

Lateralized speech perception with normal and impaired hearing

PhD thesis by
Gusztáv Lőcsei



Technical University of Denmark

2016

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 (Kgs. Lyngby, Denmark). Part of the project was carried out at the Oticon Research Centre (Snekkersten, Denmark).

The project was financed by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000 and by the Technical University of Denmark.

Supervisors

Main supervisor

Assoc. Prof. Ewen N. MacDonald
Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Co-supervisors

Assist. Prof. Sébastien Santurette
Prof. Torsten Dau
Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Abstract

Listeners with sensorineural hearing impairment often report difficulties with speech communication, predominantly in acoustic environments with multiple, spatially distributed interferers. Compensating for the listeners' elevated hearing thresholds does not necessarily resolve their listening problems. Therefore, it has been suggested that part of this difficulty resides in coding deficits that distort the internal auditory representation of stimuli above the threshold of hearing. Disrupted temporal fine structure (TFS) coding at the output of the cochlear filters is a possible source of such supra-threshold deficits. This thesis investigated the relationship between the auditory coding of supra-threshold stimuli and spatial speech perception, with a special focus on the links between TFS coding abilities and the binaural unmasking (BU) of speech in noise in lateralized listening scenarios.

Speech reception thresholds (SRTs) were measured in various noise backgrounds in headphone experiments. Target and the maskers were lateralized using interaural timing differences (ITDs). Binaural intelligibility level differences (BILD) were measured between the different lateralization conditions. The first study investigated the relationship between speech lateralization and speech intelligibility thresholds for normal-hearing (NH) listeners, whereby a strong correlation between these measures was found. Also, BILDs for a fixed amount of interferers were found to be independent of the overall number of the interferers in the acoustic scene. Two subsequent studies investigated how the sensitivity to TFS information relates to the BU abilities of elderly hearing-impaired (HI) listeners with high-frequency hearing loss. The sensitivity to monaural and binaural TFS information was assessed, by measuring frequency discrimination thresholds (FDTs) and interaural phase difference (IPD) discrimination thresholds (IPDTs) for pure tones below 1.5 kHz, respectively. The HI listeners performed worse than the NH in the speech tests and in both TFS tests, and they also showed slightly reduced BILDs. FDTs did not show any correlation with measures of speech intelligibility. In contrast, IPDTs showed a moderate correlation with BILDs. In general, the frequency range over which listeners were sensitive to IPDs was the best predictor

of their BILDs. NH and HI listeners also performed similarly when BILDs were elicited by small rather than by large ITDs, and binaural TFS coding deficits were linked to reduced BILDs to a similar degree in both cases. These results indicate a weak link between IPDTs and BU, suggesting that, HI listeners having reduced binaural TFS coding abilities (as assessed by IPDTs) can have close-to-normal BU abilities. The experiments presented in the last study examined whether these findings were a result of the HI listeners' ability to take advantage of ITDs in the high-frequency domain. BILDs were measured in conditions where the target and maskers were spatially separated by ITDs introduced below or above 1.25 kHz, or over the full frequency range of the stimuli. In contrast to the NH listeners, who were able to utilize ITDs above 1.25 kHz to a limited degree, the HI relied entirely on ITDs carried by the low-frequency stimulus content in the facilitation of BU.

The findings presented in this thesis contribute to the understanding of how supra-threshold deficits affect speech intelligibility in complex acoustic scenarios, and underline the role of robust binaural TFS coding in the facilitation of BU. Overall, HI listeners exhibited near-to-normal BILDs. Thus, the problems they experience in complex acoustical environments are likely due to monaural rather than binaural processing deficits.

Resumé

Lyttere med sensorineural hørenedsættelse har ofte vanskeligt ved talekommunikation, særligt i akustiske omgivelser med mange rumligt adskilte støjkilder. Kompensation af lytterens forhøjede høretærskel løser ikke nødvendigvis hørevanskelighederne. Det har derfor været foreslået at vanskelighederne skyldes kodningsforstyrrelser, der påvirker den indre auditive stimulus-repræsentation over høretærsklen. Forringet kodning af temporal finstruktur (TFS) ved udgangen af det indre øres filtre er en mulig kilde til disse forstyrrelser. Denne afhandling undersøger forholdet mellem auditiv kodning af stimuli over høretærsklen og rumlig taleperception med særligt henblik på forholdet mellem TFS-kodning og binaural demaskering (BU) af tale i støj i lateraliserede lyttescenarier.

Tærskler for taleforståelse (SRT) blev målt i forskellige støjbaggrunde i hovedtelefon-eksperimenter. Talekilden og støjmaskerne blev rumligt lateraliseret ved interaurale tidsforskelle (ITD). Binaural niveauforskel i forståelse (BILD) blev målt mellem de forskellige lateraliseringer. Det første studie undersøgte forholdet mellem talelateralisering og tale-forståelses-tærskler for normalt-hørende (NH) lyttere, hvor en stærk korrelation mellem disse mål blev fundet. Derudover blev det fundet at BILDs for et givet antal maskerende støjkilder var uafhængigt af det totale antal støjkilder i den akustiske scene. To efterfølgende studier undersøgte sammenhænge mellem TFS-sensitivitet og BU hos ældre hørehæmmede (HI) lyttere med højfrekvens-høretab. Sensitivitet for monaural og binaural TFS-information blev undersøgt ved at måle tærskler for frekvens-diskriminering (FDT) samt tærskler for diskriminering af interaural faseforskel (IPDT) for toner under 1.5 kHz. Hørehæmmede lyttere havde lavere taleforståelse end NH både i tale-test og i begge TFS-test og havde også reducerede BILD. FDT viste ingen korrelation med mål for taleforståelse. Derimod viste IPDT en moderat korrelation med BILD. Generalt var det frekvensområde, hvor lyttere var sensitive overfor IPD bedst til at forudsige BILD-målene. NH og HI-lyttere var også på tilsvarende niveau for BILD ved lave ITD og forringet binaural TFS-kodning var tilsvarende knyttet til reducerede BILD i begge tilfælde. Resultaterne indikerer en svag forbindelse mellem IPDT og BU, der antyder, at HI-lyttere med reduceret binaural TFS-kodning (som målt ved IPDT) kan have tæt på normale BU. Eksperimenterne i det sidste studie undersøgte, hvor-

vidt de fundne resultater kunne skyldes HI-lytters evne til at udnytte ITD ved høje frekvenser. BILD blev målt i betingelser hvor kilde og maske var rumligt adskilte med ITD enten under eller over 1.25 kHz eller med ITD over hele frekvensområdet. Modsat NH-lyttere, der i begrænset omfang gjorde brug af ITD over 1.25 kHz, så afhang facilitering af BU hos HI-lyttere udelukkende af den lavfrekvente del af stimuli.

Resultaterne præsenteret i denne afhandling bidrager til forståelsen af, hvordan forstyrrelser over høretærsklen påvirker taleforståelse i komplekse akustiske scenarier og understreger den rolle som robust binaural TFS kodning spiller for facilitering af BU. Generelt udviste HI tæt ved normale BILD. De vanskeligheder, som HI oplever i komplekse akustiske omgivelser skyldes derfor sandsynligvis hovedsageligt monaurale snarere end binaurale processerings-forstyrrelser.

Acknowledgments

It was more than four years ago when I entered the building of the Hearing Systems group at Akustikvej 352 for the first time. Years have passed, my life has clearly changed (for the better!), but there is one thing that stayed the same: the atmosphere in this group. It is the same honest, welcoming and friendly environment that was almost shocking for me the very first time.

I would like to express my sincere gratitude to my supervisors. Thank you, Ewen, for being so cheerful, positive and accessible every time I needed your guidance, and for the long hours we have spent working together. Thank you for motivating and comforting me when I felt that I reached my limits; and thank you for dealing with me the right way both when I was desperate or overexcited. Thank you, Sébastien, for the discussions, corrections and all the great ideas you have contributed to my development with. Especially thank you for welcoming me in the group in the first months. Torsten, thank you for always being so inspiring, critical and constructive. Your true and honest approach to research is only surpassed by your true and honest approach to people. I am truly happy and thankful that I had the opportunity to work with you.

I would like to thank Julie Hefting Pedersen, Søren Laugesen, Tim Schoof, Stuart Rosen and Jens Hjortkjær for the collaborations. It was a great pleasure to work with you, and all of you are a great inspiration to me. I have learned a great deal from you, both professionally and personally, the latter of which I am most grateful for.

I also want to thank Jens Cubick and Johannes Zaar for the many discussions we had. I clearly would not have been able to get this far without your friendship and support. Also thanks for the photo and Ligretto! sessions, the climbing, and

all the chats and laughter we shared. I am glad that there were plenty of these. Thank you also, Fede, for making our office possibly the least boring place in whole Denmark, and thank you, Michal, for the jokes, ideas, and cheering during the Euro Cup.

Thank you to my friends and colleagues in the Acoustic Technology and Hearing Systems group for the fantastic atmosphere at work, and especially to Caroline van Oosterhout for always helping me to deal with practicalities during my studentship. I also thank the members of the INSIPRE Framework – especially Tom Barker and Kurt Steinmetzger – all the joy they have brought into my life.

A big and special thanks goes to Annette Fugmann. I am grateful for your unlimited kindness, your friendship and all of our chats and laughter. You have not only offered me a place to stay, but a place to go home to. I feel greatly indebted for all this. Tusind tak!

Last but not least, I want to thank my family, my friends here and abroad, and Alexandra for their support and unconditional love. From you, I learn something new about life and happiness every day.

Gusztáv Lőcsei, July 14, 2016

Related publications

Journal papers

- Lócsi, G., Hefting Pedersen, J., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E. N. (**in press**). “Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners,” *Trends in Hearing*
- Lócsi, G., S., Santurette, S., Dau, T., and MacDonald, E. N. (**in preparation**). “Lateralized speech perception with small interaural time differences in normal-hearing and hearing-impaired listeners,” *to be submitted to Hearing Research*

Conference papers

- Lócsi, G., Santurette, S., Dau, T., and MacDonald, E. N. (**2014**). “Influence of acoustic complexity on spatial release from masking and lateralization,” *Proc. of Forum Acusticum*, Krakow, Poland, September 2014.
- Lócsi, G., Hefting Pedersen, J., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E. N. (**2015**). “Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners,” *Proc. of the 5th International Symposium on Auditory and Audiological Research (ISAAR): Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*, pp. 389-396, Nyborg, Denmark, August 2015.

Published abstracts

- Lócsi, G., Hefting Pedersen, J., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E. N. (2015). “Lateralized speech perception, temporal processing and cognitive function in normal hearing and hearing impaired listeners,” *7th Speech in Noise Workshop*, Copenhagen, Denmark, January 2015.
- Hefting Pedersen, J., Lócsi, G., Laugesen, S. (2015). “Relations between real versus ITD-only spatial release from masking and working memory,” *Proc. of the 3rd International Conference on Cognitive Hearing Science for Communication (CHSCOM)*, Linköping, Sweden, June 2015.
- Lócsi, G., Santurette, S., Dau, T., and MacDonald, E. N. (2016). “Binaural unmasking of speech with small interaural time differences and its relationship to binaural temporal fine structure processing,” *Proc. of the Speech Processing in Realistic Environments (SPIRE) workshop*, Groningen, Netherlands, January 2016.
- Lócsi, G., S., Santurette, S., Dau, T., and MacDonald, E. N. (2016). “Lateralized speech perception with small interaural time differences in normal-hearing and hearing-impaired listeners,” *Forschritte der Akustik DAGA 2016, 42nd German Convention on Acoustics*, Aachen, Germany, March 2016.

Contents

Abstract	v
Resumé på dansk	vii
Acknowledgments	ix
Related publications	xiii
Table of contents	xvii
List of abbreviations	xix
1 Introduction	1
2 Influence of acoustic complexity on spatial release from masking and lateralization	7
2.1 Introduction	8
2.2 Methods	9
2.2.1 Listeners	9
2.2.2 Experimental set-up and stimuli	9
2.2.3 Procedure	10
2.3 Results	13
2.3.1 SRTs and lateralization thresholds	13
2.3.2 Release from masking	14
2.3.3 Comparison of the SI and lateralization results	16
2.4 Discussion	17
2.5 Conclusion	19

3	Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners	21
3.1	Introduction	22
3.2	Methods	27
3.2.1	Participants	27
3.2.2	Temporal processing and cognitive skills	28
3.2.3	Speech perception in noise	30
3.2.4	Statistical tools	33
3.3	Results	34
3.3.1	Audiometric thresholds	34
3.3.2	Temporal processing	34
3.3.3	Cognitive abilities	36
3.3.4	Speech perception in noise	36
3.3.5	Predicting speech intelligibility	40
3.4	Discussion	43
3.5	Conclusions	49
4	Lateralized speech perception with small interaural time differences in normal-hearing and hearing-impaired listeners	51
4.1	Introduction	52
4.2	Methods	56
4.2.1	Participants	56
4.2.2	Apparatus	56
4.2.3	Binaural fine structure processing	57
4.2.4	Speech perception in noise	60
4.2.5	Statistical analysis	63
4.3	Results	63
4.3.1	Audiometric thresholds	63
4.3.2	Binaural fine structure coding	64
4.3.3	Speech perception in noise	66
4.3.4	Binaural TFS coding and unmasking of speech	69
4.4	Discussion	71

4.4.1	Summary of main findings and relations to other studies . . .	71
4.4.2	Limitations of the study	74
4.4.3	Perspectives	76
4.5	Conclusions	77
5	Contributions of low and high frequencies to binaural unmasking in hearing-impaired listeners	79
5.1	Introduction	80
5.2	Methods	82
5.2.1	Listeners	82
5.2.2	Binaural temporal fine structure coding	83
5.2.3	Speech intelligibility tests	83
5.3	Results	86
5.4	Discussion	89
5.5	Summary and conclusion	91
6	Overall discussion	93
6.1	Summary of main findings	93
6.2	Implications for lateralized speech perception	96
6.3	Limitations and perspectives	97
	Bibliography	101
	Collection volumes	113

List of abbreviations

<i>nI-mAFC</i>	<i>n</i> -Interval, <i>m</i> -Alternative, Forced-Choice
ANOVA	ANalysis Of VAriance
BILD	Binaural Intelligibility Level Difference
BMLD	Binaural Masking Level Difference
EM	Energetic Masking
ENV	ENvelope
FIR	Finite Impulse Response
FDT	Frequency Discrimination Threshold
HATS	Head And Torso Simulator
HI	Hearing-Impaired
(dB) HL	Hearing Level
HL	Hearing Loss
HTL	Hearing Threshold Level
ILD	Interaural Level Difference
IM	Informational Masking
IPD	Interaural Phase Difference
IPDT	Interaural Phase Difference detection Thresholds
ITD	Interaural Timing Difference
NH	Normal-Hearing
PTA	Pure Tone Average

RMS	Root-Mean-Square
SD	Standard Deviation
SI	Speech Intelligibility
SL	Sensation Level
SNHL	SensoriNeural Hearing Loss
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRM	Spatial Release from Masking
SRT	Speech Reception Threshold
SSN	Speech Shaped Noise
TFS	Temporal Fine Structure

1

General introduction

In everyday life, being able to hear and to understand speech are fundamental abilities that allow us to establish and maintain communication with our peers. For normal hearing (NH) listeners in quiet surroundings, speech understanding happens almost unintentionally, even in the presence of considerable background noise. People with hearing impairment (HI), however, exhibit great difficulties in speech understanding. The prevalence of hearing loss (HL) is likely to increase over the upcoming years as a result of the aging population and a greater noise exposure of the younger generations at entertainment venues or due to the extensive use of personal audio devices. To provide efficient hearing rehabilitation strategies, it is crucial to gain a better understanding of the processing mechanisms in the healthy and in the impaired auditory system, especially in the context of the perceptual consequences of hearing loss on speech perception. Traditionally, the severity of hearing loss is assessed in the clinic with an audiogram, that measures the listeners' pure-tone detection thresholds at different frequencies. Speech intelligibility (SI) performance in noise can be quantified by measuring the proportion of a spoken material (e.g. consonants, words or sentences) listeners can understand at a fixed signal-to-noise ratio (SNR), or by measuring the SNR at which listeners understand a fixed proportion of the material. Speech reception thresholds (SRT) refer to the SNR at which 50% intelligibility is obtained.

In everyday listening scenarios, sound sources are often spatially distributed. The phenomenon that listeners can focus their auditory attention on the talker of their interest (target) while suppressing other interferers (maskers) has fascinated researchers for several decades and it has been termed the "cocktail-party effect" (Cherry, 1953). The noise-robustness of NH listeners to spatially distributed interferers mainly resides in their ability to detect and exploit spatial acoustic cues

associated with the individual noise sources in the acoustic scene. As the spatial distance between the target and the maskers increases, SI performance typically improves, i.e. SRTs are reduced. This phenomenon is called spatial release from masking (SRM) and is measured as the difference in SRTs when target and maskers are spatially colocated and spatially separated. In the horizontal plane, SRM due to angular separation is mainly facilitated by better-ear listening (BE) and binaural unmasking (BU) (see e.g. Bronkhorst and Plomp, 1988). BE is a listening strategy whereby listeners achieve better SI performance by attending only to the ear that momentarily receives the signal at the more favorable SNR. As a masker gets separated from the target presented frontally, some of the high-frequency energy of the masker will be “shadowed” by the head at the contralateral ear of the listener. Interaural level differences (ILD) arise between the ears and a direct increase in the SNR appears at the contralateral ear. Listeners are also sensitive to disparities in the interaural timing differences (ITD) of the arriving sound waves. BU refers to the phenomenon whereby the effective SNR at which listeners perceive the acoustic mixture is raised by these disparities in the ITDs associated with the target and the maskers. This benefit can be expressed as the magnitude of the difference in the SRTs between the two binaural presentation modes (in dB), and is often referred to as binaural intelligibility level difference or BILD (see e.g. Levitt and Rabiner, 1967). Finally, when the target and maskers are perceptually similar, spatial separation further aids SI by facilitating the perceptual segregation and discrimination of the target from the maskers (Kidd et al., 2008). Having access to intact spatial acoustic cues is essential to solve the “cocktail party” problem.

The difficulties of HI people in speech understanding are most pronounced in acoustically complex tasks, in the presence of e.g. reverberation or multiple, spatially distributed interferers. Their problems manifest themselves both in increased SRTs and reduced SRMs. Plomp (1978) proposed that the effect of hearing loss on speech perception can be interpreted as a combination of an audibility and a distortion component. Stimulus inaudibility due to elevated hearing thresholds only limits SI performance in quiet but does not affect SRTs in noisy environments when the target speech is sufficiently audible. In contrast, hearing loss can also lead to a distortion of the internal representation of supra-threshold stimuli in

the auditory system. This affects SI in noise by increasing the SRTs even when the target is above the limit of audibility. Possible sources of such distortion deficits could include the broadening of the auditory filters (e.g. Glasberg and Moore, 1986) or degraded temporal processing abilities (Lorenzi et al., 2006; Hopkins et al., 2008; Strelcyk and Dau, 2009; Papakonstantinou et al., 2011).

From a signal processing point of view, the cochlea can be modelled as a series of overlapping band-pass filters, which decompose broadband stimuli into a series of narrowband time signals, each of which can be considered as rapid oscillations (temporal fine structure, TFS) modulated by slowly varying envelope (ENV) fluctuations (Moore, 2008). While in quiet, ENV cues can sufficiently foster robust speech perception (Shannon et al., 1995; Lorenzi et al., 2006; Lorenzi and Moore, 2008), TFS cues are thought to be critical for speech understanding in noise (Papakonstantinou et al., 2011), especially in the presence of interfering talkers (Zeng et al., 2005; Strelcyk and Dau, 2009; Lunner et al., 2012; Moore et al., 2012).

There is ample evidence that both sensorineural hearing loss (SNHL) and aging can lead to a reduced sensitivity to TFS information, and thereby impair speech perception. It has been shown that SNHL can degrade performance in monaural measures of TFS coding, including frequency discrimination thresholds of pure tones (Tyler et al., 1983; Moore and Peters, 1992) and of harmonic complexes (Moore and Peters, 1992), or low-rate frequency modulation detection thresholds of pure tones (Lacher-Fougère and Demany, 1998; Strelcyk and Dau, 2009; Santurette and Dau, 2012). Furthermore, considering binaural measures of TFS coding, both aging and SNHL affect the detection of interaural timing differences (Lacher-Fougère and Demany, 2005; Ross et al., 2007b; Hopkins and Moore, 2011; King et al., 2014) or binaural masking level differences (BMLDs, Strouse et al., 1998; Strelcyk and Dau, 2009; Papakonstantinou et al., 2011) of low-frequency pure tones. The exact role of TFS information in speech perception is, however, not yet fully understood. Currently, it appears most likely that TFS facilitates speech perception by providing acoustic cues that aid the separation of the target from the maskers (Zeng et al., 2005; Strelcyk and Dau, 2009; Lunner et al., 2012; Moore et al., 2012) and/or by aiding spatial segregation (Strelcyk and Dau, 2009; Neher et al., 2011; Neher et al., 2012).

A reduction in TFS coding abilities may negatively affect speech perception in spatial scenarios in particular, as timing differences in the low-frequency TFS carry important information both for the localization of broadband stimuli (Wightman and Kistler, 1992) and for the BU of speech (see e.g. Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988). The picture that emerges from earlier studies regarding the BU of speech in HI listeners is, however, not entirely consistent with this conjecture. For listeners with a symmetrical HL, some studies showed smaller than normal BILDs (George et al., 2012; Best et al., 2013). Other studies showed normal or close-to-normal BILDs (Bronkhorst and Plomp, 1989; Strelcyk and Dau, 2009; Goverts and Houtgast, 2010) for listeners that had a HL and/or were mid-aged or older, both of which are known to potentiate reduced binaural sensitivity to TFS.

Studies directly assessing the role of TFS coding deficits on BILDs in symmetrically impaired listeners have also reached diverging conclusions. Goverts and Houtgast (2010) used a “distortion-sensitivity” approach to demonstrate that listeners with reduced BILD were less sensitive to phase and time perturbations in the stimuli than those with normal BILDs. Strelcyk and Dau (2009) found significant correlations between BMLDs and BILDs in SSN. In contrast, Santurette and Dau (2012) obtained mixed results when they measured SRTs in SSN in a headphone experiment using head-related transfer functions. The authors found that the benefit through binaural interactions was correlated with BMLDs measured for 500 Hz and 1 kHz pure tones in noise, but not with the upper frequency limit for the detectability of a 180 degree interaural phase difference (IPD) imposed on a pure tone. The studies of Neher et al. (2011; 2012) further indicated a link between binaural measures of TFS coding and spatial speech perception in free-field listening. However, as these studies compared binaural TFS measures to SRTs, and because in free-field listening scenarios both BE and BU are potential facilitators of speech perception, it is not possible to extrapolate the role of TFS coding in BU alone from these results.

