their facilities and staff.

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.

3

Effects of expanding fast envelope fluctuations on consonant perception in hearing-impaired listeners[a]

Abstract The present study examined the perceptual consequences of three speech enhancement schemes based on multi-band nonlinear expansion of temporal envelope fluctuations between 10 and 20 Hz: (i) "idealized" envelope expansion of the speech before the addition of stationary background noise, (ii) envelope expansion of the noisy speech and (iii) envelope expansion of only those time-frequency segments of the noisy speech that exhibited signal-to-noise ratios (SNRs) above a certain limit. Linear processing was considered as a reference condition. The performance was evaluated by measuring consonant recognition and consonant confusions in normal-hearing and hearing-impaired listeners using consonant-vowel nonsense syllables presented in background noise. Envelope expansion of the noisy speech showed no significant effect on the overall consonant recognition performance relative to linear processing. In contrast, SNR-based envelope expansion of the noisy speech improved the overall consonant recognition performance equivalent to a 1 to 2-dB improvement in SNR, mainly by

[a] This chapter is based on Wiinberg, A, Zaar, J., Dau, T. (in review), *Trends in Hearing*.

improving the recognition of some of the stop consonants. The effect of the SNR-based envelope expansion was similar to the effect of envelope-expanding the clean speech before the addition of noise. SNR-based envelope expansion may therefore be valuable for speech-enhancement applications in hearing instruments.

3.1 Introduction

People with a sensorineural hearing impairment often complain about difficulties to understand speech in situations with several interfering talkers or background noise, particularly in reverberant environments. Some of these difficulties are considered to be caused by loudness recruitment, reflecting a reduced sensitivity to soft sounds and a steeper loudness growth function than observed in normal-hearing (NH) people (Fowler, 1936; Steinberg and Gardner, 1937). Modern hearing aids attempt to compensate for loudness recruitment by applying multi-band dynamic-range compression (DRC) that provides level-dependent amplification in various frequency bands, such that soft sounds are amplified more than higher-level portions of the sound. Apart from reduced audibility, cochlear hearing loss is often associated with a "distortion loss" that is considered to reflect supra-threshold processing deficits and assumed to be caused by inner hair-cell damage and/or loss of auditory-nerve fibers and synapses (e.g., Festen and Plomp, 1990; Plomp, 1978). One of the perceptual consequences of a distortion loss could be a reduced ability to capture and discriminate envelope fluctuations in a sound (e.g., Schlittenlacher and Moore, 2016; Wiinberg et al., 2015). The course of the envelope of speech in different frequency bands has been shown to be crucial for speech intelligibility (e.g., Shannon et al., 1995; Stone et al., 2008; Stone et al., 2012) and contains infor-

mation related to voicing, manner and place of articulation (Xu et al., 2005). In the case of a background noise, the modulation depth of the speech envelope becomes reduced due to the less varying noise envelope. This commonly deteriorates speech intelligibility, particularly in listeners with a hearing impairment (e.g., Stone et al., 2008; Stone et al., 2012).

It has been proposed that artificially increasing the modulation depth of the speech envelope may facilitate the extraction of speech cues and thereby improve speech intelligibility in noise (e.g., Plomp, 1988). Increasing the modulation depth of the speech envelope, without affecting the noise, would increase the signal-to-noise ratio (SNR) in the modulation domain, which has been shown to be related to speech intelligibility (Jørgensen et al., 2015). Consistent with this idea, recent speech intelligibility models based on the SNR in the modulation domain have been able to account for the effects of a large range of interferers and distortion types on speech intelligibility in NH listeners (Chabot-Leclerc et al., 2014; Chabot-Leclerc et al., 2016; Jørgensen and Dau, 2011; Jørgensen et al., 2013).

Different implementations of temporal envelope enhancement schemes have been investigated, with varying degree of success. Several studies found significant benefits from envelope expansion of the speech before the addition of noise both in NH and hearing-impaired (HI) listeners (Apoux et al., 2004; Langhans and Strube, 1982). However, the idealized processing of the speech before the addition of noise requires a priori knowledge of the clean speech signal, which cannot be assumed in practice (e.g., in hearing-aid signal processing schemes). If envelope expansion is instead applied to the noisy speech mixture, both the speech fluctuations and the intrinsic noise fluctuations are enhanced, such that no benefit in terms of the SNR in the modulation domain can be expected. In fact, consistent with this reasoning, several studies that

applied envelope expansion to the noisy speech showed no benefit or even a decreased performance relative to linear processing (Freyman and Nerbonne, 1996; Buuren et al., 1999) while others showed small benefits (e.g., Apoux et al., 2004; Clarkson and Bahgat, 1991). These results were typically consistent across NH and HI listeners when reduced audibility was compensated for by amplification. Part of the large variability regarding the benefit of envelope expansion across the different studies may have been caused by (i) differences in the details of the expansion schemes employed (e.g., the number of frequency bands, the range of modulation frequencies in which an expansion was applied, envelope thresholding, the amount of expansion, etc.), (ii) differences in the modulation spectra of the (stationary vs fluctuating noise) interferers and the speech material (e.g., sentences vs. CVs) as well as (iii) differences in the tested stimulus SNRs.

In most studies, the envelope expansion was applied to the "entire" modulation frequency range (e.g., between 0 and 500 Hz). The modulation power of long-term speech typically has a maximum around the syllabic rate, which is about 4 Hz for English, and decays thereafter with increasing modulation frequency (e.g., Plomp, 1988). Boosting modulation frequencies in this low-frequency range around the syllabic rate therefore enhances the overall dynamic range of the speech signal. Consequently, low-level speech segments are suppressed such that they may fall below the detection threshold while high-level speech segments may become uncomfortably loud, particularly for HI listeners with loudness recruitment. Therefore, audibility effects might contribute to the detrimental effects observed with some of the proposed expansion schemes. Using an alternative approach, Langhans and Strube (1982) applied expansion only at modulation frequencies above a lower cut-off modulation frequency of 2 Hz and provided dynamic range compression (DRC) for slow envelope fluc-

tuations (below 2 Hz). The idea behind this approach was that the DRC could compensate for loudness recruitment while the amplitude expansion could enhance speech envelope cues above 2 Hz. Langhans and Strube (1982) reported substantial benefits in normal-hearing (NH) listeners in terms of speech intelligibility when the processing was applied before the addition of noise. Even though this expansion scheme was successful in such idealized conditions, it might not be advantageous when applied at modulation frequencies as low as 2 Hz in the case of HI listeners with loudness recruitment since stimulus audibility might be affected. Alternatively, an enhancement of higher-frequency modulations (e.g. above 10 Hz) may increase the robustness of stop consonants and vowel onsets without compromising audibility. For example, the intelligibility of /t/ utterances has been shown to be highly correlated with the detectability of the transient in the release burst when presented in noise (Li et al., 2010; Régnier and Allen, 2008).

The present study investigated the effects of expanding modulation frequencies (in the range from 10 to 20 Hz) on consonant recognition and consonant confusions in NH and HI listeners using consonant-vowel (CV) nonsense syllables mixed with stationary Gaussian noise. It was hypothesized that the considered envelope expansion will improve the recognition of stop consonants and that detrimental effects due to the enhancement of noise fluctuations will be minimized if the envelope expansion processing is only applied to those time-frequency segments that are dominated by speech. Three different envelope expansion methods were tested: (i) “Idealized” envelope expansion of the speech before the addition of noise, (ii) envelope expansion of the noisy speech and (iii) envelope expansion of only those time-frequency segments of the noisy speech that exhibited SNRs above a certain limit. Linear processing was considered as the reference condition. Loss of audibility was compensated

for by providing individual linear frequency-dependent amplification for the HI listeners. The experimental data were analyzed with respect to overall and consonant-group specific consonant recognition scores, as well as in terms of listener-specific consonant recognition scores.

3.2 Methods

3.2.1 Listeners

Two groups of listeners participated in the experiments, a NH group and a HI group. The NH group consisted of eight adults with a median age of 26 years and ages ranging from 21 to 61 years. All had absolute thresholds below 20 dB HL for the octave frequencies between 0.125 and 8 kHz. The HI group consisted of 12 adults with symmetrical mild-to-moderately-severe sensorineural hearing losses. The median age was 72 years and the range was 50 to 80 years. The absolute thresholds for the test ear, measured using conventional audiometry, are shown in Figure 3.1. All listeners reported Danish as their first language, signed an informed consent document and were reimbursed for their efforts. Approval for the study was granted by the Scientific Ethical Committee of the Capital Region in Denmark ("De Videnskabsetiske Komitéer for Region Hovedstaden").

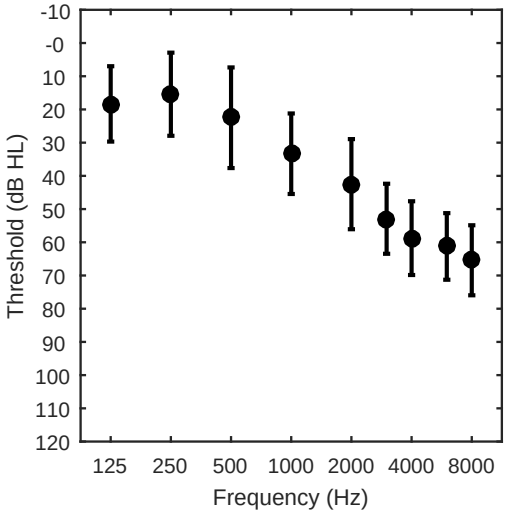


Figure 3.1: Mean absolute thresholds for the tested ear of the HI listeners, measured using conventional manual audiometry, and expressed in dB HL. Error bars represent ±1 standard deviation.

3.2.2 Stimuli

The CV syllables consisted of 15 consonants (/p, t, k, b, d, g, f, s, ʃ, v, j, l, h, m, n/) followed by the vowel /i/. Two tokens (one recording of a female talker and one of a male talker) were selected per CV from the Pitu Danish nonsense syllable speech material (Christiansen, 2011), amounting to 30 tokens overall (15 CVs × two talkers). The tokens represent a subset of the speech tokens used in a recent study on consonant perception in white noise (Zaar and Dau, 2015) which considered three recordings of each CV per talker. For each CV, the most intelligible recording of each talker was selected in the present study. The levels of the tokens were equalized using VUSOFT, a software implementation of an analog VU-meter developed by (Lobdell and Allen, 2007), such that all CVs showed the same VUSOFT peak value. This equalization strategy is mainly based on the vowel levels, thus ensuring realistic relations between the levels of the individual consonants. After equalization, the reference speech level for the SNR calculation was defined as the overall root-mean-square (RMS) level

averaged across all speech tokens.

SNR conditions of 12, 6 and 0 dB were generated by fixing the noise level and adjusting the level of the speech tokens based on the reference speech level according to the desired SNR. The speech tokens were mixed with stationary Gaussian noise such that the speech token onset was temporally positioned 400 ms after the noise onset. The stimulus duration was 1 s, including 50-ms raised-cosine onset and offset ramps for the noise. The sound pressure level (SPL) of the noise was set to 65 dB, while the overall stimulus level differed depending on the level of the speech, i.e., on the SNR. Envelope expanded signals (clean speech or noisy speech) were equalized in RMS level with the corresponding signals obtained without expansion processing. For the HI listeners, the stimuli were linearly amplified according to the NAL-R(P) frequency-dependent prescription rule based on their individual audiometric thresholds (Byrne et al., 1990). The frequency-dependent amplification was provided using a bank of seven octave-wide bandpass linear-phase, finite-impulse-response filters with center frequencies between 0.125 and 8 kHz.

3.2.3 Envelope expansion processing

The proposed multiband envelope expansion algorithm, depicted in Figure 3.2, is similar to the algorithm described in Langhans and Strube (1982). The input signal was short-time Fourier transformed by Hann-windowing the signal in time frames of 256 samples with 75% overlap between frames using a sampling rate of 44100 Hz. Each of the windowed segments was padded with 128 zeros at the beginning and the end and transformed to the spectral domain using a 512-point fast Fourier transform (FFT). The power spectral density of the resulting frequency bins was combined into 15 third-octave wide frequency bands with center frequencies between 0.323 to 8.192 kHz. The power in each band

was converted to dB SPL, and the resulting logarithmic representation of the temporal envelope was bandpass filtered over time-frames using a zero-phase fourth-order Chebyshev Type II filter (-24 dB/oct. roll-off) with 3-dB cut-off frequencies at 10 Hz and 20 Hz. The band-wise expansion gains per time-frame were computed by multiplying the bandpass-filtered envelopes with a scaling factor of 1.3. Thus, a bandpass-filtered level of 1 dB resulted in an amplification of the output level by 1.3 dB. The value of the scaling factor was based on data from Wiinberg et al. (2015). The factor was chosen such that the expansion processing restored the average modulation-depth discrimination performance of the HI listener group to that of the NH listener group at a modulation frequency of 16 Hz. The band-wise gains were converted to linear units and 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 envelope-expanded stimuli. These time segments were subsequently windowed with a Hann-window to avoid aliasing artifacts and combined using an overlap-add method to provide the processed temporal waveform.