This thesis further investigated the perception of speech in spatialized environments with a special focus on its relation to supra-threshold coding deficits in the temporal domain. Speech intelligibility in lateralized noise and also BILDs were measured in a variety of background noise conditions, and compared between

the normal and impaired-hearing population. The role of TFS cues on BU was investigated by comparing the results of the speech tests to both monaural and binaural measures of TFS coding. Furthermore, the role of ITDs in the facilitation of BU at low- and high frequency bands was evaluated in both listener groups.

Outline of the present work

Chapter 2 investigates the binaural temporal processing abilities of young NH listeners in terms of lateralized speech perception. Speech reception thresholds (SRTs) and lateralization thresholds were measured in acoustically complex environments, where both the number of maskers and their spatial distribution was varied. The study examines how the number of maskers and their distribution affects the BU of speech. Also, the relationship between speech lateralization and speech intelligibility performance is explored.

Chapter 3 presents an experiment where SRTs and BILDs were estimated for young NH and elderly HI listeners with normal-hearing at and below 1.5 kHz in attentionally complex, lateralized speech tasks. Also, cognitive and TFS coding abilities were assessed with the reading span test and with frequency discrimination and IPD detection thresholds at 250 Hz. The relationship between monaural or binaural TFS coding and the results of the speech tests are investigated under the hypothesis that, once audibility is compensated for, TFS coding abilities are predictive of SI performance.

Chapter 4 investigates the effect of the amount of lateralization on the BILDs in young NH and elderly HI listeners, and clarifies the involvement of TFS cues in the BU of speech when triggered by small and large ITDs. SRTs and BILDs were measured in speech-shaped noise and in two-talker babble. Also, pure tone IPD detection thresholds over a broad frequency range were assessed. These thresholds, and the smallest ITDs listeners were able to detect in pure tones, are compared with the amount of BILDs elicited by small or large ITDs. It is hypothesized that reduced binaural TFS coding in the HI population imposes a limitation on the amount of BU due to small ITDs, or on the minimum ITD value with which BU

can be triggered, while the magnitude of BILDs may remain close to normal when triggered by large enough ITD values.

In **Chapter 5**, the independent contributions of TFS and ENV ITDs on the BU of speech are evaluated in a similar group of listeners as in the previous chapters. The experiment was based on the study of Edmonds and Culling (2005). SRTs and BILDs were measured in two-talker babble, where both the target and the maskers were split into a low- and high-frequency region at a splitting frequency of 1.25 kHz. The contribution of different stimulus frequency regions in the BU of speech are evaluated in the two listener groups, by comparing BILDs for presentations where the maskers were lateralized to the side in the low, in the high or in both frequency regions.

Finally, **Chapter 6** summarizes the main findings of each chapter, discusses the implications of these findings across the different chapters and offers a number of suggestions for further research.

2

Influence of acoustic complexity on spatial release from masking and lateralization^a

Abstract

Speech reception thresholds (SRTs) and lateralization thresholds were measured over headphones in reversed babble noise consisting of 2, 4, 8 and 12 talkers for nine young listeners with normal hearing. The perceived locations of the target and the individual maskers were steered separately towards the left or right side of the head. For a fixed number of maskers, the distribution of interfering talkers was varied in terms of the number of maskers colocated with the target. The stimuli were spatialized by applying 0.68 ms interaural timing differences (ITD) between the ears, keeping the monaural signal-to-noise ratio constant for noise conditions with the same number of maskers but with different masker distributions. The performance between the speech intelligibility and lateralization tasks was highly correlated. In the speech intelligibility experiments, no substantial spatial-release from masking (SRM) occurred when one or more maskers were colocated with the target. Spatially separating the last 2 or 4 interferers resulted in the same SRM, independently of the overall number of maskers. The results suggest that the SRM through ITDs was independent of the overall number of interfering talkers in the acoustic scene.

^a This chapter is a revised version of Lócssei et al. (2014)

2.1 Introduction

The “cocktail party problem”, as introduced by Cherry (1953), refers to the ability of human listeners to selectively attend to a single voice in a multitude of interfering talkers. This ability partly resides in the listeners’ sensitivity to the spatial position of the different sound sources in complex acoustic environments. As interferers (or maskers) are moved away from the spatial position of the sound source of their interest (target), speech intelligibility (SI) increases, a phenomenon referred to as spatial release from masking (SRM). Many studies have focused on speech perception in spatially complex environments (e.g. Peissig and Kollmeier, 1997; Bronkhorst and Plomp, 1992; Hawley et al., 2004), establishing the dependence of SRM on both the number and the spatial distribution of the maskers.

In the horizontal plane, SRM is mainly facilitated by the disparities of the interaural level- and timing differences (ILDs and ITDs) of the arriving acoustic signals associated with each sound source. ILDs arise between the ears as the head acoustically shadows an incoming sound wave at the contralateral ear. This also means that if a target and the maskers are presented at different azimuthal angle, the signal-to-noise ratio (SNR) will differ between the ears, allowing the listener to focus on the ear receiving a more favorable SNR (better-ear listening). Similarly, ITDs also vary with spatial position due to the different travel-path lengths of the acoustic waves to the ears. Disparities in the ITDs of the target and the maskers aid SRM by indirectly increasing the SNR of the acoustic mixture through binaural unmasking (BU).

The present study investigated the binaural temporal processing abilities of young listeners with normal hearing in terms of speech perception in complex acoustic environments. Speech in noise stimuli were presented in two headphone experiments. Acoustic complexity was controlled in terms of the overall number of maskers and their spatial distribution in the scene. The stimuli were lateralized to the left or the right side of the head using ITDs only, therefore eliminating any differences arising from better-ear listening between conditions with an equal number of maskers. Various distributions where all, some or none of the maskers were lateralized towards the side of the target were tested. Speech reception thresholds

(SRTs), SRM due to BU (from now on simply SRM) and lateralization thresholds were assessed in all noise conditions. The main goals of the experiments were 1) to assess whether the overall number of maskers in the scene affects the amount of SRM from a fixed number of maskers and 2) to investigate the relationship between speech lateralization and speech intelligibility performance.

2.2 Methods

2.2.1 Listeners

A total of 9 listeners between the ages of 18 and 27 (median: 22) participated in this study. All were native Danish speakers with normal audiometric thresholds (i.e., ≤ 20 dB HL at standard audiometric frequencies) and reported no speech or language impairments. The experiments were carried out in accordance to the ethical approval granted by the Science-Ethics Committee for the Capital Region of Denmark (reference H-3-2013-004). All listeners provided informed consent and were paid for their participation.

2.2.2 Experimental set-up and stimuli

Measurements were carried out in a double-walled sound attenuating listening booth. The listening tests were implemented in MATLAB, and subjects provided their responses using a computer interface. The stimuli were presented through Sennheiser HDA200 headphones using an RME DIGI96/8 sound card at a sampling rate of 44.1 kHz. The calibration of the headphones was done using a Brüel and Kjær (B&K) 4153 artificial ear connected to a B&K 2636 sound pressure level meter, and with a B&K 4230 artificial ear calibrator. In order to achieve a flat frequency response in the headphones, equalization filters were applied to all stimuli before presentation.

Target stimuli. The target sentences used were from the Danish DANTALE II corpus (Wagener et al., 2003). The five-word sentences are spoken by a female talker and have a fixed syntactical structure (*<name> <verb> <number> <adjective>*)

<noun>), where each word is randomly chosen from 10 alternatives. The corpus is organized into 16 lists containing 10 sentences each. To reduce testing time, only the first 1.5 seconds of the sentences were used in the lateralization experiment.

Noise stimuli. To generate the noise stimuli for the conditions having 2, 4, 8 and 12 maskers, the first 2, 4, 8 and 12 male talkers from the GRID corpus (Cooke et al., 2006) were used, respectively. Preprocessing of the stimuli included up-sampling to 44.1 kHz, silent frame removal, and then time-reversal. For silent frame removal, each recording was segmented into 50-ms windows with 25 ms overlap. Silent frames were defined as frames having an energy at least 30 dB lower than the maximum energy present among the frames in the recording. For each interfering talker, a continuous stream of speech was generated by concatenating the preprocessed recordings. From each of the resulting streams, 50 non-overlapping regions with a duration of 5 s were chosen randomly. These segments were then windowed (0.1 s cosine onsets and offsets) and stored as masker trials for the speech intelligibility test. A similar procedure was used for the lateralization test, but with masker segments of 4 s duration.

2.2.3 Procedure

Both in the SI and lateralization tests, the individual maskers were presented at 50 dB SPL and the target level was adaptively varied to estimate the threshold. This means that the overall masking level increased with the number of masking streams (i.e., for 2, 4, 8 and 12 maskers, the long-term overall masker level was 53, 56, 59 and 60.8 dB SPL, respectively).

In each trial, the onset of the maskers preceded that of the target by at least 2 s. The perceived lateralized position of each stream (i.e., both maskers and target) was steered independently to the left or to the right using 0.68 ms ITDs. Here, masker streams having the same leading side as the target in the actual trial will be referred to as “colocated”, while maskers lateralized towards the lagging side of the target will be called “separated”. The lateralized position of the target sentence was randomized from trial to trial. Thus, subjects had no *a priori* knowledge about which side to attend to in each trial. A total of 16 conditions varying in the total

number and in the number of colocated maskers were tested.

The different noise conditions are denoted in the form of R_y^x in the following, where y and x indicate the overall number of maskers and the number of colocated maskers, respectively. For example, R_8^2 denotes a condition with 8 maskers, from which 2 have the same lateralized position as the target. Using the notation above, the following 16 conditions were tested: R_2^2 , R_2^1 , R_2^0 , R_4^4 , R_4^2 , R_4^0 , R_8^8 , R_8^6 , R_8^4 , R_8^2 , R_8^0 , R_{12}^{12} , R_{12}^8 , R_{12}^6 , R_{12}^4 and R_{12}^0 . In the remainder of the paper, condition groups having 2, 4, 8 and 12 interferers will be denoted as R_2^x , R_4^x , R_8^x and R_{12}^x , respectively.

Threshold estimates for conditions involving the same number of interferers were run in parallel (e.g., estimation of thresholds for R_2^2 , R_2^1 , and R_2^0 were conducted simultaneously). If these estimates had not been collected simultaneously, the listeners would have known which side to attend to, based on the distribution of maskers. For example, if only R_2^0 trials were presented, listeners would know that the target was lateralized to the side opposite that of the two maskers. The order in which condition groups were tested was based on the overall number of maskers. Thus, estimates were collected for R_2^x first and R_{12}^x last.

Speech intelligibility tests. Thresholds for 50% correct response were estimated using the standard procedure of the DANTALE II tests (Wagener et al., 2003). Thresholds were estimated using 3 lists per condition. Subjects provided their response using a computer interface, which contained a 10×5 matrix, where columns corresponded to the word categories for the sentences and each row contained one of the alternatives per category. Subjects were requested to mark the words occurring in the target sentence. The presentation level of the next target sentence was adjusted based on the number of correctly identified words. The initial presentation level was set to the estimated level of the interferers (i.e., 53, 56, 59 and 61 dB SPL for R_2^x , R_4^x , R_8^x and R_{12}^x , respectively). The order of sentence presentation within lists was randomized. Lists for each condition were chosen in a semi-randomized way. Each of the 16 lists was presented exactly 3 times during the speech intelligibility tests and none of the conditions contained the same list twice. Thresholds were calculated from the average of the last 20 presentation levels. Before the test session, a training session was performed with 5 random sentences in each condition (80 sentences overall).

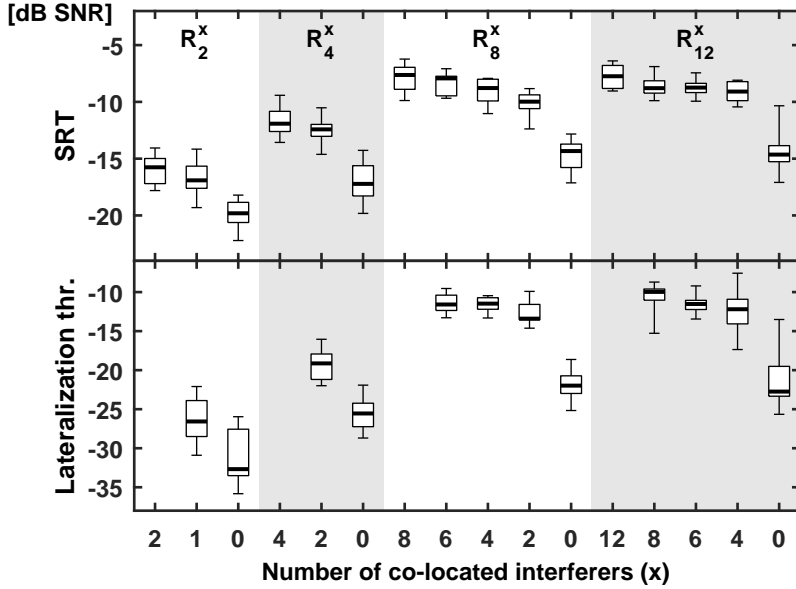


Figure 2.1: Box plots for the speech intelligibility (top) and lateralization (bottom) experiments. The whiskers extend to the most extreme data points. Condition groups using the same amount of interferers are defined at the top of the shaded areas. The number of interferers collocated with the target side (x) are denoted on the abscissa.

Lateralization tests. Thresholds for 79.4% correct response were estimated using a 1-up 3-down procedure in a 2-interval 2-alternative forced choice task. Subjects had to respond using a computer interface, and were requested to indicate which side they heard the target sentence coming from (i.e., "left" or "right"). Each condition consisted of at least 10 reversals. The first two had a step size of 5 and 2 dB respectively, while the last 8 had a step size of 1 dB. Thresholds were calculated based on the arithmetic average of the last 8 reversals. Since it was necessary to test all conditions within a group simultaneously, the threshold tracking did not stop until 10 reversals had occurred for all conditions within the group.

All listeners, except one, performed all the tests in two sessions, the first being the speech intelligibility test and the second being the lateralization test. One listener performed the speech intelligibility test in two sessions.

2.3 Results

The results of the SI and lateralization tests were expressed in dB SNR, i.e. as target threshold levels minus the estimated overall presentation levels of the masker (53, 56, 59 and 60.8 dB SPL in the 2-, 4-, 8- and 12-masker conditions). The statistical analyses were considered significant at an alpha-level of 0.05. In all of the reported repeated-measures analysis of variances (ANOVA) measures, if the assumption of sphericity was violated, the degrees of freedom were corrected with Greenhouse-Geisser estimates.

2.3.1 SRTs and lateralization thresholds

Figure 2.1 shows the speech reception (top) and lateralization thresholds (bottom) presented as box plots, grouped by the overall number of interferers in the scene (shaded areas). The results of the SI experiment generally show very low SRT values, ranging from an average of approximately -19.9 to -7.8 dB across the conditions. The average SRT for conditions having the same number of interferers (i.e., condition groups) increases gradually from approximately -17.5 to -10.3 dB as the number of maskers increases from 2 to 8, where it plateaus, yielding no further increment for 12 maskers. Within condition groups, SRTs decrease (i.e. intelligibility increases) as more and more maskers are lateralized towards the side opposite to the target. These observations were supported by a two-way ANOVA with repeated measures. Comparing the fully colocated (R_x^x), half colocated half-separated ($R_x^{x/2}$) and fully separated (R_x^0) conditions across all condition groups revealed a significant main effect of both the number of interferers ($F(1.6, 12.7) = 315.6, p < 0.001$) and the interferer distribution ($F(2, 16) = 415.6, p < 0.001$), and a significant interaction ($F(6, 48) = 13.3, p < 0.001$). *Post hoc* analyses using paired t-tests with Bonferroni correction on each condition group showed, that the R_x^0 conditions were significantly different from all the other conditions ($p < 0.001$). For 8 and 12 interferers, a repeated-measures ANOVA revealed no significant main effect for the number of interferers ($F(1, 8) = 3.76, p = 0.09$), but a significant main effect for the distribution ($F(4, 32) = 222.1, p < 0.001$) and a marginally significant

interaction ($F(1.7, 13.4) = 5.1, p = 0.03$).

With the exception of the fully colocated conditions (R_x^x), the trends in the lateralization experiments follow a similar pattern as in the SI experiments. For the fully colocated conditions, listeners could give the correct answer without actually hearing the target stimulus, but instead, noting its absence on the opposite lateralized side. This trend was only observed for conditions having 4 or more interferers. It is likely that this pattern is a result of an order effect: each of the participants completed the test procedure starting with R_2^x , and they might have become aware of this listening strategy later on in the test session.

A two-way repeated measures ANOVA was performed on the $R_x^{x/2}$ and R_x^0 conditions in all condition groups. The results showed a significant main effects for both the number of interferers ($F(3, 24) = 92.4, p < 0.001$), and distribution ($F(1, 8) = 318.6, p < 0.001$), as well as a significant interaction ($F(3, 24) = 13.1, p < 0.001$). *Post hoc* analyses using paired t-tests (with Bonferroni correction) on each condition group showed that the R_x^0 conditions were significantly different from all of the other conditions ($p < 0.001$). For 8 and 12 interferers, a significant main effect was found for the distribution ($F(3, 24) = 131.3, p < 0.001$), but not for the number of interferers ($F(1, 8) = 0.49, p = 0.5$) or for the interaction term ($F(3, 24) = 0.12, p = 0.95$).

2.3.2 Release from masking

SRM due to binaural unmasking was calculated to express the SRT benefit obtained when the distribution of the interferers was changed from a difficult layout to a more favorable one. In this case, this meant finding the SRT benefit when a fixed number of colocated masker streams were steered towards the other side of the head.

The overall SRM, calculated as the difference in SRTs between the fully colocated and fully separated conditions, increases from 4 to 6.6 dB as the number of interferers increases from 2 to 12, reaching a maximum of 6.9 dB with 8 talkers. The statistical significance of this increase in SRM with the number of interferers is supported by the significant interaction term of the corresponding ANOVA. The individual values range from a minimum of 2.4 dB with 2 interferers to a maximum

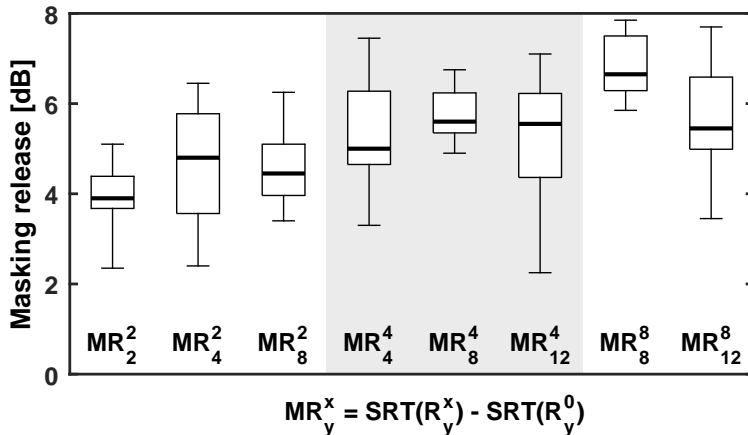


Figure 2.2: Box plots for the masking release by removing the last 2, 4 or 8 colocated interferers from the target side. MR_x^y stands for masking release between condition R_x^y and R_x^0 .

of 8 dB with 12 interferers. However, as interferers are moved away from the target side, no gradual SRM can be observed. Instead, SRT values decrease modestly as interferers are shifted away from the target side, and drop substantially, once all the maskers are separated from the target.

In Figure 2.2, SRM values are denoted as MR_y^x , and refer to the SRT benefit obtained by removing the last x colocated interferers, calculated as follows:

$$MR_y^x = \text{SRT}(R_y^x) - \text{SRT}(R_y^0). \quad (2.1)$$

There appears to be no effect of the overall number of interferers on the SRM from removing the last 2 or 4 colocated maskers. This was confirmed by paired t-tests with Bonferroni correction. There is a small difference of about 1.5 dB between SRMs when removing the last 8 talkers from the target side in the 8 and 12 interferer conditions.

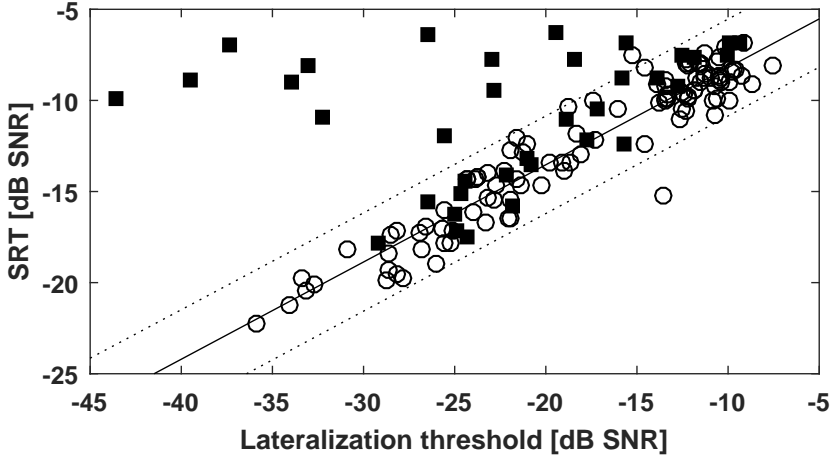


Figure 2.3: Correlation between SRTs and lateralization thresholds for each condition of each subject. Filled squares represent the conditions where all the interferers were on the target side. The solid and dashed lines indicate the regression line fitted to these data points with the 0.95% confidence bounds, respectively. Open circles indicate all the other conditions.

2.3.3 Comparison of the SI and lateralization results

Investigation of the SI and lateralization tests was done by correlation analysis. As listeners used a different paradigm for the detection task in the fully collocated conditions, these results were excluded from the analysis. Nevertheless, these are also plotted in Figure 2.3 (filled squares), which presents a scatterplot of all individuals' lateralization vs. SI results in all tested conditions. For each listener, SRT and lateralization thresholds were significantly correlated for all conditions where at least one of the maskers was collocated with the target. A good fit was obtained to the data of the individual listeners with linear regression, with r^2 values ranging between 0.80 and 0.98. The slope of the regression lines varied between 0.46 and 0.60. The regression line in Figure 2.3 was fitted to the full dataset (not counting the fully collocated data). The dashed lines indicate the 0.95% confidence bounds. This fit had a slope of 0.53 and an $r^2 = 0.90$.

2.4 Discussion

In this study, speech reception and lateralization thresholds were measured for normal hearing subjects in time-reversed babble noise. Both speech and masker stimuli were lateralized using fixed ITDs. Results showed that while increasing the number of interferers increased the SRT, it did not affect the masking release per masker stream. Also, lateralization thresholds showed a strong correlation with SRTs.

The observed patterns of the SRTs can be explained in terms of monaural and binaural effects. The monaural effect can be directly related to the number of interferers in the acoustic scene, underlying the differences in average SRTs between condition groups having the same amount of maskers. SI performance deteriorated with an increasing number of maskers, which can be explained in terms of dip listening. Running speech has a sparse spectro-temporal structure due to the inherent envelope fluctuations in different frequency channels. Since the individual masker streams were also speech-based, they offered spectro-temporal dips in which the listeners had a chance to glimpse the acoustic information carried by the target. As the number of maskers increased from 2 to 8, envelope fluctuations in the masking noise gradually disappeared, providing less opportunity for the listeners to benefit from glimpsing the target in the spectro-temporal gaps of the maskers. Apparently, the gaps in the masker streams were already significantly reduced with 8 maskers, and no disadvantage arose from adding 4 more maskers (i.e. to a total of 12 maskers), as indicated by the similar SRTs in the R_8^x and R_{12}^x condition groups.