For the SNR-based expansion scheme, a-priori information about the speech and noise components of the noisy speech mixture was used. The power of both the speech and noise components was computed for each of the fifteen frequency bands and the SNR was calculated in dB. For time-frequency segments with SNRs below -10 dB, the expansion gain was set to 0 dB. Otherwise, the expansion gain was not changed.

As listed in Table 3.1, three different envelope expansion settings were tested: Envelope expansion of the noisy speech ("Exp_{mix}"); envelope expansion applied to time-frequency segments with SNRs above -10 dB ("Exp_{SNR}") and envelope

expansion of the speech before the addition of noise ("$\text{Exp}_{\text{speech}}$").

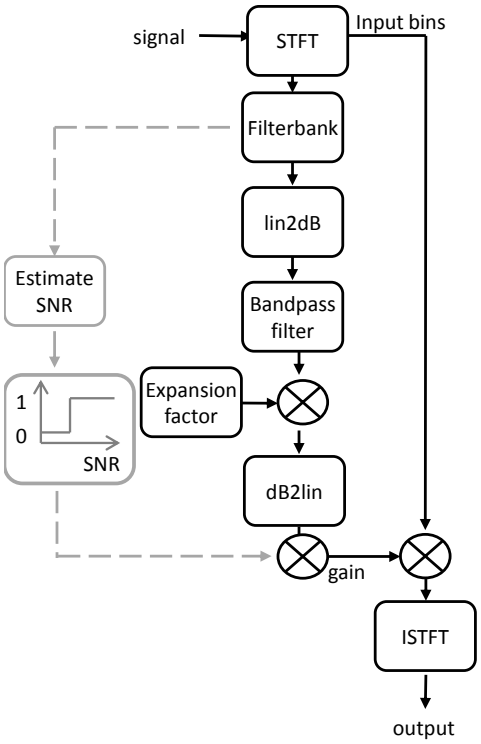


Figure 3.2: Block diagram of the proposed envelope expansion scheme. First, the signal was windowed in time segments and transformed into the frequency domain by a short-time Fourier transform (STFT). The frequency bins in each time window were combined into fifteen third-octave spaced frequency bands ("Filterbank"). The power in each band was converted to dB SPL ("lin2dB") and bandpass filtered ("Bandpass filter"). The filtered temporal envelope was then multiplied by an expansion factor of 1.3. The bandpass-filtered temporal envelope in each of the frequency bands was converted to linear units ("dB2lin") and thereafter used as gain values for the input. For the SNR-based expansion scheme, indicated in grey, the gain was set to 0 dB for time-frequency bands with SNRs below a certain limit. Finally, an inverse STFT (ISTFT) was computed to generate the final expanded signal.

Abbreviation	Processing	Expander Mode
Exp_{mix}	Noisy speech	Envelope expansion
Exp_{SNR}	Noisy speech	SNR-based envelope expansion
Exp_{speech}	Clean speech	Envelope expansion

Table 3.1: Overview of the three different envelope expansion conditions.

3.2.4 Experimental design

A control condition with speech presented in quiet was defined as "Q65". This control condition was included to evaluate whether the CV tokens were sufficiently audible in quiet at the lowest speech level occurring in the SNR conditions. The clean speech (without envelope enhancement) was therefore presented at 65 dB SPL for the NH listeners and 65 dB SPL + NAL-R(P) amplification for the HI listeners, corresponding to the speech level in the 0-dB SNR condition. The experimental sessions were split into four consecutive blocks corresponding to the four signal-processing conditions (Lin, Exp_{mix}, Exp_{SNR} and Exp_{speech}). For each of the listener groups, the order of presentation for the experimental blocks was counterbalanced using a Latin-square design to control for order effects. In order to get the listeners accustomed to the task, the "easy" control listening condition Q65 was presented first. Within each of the succeeding four experimental blocks, the three SNR conditions ranked from easy to difficult, i.e., with SNR tested in the order 12, 6 and 0 dB. For each of the SNR conditions, the 30 CV tokens were presented in random order within each of five repetition blocks. This was done to facilitate the evaluation of potential learning effects.

3.2.5 Procedure and apparatus

All signals were generated digitally in MATLAB (Version 2015b; The MathWorks. Inc., Natick Massachusetts, United States) on a PC equipped with a RME UCX Fireface sound card at a sampling rate of 44.1 kHz and with a resolution of 16

bits per sample. The stimuli were presented in a sound-attenuating booth via Sennheiser HD 650 headphones to the better ear of the listeners, as derived from the average of the audiometric thresholds at 500 Hz, 1000 Hz and 2000 Hz. The transfer function of each earpiece of the headphones was digitally equalized (101 point FIR filter) to produce a flat frequency response for frequencies between 0.100 and 10 kHz, measured with an ear simulator (B&K 4153) and a flat plate adaptor as specified in IEC 60318-1 (2009).

3.2.6 Statistical analysis

An analysis of variance (ANOVA) was conducted on a mixed-effect model to evaluate whether hearing status, SNR and processing condition had an effect on consonant recognition performance. In the mixed-effect model, listeners were nested within hearing status (normal hearing vs. impaired hearing). Listeners and repetitions were treated as random block effects, while SNR, processing condition, and hearing status were treated as fixed effects. The random-listener effect accommodates the repeated-measures design by assuming that observations from the same listener are correlated. The assumptions underlying a parametric analysis were met without transforming the dependent variable. Tukey's Honestly Significant Difference (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. The statistical analysis was performed using the lme4 and lsmeans packages in R (Bates et al., 2015; Lenth, 2016)

3.3 Results

3.3.1 Consonant recognition scores of NH and HI listeners

Figure 3.3 shows the consonant recognition scores obtained with the four different signal-processing conditions for the NH listeners (left panel) and the HI listeners (right panel) as a function of the SNR. The consonant recognition scores were calculated as the mean percent correct across all consonants, talkers, repetitions and listeners for both listener groups. For all SNRs and processing conditions, the consonant recognition scores were poorer for the HI than for the NH listeners. The scores generally increased with SNR and reached their maximum value for the quiet condition. The results showed that, for both listener groups, the two expansion conditions $\text{Exp}_{\text{speech}}$ (squares) and Exp_{SNR} (triangles) provided a small but consistent improvement relative to linear processing (asterisks) except for the $\text{Exp}_{\text{speech}}$ results for the NH listeners at 12 dB SNR where a slightly detrimental effect was found. In contrast, the condition Exp_{mix} (circles) provided a small improvement for the NH listeners but not for the HI listeners.

The outcomes of the ANOVA, summarized in Table 3.2, showed main effects of hearing status, SNR and processing condition as well as an interaction between hearing impairment and SNR. The effects of the envelope expansion schemes were largely consistent across NH and HI listeners. Post-hoc comparisons confirmed that the consonant recognition performance was improved in the $\text{Exp}_{\text{speech}}$ condition (by 1.8 percent points, p=0.008) and the Exp_{SNR} condition (by 2.1 percent points, p=0.001), relative to the linear processing condition. The standard error was 0.5 percent points in both cases. In contrast, the consonant recognition scores for the Exp_{mix} and linear processing conditions were not significantly different (p=0.99). There were no significant differences between the consonant recognition scores for the $\text{Exp}_{\text{speech}}$ and Exp_{SNR} conditions

(p=0.95), but the scores in both of these conditions were significantly higher than in the Exp_{mix} condition (p<0.01). However, there was a trend (p=0.08) towards a significant interaction between processing condition and SNR, indicating differences between Exp_{speech} and Exp_{SNR} in terms of the relative benefit provided as a function of SNR.

An alternative, more familiar, performance measure is the change in SNR corresponding to the improvement in recognition scores. The statistical analysis of the data (shown in Figure 3.3) indicates that the improvement in terms of percent correct for the Exp_{SNR} and Exp_{speech} conditions versus linear processing was roughly constant across the tested SNRs, as there was no interaction between processing condition and SNR. Psychometric functions fitted to the data points obtained with linear processing in Fig. 3.3 revealed that the recognition-score improvement for the Exp_{SNR} and Exp_{speech} conditions relative to linear processing was equivalent to a 1-dB change in SNR for the NH listeners. For the HI listeners, this improvement amounted to a 1.9-dB change in SNR. The difference in SNR improvement between the two listener groups, despite similar recognition-score improvements, was due to differences in the slopes of the respective psychometric functions, which were shallower for the HI listeners than for the NH listeners.

Figure 3.4 compares the consonant recognition scores obtained in the linear reference condition to those obtained in the three expansion conditions. To evaluate how the individual expansion schemes affect different phonetic categories, the recognition scores were averaged within the categories /p,k,t/ (blue), /b,g,d/ (green), /f,s,$\int$,v/ (red), /n,m/ (black), and /h,j,l/ (cyan). The average recognition scores obtained with the three expansion schemes (left: Exp_{mix}, middle: Exp_{SNR}, right: Exp_{speech}) are shown as a function of the average recognition scores obtained with linear processing. The results for the NH

listeners are shown in the top panels and the results for the HI listeners are shown in the bottom panels. None of the expansion schemes had a detrimental effect on the recognition scores in the NH listeners, as no points fall below the diagonals in the top panels of Fig. 3.4. As expected, the recognition scores were mainly increased for the stop consonants /p,k,t/ (blue). This improvement was largest for Exp_{SNR} (upper middle panel), slightly smaller for Exp_{mix} (upper right panel) and small for the "ideal"$\text{Exp}_{\text{speech}}$ (top right panel). In contrast to the NH listeners, the expansion schemes had a detrimental effect on the HI listeners for the consonant groups /n,m/ and /b,g,d/ (bottom panels of Figure 3.4). However, similar to the effects observed in the NH listeners, the recognition scores for /p,k,t/ (blue) were increased substantially in all expansion conditions. The effects of the expansion schemes varied strongly across the consonant groups. Interestingly, Exp_{mix} did not affect consonant recognition for the fricatives /f,s,$\int$,v/ (red dot, lower left panel), whereas Exp_{SNR} (red dot, lower middle panel) and $\text{Exp}_{\text{speech}}$ (red dot, lower right panel) provided a benefit of 5% and 4%, respectively. Overall, the SNR-based expansion Exp_{SNR} provided the largest benefits for /f,s,$\int$,v/ and the smallest detrimental effects (−2% for /n,m/ and −3% for /b,g,d/) in the HI listeners.

Df	F	Ratio	Probability
Proc. condition	(3, 1133)	7.68	< 0.001
SNR	(2, 36)	292.93	< 0.001
Hearing status	(1,18)	26.04	< 0.001
Hearing status × SNR	(2, 36)	22.38	< 0.001
Proc. condition × SNR	(6, 1127)	1.89	0.08
Proc. condition × Hearing status	(3, 1124)	2.08	0.10
Hearing status × Proc. condition × SNR	(6, 1118)	1.26	0.27

Table 3.2: Summary of the ANOVA outcomes for a mixed-effect model fitted to the consonant recognition data with a between-listener factor of hearing status, and within-listener factors of SNR and processing condition.

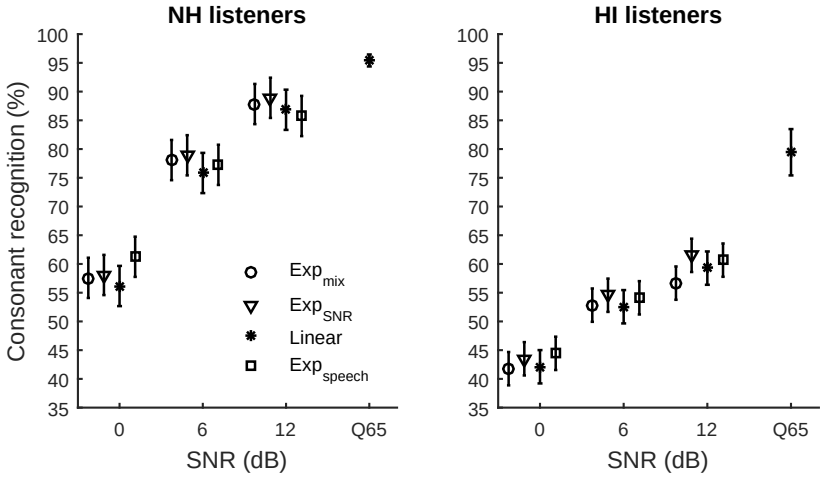


Figure 3.3: Overall consonant recognition scores for the NH listeners (left) and the HI listeners (right) as a function of the SNR for the four different signal-processing conditions. (Circles: Exp_{mix}, triangles: Exp_{SNR}, asterisks: Linear Processing, squares: $\text{Exp}_{\text{speech}}$). The error bars represent ±1 standard errors of the mean. A slight horizontal jitter was added to the data for better readability.

3.3.2 Individual listener analysis and confusion matrices

The above analysis focused on group averages, showing moderate improvements of consonant recognition scores induced by the envelope expansion on a group level. However, the individual listeners may have experienced largely different benefits from the expansion processing. To analyze the individual differences in benefit, Figure 3.5 shows a scatter plot of the across-SNR average consonant recognition performance with linear processing on the abscissa and the across-SNR average performance with Exp_{SNR} on the ordinate. Each symbol in Figure 3.5 represents the result for an individual listener (circles: NH; triangles: HI). The scatter plot reveals that the improvement in the overall recognition performance for the Exp_{SNR} conditions was mainly driven by the six listeners (four HI and two NH) for whom the expansion processing was most beneficial (on average 6.8% and 9.4%, respectively, for the Exp_{SNR} and $\text{Exp}_{\text{speech}}$ conditions). For the 14 other listeners, the expansion processing affected the consonant recognition performance by less than ±5 percent points (0.1% and

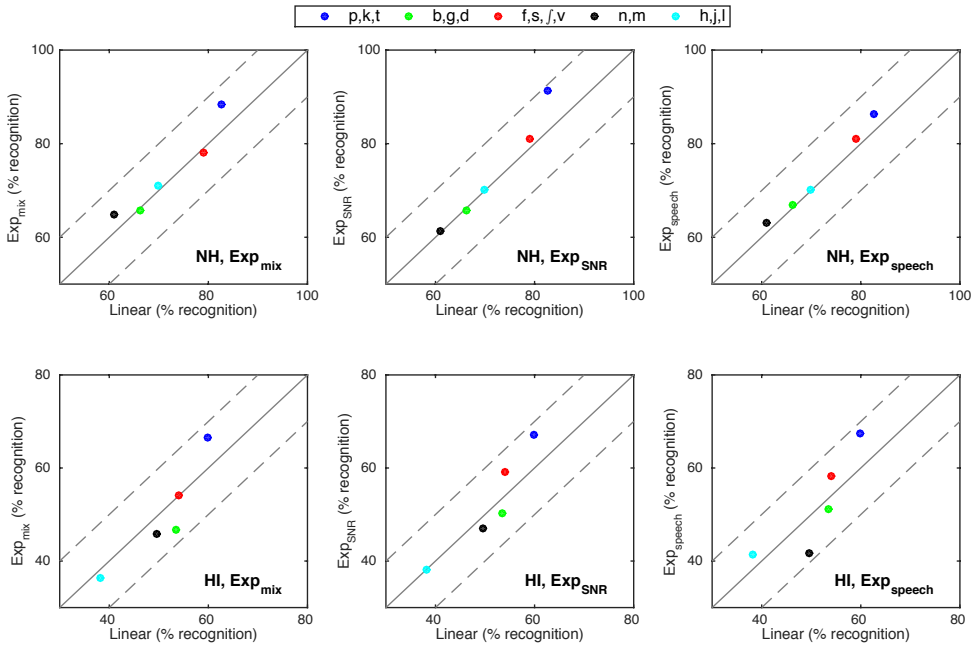


Figure 3.4: Scatter plot of consonant recognition in % measured with the linear condition vs. the three expansion conditions (from left to right: Exp_{mix}, Exp_{SNR}, and $\text{Exp}_{\text{speech}}$) in NH (top) and HI (bottom) listeners. The consonant recognition scores were averaged within the phonetic categories shown in the legend. Within each panel, the solid gray line represents equal performance while the dashed lines represent ±10% differences induced by the respective expansion scheme.

−1.0% on average for Exp_{SNR} and $\text{Exp}_{\text{speech}}$, respectively).

3.4 Discussion

Increasing the modulation depth of the speech envelope has been suggested to facilitate the extraction of speech cues and thereby improve speech intelligibility in noise. In the present study, the effects of three different envelope expansion schemes that increase the depth of envelope fluctuations between 10 and 20 Hz were tested in a consonant identification task. Envelope expansion of the noisy speech showed no significant effect on the overall consonant recognition performance relative to linear processing, neither for the NH nor the HI listeners. This finding is consistent with results from earlier studies that investigated the

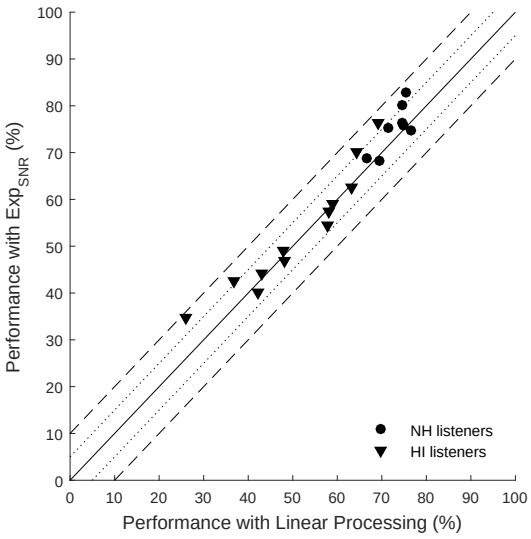


Figure 3.5: Scatter plot of consonant recognition performance with Exp_{SNR} as a function of recognition performance with linear processing. Circles and triangles show results for NH and HI listeners, respectively. For visual clarity, the data were averaged across SNR conditions. The solid, dotted and dashed lines represent equal performance, ±5 percent-point improvements, and ±10 percent-point improvements, respectively, obtained with the expansion processing relative to linear processing.

effect of expanding the envelope of noisy speech (Apoux et al., 2004; Clarkson and Bahgat, 1991; Buuren et al., 1999). While the processing improved the intelligibility of some of the plosives by enhancing the detectability of the transient release bursts, this was accompanied by an increased proportion of consonant confusions for the other consonant categories. In contrast, SNR-based envelope expansion of the noisy speech, which confined the enhancement to the time-frequency segments in which the speech power was prominent, improved the overall consonant recognition performance both for the NH and the HI listeners. Interestingly, the effect of the SNR-based envelope expansion was found to be similar to the effect of envelope-expanding the clean speech before the addition of noise.

While the expansion benefit in terms of recognition-score improvement was substantial for some listeners (about 10 percent point), the average effect

for the entire population was relatively small (about two percent points improvement of consonant recognition). Nevertheless, the observation of similar results obtained with the SNR-based processing and the clean-speech envelope expansion is promising, given that previous studies reported substantial improvements in speech perception with envelope expansion of clean speech (e.g., Apoux et al., 2004; Langhans and Strube, 1982). This suggests that the SNR-based envelope expansion scheme proposed in the present study could provide larger improvements in speech perception when combined with alternative parameter settings, such as those considered in the earlier studies. In contrast to the expansion of clean speech before the addition of noise, SNR-based envelope expansion is feasible in hearing-aid algorithms using blind SNR-estimation methods (e.g., Gerkmann and Hendriks, 2012; Martin, 2001) such that this approach might help improve speech perception in hearing-aid users.

When expressing the changes in consonant recognition performance as equivalent change in SNR, the net effect was an improvement (relative to the linear processing condition) that was about 1-dB larger for the HI listeners than for the NH listeners in the Exp_{SNR} and $\text{Exp}_{\text{speech}}$ conditions. Thus, a larger increase in SNR is required for the HI listeners than for the NH listeners to obtain the same increase of the recognition score. The benefit achieved with the expansion processing may thus be larger for HI listeners than for NH listeners. However, part of the differences between the results obtained with the NH and HI listeners may be attributed to effects of age which was not controlled for in the present study.

The reason for the relatively small (yet statistically significant) improvements in consonant recognition performance induced by the proposed envelope expansion processing may be related to the chosen parameter settings. Consistent

with the results from the present study where modulation frequencies between 10 and 20 Hz were enhanced, envelope filtering studies have demonstrated that the contribution of envelope fluctuations above 12 Hz to phoneme intelligibility is small in quiet and in stationary background noise (Drullman et al., 1994; Xu et al., 2005; Xu and Zheng, 2007). However, ceiling effects were observed in those studies and it thus remained unclear whether this finding could be reproduced if the experiment was not confounded by such effects. In contrast, in terms of sentence intelligibility, the envelope expansion of (clean) speech has been shown to provide a greater benefit when applied to a wider range of modulation frequencies. For example, Apoux et al. (2004) showed that their expansion processing of modulation frequencies in the range 0-256 Hz was more effective than in the range 0-16 Hz. Therefore, it is possible that an additional enhancement of a wider range of modulation frequencies (below 10 Hz and above 20 Hz) would increase the benefit provided by the expansion processing. However, it should be taken into account that while expanding clean speech up to modulation frequencies in the range of the fundamental frequency (about 100-200 Hz) may yield more robust periodicity information, this approach might be detrimental when applied to a noisy signal where the SNR in the modulation domain typically decreases monotonically with increasing modulation frequency. In any case, this type of expansion would still require that modulation frequencies below the syllabic rate are not enhanced to avoid compromising audibility for the HI listeners. Furthermore, expansion of slow envelope fluctuations tends to decrease the consonant-vowel intensity ratio (CVR) as the processing enhances high-intensity vowels more than low-intensity consonants (Apoux et al., 2004) which, in turn, may affect consonant recognition performance (Freyman and Nerbonne, 1989). A possible solution may be to apply expansion processing at modulation frequencies between 4 and 256 Hz in combination with amplitude

compression of the slow envelope fluctuations below 4 Hz, such that the CVR is increased as compared to the case where only expansion is applied.

The rationale for using stationary background noise rather than fluctuating background noise in the present study was to maximize the benefit provided by the expansion processing in terms of consonant recognition. This expectation was based on the results from Apoux et al. (2004) who found larger benefits provided by their expansion processing in terms of word recognition scores in stationary noise than in fluctuating noise. Furthermore, supra-processing deficits have been shown to provide stronger links to speech intelligibility in stationary noise than in fluctuating noise (e.g. Van Esch and Dreschler, 2015). Hence, the effect of the proposed expansion schemes may be smaller for fluctuating background noise maskers with a more similar modulation spectrum to the target speech. However, it should be noted that the bandpass filtering applied in the expansion algorithm corresponds to low-pass filtering in the modulation domain, such that the individual frequency bands of the noise considered for envelope expansion were in fact highly modulated. This makes the distinction between stationary and fluctuating noise less prominent than in the case of a wideband envelope expansion scheme as used in the Apoux et al. (2004) study.

It has been demonstrated that listeners can learn to adapt to artificially produced, non-linear changes of the natural auditory cues that are used for auditory perception. For example, frequency-lowering signal processing strategies have been implemented in hearing aids. Frequency lowering shifts acoustic cues from high-frequency regions to lower frequencies where audibility is typically better, thereby potentially improving the listener's access to the speech cues (see Simpson, 2009, for a review). Several studies have indicated that a period of acclimatization was necessary before frequency lowering provided benefits in

speech recognition (Ellis and Munro, 2015; Glista et al., 2012; Wolfe et al., 2011). Thus, the benefit of non-linear signal processing schemes, such as envelope expansion, may not be immediately apparent when assessed without a period of acclimatization.

The observed improvements in consonant recognition performance induced by the proposed envelope expansion schemes were found in a subgroup of the listeners, i.e., only selected listeners benefited from this type of processing. Further research is needed to clarify why these differences in benefit occur and to establish whether they can be related to individual differences in psychoacoustic measures, such as temporal envelope detection and discrimination (e.g., Schlittenlacher and Moore, 2016; Wiinberg et al., 2015) or in terms of acclimatization to the processing.

3.5 Conclusion

The present study investigated the effect of expanding envelope fluctuations between 10 and 20 Hz on consonant recognition performance in NH and HI listeners. Envelope expansion of noisy speech showed no significant effect on the overall consonant recognition performance relative to linear processing. In contrast, SNR-based envelope expansion of the noisy speech improved the overall consonant recognition performance by about 2 percent points, mainly resulting from an improved recognition of some of the stop consonants. If the change in performance was expressed in terms of equivalent change in SNR, the net effect was an improvement (relative to the linear condition) of 1 dB and 1.9 dB for the NH and HI listeners, respectively. The effect of the SNR-based envelope expansion was comparable to the effect of “idealized” envelope expansion of the clean speech before the addition of noise. The size

of the measured effects was relatively small compared to other related studies, indicating that extending the enhanced modulation-frequency range from 10-20 Hz to, e.g., 4-20 Hz might yield larger benefits. Overall, the results support the hypothesis that the detrimental effect of enhancing the noise fluctuations in the different frequency bands on speech perception is effectively reduced by SNR-based envelope expansion. Furthermore, the results suggest that, due to its practical feasibility, the proposed SNR-based envelope expansion scheme may be interesting for speech-enhancement applications in hearing aids.

Acknowledgement

We would like to thank Morten Løve Jepsen and Christoph Scheidiger for helpful discussions. This project was carried out in connection to the Centre for Applied Hearing Research (CAHR) supported by Widex, Oticon, GN Hearing and the Technical University of Denmark.

4

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 is based on Hassager, H. G., Wiinberg, A, Dau, T. (2017), *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.

4.1 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 (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 (see also Brown et al. (see also 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 compression (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.

4.2 Methods

4.2.1 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 4.1 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 exter-

nalized, 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.