The differences in SRTs within condition groups can be explained in terms of binaural effects. As summarized by Bronkhorst (2000), masking release due to BU typically ranges from 1 to 7 dB, depending on the number and spatial configuration of the talkers, and on the speech material itself. The SRMs in the current study show a good correspondence with these values, ranging from 4 to 7 dB when all maskers are separated from the target, increasing with the overall number of maskers. These results are consistent with the study of Hawley et al. (2004), who also found that SRM increases with a greater number of fluctuating interferers in the scene. The

results can be explained in terms of the framework suggested by Goverts et al. (2007), who proposed that BU processes are effective in time frames where the instantaneous SNR is relatively low. In this context, as the spectro-temporal dips in the maskers reduce with the increasing number of maskers, binaural processes can provide more masking release, since the instantaneous SNR of the sound mixture is on average for a longer period of time in the working range of BU.

Interestingly, the benefit from spatially separating the final 2 or 4 colocated maskers was the same, regardless of the total number of maskers in the noise condition. For 8 maskers a small but statistically significant difference was observed. Even though the overall number of maskers raised the average SNR the listeners needed to yield a criterion performance, it did not affect the magnitude of SRM when a fixed number of maskers was separated from the target. These findings suggest a partial independence of monaural and binaural effects in the current test setup. While the amount of temporal fluctuations affects the overall SRM listeners can achieve, BU acts equally efficiently for a fixed amount of maskers, independently of the noise floor in the actual condition.

For each individual, a strong link was observed between the lateralization and speech intelligibility performance. As expected, lateralization thresholds were lower than the SRTs in the corresponding condition, but the slope of the fitted regression line indicated that lateralization thresholds spanned over about twice as big of a range as SRTs. Threshold differences in the two tasks gradually decreased with increasing number of interferers, since the lateralization thresholds approached the SRTs. This means that the lateralization of the target became increasingly difficult with multiple interferers in the scene. In fact, differences in lateralization thresholds and SRTs were as small as about 3 dB when 8 or 12 maskers were distributed to both the left and right sides of the head. It appears therefore, that lateralization thresholds were more affected by the acoustic complexity of the tasks.

While only normal-hearing listeners were tested in the present study, the results suggest that a similar paradigm might be useful as a clinical measure. For each individual, high correlations were observed between lateralization and speech intelligibility thresholds. Thus, differences in lateralization detection might be

suitable as a proxy measure to predict the benefit of binaural unmasking on speech intelligibility. In this context, there are two potential advantages of using a lateralization detection task over a direct speech intelligibility task. First, the larger range of lateralization thresholds across conditions suggests that it may be a more sensitive clinical measure. Second, the lateralization task is language independent. Thus, there is no need to develop a native speech corpus for each language. As the spatial cues in the present study varied only in ITD, a similar paradigm, with only a few conditions, could be used to evaluate whether a hearing-impaired listener could make use of such cues. This could be useful in determining if a patient would gain any advantage with binaural hearing aids vs. two monaural devices. However, further studies involving hearing-impaired listeners are needed to test this.

2.5 Conclusion

In the present study, the effect of perceived spatial separation and the total number of masker streams was systematically investigated. In general, little spatial release from masking was observed when at least one masker remained collocated with the target. The benefit of removing the last 2 or 4 collocated maskers was found to be independent of the overall number of maskers. Finally, for each individual, lateralization and speech intelligibility thresholds were highly correlated. Thus, for situations where measuring direct speech intelligibility may be less practical, lateralization tasks could be used to predict changes in speech intelligibility.

Acknowledgments

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

3

Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners^a

Abstract

This study investigated the relationship between speech perception performance in spatially complex, lateralized listening scenarios and temporal fine structure (TFS) coding at low frequencies. Young normal-hearing (NH) and two groups of elderly hearing-impaired (HI) listeners with mild or moderate hearing loss above 1.5 kHz participated in the study. Speech reception thresholds (SRTs) were estimated in the presence of either speech-shaped noise, 2-, 4-, or 8-talker babble played reversed, or a non-reversed 2-talker masker. Target audibility was ensured by applying individualized linear gains to the stimuli, which were presented over headphones. The target and masker streams were lateralized to the same or to opposite sides of the head by introducing 0.7-ms interaural time differences between the ears. TFS coding was assessed by measuring frequency discrimination thresholds (FDTs) and interaural phase difference thresholds (IPDTs) at 250 Hz. NH listeners had clearly better SRTs than the HI listeners. However, when maskers were spatially separated from the target, the amount of SRT

^a This chapter is based on Lócssei et al. (2015).

benefit due to binaural unmasking differed only slightly between the groups. Neither the FDT, nor the IPDT tasks showed a correlation with the SRTs or with the amount of masking release due to binaural unmasking, respectively. The results suggest that, although HI listeners with normal hearing thresholds below 1.5 kHz experienced difficulties with speech understanding in spatially complex environments, these limitations were unrelated to TFS coding abilities and were only weakly associated with a reduction in binaural-unmasking benefit for spatially separated competing sources.

3.1 Introduction

Normal-hearing (NH) listeners can almost effortlessly follow a particular talker in the presence of multiple interfering acoustic sources (Cherry, 1953). Part of this robustness is due to spatial hearing, whereby listeners are able to parse different acoustic cues associated with streams located at separate spatial positions in the acoustical scene. Having access to such cues typically improves speech intelligibility when acoustic maskers are spatially separated from the target, an improvement relative to target and maskers being colocated that is referred to as spatial release from masking (SRM). SRM in the horizontal plane is mainly mediated by interaural level differences (ILDs) and interaural time differences (ITDs) (see e.g. Blauert, 1997; Bronkhorst and Plomp, 1988).

A listener's head will acoustically shadow lateral incoming sound, resulting in ILDs, which allow for "better-ear listening" to facilitate SRM. For example, if a target sound source is located in front of the listener, and the noise source is to the side, an improved signal-to-noise ratio (SNR) is observed at the ear contralateral to the noise source. Thus, increased intelligibility can be achieved with a monaural listening strategy, by attending only to the ear that has the most favorable SNR. ILDs are most prominent at frequencies above 2 kHz (Feddersen et al., 1957). ITDs occur for sound sources with lateral incidence, as such sound waves arrive delayed at the contralateral ear due to an increased travel-path length (e.g. Feddersen et al., 1957).

Pure tone ITDs can also be expressed as interaural phase differences (IPD) between the ears. ITDs play a dominant role in sound localization of relatively distant stimuli with a low-frequency content (Wightman and Kistler, 1992) and also contribute in the facilitation of speech understanding in spatial settings (Bronkhorst and Plomp, 1988). If the ITDs associated with target and masker streams are different, speech intelligibility increases as compared to situations where the ITDs are the same (e.g. Carhart et al., 1967). This facilitator of SRM is called binaural unmasking. The resulting benefit is often expressed in dB as the absolute difference in speech reception thresholds (SRTs) between the two presentation modes and referred to as the binaural intelligibility level difference (BILD). In this paper the term “SRM” is used in a general context to refer to the phenomenon that a benefit in SRTs arises as target and maskers become spatially separated, while “BILD” is used to refer to the amount of this benefit in dB, when SRM is triggered by ITDs only.

Hearing-impaired (HI) listeners experience difficulties understanding speech both in quiet and in noise. Traditionally, this effect has been considered a combination of two components: one related to the audibility of the speech stimulus and one related to the distortion of audible speech (e.g. Plomp, 1978). The audibility component manifests itself in threshold shifts for speech intelligibility in quiet that can be fully compensated for by appropriately amplifying the speech stimuli. The distortion component affects speech intelligibility in noise directly, and it has been argued that HI listeners can experience problems with understanding speech in noise arising, at least partly, from deficits in the discriminability of supra-threshold stimuli (e.g. Plomp, 1978; Dreschler and Plomp, 1985; Glasberg and Moore, 1989). The sources of such deficits could include broadening of the auditory filters (e.g., Glasberg and Moore, 1986), or degraded temporal coding (e.g. Lorenzi et al., 2006; Hopkins et al., 2008; Strelcyk and Dau, 2009; Papakonstantinou et al., 2011).

The cochlea can be modelled as a series of band-pass filters, which decomposes the incoming sound waves at the output of each cochlear filter into narrowband time-domain stimuli. These stimuli can be considered as a combination of slow envelope fluctuations (ENV) superimposed on a rapidly oscillating temporal fine structure (TFS) with frequencies close to the center frequency of each band (Moore, 2008). This TFS at the output of the cochlear filters elicits synchronized action

potentials (phase-locking) at higher stages in the auditory pathway. The neural coding of pure-tones is thought to mainly rely on phase-locking up to at least 2 kHz (Şek and Moore, 1995) and might play a role up to as high as 8 kHz (Ernst and Moore, 2012). Several studies have shown that HI listeners perform more poorly than their NH peers in tasks believed to assess monaural TFS coding, such as in frequency discrimination of pure-tones (Tyler et al., 1983; Moore and Peters, 1992) or in low-rate frequency-modulation detection (Lacher-Fougère and Demany, 1998; Strelcyk and Dau, 2009; Santurette and Dau, 2012). Furthermore, there is an accumulating body of evidence that both aging and hearing impairment degrade binaural TFS coding, as measured by the detection of interaural timing and phase differences in lateralized pure tones (Hopkins and Moore, 2011; King et al., 2014; Ross et al., 2007b) or by binaural masking level differences (Strouse et al., 1998; Papakonstantinou et al., 2011; Strelcyk and Dau, 2009).

In an attempt to map difficulties in speech perception in noisy environments to supra-threshold processing deficits, multiple studies have investigated the relationship between deficits in speech intelligibility (SI) and monaural temporal coding. While in quiet both NH and HI listeners can obtain close to normal SI performance utilizing ENV cues only (Lorenzi and Moore, 2008; Shannon et al., 1995), reduced access to TFS has been associated with deteriorated speech perception when noise is present (Lorenzi et al., 2006; Lunner et al., 2012; Papakonstantinou et al., 2011; Strelcyk and Dau, 2009). The exact role of TFS information in speech perception is, however, not fully understood to date. While earlier studies suggested that access to TFS information plays a particular role in masking release due to dip listening (e.g. Lorenzi et al., 2006), this has been debated and remains a controversial topic (Freyman et al., 2012; Oxenham and Simonson, 2009; Strelcyk and Dau, 2009). As another alternative, some studies have speculated that TFS facilitates speech perception by providing acoustic cues that aid the perceptual segregation of the target from the masker and, therefore, contributes to release from informational masking (Lunner et al., 2012).

Hearing loss has also been shown to negatively affect spatial perception of speech by reducing localization performance (Best et al., 2011; Lorenzi et al., 1999; Neher et al., 2011; Ruggles et al., 2011), increasing SRTs and reducing the amount

of SRM (Best et al., 2011; Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Neher et al., 2009; Neher et al., 2011; Peissig and Kollmeier, 1997) both in aided and unaided cases. Performance measures related to spatial perception typically vary significantly among HI listeners, even with similar audiograms (see e.g. Neher et al., 2011). In many of these studies, audibility could not entirely account for the diminished localization or speech perception performance of the HI listeners (e.g. Lorenzi et al., 1999; Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Neher et al., 2011).

Impairments in TFS coding have been associated with reduced SRTs in tasks involving spatial cues (Neher et al., 2011; Neher et al., 2012; Strelcyk and Dau, 2009). Strelcyk and Dau (2009) measured SI for full-spectrum and low-pass filtered speech in various diotic and dichotic conditions and compared it to measures of monaural and binaural TFS processing for 10 HI listeners with a sloping hearing loss above 1 kHz, but normal thresholds below. While pure tone averages (PTAs) were not correlated with SI results, speech reception thresholds in lateralized speech shaped noise (SSN) and two talker babble showed a significant correlation with measures of TFS processing, including interaural phase difference (IPD) thresholds, dichotic masked detection thresholds and frequency modulation detection thresholds. In a series of experiments, Neher et al. (2009; 2012) investigated the effect of cognitive abilities and binaural TFS processing on speech recognition in spatially complex 3-talker scenarios in HI listeners with sloping hearing loss in the high frequencies. Listeners were fitted with hearing aids to assure audibility up to about 6 kHz. Sentences were frontally presented in free field with two similar speech maskers spatially separated to the left and right. SRTs were found to be correlated with cognitive measures related to attention and to binaural TFS coding below and at 750 Hz, as measured by the TFS-LF test (Hopkins and Moore, 2010). However, these correlations became non-significant once age was controlled for, suggesting that performance on these tests was influenced by a common age-related factor. In a similar experimental setup, Neher et al. (2011) studied the relationship between cognition and TFS coding and performance in localization and speech recognition in spatial speech tests, where the competing talkers were either separated in the front-back or in the left-right dimensions. Instead of fitting their HI listeners with

hearing aids, they applied frequency-specific amplification on the stimuli to restore partial audibility. They found a significant negative correlation between SRTs and a cognitive measure assessing attention. Furthermore, Neher et al. (2011) also found an additional effect of the frequency range over which listeners were able to discriminate IPDs. Importantly, and similar to the results of Strelcyk and Dau (2009) and Papakonstantinou et al. (2011), measures of binaural TFS coding in the low frequency domain were not correlated with hearing thresholds at the same frequencies.

Even though the studies of Strelcyk and Dau (2009) and Neher et al. (2011; 2012) underline the role of binaural TFS processing in spatial speech perception, it still remains unclear under which circumstances and how robustly TFS coding facilitates speech perception in everyday listening. It appears reasonable to assume that binaural TFS coding plays a major role in binaural unmasking facilitated by ITD differences between target and maskers, and in other auditory phenomena where a combination of information from both ears might affect performance. However, while the study of Strelcyk and Dau (2009) showed a clear effect of TFS coding on speech perception in lateralized SSN, the question remains how this relationship translates into cases where more realistic background noises are applied. While Neher et al. (2009; 2011; 2012) indeed applied ecologically valid background noise in their study, stimuli were presented in free field. Such a presentation method allows for monaural listening strategies (e.g. better-ear listening), which might overshadow effects attributable to binaural TFS processing.

The current investigation complements the aforementioned studies by directly examining the relationship between monaural and binaural TFS coding and SRM attributable to binaural unmasking in isolation, without any contributions of monaural listening strategies. To assess the robustness of low-frequency TFS coding, frequency discrimination thresholds (FDTs) and interaural phase discrimination thresholds (IPDTs) for pure tones at 250 Hz were measured. SRTs were assessed in various noise conditions including stationary SSN, reversed babble noise and a two-talker masker played normally. The stimuli in the speech experiments were delivered over headphones and “spatialized” with frequency-independent ITD cues only, such that target and maskers were perceived as coming from the same

or from different lateralized positions within the head. In this way, any benefit that arises from changing the spatial distribution of target and maskers to a more favorable one, cannot be attributed to monaural effects, but can only be associated with binaural processes. It is well established that ITDs contribute to the perceived lateral position of stimuli and that SRM can be triggered by ITD cues only (Bronkhorst and Plomp, 1988; Carhart et al., 1967; Culling et al., 2004; Glyde et al., 2013). Under the assumption that SRM in this experiment will be mainly governed by ITDs in the low-frequency domain (Bronkhorst and Plomp, 1988), it was hypothesized here that listeners who have elevated pure-tone IPD thresholds will have limited capabilities to exploit ITD disparities between target and masker streams. Hence, we expected these listeners to have smaller BILDs. Thus, the current experiment investigated the relationship between binaural TFS coding (as measured by the IPDTs) and SRM, without the possible confounds from better ear listening present in other studies.

3.2 Methods

3.2.1 Participants

Nineteen elderly HI listeners participated in the study (55-85 yrs, mean: 71.7, standard deviation (SD): 7.19). As the goal was to investigate supra-threshold factors affecting speech intelligibility in the low-frequency region, the HI listeners had normal hearing or a mild hearing loss below 1.5 kHz and a mild-to-moderate hearing loss at frequencies above 1.5 kHz. The origin of hearing loss was confirmed to be sensorineural by air- and bone-conduction audiometry. Pure-tone audiometric thresholds were measured at octave frequencies between 125 and 8000 Hz, and at 750, 1500, 3000 and 6000 Hz. For each listener, the difference in hearing threshold levels (HTLs) between the ears was at most 15 dB at each tested frequency. The HI group was further divided into two age-matched subgroups: those having pure tone averages (PTA) less or equal to 40 dB HL above 1.5 kHz were classified as mildly impaired (HI_{mild} , 8 listeners) and the others were classified as moderately impaired (HI_{mod} , 11 listeners), respectively. This division was done in order to increase ho-

mogeneity of audiograms within subgroups and thus further minimizing audibility confounds at high frequencies. The mean audiometric thresholds of the NH and HI cohort are displayed in Figure 3.1. The control group consisted of ten young NH listeners (21-29 yrs, mean: 23, SD: 3.01), who had HTLs not greater than 20 dB HL, and an asymmetry across the ears not greater than 15 dB at each tested frequency.

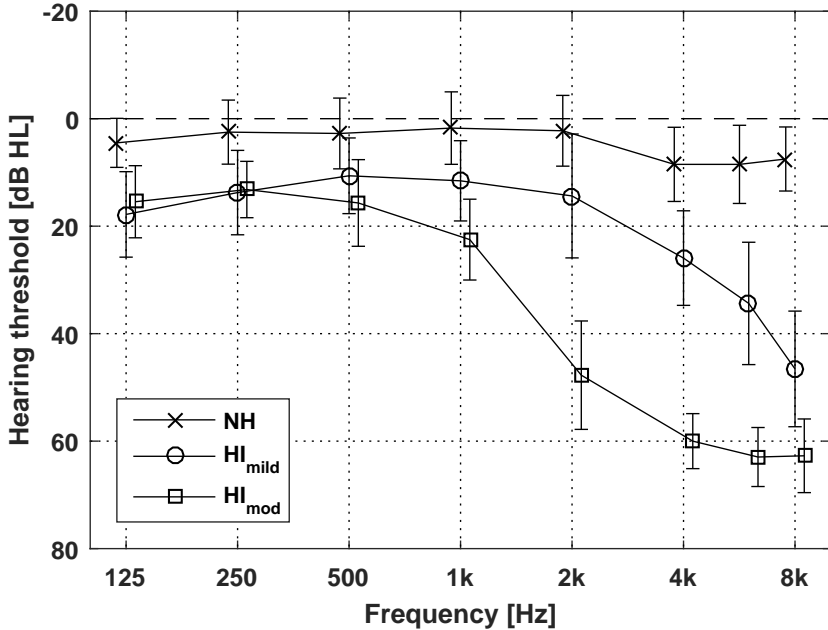


Figure 3.1: Audiometric data of the listener groups, averaged over both ears of the listeners. Horizontal bars denote ± 1 SD. The data of the NH and HI_{mod} listener groups is shifted on the x-axis for better readability.

3.2.2 Temporal processing and cognitive skills

To assess the robustness of monaural and binaural TFS coding, frequency discrimination thresholds (FDTs) and interaural phase discrimination thresholds (IPDTs) were measured at 250 Hz. Although the possibility that spectral cues might play a role in the coding of pure tones even at low frequencies cannot be fully excluded

(e.g. Plack and Oxenham, 2005), FDT thresholds show only a weak relationship with frequency selectivity (Tyler et al., 1983; Moore and Peters, 1992), suggesting the dominance of temporal cues over cues related to excitation patterns in this task. In contrast, it is generally accepted that performance in pure-tone IPD detection and discrimination tasks can be fully explained in the context of temporal coding abilities.

The FDT test was similar to that of Papakonstantinou et al. (2011). In each trial, listeners attended to three pure tones and had to indicate the target tone that had a higher frequency than the two 250-Hz references. A 3-interval 3-alternative forced-choice (3I-3AFC) paradigm was applied in combination with a multiplicative one-up two-down tracking rule. Target and reference stimuli were 500 ms long, gated by 50-ms long raised-cosine ramps, and separated by 250-ms silent gaps. The initial difference between target and reference frequency was set to 25%, and the initial step-size to a factor of 2. The step-size was reduced by a factor of 0.75 after every other reversal. The minimum step-size was 1.125, which was used for the last 8 reversals. Thresholds were calculated as the geometrical mean of these reversal points. Overall, 5 runs were performed by each subject. The final threshold was calculated as the geometric mean of the thresholds in the last 3 runs. All stimuli were presented monaurally at 65 dB SPL to the ear with the lower audiometric threshold at the test frequency. Due to time limitations, FDTs were not measured for two of the HI_{mild} and three of the HI_{mod} listeners.

The IPDT test was based on the TFS-LF test (Hopkins and Moore, 2010). Listeners were requested to select the binaurally varying target stimulus in a 2I-2AFC task. The tracking variable was changed using a multiplicative 1-up 2-down tracking rule. Both target and reference stimuli consisted of four 200-ms long pure tones presented binaurally, each gated with 20-ms long raised-cosine ramps and separated by 100-ms silent intervals. The reference and target stimuli were separated by 400-ms silent gaps. For the reference stimuli, each of the four tones were presented with the same phase across the ears. For the target stimuli, the interaural phase of the second and fourth tone was changed to $\Delta\varphi$. Initially, $\Delta\varphi$ was set to 90° . The initial step-size for the tracking variable was a factor of 3.375 and was decreased to 2.25 and 1.5 after the first and second reversals. Eight reversals were made with

this final step-size. The threshold was estimated by taking the geometrical mean of these reversal points. Listeners completed 5 threshold estimation tests and the final threshold was calculated as the geometrical mean of the last 3 runs. The stimuli were presented at 30 dB sensation level (SL).

The cognitive abilities of the listeners' were assessed using a Danish version of the reading span test (Daneman and Carpenter, 1980; Rönnerberg et al., 1989), which is designed to assess working memory by taxing memory storage and processing simultaneously. The test was administered in the visual domain, thus assuring no confounds with the status of the listeners' auditory abilities. Subjects were requested to read a series of 3-word sentences. They had to read them out aloud as they appeared word by word on a computer screen and to make a judgment about the context by saying "yes" or "no" after the last word if the sentence was meaningful or if it was absurd. The words appeared at every 0.8 seconds in each sentence and listeners had 1.75 seconds to give their response about the semantics of the sentence after the last word. After a block of 3, 4, 5 or 6 sentences, the listeners were instructed to repeat either all the first or all the last words of each sentence in the block. Subjects were encouraged to do this in the original serial order. The final score was calculated as the percentage of correctly recalled target words (disregarding the correct serial order). The test consisted of 3 blocks for each sentence length, resulting in 54 target words scored in total. To make the listeners familiar with the task, an extra block of 3 sentences was included at the beginning of the test.