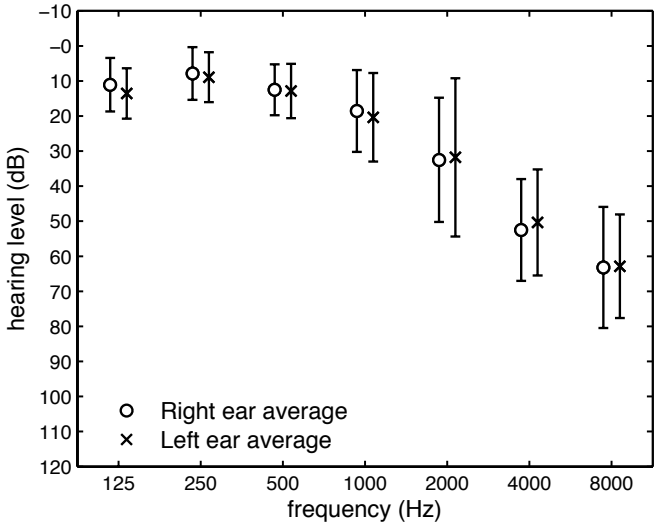


Figure 4.1: Audiometric pure-tone threshold averages for the right and left ear of the hearing-impaired listeners. The error bars represent one standard deviation of the thresholds.

4.2.2 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 4.2 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.

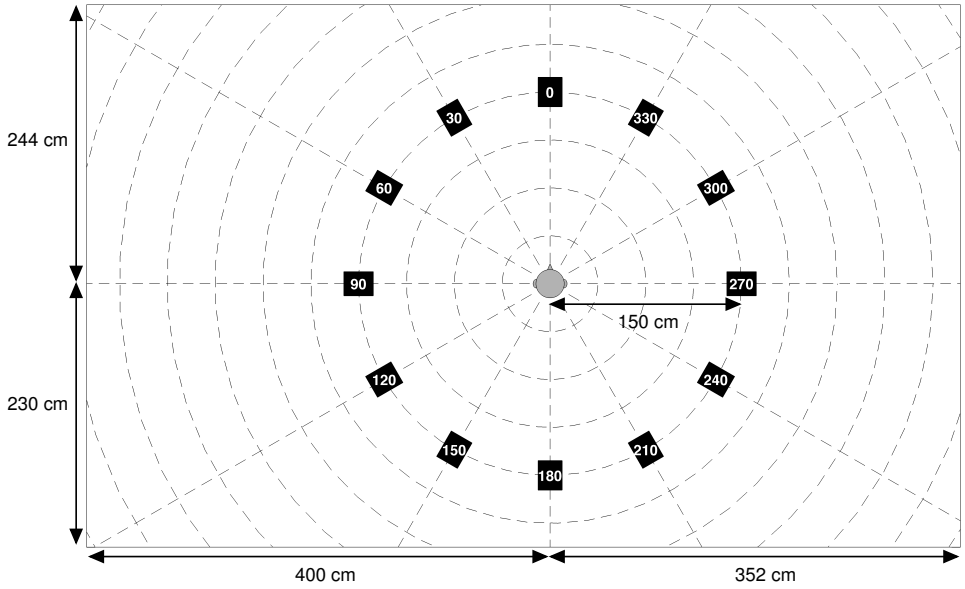


Figure 4.2: The top view of the experimental setup. The loudspeaker positions are indicated by the black squares. The grey circle in the center indicates the position of the chair, where the listener was seated. The listeners had a view direction on the loudspeaker placed at the 0° degree azimuth. The 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. 4.2 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 4.2.3). 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.

4.2.3 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, $h_{\text{hpir}}^{\text{inv}}$, 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, $h_{\text{hpir}}^{\text{inv}}$, 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 serving as training in evaluating the auditory perception on the graphical user interface.

4.2.4 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 4.2.5. Figure 4.3 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, $h_{\text{brir,l}}$, and its right part, $h_{\text{brir,r}}$. In the independent compression scheme (top panel), the input signal, s_{in}, was first convolved with $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$ and then passed through two DRC systems operating independently in each ear. In the linked compression scheme (middle panel), after

convolving with $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$ 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 $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$. 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 $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$). In practice, since the dry signal is typically not available, such a system would require a deconvolution of $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$ before compression, followed by a convolution with $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$ to provide the listener with the spatial cues.

To create the signals for the condition with linear processing, the stimuli were convolved with $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$. To compensate for the effect of the headphones, the outputs $s_{\text{out,l}}$ and $s_{\text{out,r}}$ in all conditions were convolved with $h^{\text{inv}}_{\text{hpir,l}}$ and $h^{\text{inv}}_{\text{hpir,r}}$, respectively, i.e. the left and right parts of $h^{\text{inv}}_{\text{hpir}}$. 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.

4.2.5 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

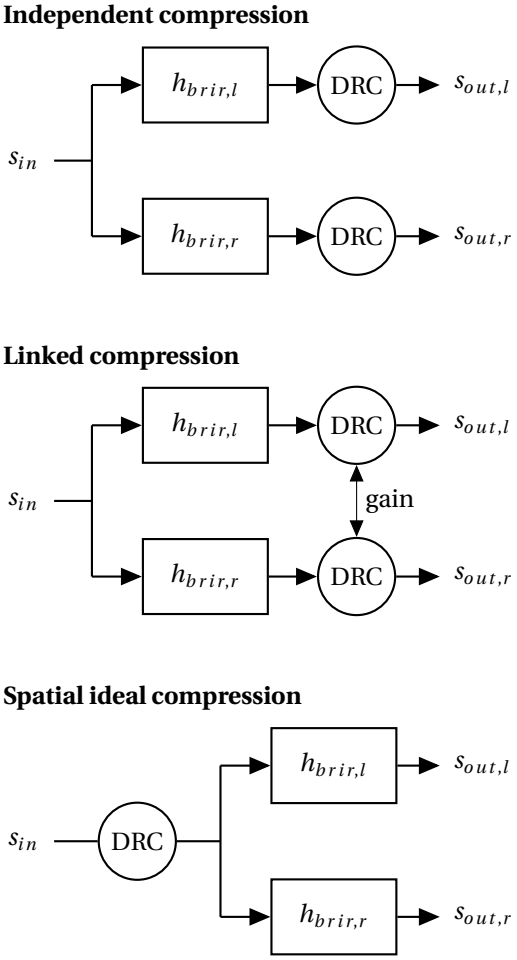


Figure 4.3: Block diagrams of the three compression conditions: Independent compression (top), linked compression (middle) and spatially ideal compression (bottom). For the independent and linked compression schemes, the dry signal, s_{in}, is convolved with the left and right BRIR, $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$, 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, $s_{\text{out,l}}$ and $s_{\text{out,r}}$, respectively.

step 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 International Electrotechnical Commission (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 Table4.1 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 4.1: The compression thresholds (CT) and compression ratios (CR) in the seven octave frequency bands.

4.2.6 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.

4.2.7 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 cues

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 envelopes were calculated by taking the absolute values of the complex outputs of the different channels. The envelopes were windowed in time using a 20 ms rectangular window and an overlap of 50%. The power of the windowed segments was calculated and converted to dB SPL. The ILD histograms were subsequently computed by subtracting the level for the left ear from the level for the right ear for those time segments where both the left- and right-ear SPL were above 0 dB SPL. The ILD distributions were estimated by applying a Gaussian kernel-smoothing window with a width of 0.9 dB on the ILD histogram.

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

$$\mathrm{IC} = \max_{\tau} \left| \frac{\sum_t s_{\mathrm{out,l}}(t+\tau) s_{\mathrm{out,r}}(t)}{\sqrt{\sum_t s^2_{\mathrm{out,l}}(t) \sum_t s^2_{\mathrm{out,r}}(t)}} \right|. \tag{4.1}$$

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, $h_{\text{brir;dir}}$, was defined as the first 2.5 ms of the impulse response, and the reverberant part, $h_{\text{brir;reverb}}$, 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 $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$ (in Fig. 4.3) were replaced by their direct parts $h_{\text{brir;dir,l}}$ and $h_{\text{brir;dir,r}}$ and the extracted gain values were applied such that the outputs $s_{\text{out;dir,l}}$ and $s_{\text{out;dir,r}}$ only contained the effect of the compression on the direct part of the signal. Correspondingly, the outputs $s_{\text{out;reverb,l}}$ and $s_{\text{out;reverb,r}}$, representing the outputs that contained the effect of the compression on the reverberant part of the signal, were obtained by replacing the impulse responses $h_{\text{brir,l}}$ and $h_{\text{brir,r}}$ with their reverberant parts $h_{\text{brir;reverb,l}}$ and $h_{\text{brir;reverb,r}}$. 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 $s_{\text{out;dry,l}}$ and $s_{\text{out;dry,r}}$ 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:

$$DRR_k = 10 \cdot \log_{10} \left(\frac{\sum_f \frac{\left|S_{\text{out;dir,k}}(f)\right|^2}{\left|S_{\text{out;dry,k}}(f)\right|^2}}{\sum_f \frac{\left|S_{\text{out;reverb,k}}(f)\right|^2}{\left|S_{\text{out;dry,k}}(f)\right|^2}} \right), \tag{4.2}$$

where $S_{\text{out;dir,k}}(f)$, $S_{\text{out;reverb,k}}(f)$, and $S_{\text{out;dry,k}}(f)$ indicate the frequency-domain versions of the time signals $s_{\text{out;dir,k}}$, $s_{\text{out;reverb,k}}$, and $s_{\text{out;dry,k}}$ 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 ($h_{\text{brir;dir,l}}$ and $h_{\text{brir;dir,r}}$) and the reverberant part ($h_{\text{brir;reverb,l}}$ and $h_{\text{brir;reverb,r}}$) of the BRIR, respectively. DRRs were calculated for the frequency range from 100 Hz to 10 kHz.

4.3 Results

Experimental data

Figure 4.4 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. 4.4), 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 substantially 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 4.5 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. 4.4). 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. 4.5) 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

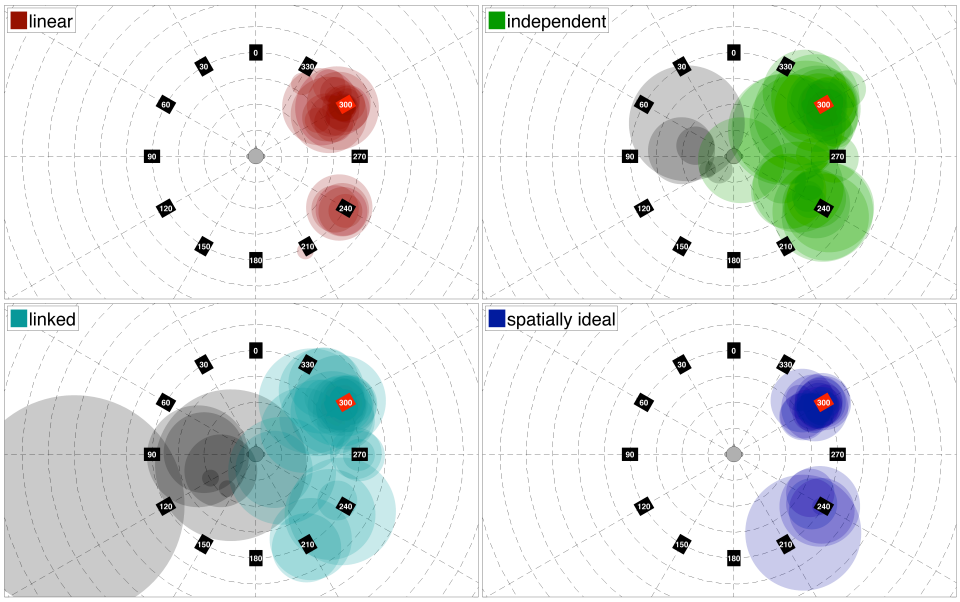


Figure 4.4: Graphical representations of the normal-hearing listeners' responses obtained with the speech stimulus virtually presented from the 300 degrees position in the listening room. The upper left panel shows the results for linear processing (reference condition). The results for independent, 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.

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.

The results obtained with the transients are shown in Fig. 4.6 for the normal-hearing listeners and Fig. 4.7 for the hearing-impaired listeners. The general pattern of results across conditions was similar to that observed for the speech

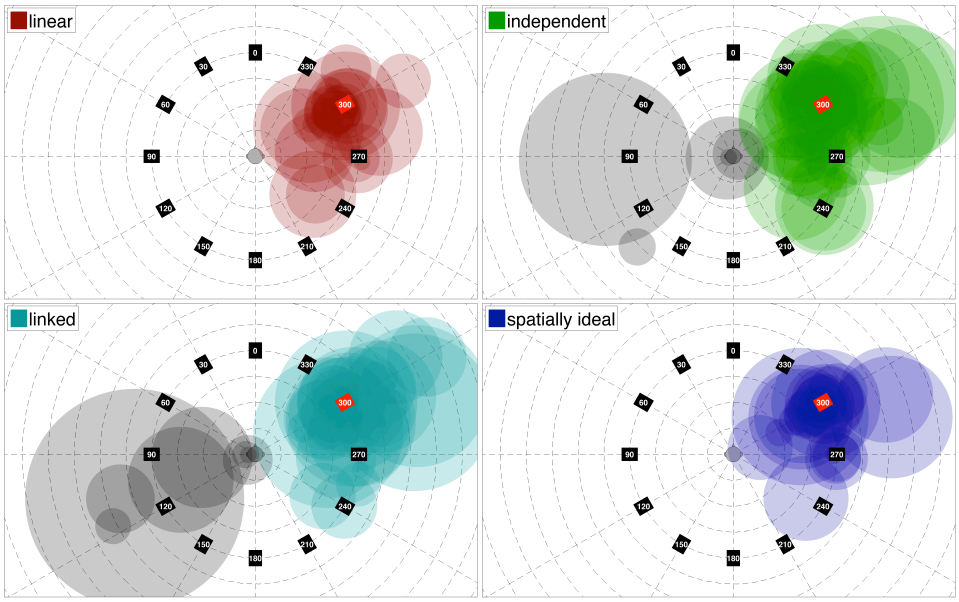


Figure 4.5: Same as Fig. 4.4, but for the hearing-impaired listeners.

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. 4.6 and 4.7 compared to those in Figs. 4.4 and 4.5. Furthermore, more image splits were documented for the transients than for speech in the independent and linked compression conditions.

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. 4.4-4.7). For the radius of the placed circles, indicating the perceived width of the sound image, the ANOVA 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 [$F(5,110) = 3.97$,

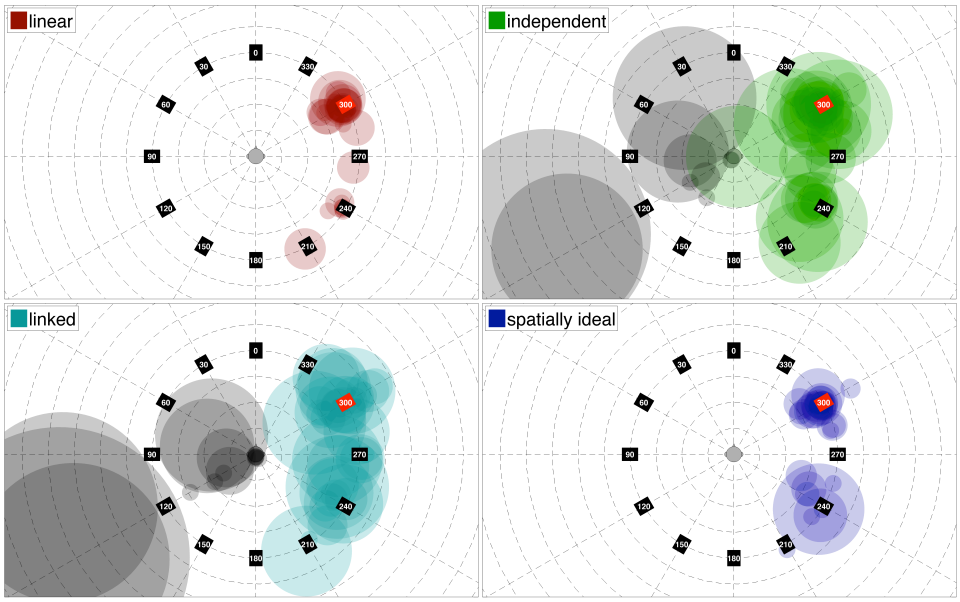


Figure 4.6: Same as Fig. 4.4, but for the normal-hearing listeners and transients.

$p \ll 0.001$]. 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

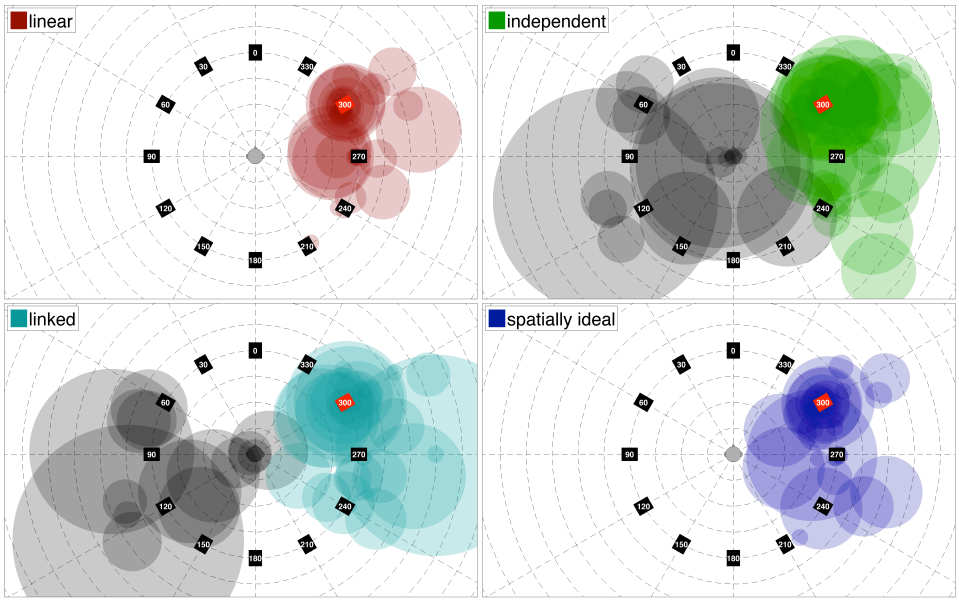


Figure 4.7: Same as Fig. 4.4, but for the hearing-impaired listeners and transients.

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.

4.3.1 Analysis of spatial cues

Figure 4.8 shows the ILD distributions for the speech (top panel) and the transients (lower panel) when virtualized from the loudspeaker positioned at 300° azimuth. 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 ILD distributions for linear processing, independent compression, linked compression and spatially ideal compression, respectively. For both stimuli, the ILDs are reduced in the independent compression condition (with a maximum at 1.5 dB) relative to the other processing conditions where the ILD statistics are similar to each other (and centred around 6 dB for the speech stimulus and 3 dB for the transients). The ILDs obtained for the transients are below those obtained for speech since the transients contain fewer time segments that are dominated by the direct sound and more segments dominated by reverberant sound energy compared to the speech stimulus.

Figure 4.9 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

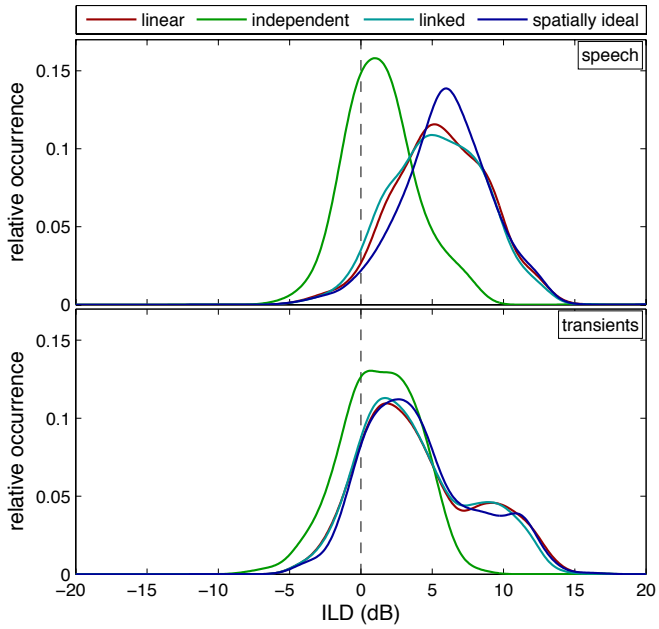


Figure 4.8: The ILD-distributions for the speech stimulus (top panel) and the transients (lower panel) when virtualized from the loudspeaker positioned at 300° azimuth. Only the results at the output of the gammatone filter tuned to 2000 Hz are shown.

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 4.10 shows temporal energy patterns for the linear processing and the three compression conditions for the speech stimulus (upper panel) and

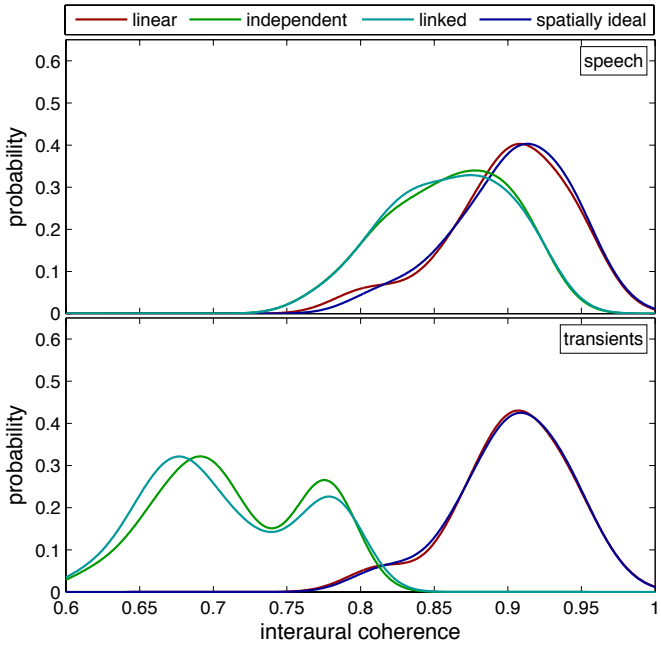


Figure 4.9: IC distributions of the ears signals, pooled across all listeners, at the output of the gammatone filter tuned to 2000 Hz. Results are shown for the speech (upper panel) and the transients (lower panel) virtualized from the frontal loudspeaker position. The red, green, light blue and dark blue functions represent the IC distribution for linear processing, independent compression, linked compression and spatially ideal compression, respectively.

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 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. 4.9 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. 4.4-4.7.

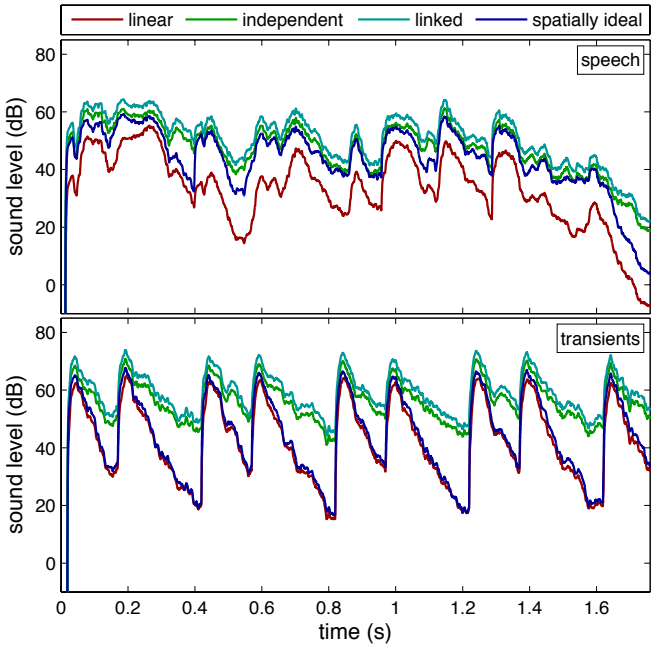


Figure 4.10: Temporal energy patterns of the speech stimulus (upper panel) and the transient stimulus (lower panel) virtualized from the frontal loudspeaker position. Only the output of the signals processed by the gammatone filter at 2000 Hz is shown. The different colors represent the different processing 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).

4.4 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 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 reason why the split images were consistently perceived from the opposite hemisphere of the primary sound image in both the linked and independent compression condition is not clear from the analysis of the interaural cues used for localization.

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 insen-

sitive 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 quiet 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 spa-

tial perception might be less prominent than the effects reported in the present study.

4.5 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.

5

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 is based on Hassager, H. G., May, T., Wiinberg, A, Dau, T. (2017), *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.

5.1 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.

5.2 Compression system

5.2.1 Algorithm overview

Figure 5.1 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. 5.2.2) 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. 5.2.3). 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 5.1.

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 5.1: 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.

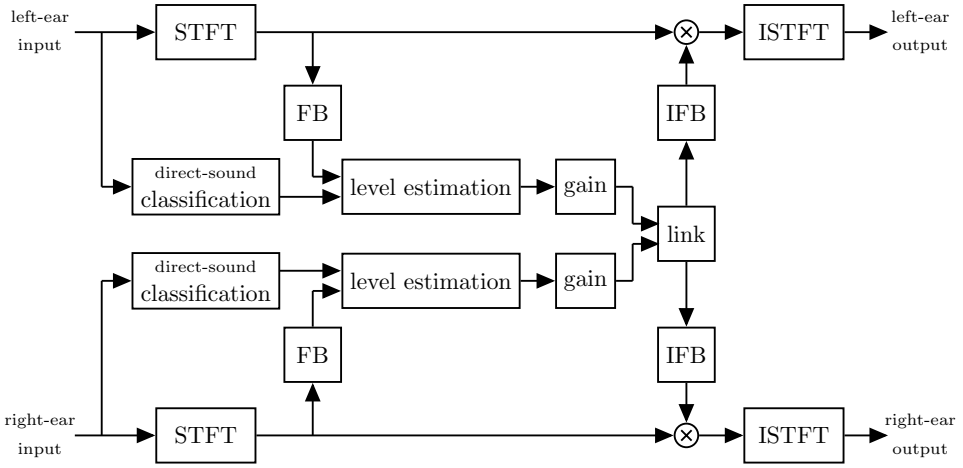


Figure 5.1: Block diagram of the proposed direct-sound driven compressor. First the left- and right-ear signals are windowed in time segments and transformed into the frequency domain by a short-time Fourier transforms (STFT). The frequency bins in each time window are combined into seven octave spaced 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 5.2 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.