3.2.3 Speech perception in noise

SRTs were measured using target sentences uttered by a female talker from the Danish DAT corpus (Nielsen et al., 2014). This open-set corpus contains low-predictability sentences with a fixed and correct grammar in a form that translates to English as "<Name> *thought about* <keyword 1> *and* <keyword 2> *yesterday*". The sentences were uttered by one of three professional female talkers. The target stream consisted of single sentences starting with the name "Dagmar", which was embedded into one of the following noise types: speech shaped noise (SSN),

reversed speech with 2, 4, or 8 streams of competing male talkers from the Grid corpus (Cooke et al., 2006) or forward speech with single sentences uttered by the 2 other female talkers from the DAT corpus. The notations S_1 , R_2 , R_4 , R_8 , and D_2 are used to denote the set of noise conditions, where SSN (S_1), reversed speech of 2, 4, or 8 competing talkers (R_2 , R_4 , R_8), or 2 interferers from the DAT corpus (D_2) are used as maskers, respectively. The subscripts indicate the number of independent streams in the masker mixture.

The different background noise types were chosen to vary the contribution of energetic vs informational masking (e.g. Kidd et al., 1994). The target and the D_2 masker sentences had the same grammatical structure and were spoken by talkers with similar voice characteristics, resulting in strong informational masking. While the reversed-speech maskers retained spectro-temporal fluctuations characteristic to running speech, informational masking was substantially reduced in these cases due to the complete lack of semantic content and the different voice characteristics. With the reduction of spectro-temporal fluctuations, the energetic masking component became dominant and was most pronounced with the S_1 masker, which was at the same time perceptually highly discernible from the target, offering minimal informational masking.

The maskers in the S_1 , R_2 , R_4 , and R_8 conditions were spectrally shaped to have the same long-term average spectrum as the target talker. For the S_1 conditions, 50 tokens of 5 seconds were generated. The actual masker tokens in the S_1 conditions were randomly selected from these on each trial. For the R_2 , R_4 , and R_8 conditions, continuous streams of sentences were generated from each of the first eight male talkers from the Grid corpus. Low-energy intervals were removed and the resulting recordings were time-reversed. 50 non-overlapping tokens of 5 seconds were selected from each of these talkers. When generating masker tokens, single random tokens were drawn from the pre-generated pool of tokens for each of the first 2, 4, or 8 Grid talkers, which were then mixed. Similarly to the S_1 conditions, this was done trial-by-trial. Finally, in the D_2 conditions, randomly selected full sentences were used as maskers. In the SSN and reversed speech conditions, maskers started 1 s before the onset of the target sentence and ended with the target sentence. The D_2 maskers started at the same time as the target.

Both target and maskers were presented as coming from a lateral position towards the left or right side of the head, which was achieved by introducing 0.7-ms ITDs between the ears for each of these streams. For each masker condition, the target and maskers could be lateralized to either side independently of each other. The spatial distribution of the masker streams compared to the side of the target was varied systematically. The terms “fully colocated” and “fully separated” refer to masker distributions where all or none of the masker streams were lateralized towards the side of the target, respectively. Conditions where only a subset of the maskers was colocated with the target were also tested, in order to investigate how various spatial distributions affect BILDs with a fixed number of sources in the different listener groups. When referring to a specific spatial distribution within noise conditions, the number of masker streams colocated with the target side will be displayed in the superscript. All possible spatial distributions were tested in the S_1 , R_2 and D_2 noise conditions (S_1^1, S_1^0 and R_2^2, R_2^1, R_2^0 and D_2^2, D_2^1, D_2^0). In the R_4 and R_8 conditions the number of maskers lateralized towards the target side was varied in twos (R_4^4, R_4^2, R_4^0 and $R_8^8, R_8^6, R_8^4, R_8^2, R_8^0$). The side to which the target was lateralized was randomized trial-by-trial. Spatial conditions with each masker type were clustered into separate blocks and the SRT tracking procedure for the different spatial conditions within these blocks were interleaved. For the R_8 maskers, two blocks were made with conditions R_8^0, R_8^4, R_8^8 and R_8^2, R_8^6 , respectively. This means that at the beginning of each trial, listeners had no prior knowledge of which side to attend to, and they needed to actively “tune in” to the acoustic scenario in order to correctly recognize the keywords.

The stimuli were presented over headphones. The target sentences were first scaled to a nominal sound pressure level (SPL) of 63.5 dB free field and mixed with the maskers at the desired SNR. The stimuli were then processed by a 512 order finite impulse response (FIR) filter. Besides compensating for the frequency response of the electro-acoustic equipment, this filter simulated the frequency response of the outer ear in a diffuse-field listening scenario by implementing the diffuse-field-to-eardrum transfer function, as defined in Moore et al. (2008), and also compensated for the loss of stimulus audibility. The elevated hearing thresholds of the HI listeners were compensated for by applying frequency depen-

dent linear gains based on the individual audiograms and the long-term average spectrum of the target speech in a similar way as in the studies of Neher et al. (2011) and Nielsen et al. (2014). The audibility criterion was set such that the long-term root mean square (RMS) values of the target speech evaluated in 1/3-octave frequency bands were presented 13.5 dB above threshold at and below 3 kHz. This was reduced to 2.5 dB at 8 kHz by logarithmic interpolation at the intermediate frequencies. Finally, the stimuli were bandpass-filtered between 200 Hz and 10 kHz prior to presentation. SRTs corresponding to the 50% sentence correct values were tracked by adapting the masker level in 2 dB steps. SRTs were estimated over one list in each condition, calculated as the average of the presentation levels associated with sentences 5 to 21 (the last one being the level of the hypothetical 21st sentence). The speech tests were performed in two sessions and listeners were trained on 3 lists before each visit. We tested the S₁, R₂, and R₄ conditions during the first and the R₈ and D₂ conditions during the second visit. Within each visit, the presentation order of the conditions was balanced as much as possible across listeners using a Latin square design. List numbers used for the target sentences were balanced between conditions with the same technique.

3.2.4 Statistical tools

For all statistical tests below the Type I error rate was fixed at 0.05. The group means in age, PTAs, and in measures assessing temporal and cognitive abilities were compared using one-way analysis of variance (ANOVA) models with pairwise comparisons as *post hoc* tests with the Tukey's HSD method for multiple comparisons, unless stated otherwise. The results of the SI experiments were analyzed with mixed-design ANOVA models, subjecting listeners within groups to repeated measures. The degrees of freedom were adjusted with Greenhouse-Geisser correction where the assumption of sphericity was violated. Multiple comparisons in these analyses used the Bonferroni correction to control the family-wise error rate.

3.3 Results

3.3.1 Audiometric thresholds

All listeners were selected to have normal or close-to-normal hearing thresholds up to 1.5 kHz, while above 1.5 kHz the HI listeners had HL up to moderate levels. However, both HTLs averaged up to 1.5 kHz (PTA_{low}), above 1.5 kHz (PTA_{high}) and averaged at octave frequencies from 0.25 to 4 kHz (PTA_{oct}) were significantly different between the three listener groups. This was confirmed by one-way ANOVAs (PTA_{low} : $F(2, 26) = 24.34, p < 0.001$; PTA_{high} : $F(2, 26) = 213.98, p < 0.001$; PTA_{oct} : $F(2, 26) = 79.07, p < 0.001$) and *post hoc* analyses ($p < 0.05$ in all cases). A one-way ANOVA have been conducted on the hearing threshold levels at 250 Hz (HTL_{250}), at which temporal processing abilities were assessed. The effect of listener group was significant ($F(2, 26) = 11.09, p < 0.001$). Pairwise *post hoc* tests revealed that NH group significantly differed from both of the HI groups ($p = 0.001$), while the thresholds of the HI groups did not show any significant difference.

3.3.2 Temporal processing

The results from the FDT and IPDT experiments are displayed in Figure 3.2 for the NH (white), HI_{mild} (light gray) and HI_{mod} (dark gray) listener groups. The data analysis was performed on the log-transformed FDT and IPDT scores, as the data were more normally distributed this way (Anderson-Darling test). This is in line with earlier studies (Strelcyk and Dau, 2009; Lacher-Fougère and Demany, 2005). Accordingly, the ordinate of these figures are also logarithmic.

The left panel of Figure 2 shows the results of the FDT test for the three listener groups. The FDT scores are expressed in percentage as the frequency difference between the deviant stimulus and the 250 Hz reference. The results for both the NH and HI listeners are consistent with earlier studies (e.g., Papakonstantinou et al., 2011). On average, NH listeners performed better than the HI listeners (two-tailed t-test, $p = 0.014$). Interestingly, the HI_{mild} listeners performed worse than the HI_{mod} group. These observations were confirmed by a one-way ANOVA with *listener group* as between subject factor. The effect of *listener group* was

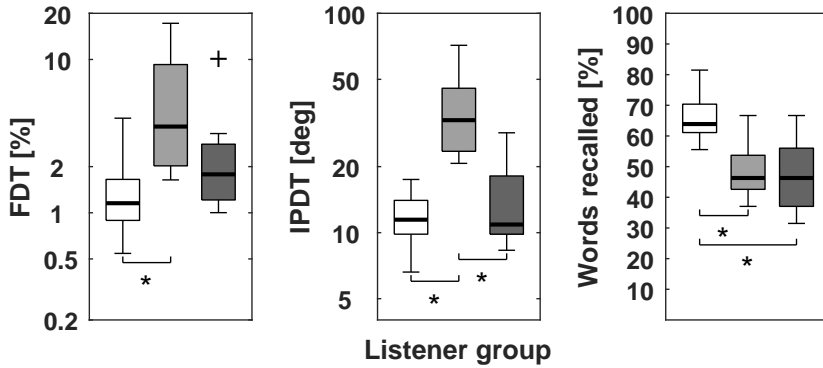


Figure 3.2: Box plots illustrating the FDT, IPDT and Reading span test results. White, light gray and dark gray boxes stand for the data of the NH, HI_{mild} and HI_{mod} groups, respectively. The thick black lines denote the medians and the boxes extend to the 25th and 75th percentiles. The thin lines extend to the most extreme data points within 1.5 interquartile range from the 25th and 75th percentiles and + indicates outlier data. Asterisks denote statistically significant differences in means.

significant ($F(2, 21) = 5.67, p = 0.011$), and *post hoc* analyses revealed that the only significant difference in group means was the one between the NH and HI_{mild} group ($p = 0.008$).

The results from the IPDT test are displayed in the middle panel of Figure 2. The performance of the HI group spanned a wider range than the data of the NH group; some HI listeners show similar thresholds as NH listeners, while some show deficits in detecting IPDs. This observation is in line with earlier studies applying a similar experimental paradigm (Hopkins and Moore, 2011; King et al., 2014). The difference between the NH and HI group means was significant (two-tailed t-test, $p = .019$), but similar to the tendencies in the FDT test, this difference was driven by the elevated IPDT thresholds of the HI_{mild} group. This was supported by a one-way ANOVA, which revealed a significant effect of *listener group* on the IPDT thresholds ($F(2, 26) = 21.73, p < 0.001$). *Post hoc* analyses revealed that the differences in mean thresholds between the NH and HI_{mild} and between the HI_{mild} and HI_{mod} groups were both statistically significant ($p < 0.001$).

3.3.3 Cognitive abilities

The results from the reading span test are shown in the right panel of Figure 2. On average, the NH group recalled 65.7% of all words presented (35.5 words recalled) while the HI groups recalled only 47.9% (25.9 words recalled). A one-way ANOVA showed a significant effect of listener group on the reading span scores ($F(2,26) = 10.35, p < 0.001$). *Post hoc* tests confirmed that the NH listeners performed significantly better than both the HI_{mild} ($p = 0.004$) and HI_{mod} ($p = 0.001$) listener groups, while the difference between the two HI groups remained non-significant. These results are consistent with earlier studies showing an age-related decline of working memory (e.g., Schoof and Rosen, 2014).

3.3.4 Speech perception in noise

In Figure 3.3 the horizontal black bars and the corresponding boxes around them show the mean SRTs and ± 1 SD of the NH (white), HI_{mild} (light gray) and HI_{mod} (dark gray) listener groups in all of the tested conditions. The shaded panels mark condition groups where the same type of background noise was utilized. When moving along the abscissa from left to right within each panel, the spatial distribution of the maskers changes gradually from all colocated to all separated from the side of the target. On average, the HI_{mod} listeners performed worse than the HI_{mild} listeners, who showed degraded performance compared to NH in most of the tested conditions. While for the NH listener group there was a considerable variation in mean SRTs between condition groups, there was no such tendency in the HI_{mod} group. Instead, within each condition there was a greater spread of individual thresholds for HI than for NH listeners.

The average SRT associated with each of the noise types was estimated by calculating the mean of the SRTs in the fully colocated and fully separated masker distributions (Table 3.1). The other spatial distributions were left out from the calculation, as those would bias the SRTs towards higher SNRs for those noise types that had a higher number of spatial distributions. Group differences were smallest in the steady-state masker conditions (S_1) and gradually increased as more spectro-temporal fluctuations appear in the background noise. The NH

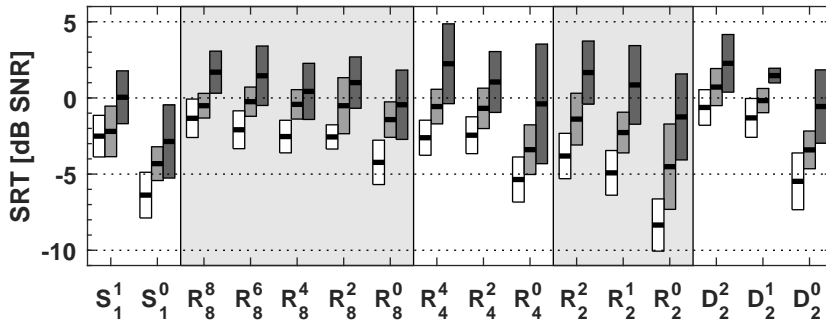


Figure 3.3: SRTs for NH (white), HI_{mild} (light gray) and HI_{mod} (dark gray) listeners. Horizontal black bars denote group means and the boxes represent ± 1 SD. The white/gray areas in the background denote condition groups with the same masker type. Condition group notations: S_x^y : speech shaped noise; R_x^y : reversed speech maskers; D_x^y : forward speech maskers; x denotes the total number of masker streams in the tested condition and y indicates the number of maskers lateralized to the side of the target.

listeners yielded the lowest SRTs in the R_2 conditions, while for the HI listeners the best SRTs were achieved in the S_1 conditions. Despite the inherent spectro-temporal fluctuations in the R_8 backgrounds, all groups had elevated thresholds as compared to the stationary S_1 conditions. While NH listeners performed better as the number of reversed interferers decreased from 8 to 4 to 2, HI listeners performed similarly in all of these conditions. This is consistent with the results of earlier studies showing that HI listeners have smaller masking release due to spectro-temporal fluctuations than NH (Christiansen and Dau, 2012; Festen and Plomp, 1990; Strelcyk and Dau, 2009). Spatially separating maskers from the target increased intelligibility performance within each listener group. This benefit was most pronounced once all masker streams were presented spatially separated from the target.

In order to test the statistical significance of the abovementioned observations, mixed ANOVAs were performed on the SRTs and BILDs in the fully collocated and fully separated conditions. Figure 3.4 shows the SRTs only in these distributions (top panel), with the BILDs in each noise condition calculated as the difference in SRTs between these two lateralized conditions (bottom panel). A mixed ANOVA with SRTs as the dependent variable, *noise type* (S_1 , R_8 , R_4 , R_2

Noise type	Average SRT [dB SNR]		
	Listener group		
	NH	HI _{mild}	HI _{mod}
S ₁	-4.44	-3.25	-1.41
R ₈	-2.78	-0.97	0.62
R ₄	-3.98	-1.98	0.93
R ₂	-6.08	-2.95	0.21
D ₂	-3.05	-1.35	0.85

Table 3.1: SRTs of the listener groups averaged over the fully collocated and fully separated masker distributions for the five noise types. Condition group notations as in Figure 3.3

and D₂) and *lateralization* as within-subject and *listener group* (NH, HI_{mild} and HI_{mod}) as between-subject factors showed a main effect of *lateralization* ($F(1, 26) = 311.41, p < 0.001$), *noise type* ($F(3.11, 80.95) = 28.02, p < 0.001$) and *listener group* ($F(2, 26) = 24.171, p < 0.001$). The interaction was significant between *noise type* and *listener group* ($F(6.23, 80.9) = 5.03, p < 0.001$). Bonferroni corrected paired t-tests within listener groups showed that the SRTs were significantly greater in the R₈ than in the S₁ condition for all of the listener groups ($p < 0.0125$ in each case), indicating that substituting the SSN noise masker with the reversed 8-talker babble resulted in increased masking, despite the inherent spectro-temporal fluctuations of the masker. In contrast, differences between the S₁ and R₂ noise condition were lower than zero for the NH ($p = 0.006$), not significantly different from zero for the HI_{mild} ($p = 0.6$), and greater than zero for the HI_{mod} listeners ($p = 0.002$). This supports the observation that NH listeners' performance improves as the number of interfering streams decreases from 8 to 2, perhaps due to the increasing spectro-temporal gaps present in the masker with fewer streams. However, this release from masking was reduced for the HI_{mild} listeners and completely absent for the HI_{mod} listeners.

As the interactions were also significant between *lateralization* and *listener group* ($F(2, 26) = 4.91, p = 0.016$), and *lateralization* and *noise type* ($F(3.08, 6.15) = 4.57, p = 0.005$), a second mixed ANOVA was conducted on the BILD values with *noise type* as within and *listener group* as between subject factors. Consistent with the ANOVA conducted on the SRTs, this analysis showed a main effect of *listener*

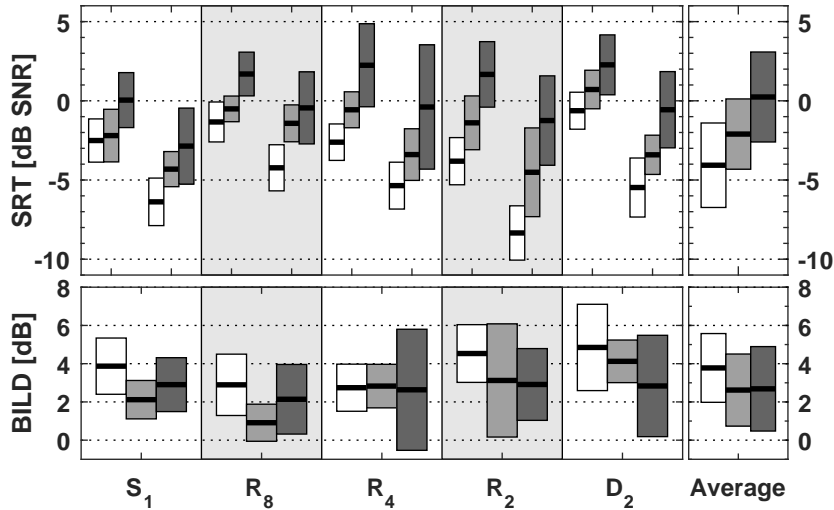


Figure 3.4: Top panel: SRTs in the fully colocated and fully separated target-masker distributions (left and right part of each block, respectively, repeated from Figure 3.3). Bottom panel: BILDs calculated as the difference between the colocated and separated SRTs. The horizontal black lines with the white, light gray, and dark gray boxes stand for the data of the NH, HI_{mild}, and HI_{mod} groups (mean and SD). The white/gray areas in the background denote condition groups with the same masker type. Condition group notations: S₁: speech shaped noise masker; R₈, R₄, R₂: reversed speech masker consisting of 8, 4 or 2 competing talkers; D₂: forward speech masker consisting of 2 competing talkers.

group ($F(2, 26) = 4.91, p = 0.016$) and noise type ($F(3.08, 79.96) = 4.57, p = 0.005$), and no significant interaction. Pairwise comparisons with Bonferroni correction revealed that the BILD in the R₈ condition was significantly lower than the BILD in the S₁ and D₂ conditions ($p < 0.005$ in each case). Compared to the NH listeners, the average BILD was lower by about 1 dB for both the HI_{mild} and HI_{mod} listeners ($p < 0.017$).

In Figure 3.5, the BILDs are shown in all of the multiple-talker masker conditions, as a function of the number of colocated maskers with the target side. The panels from top to bottom indicate the results for the NH, HI_{mild}, and HI_{mod} listeners, respectively. In each of the listener groups, the amount of masking release decreases rapidly as soon as even a single noise source is added to the target side. Colocating additional maskers with the target has only a minimal effect (about

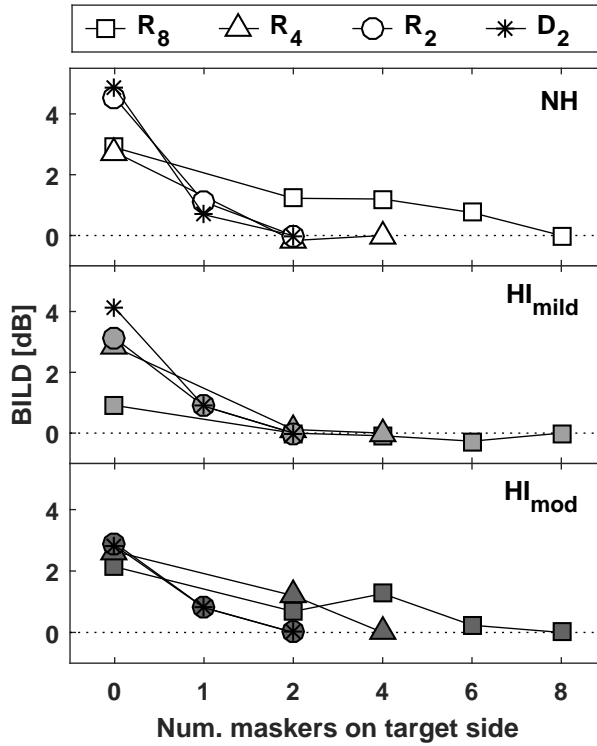


Figure 3.5: BILDs as a function of number of interfering talkers colocalized with the target side in the different interferer conditions. The top, middle and bottom panels illustrate the results of the NH, HI_{mild}, and HI_{mod} listener groups. Condition group notations: R₈, R₄, R₂: reversed speech masker of 8, 4 or 2 competing talkers; D₂: forward speech masker of 2 competing talkers.

1 dB) on the masking release values in all of the listener groups. One-sample t-tests with Bonferroni correction within listener groups showed that BILDs did not differ significantly from 0 as long even one masker was presented from the side of the target ($p > 0.005$).

3.3.5 Predicting speech intelligibility

As mentioned before, the main goal of the present study was to investigate the role of monaural and binaural TFS coding on lateralized speech perception. The

presence of any interdependencies between these two domains was checked by calculating and analyzing Pearson's correlation coefficients. The aforementioned statistics were only examined on a limited set of variable combinations, based on prior assumptions about the roles of monaural and binaural TFS coding in such listening scenarios. Correlations in the HI listener groups were investigated with the HI_{mild} and HI_{mod} groups taken separately and collapsed. If not stated otherwise, results regarding correlations reported below generalize to both of the cases where HI groups were treated separately or collapsed.

Before analyzing the predictability of SI with measures of auditory processing, the interdependencies between the predictor measures were assessed. First the effect of aging and elevated HTLs on the FDT and IPDT test results in the HI listener groups were tested. The correlations were all non-significant. The correlation between age and PTA_{oct} was also non-significant. In the HI groups, no age-related decline of working memory was found, and cognitive abilities were not correlated with the results of the tests assessing TFS coding.

Figure 3.6 presents performance measures from the SI tests compared to PTA_{oct}, FDT, or IPDT for the NH (black dots), HI_{mild} (light gray diamonds), and HI_{mod} listeners (dark gray squares). The top left panel shows the SRTs averaged over all noise conditions in the fully collocated and fully separated target-masker distributions (SRT_{avg}) as a function of PTA_{oct}. The correlation between PTA_{oct} and SRT_{avg} was significant when the HI_{mild} and HI_{mod} groups were pooled together ($r(17) = 0.55, p = 0.015$). The slope of the regression line was 0.11, showing that, on average, a 9 dB increment in PTA_{oct} was associated with about 1 dB increment in SRT. This correlation was not significant when the HI_{mild} and HI_{mod} groups were considered separately. Importantly, this suggests that the potential effect of audibility on the SI results was minimized in the two subgroups.

Multiple studies have suggested that robust TFS coding aids segregation of speech from background noise (Strelcyk and Dau, 2009; Lunner et al., 2012), e.g. by providing cues for the perceptual separation of the target from the interferers (Lunner et al., 2012). Therefore, it was hypothesized that deficits in monaural TFS coding affect SI in noise directly by increasing SRTs. Accordingly, the correlations between the FDT scores and the SRT scores averaged over all fully collocated noise

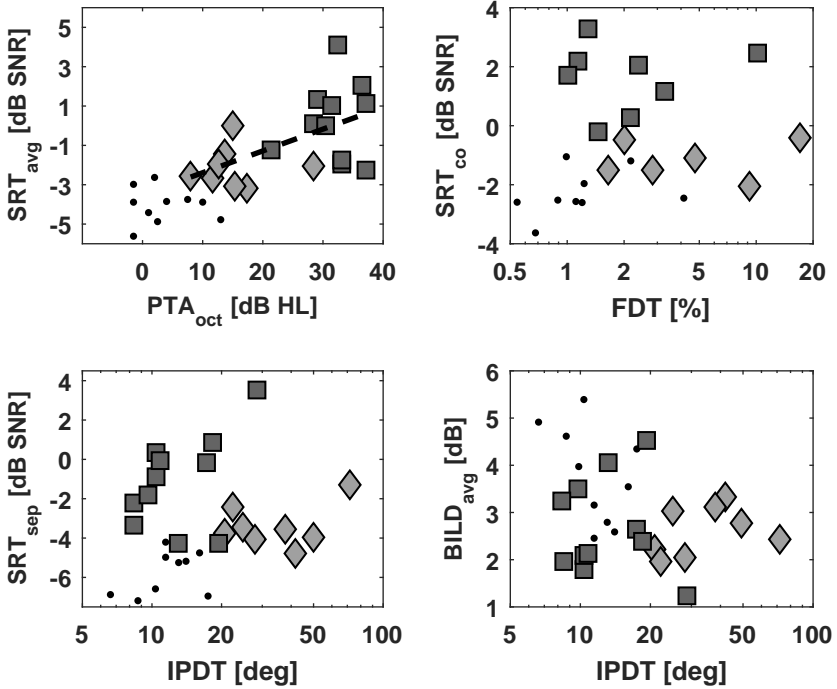


Figure 3.6: Scatter plots between predictor measures and speech perception results for the NH (black dots), HI_{mild} (light gray diamonds) and HI_{mod} (dark gray squares) listeners. Top left: SRT_{avg} as a function of PTA_{oct} . The dashed line stands for the fitted regression line on the data of the HI_{mild} and HI_{mod} listeners. Top right: SRT_{co} as a function of FDT. Bottom left: SRT_{sep} as a function of IPDT. Bottom right: $BILD_{avg}$ as a function of IPDT.

conditions (SRT_{co}) were tested. Since target-masker similarity was greatest in the D_2 condition, FDTs were also compared with SRTs obtained in the D_2^2 condition. Only SRTs in the fully colocated conditions were used, as a difference in spatial position of sound sources might serve as a cue for streaming, which is likely to be linked to binaural TFS processing abilities in the current setup. Consequently, the relations were examined between IPDT and SRT averaged over all the fully separated noise conditions (SRT_{sep}), and also with BILD values averaged over all noise conditions ($BILD_{avg}$). It was hypothesized that listeners who have elevated pure tone IPDTs will have limited capabilities to exploit ITD disparities between

target and masker streams, and thus have elevated SRTs when target and maskers are spatially separated. In turn, this would also affect spatial release from masking by reducing the magnitude of the BILDs. The top right panel of Figure 3.6 shows the SRT_{co} values as a function of FDTs. No correlation was found between these two measures, nor between the SRTs in the D_2^2 condition and the FDTs, even once PTA_{oct} was controlled for, which contradicted our hypothesis and some previous results (Papakonstantinou et al., 2011). The scatter plot of the SRT_{sep} and $BILD_{avg}$ values as a function of IPDT are shown in the bottom left and bottom right panel of Figure 6. The only significant correlation was between the IPDT and SRT_{sep} scores when the entire HI group was considered ($r(17) = -0.64, p = 0.014$). Contrary to our hypothesis, the negative correlation coefficient indicated that those listeners who had lower IPDT thresholds (and thus displayed more robust binaural TFS coding abilities) were the ones suffering from difficulties in the lateralized speech perception tasks. As the HI_{mod} listeners had higher PTAs but also better IPDT thresholds than those in the HI_{mild} group, and because PTA_{oct} was correlated with SRT_{avg} values, we re-ran the analysis by controlling for PTA_{oct} . The correlation between IPDT and SRT_{sep} became non-significant, thus suggesting that the negative correlation without controlling for PTA_{oct} is in fact driven by the distribution of IPDT scores and the difference in hearing thresholds between the HI_{mild} and HI_{mod} listener groups.

3.4 Discussion

The aim of the current study was to clarify the relationship between monaural and binaural TFS coding in the low-frequency domain and speech perception in spatially complex acoustic scenarios. Under the assumption that a reduction in the ability to code binaural TFS information limits SRM by affecting the amount of binaural unmasking, stimuli were presented over headphones and were spatialized by applying ITDs only. Thus, contributions of better ear listening to SRM were eliminated. It was hypothesized that diminished binaural TFS coding, as assessed by measuring IPDTs, would be associated with reduced BILDs or elevated SRTs in

conditions where target and maskers are separated by lateralization. Furthermore, FDTs were measured to quantify the robustness of monaural TFS coding and to test a hypothesized association between increased FDTs and increased SRTs in the collocated target-masker conditions. Individualized linear gains were applied to all speech stimuli to reduce the effect of stimulus inaudibility at high frequencies. To further reduce the effect of inter-individual differences of audibility, the results were also investigated in two homogenous subgroups of the HI listeners in terms of their audiograms.

On average, HI listeners performed worse in both monaural and binaural measures of TFS coding. However, the analysis of the HI subgroups revealed that this difference was associated with elevated FDT and IPDT thresholds of the listeners in the HI_{mild} group. Listeners in the HI_{mod} group performed similarly to those in the NH group. A significant overlap between the spread of data of NH and HI has been observed earlier as well (Hopkins and Moore, 2011; Papakonstantinou et al., 2011). While previous studies associated both aging and elevated hearing thresholds with impoverished binaural temporal coding (King et al., 2014), this pattern of differences in the FDT and IPDT tests between the HI_{mild} and HI_{mod} listeners is surprising considering that these groups were age-matched and had the same hearing threshold levels at 250 Hz, both as averaged between ears and considering the better ear only. It is noteworthy that asymmetry between audiometric thresholds at 250 Hz were higher for the HI_{mild} than for the HI_{mod} listeners ($t(17) = 2.5, p = 0.023$). As stimuli in the IPDT experiment were presented at equal sensation levels at the ears, one could speculate that the group differences in the IPDT results within the HI panel are in fact a result of shifted lateralization due to differences in the absolute presentation levels between the ears. Nonetheless, this explanation seems unlikely. First of all, while this difference in asymmetry was statistically significant, it was rather small (about 3.5 dB). Furthermore, listeners were requested to detect a change in lateralized position, and not to identify an absolute position. As most of the HI listeners have participated in psychoacoustic experiments previously, it has been tested whether prior experience might explain the observed tendencies between HI listener subgroups. Experience was quantified as the number of times each listener had participated in psychoacoustic experiments

over the past 2 years. In both HI groups, 3 listeners had no prior experience with psychoacoustic tests, while the rest had participated in up to as many as 10 visits. There was no significant difference in average number of visits between HI_{mild} and HI_{mod} ($t(17) = -1.49, p = 0.165$) and no significant correlation between number of visits and the FDT or IPDT results ($p > 0.1$ in both cases). Given the relatively small number of listeners in each group, it might be that this distribution of the data was merely the result of partitioning the HI group into two subgroups.

The HI listeners showed elevated SRTs as compared with the NH group. Consistent with earlier studies (Festen and Plomp, 1990), the differences between listener groups were relatively small in continuous but greater in fluctuating background noise, ranging from 3 to 6 dB in the former and in the latter case, respectively. Generally, as spectro-temporal fluctuations increase in the masker, the energetic masking between target and masker decreases monotonically, allowing for listening in the dips. Nonetheless, for all listener groups, the most challenging scenario was the R₈ condition, yielding higher SRTs than the conditions with the S₁ masker in each of the listener groups. This difference, however, cannot be explained based on energetic masking in the classical sense, as the R₈ maskers have a sparser spectro-temporal structure than the S₁ maskers. It is more likely that these differences arise from susceptibility to modulation masking (Houtgast, 1989; Takahashi and Bacon, 1992). The differences in average SRTs between the R₈ and R₂ masker conditions were relatively small (about 3.3 dB) even for the NH listeners. This could be partly attributed to the removal of low-energy intervals in the Grid maskers, which reduced the amount of inherent fluctuations already in the R₂ masker condition. Nonetheless, the results clearly show that HI listeners had difficulties with understanding speech in modulated noise compared to the NH group, which is consistent with earlier reports (Christiansen and Dau, 2012; Festen and Plomp, 1990; Strelcyk and Dau, 2009).

The amount of masking release due to spatial separation was comparable for the NH and HI groups, about 4 and 3 dB, respectively. This means that while listener groups differed significantly in performance when considering the SRTs, the binaural benefit they gained due to ITD differences between target and maskers was similar. Thus, it appears that, in the current experiments, the performance of

the HI listeners was limited by monaural rather than by binaural factors. These results are in line with earlier studies showing nearly normal amount of BILDs for HI listeners (Bronkhorst and Plomp, 1989; Bronkhorst, 2000; Santurette and Dau, 2012).

As regarding speech perception performance and monaural TFS coding, no support was found for a link in the current study. Reduced FDTs showed no association with increased SRTs averaged in the collocated conditions. Similarly, a relationship between FDTs and the SRTs in the D_2^2 condition was absent, where access to TFS structure might be of particular importance, as it can aid the cueing of the target voice by providing information about, e.g., its formant structure. It appears therefore, that performance in the speech intelligibility tasks was not limited by monaural temporal processing abilities of the listeners, at least not as measured by FDTs. It has to be emphasized that uncertainties exist regarding the way and extent to which TFS is utilized in monaural speech processing. There is accumulating evidence that, in contrast to earlier suggestions, TFS is not involved in masking release due to temporal fluctuation (Freyman et al., 2012; Oxenham and Simonson, 2009; Strelcyk and Dau, 2009), but it rather facilitates speech understanding in noise by providing cues for the perceptual segregation of target from the maskers (Strelcyk and Dau, 2009; Lunner et al., 2012). This conjecture nonetheless needs further investigation, especially that there are indications that TFS coding deficits can be associated with degraded speech intelligibility even in listening tests utilizing highly discernible target and masker (Papakonstantinou et al., 2011).

While based on the work of Bronkhorst and Plomp (1988) it was assumed that low-frequency IPDTs will be predictive of the size of BILDs in the current setup, no support was found for this hypothesis. One possible reason for the lack of a clear relationship between the measured IPDT thresholds and BILDs could be that, while in the former case binaural TFS coding abilities were assessed at a single frequency, binaural unmasking of speech is being effectuated over a broad range of frequencies. Edmonds and Culling (2005) showed that limiting the frequency range at which listeners have access to ITDs decreases BILDs. While ITDs above the frequency range at which listeners are sensitive to TFS ITDs also contribute

to binaural unmasking, most likely in the form of ENV ITDs, the contribution of low-frequency TFS ITDs is greater than that of high-frequency ENV ITDs, at least for NH listeners (e.g. Edmonds and Culling, 2005). Since the upper frequency limit of sensitivity to TFS ITDs reduces with progressing age (e.g. Ross et al., 2007b), it is possible that BILDs listeners can obtain in a particular listening scenario are more affected by the frequency range over which they can detect TFS ITDs. Therefore, it appears possible that IPDT measures also at higher frequencies or a measure of the frequency range at which listeners were sensitive to such differences would have been more predictive of the obtained BILDs (cf. Neher et al., 2011).

It cannot be excluded that listeners with reduced low-frequency TFS ITD sensitivity rely and utilize ITD cues in the high-frequency domain to cue the target and the maskers, and therefore to facilitate SRM to a greater degree than NH listeners. In fact, some studies suggest that sensorineural hearing loss can lead to an enhancement of temporal ENV coding, due to e.g. the broadening of the auditory filters and reduced cochlear compression (Henry et al., 2014; Bianchi, Fereczkowski, Zaar, Santurette & Dau, in press). If the high-frequency hearing loss of the HI listeners in the current study was coupled with broader auditory filters and reduced compression, e.g. as a result of outer-hair cell damage, it is possible that the ENV representation of the speech stimuli for these listeners was enhanced, resulting in an ENV structure with greater modulation depths. It has been shown that for amplitude-modulated high-frequency pure-tones, sensitivity to threshold ENV ITDs decreases with increasing modulation depth (e.g. Bernstein and Trahiotis, 2009). In this view, the possibility arises that some of the HI listeners rely more on the ENV ITD cues at high frequencies than on low-frequency TFS ITDs when facilitating SRM, which might explain why low-frequency IPDTs were not correlated with the BILDs.

Another and perhaps the most likely reason for the lack of any clear relationship between binaural TFS processing and SRM could be that relatively large ITDs were used to elicit different spatial positions. While the effect of aging and hearing loss on the detection of IPDs in TFS is transparent in several studies, it appears that most of the HI listeners retain their ability to detect binaural delays of the magnitude utilized in the current study (see e.g. Hopkins and Moore, 2011; King

et al., 2014). These time differences were also clearly detectable at 250 Hz to almost all of the HI listeners tested in the current study. For maskers which induce strong informational masking, small spatial separations between the target and the maskers can provide strong segregation cues and trigger substantial SRM, even in listening scenarios where any benefits due to better-ear listening are greatly reduced (see e.g. Marrone et al., 2008c; Swaminathan et al., 2015). Large ITD separations might have enabled these segregation cues to come into operation for all of our listeners. In this view, the effect of reduced binaural TFS coding on SRM might be more pronounced when the ITD differences between the target and maskers are relatively small, providing some but not all listeners the segregation cues to facilitate SRM. All in all, it appears that in the current experiments performance was not limited by TFS processing abilities.

It should be mentioned that in the speech experiments, the side of the target as well as the different spatial distributions were alternated randomly on a trial-by-trial fashion, making it impossible for the listeners to follow a listening strategy where one focuses on a pre-defined spatial position. It is likely that such a presentation method makes performing the task attentionally taxing, and thus limits performance at an attentional level. If performance in the speech tests was indeed limited by attentional factors, then it would be expected that the relationships between both FDTs and SRM, and IPDTs and SRM would be affected by this. As attentional abilities were not measured, it is not possible to assess the impact of these on the speech tasks.

Last, SRTs were positively correlated with audiometric thresholds. It is speculated that these differences in the SRTs arose to some extent from impairment factors not directly related to reduced audibility, as these have been partly compensated for. It might be more likely that the correlation between SRT_{avg} and PTA_{oct} is, in fact, at least partly, attributed to the broadening of the auditory filters at higher sound pressure levels, which has been shown to affect SRTs (Studebaker et al., 1999). As the two HI groups were divided based on the extent of their hearing loss in the high frequency domain, they also received different amounts of amplification leading to a difference in presentation levels.

3.5 Conclusions

Consistent with earlier studies (Neher et al., 2011), the results of the speech experiments revealed that HI listeners experience difficulties in spatial listening tasks. The difficulties were more pronounced in fluctuating background noise than in steady-state noise. However, in contrast to earlier studies (Papakonstantinou et al., 2011), between-subject differences in the HI group could not be explained by TFS coding as measured by FDTs, but by average audiometric thresholds. It is likely that the correlations between SRTs and PTAs can be, at least partly, attributed to factors other than audibility (such as broader auditory filters at higher presentation levels), as the audibility of the target stimuli was individually compensated for. BILDs were smaller for the HI than for the NH listeners, but only by about 1 dB. Low-frequency IPDTs did not correlate with BILDs. BILDs in an experimental paradigm applying smaller ITDs to separate target from maskers would be more limited by elevated IPDTs and may thus be a more sensitive measure to investigate the effect of binaural TFS processing on spatial speech perception.

Acknowledgements

Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000. The authors thank Andrew Oxenham and two anonymous reviewers for their constructive comments on an earlier version of this manuscript.

4

Lateralized speech perception with small interaural time differences in normal-hearing and hearing-impaired listeners^a

Abstract

Both aging and sensorineural hearing loss affect negatively the listeners' sensitivity to binaural temporal fine structure (TFS) information. Despite this, spatial release from masking (SRM) elicited by interaural timing differences (ITDs) only can be almost normal for listeners with symmetrical hearing loss. This study investigated whether elderly hearing-impaired (HI) listeners also experience similar SRMs as young normal-hearing (NH) listeners, when SRMs are elicited by small ITDs. Also, correlations between measures of binaural TFS coding and the magnitude of SRM were investigated. Speech reception thresholds (SRTs) and SRM due to ITDs were measured in a headphone experiment for 10 young normal-hearing (NH) and 10 older hearing-impaired (HI) listeners, who had normal or close-to-normal hearing below 1.5 kHz. Diotic target sentences were presented in diotic or dichotic speech-shaped noise (SSN) or two-talker babble (TT) maskers. In the dichotic conditions, maskers were lateralized by delaying the

^a This chapter is based on Lócssei, G., S., Santurette, S., Dau, T., and MacDonald, E. N. (in preparation for Hearing Research).

masker waveforms in the left headphone channel. Multiple magnitudes of masker ITDs were tested in both noise conditions. Binaural TFS coding abilities were assessed by measuring pure-tone interaural phase difference (IPD) detection thresholds at multiple frequencies between 250 Hz and 1.5 kHz. The amount of SRM was only different between listener groups when the TT maskers were separated by large ITDs from the target, and even in this case, group differences were relatively small. Both when elicited by small and large ITDs, SRMs were moderately correlated with the highest frequency at which HI listeners were able to detect TFS IPDs in pure tones. The results suggest that reduced sensitivity to binaural TFS affects SRM elicited by small and large ITDs simultaneously.

4.1 Introduction

In everyday life, listeners often have to cope with communication scenarios where the speech stream of their interest (target) is embedded in a background of irrelevant acoustic interferers (maskers). Access to spatial acoustic cues associated with the target and with the maskers fosters speech perception in complex listening tasks. As target and maskers get separated in space, speech intelligibility (SI) typically improves, i.e. listeners exhibit lower signal-to-noise ratios (SNR) at a criterion performance (Peissig and Kollmeier, 1997). This has been termed spatial release from masking (SRM). In the horizontal plane, SRM is mainly facilitated by better-ear listening and binaural unmasking (BU). Better-ear listening is a monaural phenomenon, whereby listeners improve speech perception by listening solely with the ear that has a more favorable SNR; spatially separating the masker from a frontally-presented target normally results in an increased SNR at one of the listener's ears, where the head acoustically shadows the masker. Binaural unmasking refers to the phenomenon where disparities in the interaural timing differences (ITDs) or interaural phase differences (IPDs) associated with the target and the masker improve SI performance. For young normal-hearing (NH) listeners, the

binaural unmasking of speech mostly depends on the ITD or IPD disparities at low frequencies (Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005), below about 1.5 kHz, while contributions of such interaural differences at higher frequencies are small (Edmonds and Culling, 2005). The amount of SRM resulting from BU can be expressed as the magnitude of the difference in SRTs between two binaural presentation modes (in dB), and is often referred to as binaural intelligibility level difference or BILD (see e.g. Levitt and Rabiner, 1967).

Compared to their NH peers, hearing-impaired (HI) listeners suffer from difficulties when understanding speech in spatially complex acoustic scenarios, as reflected in elevated speech reception thresholds (SRTs) and in a reduction of the amount of SRM they achieve. The effect of hearing loss on speech intelligibility is often described as a combination of an audibility component and of a distortion component (Plomp, 1978). First, hearing loss typically results in elevated hearing threshold levels (HTLs). This affects stimulus audibility directly, resulting in elevated SRTs in quiet. Second, speech perception difficulties due to threshold shifts are often accompanied by other impairment factors, which distort the internal representation of supra-threshold stimuli in the auditory system, typically affecting SRTs in noise. Sensorineural hearing loss (SNHL) can introduce such “distortion loss” due to, e.g., the broadening of the auditory filters and the reduction of cochlear compression (Glasberg and Moore, 1986; Oxenham and Bacon, 2003) or degraded temporal fine structure^b (TFS) processing (Strelcyk and Dau, 2009; Papakonstantinou et al., 2011; Moore et al., 2012). A large part of the reduction in SRM observed in HI listeners can be attributed to their inability to take full advantage from better-ear listening due to stimulus inaudibility (see e.g. Bronkhorst and Plomp, 1989). Since the acoustic head shadow effect is most pronounced above 2 kHz, high-frequency hearing impairment substantially reduces the amount of SRM. Regarding binaural unmasking, listeners with an asymmetric hearing loss tend to show a greater reduction in BILDs than listeners with a symmetrical impairment (Jerger et al., 1984). For listeners with a symmetrical hearing loss, some studies showed smaller than normal BILDs (George et al., 2012; Best et al., 2013), while other studies showed normal or close-to-normal BILDs (Bronkhorst and Plomp, 1989; Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Strelcyk and Dau,

2009; Goverts and Houtgast, 2010). Thus, despite elevated audiometric thresholds, many listeners retain their ability to detect and utilize the binaural acoustic cues that foster the unmasking of speech in such scenarios.