5.2.2 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

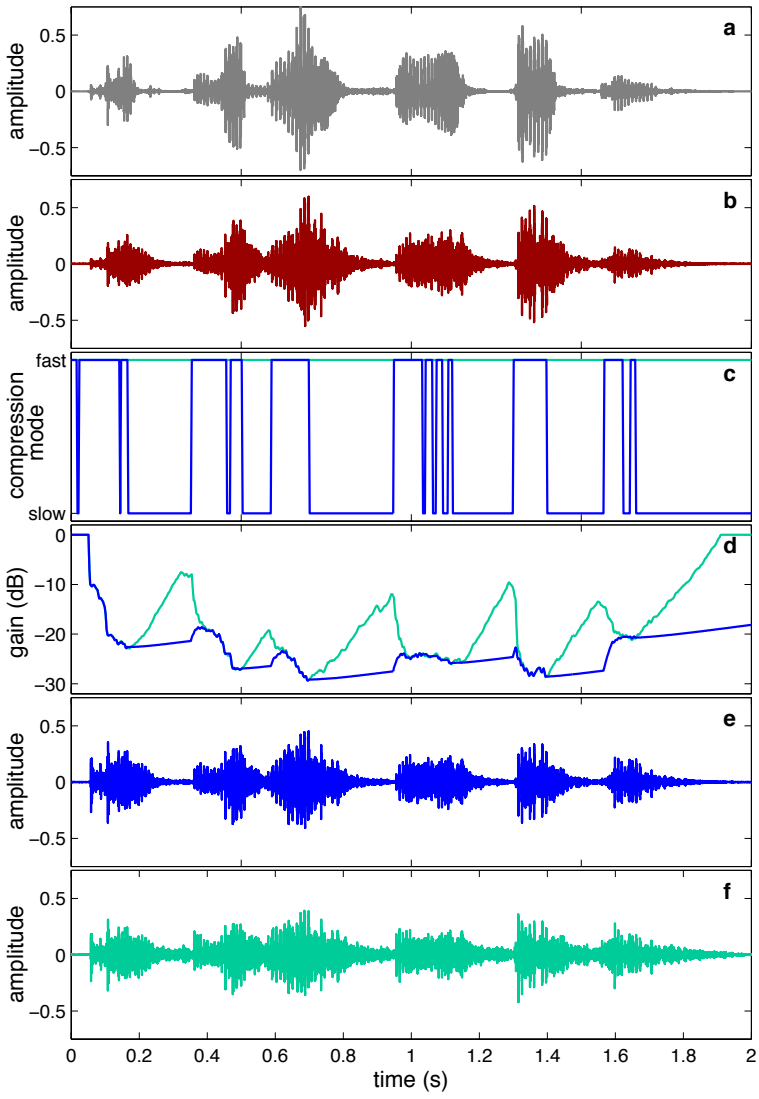


Figure 5.2: Example illustrating a bandpass filtered HINT sentence extracted at the center frequency of 1000 Hz. (a) Anechoic sentence, (b) reverberant sentence, (c) the blind binary classification (blue) where a value of one indicates direct-sound activity, (d) the corresponding gain function for conventional 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:

$$D_s(t,f) = \lambda \cdot D_s(t-1,f) + (1-\lambda) \cdot |D(t,f)|^2 \tag{5.1}$$

and

$$R_s(t,f) = \lambda \cdot R_s(t-1,f) + (1-\lambda) \cdot |R(t,f)|^2, \tag{5.2}$$

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:

$$SRR(t,f) = 10 \cdot \log_{10}\left(\frac{D_s(t,f)}{R_s(t,f)}\right) \tag{5.3}$$

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:

$$C^{SRR}(t,f) = \begin{cases} 1, & SRR(t,f) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.4}$$

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:

$$F(t,f) = 10 \cdot \log_{10}\left(\frac{\sigma^2(|S(t,f)|^{\alpha})}{\sigma^2(|S(t,f)|)}\right) \tag{5.5}$$

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:

$$C^{Blind}(t,f) = \begin{cases} 1, & F(t,f) > T(t,f), \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

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 5.2.2 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 5.2. 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 5.2: The blind classification performance in terms of the H, HA, and H-FA for the seven octave frequency channels.

5.2.3 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:

$$X_s(t,f) = c \cdot X_s(t-1,f) + (1-c) \cdot |X(t,f)|^2, \tag{5.7}$$

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:

$$c = \begin{cases} c_{\text{attack}}^{\text{fast}} & \text{when} \quad |X(t,f)|^2 \geq X_s(t-1,f) \quad \text{and} \quad C(t,f)=1 \\ c_{\text{release}}^{\text{fast}} & \text{when} \quad |X(t,f)|^2 < X_s(t-1,f) \quad \text{and} \quad C(t,f)=1 \\ c_{\text{attack}}^{\text{slow}} & \text{when} \quad |X(t,f)|^2 \geq X_s(t-1,f) \quad \text{and} \quad C(t,f)=0 \\ c_{\text{release}}^{\text{slow}} & \text{when} \quad |X(t,f)|^2 < X_s(t-1,f) \quad \text{and} \quad C(t,f)=0 \end{cases} \tag{5.8}$$

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 International Electrotechnical Commis-

sion (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.

5.3 Methods

5.3.1 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.

5.3.2 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 5.3 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. 5.3

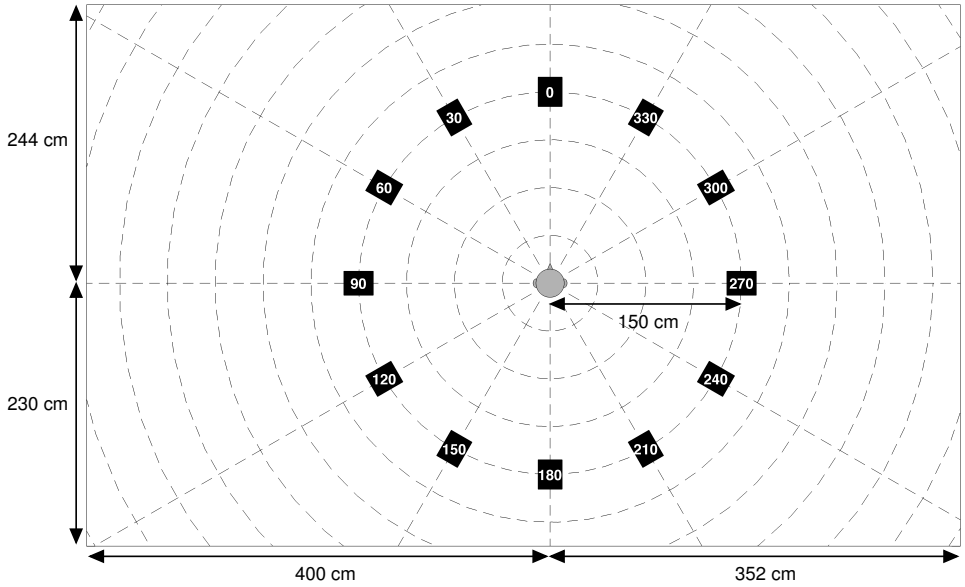


Figure 5.3: The top view of the experimental setup. The loudspeaker positions are indicated by the black squares. The grey circle in the center indicates the position of the chair, where the listener was seated. The listeners had a view direction on the loudspeaker placed at the 0°degree azimuth. The 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 5.3.3). 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.

5.3.3 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, $h_{\text{hpir}}^{\text{inv}}$, 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, $h_{\text{hpir}}^{\text{inv}}$, 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).

5.3.4 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 5.3, 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 $h_{\text{hpir}}^{\text{inv}}$, 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 5.3: Overview of the different processing conditions involving compression.

5.3.5 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.

5.3.6 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 $s_{\mathrm{out,l}}$ and $s_{\mathrm{out,r}}$ occurring over an interval of $|\tau| \leq 1$ ms (e.g., Blauert and Lindemann, 1986; Hartmann et al., 2005):

$$\mathrm{IC} = \max_{\tau} \left| \frac{\sum_t s_{\mathrm{out,l}}(t+\tau) s_{\mathrm{out,r}}(t)}{\sqrt{\sum_t s^2_{\mathrm{out,l}}(t) \sum_t s^2_{\mathrm{out,r}}(t)}} \right| . \tag{5.9}$$

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 (Glas-

berg 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.

5.4 Results

5.4.1 Experimental data

Figures 5.4 and 5.5 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. 5.4, 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. 5.4), 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 5.5 shows the corresponding results for independent compression. The general pattern of results was similar to that found for linked compression (from Fig. 5.4). 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. 5.4). The reported image splits were in both direct-sound

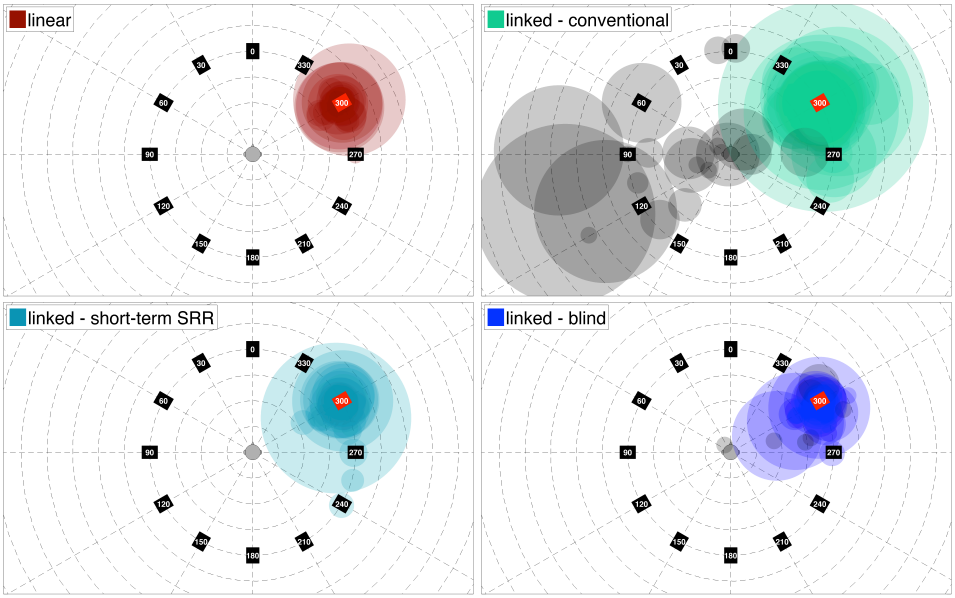


Figure 5.4: Graphical representations of the listeners' responses obtained with the speech virtually presented from the 300 degrees position in the listening room. The upper left panel shows the results for linear processing (reference condition). The results for conventional linked compression, direct-sound 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.

driven compression conditions placed around the position of the head. The 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

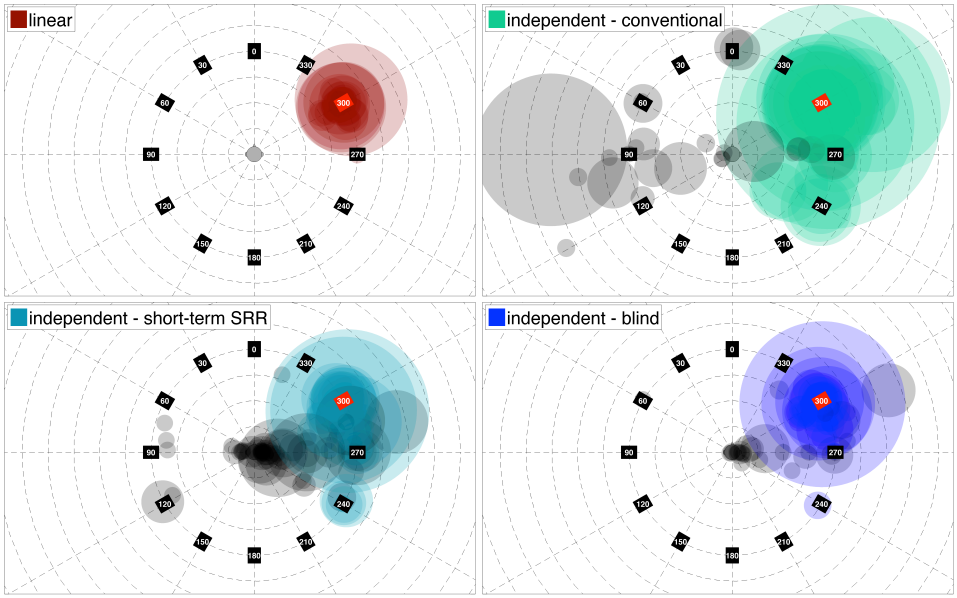


Figure 5.5: Same as Fig. 5.4, but for the independent compression conditions.

cm and 37.0 cm for the linked and the independent compression condition, 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.

5.4.2 Analysis of spatial cues

Figure 5.6 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).