The finding that HI listeners often show similar BILDs as NH seems surprising given the evidence that SNHL and aging affect binaural temporal coding abilities, such as the detection of IPDs (Ross et al., 2007b; Hopkins and Moore, 2011; King et al., 2014) and binaural masking level differences (BMLD) (Hall et al., 1984; Jerger et al., 1984; Staffel et al., 1990). With progressing age, the upper frequency limit at which listeners can detect binaural differences in TFS have been demonstrated to decline (Ross et al., 2007b) while detection thresholds below this upper limit increase (Moore et al., 2012; King et al., 2014). Furthermore, SNHL can lead to reduced binaural TFS detection abilities independent of aging (King et al., 2014). It is generally assumed that, when using pure tones, both the sensitivity to IPDs and the amount of BMLDs rely on the robustness of binaural TFS coding (Hall et al., 1984; Santurette and Dau, 2012), and it is a sensible assumption that BILD is facilitated, at least partly, by the same underlying neural mechanisms.

Studies about the role of supra-threshold coding deficits on BILDs in symmetrically impaired listeners have reached diverging conclusions. Goverts and Houtgast (2010) used a “distortion-sensitivity” approach to investigate the effect of supra-threshold deficits on BILDs and found that a reduction in BILD was associated with coding deficits in the phase and time representations of the stimuli. In contrast, Santurette and Dau (2012) measured BILDs in SSN in a headphone experiment using head-related transfer functions. They found that the BILDs were correlated with BMLDs measured at 500 Hz and 1 kHz, but not with the upper frequency limit for the detectability of a 180 degree IPD imposed on a pure tone. Furthermore, in a more recent study, Lócssei et al. (2015) did not find any correlation between IPD detection thresholds at 250 Hz in quiet and average BILDs obtained in a variety of

^b Broadband signals entering the cochlea will be decomposed by the auditory filters into a series of narrow-band signals with different center-frequencies, each corresponding to a unique point on the basilar membrane. Such narrow-band signals can be considered as slowly-varying fluctuations (envelope, ENV) superimposed on rapid oscillations (temporal fine structure, TFS) with a rate close to the center frequency of the corresponding band-pass filter (Moore, 2008).

background noise conditions.

Most studies investigating BILDs in NH and HI listeners utilized relatively large ITDs (of typically more than 500 ms) or a complete phase inversion to trigger BU. It is questionable whether such presentation methods are adequate to capture deficits in SI related to binaural TFS coding. Even though aging and SNHL affect the sensitivity to interaural delays in TFS, this reduction in sensitivity might still allow many listeners to detect low-frequency ITDs associated with laterally incoming sound waves in real life. In the studies of Lacher-Fougère and Demany (2005) and Hopkins and Moore (2011), IPD thresholds with 500 Hz pure-tone carriers were below 40 degrees for many of the HI listeners, corresponding to ITD thresholds of below about 0.2 ms. Assuming that BILDs are mainly governed by low-frequency carrier ITDs, the reason why earlier studies failed to show a clear reduction in BILDs in HI listeners could be that relatively large ITDs were applied when separating the target from the maskers. This could also explain why only a weak correlation was found between BILDs and binaural measures of TFS processing in earlier studies. In this view, a reduced BILD could be expected when BU is established by smaller ITDs, which are at or below the detection thresholds of some individuals.

The present study further investigated how binaural unmasking is affected in elderly listeners with a symmetrical high-frequency SNHL, and whether a reduction in BILD can be related to supra-threshold processing deficits, as represented by deficits in binaural TFS coding. SRTs were measured with diotic speech embedded in speech-shaped noise (SSN) or two-talker maskers which were either presented diotically or dichotically. Stimuli were presented over headphones and, when presented dichotically, the maskers were lateralized to the side of the head by introducing frequency-independent ITDs. The main research question was whether BILDs are affected differently by the amount of lateralization in the two listener groups. The hypothesis was that deficits in binaural unmasking abilities, as measured by BILDs, should be more prominent when triggered by small ITDs than by large ITDs. Furthermore, it was hypothesized that a loss of binaural sensitivity to TFS should impose a limitation on binaural unmasking, in particular when BILDs are triggered by small rather than by large ITDs. BILDs were measured for speech stimuli embedded in noise and separated by either large or small ITDs for a group

of young NH listeners and elderly HI listeners. In order to assess binaural TFS coding at low frequencies as a possible bottleneck for facilitating BILDs, TFS IPD thresholds were measured in pure-tone carriers over a range of frequencies. BILDs in the large and small ITD conditions were compared between the listener groups and the resulting IPD threshold profiles were contrasted with the size of BILDs in both cases.

4.2 Methods

4.2.1 Participants

Ten young NH (20-27 years, mean: 23, standard deviation (SD): 2.31) and 10 elderly HI (50-76 years, mean: 66.9, SD: 7.48) listeners participated in the study. For each listener, HTLs were measured with pure-tone audiometry at octave frequencies between 125 Hz and 8 kHz, and at octave frequencies between 750 Hz and 6 kHz. All NH participants had HTLs no greater than 20 dB HL at each of the tested frequencies. Listeners in the HI subgroup had HTLs below 25 dB HL at and below 1.5 kHz and a mild-to-moderate hearing loss at higher frequencies, except for two listeners (*j* and *h*), who had slightly elevated HTLs at 1.5 kHz. Age, gender, and audiometric thresholds averaged between the ears, are displayed in Table 4.1, ordered by increasing pure-tone threshold averages between 250 Hz and 1.5 kHz (PTA_{low}). For most listeners, HTLs between the ears did not differ by more than 10 dB at each measured frequency, and air-bone gaps were no greater than 15 dB at octave frequencies between 500 Hz and 4 kHz. The listeners completed the experiments through 3 visits. Each participant signed a consent form approved by the Science-Ethics Committee of the Capital Region of Denmark and received financial compensation for their participation.

4.2.2 Apparatus

The listeners were seated in a double-walled, sound attenuated booth, and the experiments were carried out using headphones. The stimuli were presented

#	Sex	Age	Audiometric thresholds averaged between the ears [dB HL]										Pure-tone averages			
			125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k	PTA _{low}	PTA _{high}	PTA _{oct}
a	m	60	7.5	0	5	5	0	0	7.5	25	32.5	27.5	35 ⁺	2	25.5	9
b	m	50	2.5	0	5	5	0	0	10	25	37.5	35	45	2	30.5	10.5
c	m	67	22.5	15	7.5	2.5	2.5	0	2.5	22.5	30	32.5*	65	5.5	30.5	11.5
d	f	65	7.5	2.5	5	10	10	7.5	7.5	12.5	22.5	32.5*	45	7	24	9.5
e	f	72	10	5	7.5	5	7.5	15	22.5	40	32.5	52.5	67.5	8	43	15
f	f	66	15	15	12.5	15	12.5	20	37.5*	50	45	50	65	15	49.5	24.5
g	f	72	22.5	12.5	15	15	17.5	17.5*	20 ⁺	25	27.5	35	52.5	15.5	32	18.5
h	f	69	5	5	12.5	17.5	22.5	32.5	40	47.5	52.5	55	60	18	51	26.5
i	m	76	15	12.5	20	20	15	25	25	45	57.5	52.5	67.5	18.5	49.5	26
j	f	72	15	25	22.5	20	20	27.5	32.5	32.5	40	45	65	23	43	28

Table 4.1: Gender, age and audiometric data of the HI listeners. From left to right: subject ID, gender (f for female and m for male), age, audiometric thresholds, pure-tone averages (audiometric thresholds averaged between 250 Hz and 1.5 kHz [PTA_{low}], above 1.5 kHz [PTA_{high}], or at octave frequencies between 250 Hz and 4 kHz [PTA_{oct}]). Signs in the subscripts indicate thresholds where the asymmetry between the ears was 15 dB (*) or 20 dB (+). In all other cases, thresholds differed at most by 10 dB between the ears.

at a sampling frequency of 48 kHz using a personal computer running Matlab, connected to an RME Fireface UCX soundcard and finally delivered through an SPL Phonitor Mini headphone amplifier to Sennheiser HDA200 circumaural headphones. During the IPD threshold tests, the listeners were seated alone in the booth and used a 21.5" touch-screen, a keyboard or a mouse to provide their responses. During the speech tests the experimenter was present, controlling and scoring the experiments using the screen or the mouse. In these cases, the listeners were seated on the other side of the screen, facing the experimenter.

4.2.3 Binaural fine structure processing

The experiments assessing sensitivity to binaural TFS information were inspired by the measures presented by Ross et al. (2007a) and Hopkins and Moore (2010). The task of the listeners was to detect IPDs of pulsating pure-tones at different frequencies. First, the frequency range was assessed, at which an IPD of 180 degrees could be detected (IPD_{fr}). Thereafter, IPD thresholds at fixed frequencies ranging from 250 Hz up to IPD_{fr} were measured in 250 Hz steps, denoted as IPD_{fr} (or specifically as IPD₂₅₀, IPD₅₀₀, etc. when referring to the threshold at a specific carrier frequency). The experiments were carried out during the first visit, after the assessment of the listeners' audiogram.

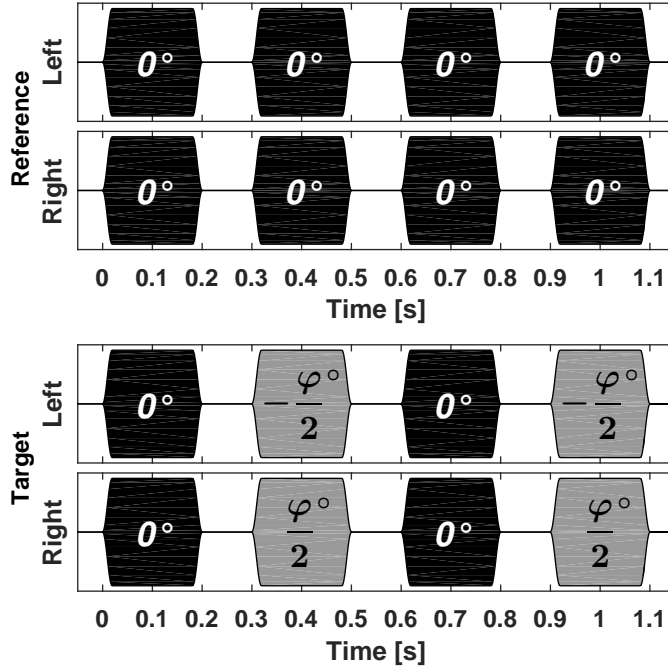


Figure 4.1: Example of stimulus intervals in the IPD threshold tests. The black and gray shaded areas denote tone bursts that were presented diotically (0° IPD between the ears) or dichotically (φ° IPD between the ears), respectively. Top: reference intervals contain 4 tone bursts presented diotically. Bottom: target intervals alternate between diotic and dichotic tone bursts.

IPD thresholds were estimated using a 3-interval 3-alternative forced-choice paradigm in combination with a multiplicative weighted up-down procedure (Kaernbach, 1991). Reference and target intervals are illustrated in Figure 4.1. Each interval contained a sequence of four 200-ms pure tones presented at the same frequency f_c , separated by 100-ms silent gaps. The gaps between presentation intervals were 400 ms long. In the reference intervals, all of the tones were presented diotically, i.e. with an identical starting phase, yielding a focal perceived position inside the head. In the target interval, the 1st and 3rd tones were presented with zero IPD, and the 2nd and 4th tones with a starting phase of $-\frac{\Delta\varphi}{2}$ and $\frac{\Delta\varphi}{2}$ in the left and right channels, respectively, yielding a total IPD of $\Delta\varphi$. Within each interval, the tones were gated synchronously between the ears with 20-ms long

raised-cosine ramps. The 70.7% point was estimated on the psychometric functions with the weighted up-down procedure. The stimuli were delivered at 30 dB sensation level (SL) as estimated by the individual audiometric thresholds. For carrier frequencies between the measured audiometric frequencies, the pure-tone thresholds were estimated from the audiogram by logarithmic interpolation.

In the IPD_{fr} test, the frequency f_c was the tracking variable while $\Delta\varphi$ was fixed at 180 degrees. The initial value of f_c was set to 1 kHz and 500 Hz for the NH and HI listeners, respectively. A test run consisted of 8 reversals in total. The step-size was set to 0.2 octaves for the first two reversals, which reduced to 0.1 octaves for further 6 reversals. Thresholds were estimated as the geometric mean of the last 6 reversal points. In the IPD_{lf} tests, the magnitude of $\Delta\varphi$ was adapted to track the detection thresholds at fixed frequencies. Thresholds were estimated starting from 250 Hz up to the IPD_{fr} threshold, in steps of 250 Hz. The starting value of $\Delta\varphi$ was frequency dependent and different for the two listener groups. For the NH listeners, it was set to 45 degrees at and below 750 Hz and 90 degrees above this frequency, while for the HI listeners it started at 45 degrees at 250 Hz, 90 degrees at 500 Hz and 180 at all other frequencies. The step size was initially set to 1.25^3 , which changed to 1.25^2 after the second reversal and further decreased to 1.25 after two more reversals. Six reversals were taken with this final step-size and thresholds were calculated as the geometrical mean of the reversal points. The listeners first received a verbal introduction to the test and a few example stimuli were presented with 180 degree IPDs. After this, the listeners completed at least 5 runs in each condition. The final thresholds were calculated as the geometrical average of the last 3 runs.

The results of the IPD experiments yielded the estimation of individual IPD discrimination threshold curves sampled over the frequency range in which listeners were sensitive to such differences. However, since these IPD profiles contain multiple points, a comparison with the results in the speech experiments is not straightforward, especially as the number of sampling intervals differs between listeners. Therefore, individual threshold curves were characterized by only two parameters: 1) the IPD_{fr} threshold and 2) the minimum interaural time difference (ITD_{min}) listeners were sensitive to over all the tested frequencies. While IPD dis-

crimination thresholds grow monotonically with increasing frequency (see e.g. Hopkins and Moore, 2011; Lacher-Fougère and Demany, 2005), these threshold curves become a non-monotonic function of frequency when expressed as ITDs, with best thresholds between 750 Hz and 1000 Hz for NH (Brughera et al., 2013; Hartmann et al., 2013). The ITD_{\min} thresholds were determined for each listener by converting their IPD_{if} thresholds to corresponding time-delays and by taking the minimum value across frequency.

4.2.4 Speech perception in noise

Speech material

In the speech tests, the DAT corpus (Nielsen et al., 2014) was used. This open-set corpus contains grammatically correct, low-predictability sentences in the form of “<Name> *thought about* <keyword 1> *and* <keyword 2> *yesterday*”. The <Name> field corresponds to either “Dagmar”, “Asta” or “Tine”, which serve as call signs that facilitate cueing the sentences. The sentences are uttered by one of three professional female talkers with similar voice characteristics, each assigned to the complete set of sentences with the same call sign. For each call sign the corpus was organized into 3 training and 10 test lists of 20 sentences.

Speech interferers

SI performance was evaluated both in SSN and in an interfering two-talker background (TT). In the SSN condition, the “Dagmar” sentences were used as target material, and the long-term average spectrum of the noise was matched to that of the “Dagmar” sentences. To avoid repeating any lists within the experiment, the “Asta” sentences were used in the TT condition. In these cases, sentences spoken by the two other talkers were applied as maskers. As the three talkers in the DAT corpus have similar voice characteristics (Nielsen et al., 2014), no spectral matching was applied between target and maskers in the TT conditions. The SSN tokens were semi-randomly chosen from a pool of fifty 5-second noise samples, which were then truncated to match with the length of the target sentence. The

TT maskers started at the same time as the target but could end earlier or later than the target. The target sentences were always presented diotically while the maskers were delivered in one of the following lateralization settings: 1) diotic presentation, colocated with the target (SSN_{co} and TT_{co}), 2) lateralized to the side through large ITDs (SSN_{lrg} and TT_{lrg}), or 3) lateralized to the side through small ITDs (SSN_{sm} and TT_{sm}). In each case, lateralization was realized by applying a frequency-independent time delay on the full masker waveforms in the left channel of the headphones, which shifted the perceived lateral position of the maskers towards the right side of the head. The listeners were requested to listen to the target sentences starting with the pre-defined call sign and to repeat the two keywords as accurately as possible.

Measurement procedure

The beginning of each sentence presentation was indicated by a 100-ms 1-kHz beep tone followed by a 500-ms silent gap. Target and maskers then started at the same time, gated by 50-ms raised-cosine ramps. In all but the TT_{sm} condition, SI performance was assessed by measuring sentence-correct SRTs. The level of the target was fixed and the masker level was adapted in 2-dB steps while keeping its ITD constant, which was either set to 0 ms in the diotic masker condition, or to 0.68 ms or 0.27 ms in the dichotic conditions with large or small ITDs, respectively. The SRTs for single lists were calculated by subtracting the overall masker presentation levels, averaged from the 5th sentence to the hypothetical 21st sentence, from the presentation level of the target. The resulting thresholds were thus expressed in SNR. In the TT_{sm} condition, however, instead of measuring SRTs at a fixed ITD, the 50% sentence-correct point was tracked as a function of ITD at a fixed SNR. Since relatively large SRT drops can be observed even with small spatial separations when target and maskers are highly confusable (Marrone et al., 2008a; Kidd et al., 2010; Swaminathan et al., 2015), this presentation method was expected to be a sensitive way to assess the lower ITD limit at which listeners can benefit from BU. This SNR criterion was set to be 3 dB lower than the SRT in the colocated target-masker condition, corresponding to a BILD of 3 dB. Maskers were initially

lateralized to the right by 0.35-ms ITDs. This value was reduced if both keywords were identified correctly or increased otherwise. The available ITDs were restricted to the range of 0 to 0.68 ms. The amount of change in ITD was initially set to 62 μ s, which changed to 41 μ s after the 5th sentence. Thresholds were calculated as the average of the ITD values used from the 10th to the hypothetical 21st sentence. Three thresholds were estimated in each of the 6 noise conditions, and the final thresholds were calculated as the arithmetic average of these 3 thresholds.

In order to accentuate individual differences, listeners were subjected to exactly the same experimental protocol, including the order of conditions and lists tested or the identity of sentences and noise tokens. Performance in the TT and SSN conditions was assessed during separate visits. As the target in the SSN conditions was identical to one of the TT maskers, first the TT conditions were evaluated to omit masker familiarity in the latter tests. SRTs over 9 lists were evaluated during each visit, using 3 lists in each condition. The listeners were familiarized with the test material using the 3 DAT training lists at the beginning of each visit, containing two lists with colocated and one with separated target-masker conditions. To further minimize any remaining training effects, during the first 6 runs in the testing phase, the conditions where target and maskers are colocated or separated by large ITDs were tested in an alternating manner (run 1, 3 and 5: colocated conditions). The conditions with small ITD separation were tested during the last 3 runs. Care was taken to distribute the test lists between conditions with about the same expected average SRT, based on the average intelligibility of the lists (Nielsen et al., 2014). The maskers were also fixed. Thus, except for the individual amplification schemes, listeners received exactly the same stimuli in the same order.

Stimulus calibration

The headphones were calibrated to yield a flat response as measured by a Brüel & Kjær 4180 Head and Torso Simulator (HATS). Audibility of the stimuli was restored by applying individualized linear gains based on the individual listeners' audiogram and on the long-term average spectrum of the "Dagmar" sentences. The

audibility criterion was set to be 15 dB above the individual hearing thresholds for one-third octave bands between 100 Hz and 3 kHz, which was reduced to 12, 8 and 0 dB at 4, 6 and 8 kHz. Then, the target stimulus was scaled to a nominal level of 65 dB SPL when measured at the eardrums of the HATS and mixed with the scaled maskers. The individualized gains were applied to this mixture using 512 order finite impulse response filters, amplifying both target and maskers. These filters also compensated for the headphone frequency response. At each presentation, the stimuli were band-pass filtered between 200 Hz and 10 kHz. Presentation levels were limited to 94 dBA and if the estimated overall presentation level of a stimulus exceeded this, it was downscaled in 2-dB steps.

4.2.5 Statistical analysis

The alpha-level in all of the statistical tests mentioned below was fixed at 0.05. This level was adjusted in the case of multiple comparisons using a sequentially-rejective Holm-Bonferroni correction, to control for the inflated family-wise error rate. The speech tests were analyzed using mixed-design analysis of variance (ANOVA) models. Further details of the statistical analyses are provided in the respective descriptions of the results in Section 4.3.

4.3 Results

4.3.1 Audiometric thresholds

Despite the efforts to match the NH and HI groups in hearing levels at the low frequencies, there remained a difference of almost 10 dB between the groups in terms of PTA_{low} ($p = 0.004$). Thus, even though the majority of the HI listeners had normal hearing levels at most of the tested frequencies below 1.5 kHz, a possible contribution of elevated HTLs in this frequency region on a group level cannot be excluded.

4.3.2 Binaural fine structure coding

The IPD thresholds are shown in Figure 4.2 and in the left panel of Figure 4.3 for both the NH and HI listeners. The solid horizontal black lines denote the group means and the corresponding boxes represent ± 1 SD. Before the statistical analysis, the data were log-transformed, yielding a close-to-normal distribution of the thresholds. This is consistent with previous reports in the literature (Lacher-Fougère and Demany, 2005; Strelcyk and Dau, 2009; Hopkins and Moore, 2011; King et al., 2014). The average thresholds of the HI group were generally worse than those obtained in the NH group. While most of the HI listeners performed poorer than normal in each of the tested conditions, listeners *a*, *b* and *i* had thresholds in the same range as the NH listeners, except for the highest frequency tested (Figure 4.2). While all of the NH listeners reached IPD_{fr} thresholds above 1 kHz, 9 out of the 10 HI listeners had thresholds below that (Figure 4.3). The findings are consistent with studies showing a relatively large spread of IPD thresholds of elderly and hearing-impaired listeners, ranging from normal to very abnormal (e.g. Neher et al., 2012). In general, the data are in good agreement with earlier reports on IPD discrimination thresholds in NH and elderly HI listeners (Ross et al., 2007a; Ross et al., 2007b; Hopkins and Moore, 2011; Neher et al., 2012).

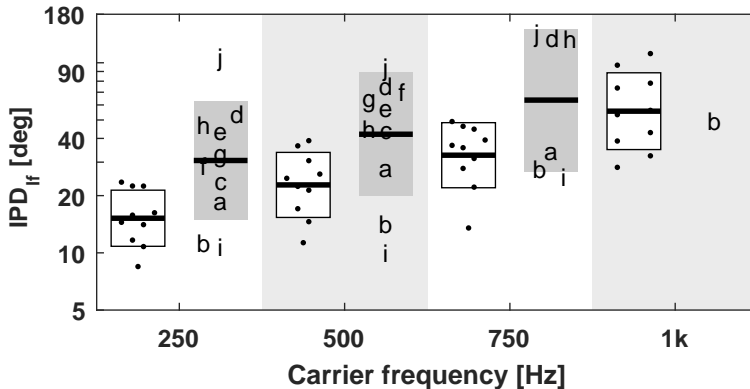


Figure 4.2: Results in the IPD_{fr} experiments for the NH (dots) and HI (letters) listener groups. Black horizontal lines mark group means and the boxes denote ± 1 SD of the corresponding groups. The shading of the background is according to the conditions with different carrier-frequencies.

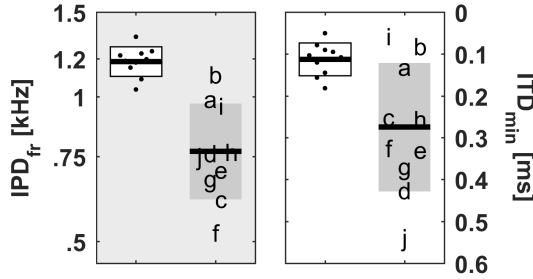


Figure 4.3: Results in the IPD_{fr} and ITD_{min} experiments of the NH (dots) and HI (letters) listeners. Black horizontal lines mark group means and the boxes denote ± 1 SD of the corresponding groups. Note that the y-axis in the right panel is reversed, so that data points located further towards the top of each panel represent better performance.

Independent t-tests were performed on the IPD_{fr} and IPD_{lf} thresholds up to 750 Hz to evaluate the significance of the results described above. The analysis revealed significant differences between the group means of the IPD₂₅₀ ($t(18) = -2.79, p = 0.012$), IPD₅₀₀ ($t(18) = -2.30, p = 0.033$) and IPD_{fr} ($t(18) = 5.67, p < 0.001$), but not regarding the IPD₇₅₀ tests ($t(6.275) = -1.79, p = 0.122$). A sequentially rejective Holm-Bonferroni correction was applied to correct for multiple comparisons, which only affected the statistics of the IPD₅₀₀ test, rendering the differences between the means non-significant ($p < .05$ before but $p > .025$ after correction). All in all, significant differences were confirmed between the two groups in the IPD₂₅₀ and IPD_{fr} tests. However, in the IPD_{lf} tests at frequencies at or above 750 Hz, thresholds were only measured for a subset of the HI listeners, and the other listeners were not included when comparing group differences, biasing the group means towards lower values than the true group average. Most HI listeners were, in fact, not sensitive to any changes in IPDs at these frequencies, which is also clearly reflected in the IPD_{fr} test results. Therefore, it appears that the HI listeners performed consistently worse than NH in these tasks at all tested frequencies. This is also supported by the significant difference between the groups with respect to the ITD_{min} thresholds ($t(10.16) = -3.234, p = 0.009$), which amounted to 0.11 ms for NH and 0.27 ms for HI (see Figure 4.3). While 6 out of the 10 listeners in the NH group had the ITD_{min} at 750 Hz or 1 kHz, 9 out of the 10 HI listeners had

the ITD_{min} thresholds at 250 or 500 Hz. In the HI group, the ITD_{min} thresholds did not correlate significantly with the IPD_{fr} thresholds ($r = -0.61, p = 0.064$), thus both measures were tested as possible predictors of BILD. Pearson's correlations revealed that the IPD_{fr} thresholds were significantly correlated with the IPD_{750} ($r = -0.948, p = 0.004$) and IPD_{500} ($r = -0.74, p = 0.014$) thresholds, but not with the IPD_{250} ($r = -0.485, p = 0.16$) thresholds, even after correcting for multiple comparisons.

Within the HI group, no correlation was found between the IPD_{fr} thresholds and age or the HTLs averaged between the ears at the corresponding test frequencies. A correlation was also absent between IPD_{fr} or ITD_{min} and age or PTA_{low} .

4.3.3 Speech perception in noise

Figure 4.4 shows the SRTs for the NH (dots) and the HI listeners (numbers) obtained in the fixed-ITD conditions. The obtained SRTs are comparable with the results of Nielsen et al. (2014) and Lócssei et al. (2015), who used the same speech material in similar settings.

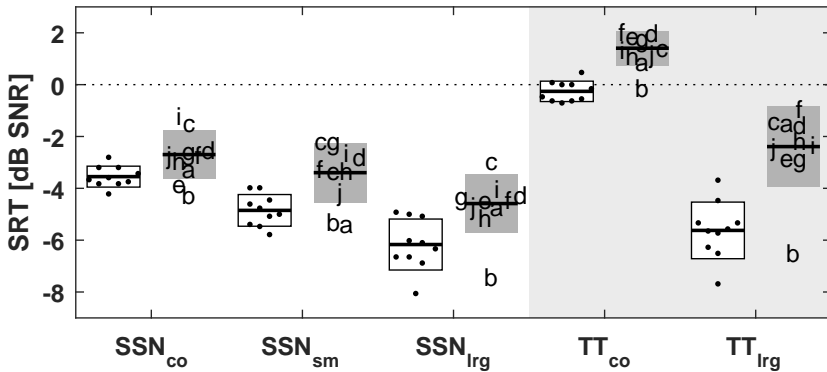


Figure 4.4: SRTs in SSN and two-talker babble (TT) for NH (dots) and HI (letters) listeners. Solid black horizontal lines mark group means and the boxes denote ± 1 SD. The background shadings mark condition groups using the same type of background noise. In each condition the target was presented diotically. The different test conditions are denoted on the x-axis. Subscripts indicate the ITD configuration of the masker: **co**: diotic presentation, collocated with the masker; **sm**: masker lateralized with a small ITD (0.27 ms); **lrg**: masker lateralized with a large ITD (0.68 ms).

A mixed-design ANOVA was conducted on a subset of the SRT data, including the SSN_{co} , SSN_{lrg} , TT_{co} and TT_{lrg} conditions. The model contained the SRTs as the dependent variable, and used *noise type* (SSN or TT) and *lateralization* ('co' or 'lrg') as within-subject factors and *listener group* (NH or HI) as between-subject factors. The statistical analysis revealed that all main effects were highly significant along with all two-way interactions. On a group level, NH listeners performed better in all of the tested conditions than the HI listeners, yielding lower SRTs ($F(1, 18) = 30.42, p < 0.001$). Thresholds in the SSN conditions were lower than in the TT conditions indicating that, on average, listeners had more difficulties understanding the target sentences in TT noise than in SSN ($F(1, 18) = 309.4, p < 0.001$). Compared to the NH group, HI listeners had more pronounced difficulties in the TT than in the SSN conditions, which is supported by the significant interaction term between *noise type* and *listener group* ($F(1, 18) = 18.32, p < 0.001$). Also, listeners showed a clear benefit in SRTs when maskers got lateralized, which is indicated by the drop of SRTs within the shaded areas as going from left to right ($F(1, 18) = 311.77, p < 0.001$). Differences in SRTs between listener groups were generally smaller when target and maskers were colocated, and slightly increased as the maskers were gradually lateralized towards the side. With the maskers colocated with the target, thresholds in the TT condition were higher than in the SSN condition for both listener groups. However, this difference between noise conditions decreased once the target and the maskers became spatially separated. The statistical significance of this effect is supported by the interaction term between *noise type* and *lateralization* ($F(1, 18) = 73.77, p < 0.001$). Furthermore, the NH listeners yielded similar thresholds in the SSN_{lrg} and in the TT_{lrg} conditions, but the difference in these SRTs was significant in the HI group (two-tailed paired t-test, $t(9) = -7.93, p < 0.001$).

Figure 4.5 illustrates the measures characterizing the amount of BU of speech in the SI tests. In the left panel, the BILDs due to masker lateralization are shown. The right panel indicates the results obtained in the TT_{sm} condition, where the ITD thresholds were assessed at a fixed SNR that was set to be 3 dB below the individual SRTs in the TT_{co} condition. Again, these thresholds indicate the amount of lateralization imposed on the TT-masker the listeners needed to obtain a BILD of

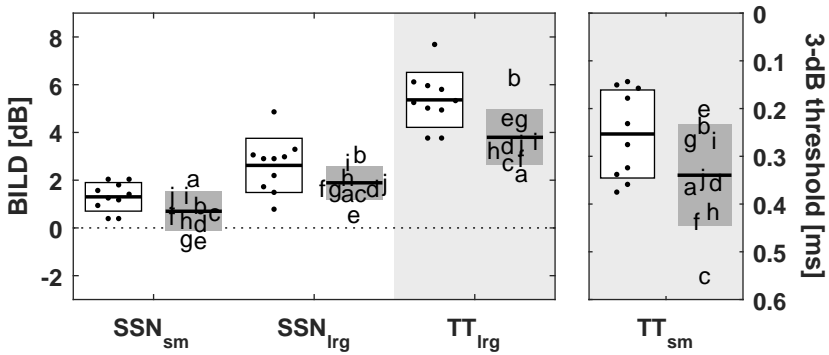


Figure 4.5: BILDs at fixed ITDs in SSN and two-talker babble (TT) and the ITD threshold needed to yield a fixed 3-dB BILD in the TT noise (i.e. a 3 dB decrease in SRTs as compared to the TT_{co} condition). Solid horizontal black lines and the boxes around denote group means and ± 1 SD for the NH (dots) and HI individuals (letters). Background shadings mark condition groups with the same noise type. Note that the first 3 conditions to the left are expressed in dB, while the last condition in ms. Condition notations are the same as in Figure 4.4.

3 dB. Note that the ordinate is reversed in the right panel, such that values towards the top of the figure indicate better performance in both the left and right panels. In general, the NH listeners showed a slightly better performance than the HI listeners in all conditions. When considering the BILDs at fixed ITDs, the statistical significance of this trend was supported by the significant interaction term between *lateralization* and *listener group* ($F(1, 18) = 8.81, p = 0.008$) in the ANOVA model conducted on the SRTs. It is clear that most of the listeners benefitted from masker lateralization in all of the tested conditions. While BILDs were small in the SSN_{sm} condition, they increased as the ITD magnitudes of the maskers increased from 0.27 to 0.68 ms. The benefit was greatest in the TT_{lrg} condition, where it reached 5.4 dB and 3.8 dB for the NH and HI listeners, respectively. Not only did the NH listeners show greater BILDs in the conditions with fixed ITDs, but they also yielded a 3 dB BILD at smaller ITDs than the HI listeners. Nonetheless, independent t-tests on the BILD data indicated that the only case where differences were significantly different between the two listener groups was the TT_{lrg} condition ($t(18) = 3.03, p = 0.007$).

Hearing threshold levels averaged over octave-frequencies between 250 Hz and 4 kHz (PTA_{oct}) were not correlated with average performance in the SI tests,

calculated as the mean of the SRTs obtained in the SSN_{co} , SSN_{lrg} , TT_{co} and TT_{lrg} conditions, nor with the amount of BILDs averaged in the TT_{lrg} and SSN_{lrg} conditions. These results suggest that the performance was not limited in these tests by audibility or impairment factors related to elevated hearing thresholds *per se*.

All in all, as regarding SRTs in noise, the majority of the HI listeners performed worse than the NH listeners in all of the tested conditions. The only exception was listener *b*, who can be considered as a “good performer” even when compared to the NH listeners. In contrast, the listeners’ ability to utilize binaural unmasking to aid speech intelligibility in lateralized conditions appeared to be only mildly affected in the HI group. While NH listeners generally showed slightly greater BILDs in all tested conditions, as well as sharper spatial tuning abilities, the only case where group differences reached significance was when two interfering talkers were separated from the target by large ITDs. Group differences were not pronounced with regard to spatial tuning abilities.

4.3.4 Binaural TFS coding and unmasking of speech

The primary goal of the study was to investigate the relationship between binaural sensitivity to TFS and the BU of speech. In order to do so, Pearson’s correlations were calculated between each of the four measures of BU and the ITD_{min} or IPD_{fr} results within the group of HI listeners. The reason for testing the correlations in the HI group was that, besides the differences in the TFS processing abilities of the NH and HI groups, there were significant group differences in age and audiometric thresholds. Even though the correlations were significant when the NH and HI groups were collapsed, these correlations disappeared when both age and elevated audiometric thresholds were controlled for. By using only the HI listeners in these analyses, the relationship between the binaural TFS coding abilities and the binaural unmasking of speech was evaluated on a more direct way.

The obtained correlation coefficients with the corresponding p-values are displayed in Table 4.2. Even though none of the correlations reached significance, the direction of the correlations was in all cases consistent with the hypothesis

Pearson's correlation			
Variables		r	p
ITD _{min} vs.	SSN _{sm}	-0.43	0.23
	SSN _{lrg}	-0.49	0.15
	TT _{sm}	0.13	0.71
	TT _{lrg}	-0.15	0.68
IPD _{fr} vs.	SSN _{sm}	0.54	0.1
	SSN _{lrg}	0.58	0.08
	TT _{sm}	-0.51	0.13
	TT _{lrg}	0.38	0.28

Table 4.2: Pearson's product-moment correlations between the measures of binaural TFS processing abilities and BILDs within the HI listener group. Condition notations are the same as in Figure 4.4.

that the listeners with better TFS coding abilities yield greater BILDs and yield a 3-dB benefit at smaller ITDs. First of all, the IPD_{fr} test showed a greater correlation than the ITD_{min} scores with all of the measures of BU of speech, suggesting that the facilitation of BUs is more related to the frequency range at which listeners are sensitive to TFS phase-information than to the acuity of their sensitivity in absolute terms in a narrow frequency range. Figure 4.6 shows the scatter plots of the IPD_{fr} thresholds and the measures of BU in the speech tests. The figure contains the results of both the NH and the HI listeners, which are displayed as dots and numbers, respectively. Interestingly, the correlation analysis revealed that BU depended on binaural TFS sensitivity to a similar degree when target and maskers were separated by a small ITD as when they were separated by large ITDs, which contradicted the initial hypothesis. Finally, the correlations were not smaller when the results in the SSN rather than in the TT conditions were compared to the measures of binaural TFS processing, even though BILDs covered only a relatively small range in the SSN conditions.

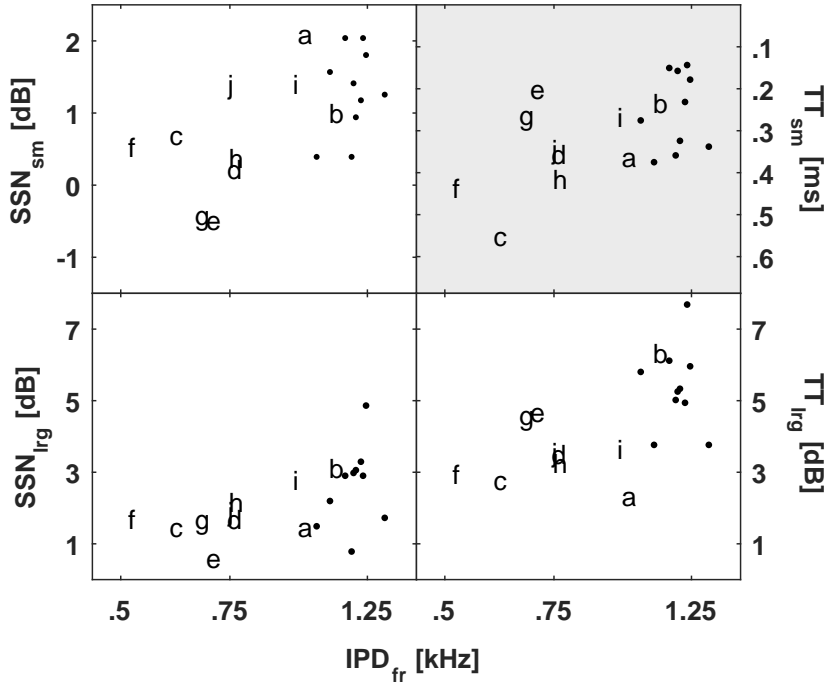


Figure 4.6: Scatter plots between the individual IPD_{fr} thresholds and the BILDs (white background) or the ITD thresholds to yield a 3-dB BILD (gray background). The corresponding noise conditions are marked on the labels of the ordinates and are consistent with the notations in Figure 4.4. Dots and letters denote the results of the NH and HI listeners, respectively. Note that the y-axis in the top right panel is reversed.

4.4 Discussion

4.4.1 Summary of main findings and relations to other studies

The results of the IPD_{lf} and IPD_{fr} tests were consistent with those found in earlier studies applying similar test paradigms (e.g. Ross et al., 2007b; Hopkins and Moore, 2011; Neher et al., 2011; King et al., 2014), showing that the HI listeners performed worse than their NH peers when considering the IPD_{fr} , IPD_{lf} and ITD_{min} thresholds. In contrast to studies suggesting an effect of both age and elevated audiometric thresholds on IPD detection performance (e.g. Ross et al., 2007b; King

et al., 2014), no correlations were found between IPD_{fr} or ITD_{min} and age or PTA_{low} . These inconsistent findings are likely a result of the small spread both in age and audiometric thresholds of the HI listeners in the current study. Nonetheless, the threshold differences between the NH and HI groups can most likely be attributed to age-related physiological changes in the auditory pathway, as differences in hearing thresholds between the NH and HI listener groups below 1.5 kHz were relatively small. In the tested HI listeners, a reduction in the frequency range where TFS IPDs could be detected was not necessarily linked to a reduction of TFS IPD thresholds at the lowest tested frequency. The dissociation was less pronounced when the IPD_{fr} thresholds were compared to the ITD_{min} thresholds and absent when compared to the IPD_{750} or IPD_{500} thresholds. The lack of correlation between the IPD_{fr} and IPD_{250} thresholds should be considered with some caution, due to the small sample size of the listener groups applied in the current study. It is possible that an age-related reduction in binaural TFS processing abilities occurs at very low and higher frequencies simultaneously, e.g. as a result of an age-related neural degeneration of spiral-ganglion cells, the extent of which was shown to be similar across cochlear partitions (Makary et al., 2011).

In the speech experiments, all listeners yielded lower average SRT values when the target sentences were presented in SSN than when presented in TT noise. The SRTs obtained in the TT_{co} condition, with average values of about 0 and 1.5 dB SNR for the NH and HI groups, respectively, were consistent with earlier reports measuring SRTs in similar scenarios (Brungart, 2001; Marrone et al., 2008b; Kidd et al., 2010). While listeners in both groups showed a clear benefit when the maskers were lateralized to the side, indicating the presence of an active BU mechanism, the amount of BILDs differed slightly between the two groups, and this difference only reached significance in the TT_{lrg} condition. Therefore, the results obtained in the TT conditions do not support the hypothesis that the HI listeners' processing deficits in BU are more pronounced when triggered by small rather than by large ITDs. Rather, the deficits in the BU of speech manifested themselves mainly by reducing the overall benefit HI listeners could yield when target and maskers were separated by large ITDs. Nonetheless, the SRTs obtained in the SSN_{lrg} and TT_{lrg} conditions suggest the possibility that BILDs in the TT_{lrg} condition were,

at least partly, affected by monaural deficits in temporal processing. The SRTs in the TT conditions were different from that in the SSN conditions as the TT interferers act as informational maskers (Kidd et al., 2008) and also because they offer modulation masking (Houtgast, 1989; Takahashi and Bacon, 1992) due to the inherent spectro-temporal fluctuations characteristic to running speech. While the NH listeners yielded similar SRTs in the TT_{lrg} and SSN_{lrg} conditions, the HI listeners had about 2 dB higher SRTs in the TT_{lrg} condition than in the SSN_{lrg} condition. However, informational masking is substantially reduced when target and maskers are spatially separated (e.g. Arbogast et al., 2002; Kidd et al., 2008). Therefore, the performance in the TT_{lrg} condition can be assumed to be limited by factors other than informational masking (c.f. Best et al., 2013). Several studies have shown that HI listeners are more susceptible to modulation masking than NH listeners, which manifests itself in less-than-normal fluctuating-masker benefit when modulations are imposed on a stationary masker (Festen and Plomp, 1990; Strelcyk and Dau, 2009). Therefore, it is possible that, compared to the NH listeners, the HI listeners would have elevated thresholds in the TT_{lrg} condition due to their susceptibility to modulation masking, even if they had intact binaural processing abilities. The extent to which such monaural factors might have contributed to the reduced BILDs in the current study is nonetheless difficult to evaluate, as it is likely that both informational and modulation masking are involved in the TT_{co} and TT_{lrg} conditions.

The IPD_{fr} thresholds generally showed a greater correlation with all of the BILD measures than the ITD_{min} thresholds. This suggests that, if BILDs are indeed limited by binaural TFS processing abilities, the critical factor is rather the frequency range at which listeners have access to such cues than the best sensitivity within a narrow range of frequencies. In fact, Levitt and Rabiner (1967) argued that the facilitation of BILDs might require access to ITD information over a broad range of frequencies, while good sensitivity to relevant acoustic cues in a narrow frequency range might assure good performance in masking release in a signal detection task. The correlation coefficients were of similar magnitude when BILD measures were elicited by small as by large ITDs. These results are in contrast to the initial hypothesis that deficits in binaural TFS processing might affect BU of speech

elicited by small ITDs. The results suggest that binaural TFS processing deficits affect BU to a similar degree at small and large ITDs, and imposes limitations on the listeners' spatial tuning abilities and on the amount of maximum achievable BILD synchronously. Neher et al. (2011) suggested that binaural TFS information might aid spatial speech perception by providing acoustic cues for the perceptual separation of the target and the maskers, which would facilitate directing auditory attention towards the target. Should that be the case, the current results indicate that access to TFS information might equally be important when target and maskers are separated by small and large ITDs.

All in all, binaural TFS processing abilities were only weak predictors of spatial speech processing abilities in the current study. The measured correlation coefficients were similar to those found by Neher et al. (2011,2012), where IPD detection thresholds were compared to SRTs obtained in spatial SI tests. In those studies, SI was measured in free-field conditions, which allowed the listeners to use listening strategies unrelated to BU, such as e.g. better-ear listening. Even though better-ear listening was eliminated in the current study, binaural sensitivity to TFS information, as measured by IPD detection thresholds in pure tones, could explain only a comparably small amount of variability (about 30%) in the speech data as in the results of these previous studies.

4.4.2 Limitations of the study

Only 10 elderly HI listeners participated in the experiments of the current study. Although a relationship between binaural TFS processing abilities and lateralized speech perception in noise have been reported in studies using similar sample sizes (e.g. Strelcyk and Dau, 2009), such experimental designs are lacking the statistical power of detecting small effect sizes. This might especially hold for the current study, as BILDs only rarely exceed 7 dB even for NH listeners, which might further complicate the detection of any correlations due to the inherent noise in the speech data. The current study attempted to ameliorate this issue by testing each listener on exactly the same material in the same order. As the correlations consistently approached significance without yielding it, it is likely that a genuine

moderate correlation persists between the tested measures. This might be revealed by increasing the sample size of the HI group.

It is also questionable whether the IPD_{fr} and ITD_{min} thresholds indeed measure independent aspects of interaural processing abilities. Multiple studies have shown that, for NH listeners, TFS ITD thresholds are lowest between 750 Hz and 1 kHz, and that TFS ITDs decline rapidly above about 1.25 kHz (Zwislocki and Feldman, 1956; Brughera et al., 2013), which is only about a half an octave higher than the frequency range of best performance in the same task. Therefore, it is plausible that a decrease in the IPD_{fr} thresholds will inherently affect the ITD_{min} thresholds. In this view, the fact that the correlation between ITD_{min} and IPD_{fr} did not yield significance might also be due to the coarse sampling of IPD_{fr} thresholds in the frequency domain, which might result in an inaccurate estimation of the true ITD_{min} values.