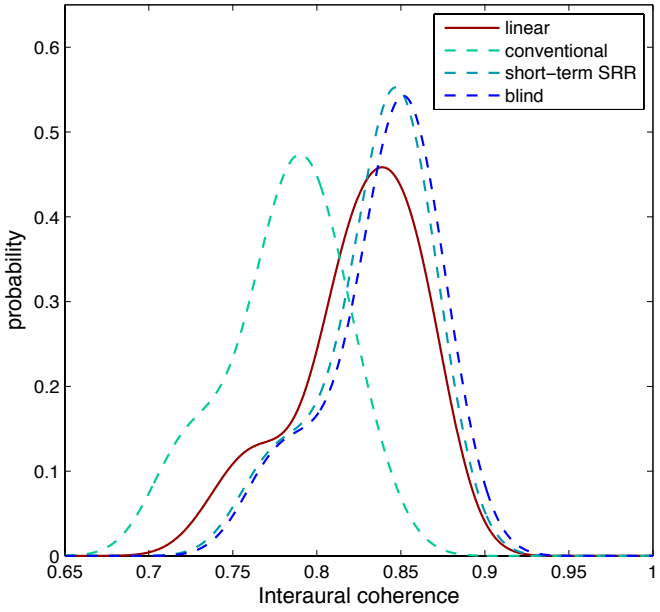


Figure 5.6: IC distributions of the ears signals, pooled across all listeners, at the output of the gammatone filter tuned to 1000 Hz. Results are shown for the speech virtualized from the frontal loudspeaker position. The solid red, dashed light green, dashed light blue and dashed blue curves represent 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.

5.5 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 5.2). 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.

5.6 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.

6

General discussion

In this thesis, behavioral methods were used to study the effects of non-linear hearing-aid amplification strategies on hearing-impaired (HI) listeners' spatial perception, temporal resolution and speech perception. The objectives were to better understand the consequences of conventional hearing-aid compression on the different behavioral outcome measures and to analyze the effects of various modifications of this amplification strategy. In the following section, the main findings are summarized.

6.1 Summary of main findings

The experiments in Chapter 2 showed that properly set fast-acting hearing-aid compression can restore the HI listeners' modulation detection thresholds to the level observed in normal-hearing (NH) listeners. Despite this "normalization" of the modulation detection thresholds, hearing-aid compression does not seem to provide a benefit for speech intelligibility. Furthermore, hearing-aid compression may not be able to restore modulation depth discrimination thresholds to the level of "normal" processing suggesting that the two measures of sensitivity to amplitude modulations – modulation detection vs. modulation depth discrimination - represent different limits of resolution. The results obtained in the NH listeners showed that the ability to discriminate modulation depths is correlated with speech intelligibility in stationary noise. In HI listeners, additional deficits of supra-threshold processing beyond degraded modulation

depth discrimination may affect speech intelligibility.

The results of the study in chapter 2 inspired the development of a speech enhancement system based on a multiband non-linear expansion of temporal envelope fluctuations between 10 and 20 Hz. This system was designed to compensate for the degraded modulation-depth discrimination sensitivity observed in the hearing-impaired listeners. The idea was that this could facilitate the recognition of transient speech cues. The effects of three speech enhancement schemes on consonant recognition were evaluated in Chapter 3. The results revealed that SNR-based envelope expansion of the noisy speech improved the overall consonant recognition scores relative to linear processing, mainly by improving the recognition of some of the considered stop consonants. The effect of the SNR-based envelope expansion was similar to the effect of envelope-expanding the clean speech before the addition of noise. While the average effect was relatively small (about two percent points improvement of consonant recognition), the observation of similar results obtained with the SNR-based processing and the clean-speech envelope expansion was promising, given that previous studies reported substantial improvements in speech perception with envelope expansion of clean speech (e.g., Apoux et al., 2001; Langhans and Strube, 1982; Lorenzi et al., 1999). This suggests that SNR-based envelope expansion may provide larger improvements in speech perception when combined with alternative parameter settings, such as those considered in the earlier studies.

The effects of three fast-acting hearing-aid compression schemes on spatial perception were evaluated in Chapter 4: Independent compression at each ear, linked compression between the two ears, and "spatially ideal" compression operating solely on the dry source signal. The results revealed that both independent and linked fast-acting compression strongly distorted the spatial

perception of sounds in a reverberant acoustic environment. Both compression strategies increased the diffusiveness of the perceived sound and led to broader, sometimes internalized, sound images as well as sound-image splits. It was found that the observed spatial distortions were mainly resulting from the enhancement of the reflected sound components relative to the direct sound. It was concluded that fast-acting 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.

The findings in Chapter 4 initiated the study presented in Chapter 5. Here, a direct-sound driven compression scheme was proposed. The direct-sound driven compression scheme only applied fast-acting compression to components of the stimulus that were dominated by the direct sound while the stimulus components dominated by reverberation were processed linearly via longer time constants. Effectively, this scheme preserves the energy ratio of the direct sound to the reverberant sound. Both a "ideal" classification method based on a priori knowledge of the binaural room impulse response, and a blind classification method proposed by Hazrati et al. (2013), were used to classify the T-F units dominated by the direct sound and those dominated by reverberation. The results demonstrated that the proposed direct-sound driven compression scheme preserved the listener's spatial impression when the compression was linked between the two ears. In contrast, the direct-sound driven compression in the condition where independent compression was applied in the two ears created a sense of movement of the sound between the two ears. This indicates that the DRC processing needs to maintain the natural interaural cues (IC and ILDs) in order to preserve the listener's spatial impression.

6.2 Implications and perspectives

An intuitive method of compensating for loss of cochlear compression is to develop a hearing-aid compression system which imitates the missing compression of the cochlear as closely as possible, such that the inner hair cells receive the same input as if the outer hair cells were not damaged. Because outer hair cells react very fast to changes in the input level, fast-acting gain adaptation is required to imitate their function. This has been the rationale for using fast-acting compression in hearing aids (e.g., Allen, 1996). Such approaches may be characterized as loudness normalization, since a key objective is to restore loudness perception back to normal. In fact, it has been shown that a fast-acting multi-band compression system with attack and release time constants of 10 ms can restore normal loudness summation and differential loudness growth in HI listeners (Strelcyk et al., 2012). It has also been demonstrated that measures of auditory temporal resolution that are assumed to be affected in HI listeners by recruitment, such as gap detection, forward masking and amplitude modulation detection (see Chapter 2) can at least partly be restored to normal by fast-acting compression with time constants less than 60 ms (e.g.,Moore et al., 2001; Brennan et al., 2015; Kowalewski et al., 2015). Based on these perceptual results, it could be assumed that loudness restoring fast-acting hearing-aid compression can improve speech intelligibility in everyday sound environments. In support of this assumption, Gatehouse et al. (2006) showed that in self-reports of everyday listening, compression processing with fast time constants is, on average, rated to be better for speech intelligibility than linear amplification and fitting procedures with slow time constants. However, in other studies it has been argued that compression processing with fast time constants can be detrimental for speech intelligibility because the processing reduces the spectral and temporal contrasts in the speech (e.g., Plomp, 1988). In Chapter 2,

Plomp's hypothesis was further tested and it was found that, in conditions with stationary noise maskers, fast-acting compression indeed has a small detrimental effect on speech intelligibility. However, it should be noted that the tested compression system was not adapted to the individual HI listeners hearing loss, and thereby did not provide loudness normalization. This was also the case in other studies showing a detrimental effect of fast-acting compression on speech intelligibility (Drullman and Smoorenburg, 1997; Reinhart et al., 2016; Noordhoek and Drullman, 1997). Hence, these results do not necessarily justify a recommendation for linear amplification instead of loudness normalizing fast-acting compression schemes. Rather, the results indicate that fast-acting compression may degrade speech intelligibility when the compression settings are not adapted to the individual listener's hearing loss.

Despite properly set fast-acting compression may provide loudness normalization and restore aspects of temporal resolution back to normal, several hearing aids make use of slow-acting compression schemes (e.g., Ludvigsen, 2001). One of the reasons is that HI listeners find listening with slow-acting compression hearing aids more pleasant and better sounding than listening with fast-acting compression hearing aids (see Kates2010 for a review). These findings have been associated with inter-modulation and harmonic distortion products that arise when the compression processing flattens the signal envelope. However, some of these findings may also be associated with the finding in Chapter 4 that fast-acting compression distorts the natural spatial perception of the HI listeners.

Bibliography

Allen, J. B. (1996). “Derecruitment by multiband compression in hearing aids”. In: *Psychoacoustics, Speech and Hearing Aids*. World Scientific. Chap. III, pp. 141–152.

Apoux, F., N. Tribut, X. Debruille, and C. Lorenzi (2004). “Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners”. In: *Hearing Research* 189.1-2, pp. 13–24.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1, https://cran.r–project.org/package=lmerTest.

Bianchi, F., M. Fereczkowski, J. Zaar, S. Santurette, and T. Dau (2016). “Complex-Tone Pitch Discrimination in Listeners With Sensorineural Hearing Loss”. In: *Trends in Hearing* 20.0, pp. 1–15.

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.

Boothroyd, A., N. Springer, L. Smith, and J. Schulman (1988). “Amplitude compression and profound hearing loss”. In: *Journal of Speech and Hearing Research* 31.3, pp. 362–376.

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.

Brennan, M. A., F. J. Gallun, P. E. Souza, and C. Stecker (2013). "Temporal Resolution with a Prescriptive Fitting Formula." In: *American journal of audiology* 22.December, pp. 216–226.

Brennan, M. A., R. W. McCreery, and W. Jesteadt (2015). "The influence of hearing-aid compression on forward-masked thresholds for adults with hearing loss". In: *Journal of the Acoustical Society of America* 138.4, pp. 2589–2597.

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.

Buuren, R. a. van, J. M. Festen, and T Houtgast (1999). "Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality." In: *The Journal of the Acoustical Society of America* 105.5, pp. 2903–13.

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.

Chabot-Leclerc, A., S. Jørgensen, and T. Dau (2014). "The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction". In: *The Journal of the Acoustical Society of America* 135.6, pp. 3502–3512.

Chabot-Leclerc, A., E. N. MacDonald, and T. Dau (2016). "Predicting binaural speech intelligibility using the signal-to-noise ratio in the envelope power spectrum domain". In: *The Journal of the Acoustical Society of America* 140.1, pp. 192–205.

Christiansen, C., E. N. MacDonald, and T. Dau (2013). "Contribution of envelope periodicity to release from speech-on-speech masking." In: *The Journal of the Acoustical Society of America* 134.3, pp. 2197–204.

Christiansen, T. U. (2011). "Objective Evaluation of Consonant-Vowel pairs produced by Native Speakers of Danish". In: *Forum Acusticum*.

Clarkson, P. M. and S. F. Bahgat (1991). "Envelope expansion methods for speech enhancement." In: *The Journal of the Acoustical Society of America* 89.3, pp. 1378–82.

Dau, T, D Puschel, and A Kohlrausch (1996a). "A quantitative model of the "effective" signal processing in the auditory system .2. Simulations and measurements". In: *Journal of the Acoustical Society of America* 99.6, pp. 3623–3631.

Dau, T, D Püschel, and a Kohlrausch (1996b). “A quantitative model of the "effective" signal processing in the auditory system. I. Model structure.” In: *The Journal of the Acoustical Society of America* 99.6, pp. 3615–3622.

Dau, T, B Kollmeier, and A. Kohlrausch (1997). “Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration.” In: *The Journal of the Acoustical Society of America* 102.5 Pt 1, pp. 2906–2919.

Dau, T., J Verhey, and A. Kohlrausch (1999). “Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers.” In: *Journal of the Acoustical Society of America* 106.5, pp. 2752–2760.

Dreschler, W. A., H Verschuure, C Ludvigsen, and S Westermann (2001). “ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment.” In: *Audiology* 40.3, pp. 148–157.

Drullman, R and G. F. Smoorenburg (1997). “Audio-visual perception of compressed speech by profoundly hearing-impaired subjects.” In: *Audiology : official organ of the International Society of Audiology* 36.3, pp. 165–177.

Drullman, R, J. M. Festen, and R Plomp (1994). “Effect of temporal envelope smearing on speech reception.” In: *The Journal of the Acoustical Society of America* 95.2, pp. 1053–64.

Edwards, B. (2004). “Hearing Aids and Hearing Impairment”. In: *Speech Processing in the Auditory System*. Vol. 18. New York: Springer-Verlag, pp. 339–421.

Elhilali, M., T. Chi, and S. A. Shamma (2003). “A spectro-temporal modulation index (STMI) for assessment of speech intelligibility”. In: *Speech Communication* 41.2-3, pp. 331–348.

Ellis, R. J. and K. J. Munro (2015). “Benefit from, and acclimatization to, frequency compression hearing aids in experienced adult hearing-aid users”. In: *International Journal of Audiology* 54.1, pp. 37–47.

Ewert, S. D. and T. Dau (2000). "Characterizing frequency selectivity for envelope fluctuations." In: *Journal of the Acoustical Society of America* 108.3 Pt 1, pp. 1181–96.

Ewert, S. D. and T. Dau (2004). "External and internal limitations in amplitude-modulation processing". In: *The Journal of the Acoustical Society of America* 116.1, pp. 478–490.

Feng, Y., S. Yin, M. Kiefte, and J. Wang (2010). "Temporal resolution in regions of normal hearing and speech perception in noise for adults with sloping high-frequency hearing loss." In: *Ear and hearing* 31.1, pp. 115–125.

Festen, J. M. and R Plomp (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing." In: *The Journal of the Acoustical Society of America* 88.4, pp. 1725–1736.

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.

Freyman, R. L. and G. P. Nerbonne (1989). "The Importance of Consonant-Vowel Intensity Ratio in the Intelligibility of Voiceless Consonants". In: *Journal of Speech Language and Hearing Research* 32.3, pp. 524–535.

Freyman, R. L. and G. P. Nerbonne (1996). "Consonant Confusions in Amplitude-Expanded Speech". In: *Journal of Speech Language and Hearing Research* 39.6, pp. 1124–1137.

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.

Gatehouse, S., G. Naylor, and C. Elberling (2006). "Linear and nonlinear hearing aid fittings – 1. Patterns of benefit". In: *International Journal of Audiology* 45.3, pp. 130–152.

Gerkmann, T. and R. C. Hendriks (2012). “Unbiased MMSE-based noise power estimation with low complexity and low tracking delay”. In: *IEEE Transactions on Audio, Speech and Language Processing* 20.4, pp. 1383–1393.

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

Glasberg, B. R., B. C. J. Moore, and R. W. Peters (2001). “The influence of external and internal noise on the detection of increments and decrements in the level of sinusoids”. In: *Hearing Research* 155.1-2, pp. 41–53.

Glista, D., S. Scollie, and J. Sulkers (2012). “Perceptual acclimatization post nonlinear frequency compression hearing aid fitting in older children”. In: *Journal of Speech, Language, and Hearing Research* 55.6, pp. 1765–1787.

Gnewikow, D., T. Ricketts, G. W. Bratt, and L. C. Mutchler (2009). “Real-world benefit from directional microphone hearing aids.” In: *Journal of rehabilitation research and development* 46.5, pp. 603–618.

Grose, J. H., H. L. Porter, E. Buss, and J. W. Hall (2016). “Cochlear hearing loss and the detection of sinusoidal versus random amplitude modulation”. In: *The Journal of the Acoustical Society of America* 140.2, EL184–EL190.

Hartmann, W. M. 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., 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.

Houtgast, T. and H. J. M. Steeneken (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria". In: *The Journal of the Acoustical Society of America* 77.3, pp. 1069–1077.

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.

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

Jørgensen, S. and T. Dau (2011). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing". In: *The Journal of the Acoustical Society of America* 130.3, pp. 1475–1487.

Jørgensen, S., S. D. Ewert, and T. Dau (2013). "A multi-resolution envelope-power based model for speech intelligibility." In: *The Journal of the Acoustical Society of America* 134.1, pp. 436–46.

Jørgensen, S., R. Decorsière, and T. Dau (2015). "Effects of manipulating the signal-to-noise envelope power ratio on speech intelligibility". In: *The Journal of the Acoustical Society of America* 137.3, pp. 1401–1410.

Kale, S. and M. G. Heinz (2010). "Envelope coding in auditory nerve fibers following noise-induced hearing loss". In: *JARO - Journal of the Association for Research in Otolaryngology* 11.4, pp. 657–673.

Kates, J. M. (2008). *Digital hearing aids*. San Diego, CA: Plural Publishing, 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.

Keidser, G. et al. (2006). "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers." In: *International Journal of Audiology* 45.10, pp. 563–579.

Kohlrausch, A., R. Fassel, and T. Dau (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers". In: *The Journal of the Acoustical Society of America* 108.2, pp. 723–734.

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.

Kowalewski, B., E. Macdonald, O. Strelcyk, and T. Dau (2015). "Auditory-model based assessment of the effects of hearing loss and hearing-aid compression on spectral and temporal resolution". In: *Proceedings of the International*

Symposium on Auditory and Audiological Research 5.December, pp. 173–180.

Langhans, T and H. Strube (1982). "Speech enhancement by nonlinear multiband envelope filtering". In: *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 7. Institute of Electrical and Electronics Engineers, pp. 156–159.

Lee, J. and S. Bacon (1997). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration". In: *J Acoust Soc Am* 101.6, pp. 3688–3693.

Lenth, R. V. (2016). "Least-Squares Means: The R Package lsmeans". In: *Journal of Statistical Software* 69.1, pp. 1–43.

Levitt, H (1971). "Transformed Up-Down Methods in Psychoacoustics". In: *The Journal of the Acoustical society of America*, pp. 467–477.

Li, F., A. Menon, and J. B. Allen (2010). "A psychoacoustic method to find the perceptual cues of stop consonants in natural speech". In: *The Journal of the Acoustical Society of America* 127.4, pp. 2599–2610.

Lobdell, B. E. and J. B. Allen (2007). "A model of the VU (volume-unit) meter, with speech applications." In: *The Journal of the Acoustical Society of America* 121.1, pp. 279–85.

Lopez-Poveda, E. A. and P. Barrios (2013). "Perception of stochastically undersampled sound waveforms: A model of auditory deafferentation". In: *Frontiers in Neuroscience* 7.124, pp. 1–13.

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.

Martin, R (2001). “Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics”. In: *IEEE Signal Proc. Letters* 9.5, pp. 504–512.

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.

Metselaar, M., B. Maat, P. Krijnen, H. Verschuure, W. Dreschler, and L. Feenstra (2008). “Comparison of speech intelligibility in quiet and in noise after hearing aid fitting according to a purely prescriptive and a comparative fitting procedure.” In: *European Archives of Oto-Rhino-Laryngology* 265.9, pp. 1113–20.

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. 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.

Moore, B. C. J. 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., B. R. Glasberg, J. I. Alcántara, S Launer, and V Kuehnel (2001). "Effects of slow- and fast-acting compression on the detection of gaps in narrow bands of noise." In: *British journal of audiology* 35.6, pp. 365–74.

Moore, B. C. J. and R. D. Patterson (1986). "Auditory filters and excitation patterns as representations of frequency resolution". In: *Frequency Selectivity in Hearing*. Ed. by B. C. J. Moore. Academic, London.

Moore, B. C. J., M. Wojtczak, and D. A. Vickers (1996). "Effect of loudness recruitment on the perception of amplitude modulation". In: *The Journal of the Acoustical Society of America* 100.1, pp. 481–489.

Musa-Shufani, S., M. Walger, H. von Wedel, and H. Meister (2006). "Influence of dynamic compression on directional hearing in the horizontal plane." In: *Ear and hearing* 27.3, pp. 279–85.

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

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.

Noordhoek, I. M. and R Drullman (1997). "Effect of reducing temporal intensity modulations on sentence intelligibility." In: *The Journal of the Acoustical Society of America* 101.1, pp. 498–502.

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

Plomp, R. (1978). “Auditory handicap of hearing impairment and the limited benefit of hearing aids”. In: *The Journal of the Acoustical Society of America* 63.2, pp. 533–549.

Plomp, R. (1988). “The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function.” In: *The Journal of the Acoustical Society of America* 83.6, pp. 2322–2327.

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.

Régnier, M. S. and J. B. Allen (2008). “A method to identify noise-robust perceptual features: application for consonant /t/”. In: *Journal of the Acoustical Society of America* 123.5, pp. 2801–2814.

Reinhart, P. N., P. E. Souza, N. K. Srinivasan, and F. J. Gallun (2016). “Effects of Reverberation and Compression on Consonant Identification in Individuals with Hearing Impairment”. In: *Ear and Hearing* 37.2, pp. 144–152.

Schlittenlacher, J. and B. C. J. Moore (2016). “Discrimination of amplitude-modulation depth by subjects with normal and impaired hearing”. In: *The Journal of the Acoustical Society of America* 140.5, pp. 3487–3495.

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.

Shannon, R., F. Zeng, and V Kamath (1995). “Speech recognition with primarily temporal cues”. In: *Science* 270, pp. 303–304.

Sheft, S. and W. A. Yost (1990). “Temporal integration in amplitude modulation detection.” In: *The Journal of the Acoustical Society of America* 88.2, pp. 796–805.

Simpson, A. (2009). “Frequency-Lowering Devices for Managing High-Frequency Hearing Loss: A Review”. In: *Trends in Amplification* 13.2, pp. 87–106.

Souza, P. E. and C. W. Turner (1996). “Effect of single-channel compression on temporal speech information.” In: *Journal of speech and hearing research* 39.5, pp. 901–11.

Souza, P. E. and C. W. Turner (1998). *Multichannel compression, temporal cues, and audibility.*

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.

Stone, M. A. and B. C. J. Moore (1992). “Syllabic compression: Effective compression ratios for signals modulated at different rates”. en. In: *British Journal of Audiology* 26.6, pp. 351–361.

Stone, M. A., C. Füllgrabe, and B. C. J. Moore (2008). “Benefit of high-rate envelope cues in vocoder processing: effect of number of channels and spectral region.” In: *The Journal of the Acoustical Society of America* 124.4, pp. 2272–82.

Stone, M. A., K. Anton, and B. C. J. Moore (2012). “Use of high-rate envelope speech cues and their perceptually relevant dynamic range for the hearing impaired”. In: *The Journal of the Acoustical Society of America* 132.2, pp. 1141–1151.

Strelcyk, O., N. Nooraei, S. Kalluri, and B. Edwards (2012). “Restoration of loudness summation and differential loudness growth in hearing-impaired lis-

teners". In: *The Journal of the Acoustical Society of America* 132.4, pp. 2557–2568.

Takahashi, G. A. and S. P. Bacon (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly." In: *Journal of speech and hearing research* 35.6, pp. 1410–21.

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 Esch, T. E. M. and W. A. Dreschler (2015). "Relations Between the Intelligibility of Speech in Noise and Psychophysical Measures of Hearing Measured in Four Languages Using the Auditory Profile Test Battery". In: *Trends in Hearing* 19.0, pp. 1–12.

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.

Van den Bogaert, T., T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters (2006). "Horizontal localization with bilateral hearing aids: without is better than with." In: *The Journal of the Acoustical Society of America* 119.1, pp. 515–526.

Villchur, E. (1973). "Signal processing to improve speech intelligibility in perceptive deafness". In: *The Journal of the Acoustical Society of America* 53.6, pp. 1646–1657.

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.

Wiinberg, A., M. L. Jepsen, B. Epp, and T. Dau (2015). "Effects of dynamic-range compression on temporal acuity". In: *the 5th International Symposium on Auditory and Audiological Research (ISAAR-2015)*. Nyborg, Denmark.

Wolfe, J. et al. (2011). "Long-term effects of non-linear frequency compression for children with moderate hearing loss". In: *International Journal of Audiology* 50.6, pp. 396–404.

Xu, L. and Y. Zheng (2007). "Spectral and temporal cues for phoneme recognition in noise". In: *The Journal of the Acoustical Society of America* 122.3, pp. 1758–1764.

Xu, L., C. S. Thompson, and B. E. Pfingst (2005). "Relative contributions of spectral and temporal cues for phoneme recognition". In: *The Journal of the Acoustical Society of America* 117.5, pp. 3255–3267.

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

Yund, E. W. (1995). “Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise”. In: *The Journal of the Acoustical Society of America* 97.2, pp. 1206–1223.

Zaar, J. and T. Dau (2015). “Sources of variability in consonant perception of normal-hearing listeners”. In: *The Journal of the Acoustical Society of America* 138.3, pp. 1253–1267.

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*, Pitch representations in the impaired auditory system and implications for music perception, 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 Löcsei*, Lateralized speech perception with normal and impaired hearing, 2016.

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation, 2017.

Vol. 27: *Henrik Gerd Hassager*, Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing, 2017.

Vol. 28: *Richard Ian McWalter*, Analysis of the auditory system via synthesis of natural sounds, speech and music, 2017.

Vol. 29: *Jens Cubick*, Characterizing the auditory cues for the processing and perception of spatial sounds, 2017.

Vol. 30: *Gerard Encina-Llamas*, Characterizing cochlear hearing impairment using advanced electrophysiological methods, 2017.

Vol. 31: *Christoph Scheidiger*, Assessing speech intelligibility in hearing impaired listeners, 2017.

The end.

To be continued...

One of the most common reported complaints by people with sensorineural hearing loss is abnormal difficulties in understanding speech in complex acoustic environments. In an attempt to compensate for these difficulties, many modern hearing aids use hearing aid compression. Here, the effect of hearing aid compression on spatial perception, temporal resolution and speech perception were investigated in normal-hearing (NH) and hearing-impaired (HI) listeners. Based on the findings, a hearing aid compression system that preserve the spatial impression by adaptively selecting appropriate time constants was developed and evaluated in a reverberant listening room. Also, a speech enhancement system that restored aspects of hearing-impaired listeners' temporal resolution back toward the level observed in the normal-hearing listeners was developed and evaluated on consonant recognition. Overall, the outcomes of this thesis have implications for how hearing aids should compensate for sensorineural hearing loss.

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