The limitations of the experimental paradigm utilized in the TT_{sm} condition deserve some further attention. This condition assessed the sharpness of spatial tuning due to BU by measuring the amount of ITDs by which target and maskers had to be separated in order to give raise to a BILD of 3 dB. First, assuming that the magnitude of the BILD monotonically increases with increasing ITD, this paradigm is only plausible if one assumes that listeners can obtain a 3 dB benefit at the largest ITDs applied. While this was clearly the case for the NH listeners, who showed a BILD of at least 3.7 dB, and about 5.4 dB on average, three listeners from the HI group (listener *a*, *c* and *f*) had a BILD of below 3 dB in the TT_{lrg} condition. Theoretically, for these listeners, the thresholds in the TT_{sm} conditions should be greater than 0.68 ms. Thus, even though these listeners had the greatest thresholds in the TT_{sm} condition, their results should be treated with caution. Furthermore, the average BILDs of the HI listeners in the TT_{lrg} condition was about 4 dB, while the thresholds in the TT_{sm} condition were assessed for a fixed BILD of 3 dB. This means that the differences in performance criteria between these two conditions were relatively small. A possible modification of the existing paradigm to alleviate these issues would be to use identical talkers for the target and the maskers, which would likely increase the BILDs for all listeners. Finally, the current paradigm used a linear step size in the ITD-tracking procedure. In contrast, multiple studies

have shown that pure-tone IPD detection thresholds are typically log-normally distributed between subjects (e.g. Lacher-Fougère and Demany, 2005; Strelcyk and Dau, 2009). Assuming a relationship between BU in the speech tests and the elevation of IPD detection thresholds, one can argue that the adjustment of the ITD values on a log-scale in the TT_{sm} condition would be a more appropriate method, as it would allow the detection of more subtle differences between subjects at lower ITDs.

4.4.3 Perspectives

Even though the current study revealed only a weak correlation between IPD detection thresholds and BILDs, it is clear that robust binaural temporal coding underlies the facilitation of the BU of speech in noise. Assuming that TFS information plays a critical role in this, the question arises whether IPD or ITD detection thresholds of pure tones are indeed the most relevant measures to relate to BILDs. From a clinical perspective, IPD detection threshold tests are appealing as they are conceptually simple, time efficient and there is no need for extensive training to yield relatively stable thresholds (Hopkins and Moore, 2010). However, IPD detection thresholds estimate the minimum time difference across the ears that give rise to a lateralized stimulus percept, but no estimate about the extent of laterality such stimuli can yield, which may be a more relevant feature of the auditory system when considering BILDs. It would be interesting to test how the extent of laterality of stimuli at fixed IPDs changes with aging or progressing HL and how it relates to binaural speech processing. Finally, since the BU of speech is facilitated over a broad frequency range, where binaural processes in different auditory channels are thought to operate independently, combined measures of binaural temporal sensitivity over multiple frequencies may provide a better framework when assessing the relationship between TFS coding and BU processes.

It is also possible that some of the variance in the measured BILDs were related to the listeners' sensitivity to envelope (ENV) ITDs. It has been shown that frequencies above 1.5 kHz can also carry ITD information that NH listeners can utilize to facilitate BU (Edmonds and Culling, 2005). The role of ENV ITDs cues at

high frequencies might even be more pronounced in the facilitation of BILDs in HI listeners, assuming that the ENV representation of the speech stimuli is enhanced in the region of their hearing loss due to e.g. the broadening of the auditory filters or reduced cochlear compression (Henry et al., 2014; Bianchi et al., 2015). For example, increasing the modulation depth of amplitude-modulated high-frequency pure tones has been shown to increase sensitivity to ENV ITD cues (Bernstein and Trahiotis, 2009). Therefore, it cannot be excluded that in the current study the measured BILDs partly relied on the listener's ability to detect and exploit binaural timing differences in the ENV domain, which warrants further research.

4.5 Conclusions

HI listeners showed a reduction in binaural TFS coding abilities compared to NH listeners, as reflected in a reduction of the IPD_{fr} and an increase of the ITD_{min} thresholds. The relation between the IPD_{lf} and IPD_{fr} thresholds was less pronounced than the relation between the IPD_{fr} and ITD_{min} thresholds. It is likely that a reduction in the upper frequency limit at which listeners can detect IPD differences is linked to the minimum TFS ITD listeners are sensitive to. Although deficits were observed in speech perception abilities in SSN and two-talker babble in terms of SRTs, HI listeners could utilize ITDs to a similar degree as NH listeners to facilitate the binaural unmasking of speech. A slight difference was observed between the group means when target and maskers were separated from each other by large ITDs, but not when separated by small ITDs. Therefore, HI listeners did not experience greater difficulties in terms of reduced BILDs when spatial differences between target and maskers were induced by small ITDs. Within the HI group, correlation analyses showed only a weak link between the magnitude of BILDs and the IPD_{fr} thresholds. The experiments also showed that BILDs elicited by small and large ITDs are similarly affected by hearing impairment, i.e. a reduction of the greatest BILD a listener can obtain is associated with a reduction in spatial tuning abilities of the listener.

Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

5

Contributions of low and high frequencies to binaural unmasking in hearing-impaired listeners

Abstract

This study investigated the contribution of interaural timing differences (ITDs) in different frequency regions to binaural unmasking (BU) of speech. Speech reception thresholds (SRTs) and binaural intelligibility level differences (BILDs) were measured in two-talker babble in 6 young normal-hearing (NH) and 9 elderly hearing-impaired (HI) listeners with normal or close-to-normal hearing at and below 1.5 kHz. Target sentences were presented diotically, embedded into a stream of diotic or dichotic maskers. Both target and masker sentences were split into frequency regions above and below 1.25 kHz. In the dichotic listening conditions, the maskers were lateralized to the left side by introducing 0.68-ms ITDs in either the low-frequency band, the high-frequency band, or both bands simultaneously. BILDs were found to be similar in both listener groups when the ITDs were imposed in the low-frequency domain only. ITDs in the high-frequency band alone did not produce any BILD in any of the groups. However, when ITDs were imposed in both frequency bands, the NH listeners yielded significantly greater BILDs than the HI listeners. The results suggest that, on a group level, HI listeners relied solely on ITDs in the low-frequency domain while NH listeners were able to utilize ENV ITDs above 1.25 kHz

to facilitate the BU of speech.

5.1 Introduction

In everyday life, we are surrounded by acoustically rich environments where speech and noise are often present simultaneously. While normal-hearing (NH) listeners can follow the talker of their choice almost effortlessly in a conversation between multiple peers, speech understanding is challenging for most HI listeners in noisy acoustic environments. In complex listening scenarios, spatial hearing helps to perceptually separate the target talker from the surrounding interferers (maskers) based on the acoustic cues associated with each stream in the auditory scene. In the horizontal plane, these cues mainly consist of interaural level differences (ILDs) and interaural timing differences (ITD). As the spatial separation between the target and the masker increases, spatial release from masking (SRM) occurs, i.e. speech reception thresholds (SRTs) drop corresponding to an increased speech intelligibility (SI) performance. Both ILDs and ITDs contribute to this SRM benefit, the first through a phenomenon called better-ear listening, the second through binaural unmasking (BU). The benefit in SRM from BU only, measured in dB, will be referred here to as binaural intelligibility level difference (BILD).

It is well established that listeners are sensitive to spatial information carried by ITD cues both in the temporal fine-structure (TFS) and in the envelope (ENV) of the acoustic stimuli (e.g. Henning and Ashton, 1981; Bernstein and Trahiotis, 1985; Lacher-Fougère and Demany, 2005). For broadband stimuli, low-frequency TFS cues have been shown to be more dominant than high-frequency ENV cues for localization (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). The dominant role of TFS in localization is further supported by the findings of Smith et al. (2002), who manipulated independently the TFS and ENV ITD information of synthetic speech stimuli and found that the perceived stimulus laterality depended mainly on TFS ITDs. Similarly, studies investigating the BU of speech in noise have typically found that BILDs are determined by ITDs in the low-frequency domain (Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and

Culling, 2005), suggesting that TFS ITDs carry critical information for the BU of speech as well.

Several studies have shown that both aging and hearing impairment negatively affect binaural temporal coding abilities, reducing the sensitivity to detect interaural temporal differences both in the TFS and ENV of ongoing stimuli (Lacher-Fougère and Demany, 2005; Hopkins and Moore, 2011; Santurette and Dau, 2012; King et al., 2014). A loss of sensitivity to the acoustic cues that are thought to be important for the facilitation of BU may also negatively affect BILDs. Indeed, Goverts and Houtgast (2010), who used a distortion sensitivity approach to investigate the relationship between supra-threshold deficits and BILDs in HI listeners, found that listeners with reduced BILDs were less sensitive to dichotic temporal and phase distortions in the test signals than those with normal BILDs. However, other studies (Neher et al., 2011; Neher et al., 2012; Santurette and Dau, 2012) showed only a moderate or no correlation between SI scores in spatial settings and behavioural measures of TFS ITD sensitivity in HI listeners. Similarly, the results from Chapter 4 of this thesis indicated only a weak link between the sensitivity to binaural TFS information and BILDs in speech-in-noise tests.

A possible explanation for the weak relationship between TFS ITD sensitivity and BILDs in various conditions can be that HI listeners utilize delays in the ENV of the stimuli that contribute to BILDs. Edmonds and Culling (2005) investigated how ITDs in isolated frequency bands contributed to BILDs in young NH listeners. Their results indicated that ITDs in the frequency regions both below and above 1.5 kHz provided some masking release, but also that the full advantage was only achieved when ITDs were present over the full spectrum. Therefore, it appears that NH listeners can also exploit ENV ITDs at higher frequencies to aid speech perception. If this was the case for the HI listeners tested in the previous chapters, this could explain why TFS ITDs were of only limited use when predicting BILDs. It is possible that the degree to which the HI listeners relied on TFS and ENV ITD cues was altered compared to the NH listeners. Recent studies have suggested that sensorineural hearing loss may lead to an enhancement of envelope coding at peripheral stages, due to e.g. broader auditory filters or reduced cochlear compression (e.g. Henry et al., 2014; Bianchi et al., 2015). These physiological changes could increase the

relative importance of ENV cues compared to TFS cues in the facilitation of BILDs. Alternatively, older listeners may rely less on ENV ITD cues at high-frequencies than young listeners, due to a general age-related decline in binaural temporal processing abilities that affects both TFS and ENV processing (He et al., 2008; King et al., 2014).

To the best knowledge of the author, it has not yet been investigated whether older HI listeners can utilize high-frequency ENV ITDs for the unmasking of speech in noise in a similar way as young NH listeners. In the current study, the contribution of TFS and ENV ITDs in different frequency regions to BILD was evaluated in young NH and older HI listeners. BILDs were measured in a speech-on-speech task. The target and the interferers were divided into two independent low- and high-frequency regions, and ITDs were imposed on the interferers in the low, high, or both frequency domains.

5.2 Methods

5.2.1 Listeners

6 young NH (mean: 24.2, standard deviation (SD): 2.2) and 9 older HI (mean: 69.6, SD: 5.5) participated in the study. All of the NH listeners and 7 of the HI listeners also participated in the study described in the previous chapter. For each listener, hearing threshold levels (HTLs) were measured at octave frequencies between 125 Hz and 8 kHz and between 750 Hz and 6 kHz. The NH listeners had normal audiometric thresholds (i.e. ≤ 20 dB HL) at the measured audiometric frequencies. Most of the HI listeners had normal hearing below 1.5 kHz, but a mild-to-moderate hearing loss at frequencies above 1.5 kHz. In all listeners, the hearing thresholds between the ears differed by at most 15 dB at each tested audiometric frequency. The average hearing thresholds for the HI listeners are displayed in Table 5.1. All listeners provided written consent and received compensation for their efforts. All but one listener were tested over a single visit lasting between 2 and 3 hours. One NH listener was tested over two visits.

#	Sex	Age	Audiometric thresholds averaged between the ears [dB HL]										Pure-tone averages			
			125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k	PTA _{low}	PTA _{high}	PTA _{oct}
a	f	58	5	0	2.5	2.5	2.5	-5	7.5	17.5	30	35	17.5	.5	21.5	8.5
b	f	66	12.5	7.5	7.5	10	10	5	10	10	25	25	40	8	22	12
c	f	75	22.5	15	10	7.5	5	5	15	15	22.5	17.5	50	8.5	24	13.5
d	m	68	20	12.5	10	7.5	7.5	5	5	22.5	32.5	37.5	67.5*	8.5	33	13.5
e	f	72	10	5	10	7.5	5	15	22.5*	40	30	45	72.5	8.5	42	14.5
f	f	67	17.5*	10	12.5	7.5	7.5	12.5*	35	50	55	52.5	50	10	48.5	24
g	f	72	22.5	12.5	20	15	20	22.5*	22.5*	27.5	27.5	30	52.5	18	32	20.5
h	m	76	15	12.5	20	20	15	25	25	45	57.5	52.5	67.5	18.5	49.5	26
i	f	72	17.5	25	22.5	20	20	22.5	27.5*	37.5*	40	52.5*	65	22	44.5	27

Table 5.1: Data of the HI listeners, including (from left to right): listener ID, gender (f: female; m: male), age, average audiometric thresholds across both ears measured with air-conduction, and pure-tone averages (averages of the audiometric thresholds between 250 Hz and 1.5 kHz [PTA_{low}], above 1.5 kHz [PTA_{high}], or at octave frequencies between 250 Hz and 4 kHz [PTA_{oct}]). Differences in audiometric thresholds across the left and right ears were not greater than 10 dB, except for the cases marked by asterisks.

5.2.2 Binaural temporal fine structure coding

The listeners' sensitivity to binaural TFS information was assessed by measuring the upper frequency limit at which listeners were able to detect an interaural phase shift of 180° (IPD_{fr}). A modified version of the corresponding experiment described in Chapter 4 was used for this purpose. The task was the same 3-interval 3-alternative forced-choice paradigm, combined with a multiplicative weighted up-down tracking procedure (Kaernbach, 1991) that estimated the 70.7% point on the psychometric function. Reference and target intervals were presented at 30 dB sensation level (SL) and contained 4 tone bursts presented diotically or alternated between the diotic and dichotic presentation modes. The tone burst were 300 ms long, gated with 50 ms raised-cosine ramps and separated by 100 ms silent gaps. The intervals were separated by 400 ms silent gaps. The parameters of the up-down procedure were the same as described in Chapter 4. Six thresholds were evaluated for each listener, and the final threshold was calculated as the geometric mean of the last 3 thresholds.

5.2.3 Speech intelligibility tests

Similar to the speech experiments described in Chapters 3 and 4, speech intelligibility was assessed using the open-set DAT corpus (Nielsen et al., 2014). The

“Dagmar” sentences were presented as target material against a two-talker masker (TT), which consisted of sentence pairs spoken by the two other talkers of the same corpus. As the three talkers have similar voice characteristics, no spectral matching was applied between the target and the maskers.

The start of each sentence was preceded by a 100-ms 1-kHz pure tone. The target and masker sentences started synchronously 500 ms after this warning tone, and were gated by 50-ms raised-cosine ramps. In contrast to the standard DAT procedure applied in Nielsen et al. (2014), the maskers were not time-stretched to match the sentence endings of the targets. The target sentences were always presented diotically. In the reference condition, the maskers were presented diotically, i.e. colocalized with the target (TT_{co}). In the remaining conditions, the maskers were lateralized towards the right side by imposing a 0.68 ms timing delay in the left channel. This delay was either imposed on the full spectrum (TT_{bb} for “broad band”), the low spectral region (TT_{lp}, “low pass”), or the high spectral region (TT_{hp}, “high pass”) of the maskers. In order to manipulate the ITD relations in the low- and high-frequency parts of the maskers independently, low-pass and high-pass filtered version of the original stimuli were created for both headphone channels prior to presentation, and the time delays were applied to the left-ear channel in the corresponding frequency regions. The resulting low- and high frequency time signals were then added in each channel and presented to the listener. In the case of the target sentences, no time delays were applied. The frequency bands were created using 2048 order finite impulse response filters. The cutoff frequencies of the low-pass and high-pass filters were set to 1173 Hz and 1332 Hz, respectively, corresponding to a 1 equivalent rectangular band (ERB) notch centered at 1.25 kHz between the low-pass filtered and high-pass filtered parts. Filter slopes were greater than 500 dB/oct in both cases in order to prevent any interactions between the two spectral regions.

In each condition, sentence correct SRTs were measured by varying the level of the maskers in 2 dB steps. The initial signal-to-noise (SNR) ratio was set to 3 dB in the TT_{co} and to 0 dB in all the other conditions. When calculating the SRTs for each list, the presentation levels of the maskers from the 5th to the hypothetical 21st sentence were averaged and subtracted from the presentation level of the target.

SRTs were measured over 3 lists in each condition and the final SRT value for each condition was calculated as the average of these, expressed in SNR. Overall, 12 lists were used in the testing phase, and 3 additional lists in the training phase, two of which were presented in the TT_{co} condition and one in the TT_{bb} condition. As the DAT corpus only includes 10 test and 3 training lists, two of the original training lists (List 1 and 2) were used as test lists as well. List 3 from the training lists was used for training, which was complemented by two more custom-made lists. These lists were generated by taking the first 4 sentences from the test lists, and organizing them into two lists of 20 sentences each. Thus, the first 4 sentences in each test block were presented to the listeners during the training phase. Both the used target lists and the order of the different conditions were balanced as much as possible within the NH and HI listener groups using Latin square designs. The masker tokens were paired according to their list and sentence number, and in each trial a random masker-pair was chosen.

The frequency range of the stimuli was restricted to between 100 Hz and 10 kHz. A 512 order finite impulse response filter was used to compensate for the frequency response of the electro-acoustic equipment at the eardrums of a Brüel & Kjær 4180 Head and Torso Simulator (HATS), and to simulate the frequency response of the outer ear in a diffuse-field listening scenario (Moore et al., 2008). This filter also compensated for the loss of stimulus audibility based on the hearing thresholds of the individuals and the long-term average spectrum of the target speech. The same audibility criterion was used as in the previous chapter: 15 dB at and below 3 kHz, which was reduced to 12 dB, 8 dB and 0 dB at 4, 6 and at 8 kHz and above. The target sentences were presented at a nominal level of 65 dB sound pressure level (SPL) “free field”. The presentation levels were limited to at most 94 dBA. If the estimated level of a trial exceeded this level, it was scaled down in 2 dB steps to be below this upper limit.

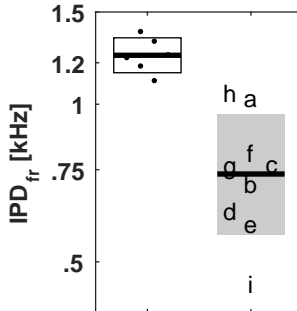


Figure 5.1: The upper frequency limit for detecting a 180° interaural phase difference in the TFS of a pure tone. The dark horizontal lines with the white and gray boxes stand for the mean and ± 1 SD of the NH and HI listener groups, respectively. Dots and letters denote individual thresholds within the corresponding groups.

5.3 Results

While the HI listeners had audiometric thresholds below 20 dB HL at most of the tested frequencies below 1.5 kHz, the group difference in PTA_{low} , calculated as the average of HTLs measured between 250 Hz and 1.5 kHz, was statistically significant (independent t-test, $t(13) = 2.44, p = 0.030$). Thus, when comparing the results between the listener groups, an effect of hearing loss cannot be excluded.

The results of the IPD_{fr} experiments are shown in Figure 5.1. Horizontal black lines denote the group means and the white and gray boxes indicate ± 1 SD of the NH and HI listener groups, respectively. For analysis purposes, the thresholds were log-transformed, as the distribution of IPD detection thresholds is typically log-normal (see e.g. Lacher-Fougère and Demany, 2005; Strelcyk and Dau, 2009). There was a clear difference between groups: the IPD_{fr} thresholds of the NH listeners averaged around 1240 Hz while HI listeners performed consistently worse, with a group average of 736 Hz. This difference was statistically significant ($t(13) = -4.65, p < 0.001$). The lowest and highest thresholds in the HI group were 456 Hz and 1056 Hz, showing a large spread of how the individual listeners performed. Within the HI group, neither age nor PTA_{low} was correlated with the IPD_{fr} thresholds. The group means of the IPD_{fr} thresholds found in this study were the same as those reported in Chapter 4, as confirmed by independent t-tests ($p > 0.05$). The differences in the experimental paradigms applied in the two studies did not affect the outcomes of the tests, as paired t-tests conducted on the results obtained in the two experiments for those listeners who participated in both experiments were

not significant ($p > 0.05$ for both groups). The obtained results are consistent with earlier studies (Ross et al., 2007a; Neher et al., 2011; Lőcsei et al., 2015). The HI listeners showed similar binaural TFS processing abilities as the HI listeners tested in Chapter 4.

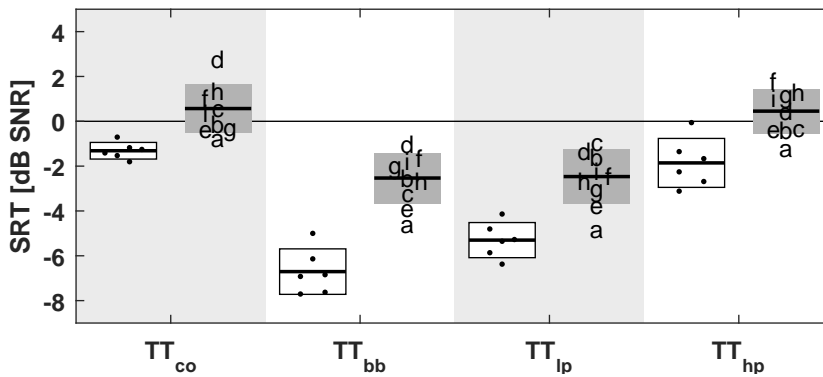


Figure 5.2: Speech reception thresholds of the NH and HI listeners in the speech intelligibility tests. The target was always presented diotically. The maskers were presented either diotically (TT_{co}), or lateralized to the right side in the low or high frequency domains (TT_{lp} or TT_{hp}), or over the full frequency domain (TT_{bb}). The dark horizontal lines with the white and gray boxes stand for the mean and ± 1 SD of the NH and HI listener groups, respectively. Dots and letters denote individual thresholds within the corresponding groups.

Figure 5.2 shows the group means and standard deviations of the SRTs for the two listener groups (NH: white boxes, HI: gray boxes) in the SI experiments. SRTs in the TT_{bb} and TT_{lp} conditions were lower than in the TT_{co} condition, indicating a masking release when ITDs were imposed at least on the low-frequency part of the maskers. In contrast, SRTs were similar in the TT_{co} and TT_{hp} conditions for both listener groups. While for the NH listeners SRTs were slightly higher in the TT_{lp} condition than in the TT_{bb} condition, HI listeners performed similarly in the two conditions.

The BILDs were calculated as the difference between SRTs in the TT_{co} and all the other conditions. The group means and standard deviations are shown in Figure 5.3. Group differences in BILDs were highest in the TT_{bb} condition, amounting to about 2 dB, which is slightly larger than what was reported in the earlier chapters. Group differences were less pronounced in the TT_{lp} condition, as

the magnitude of BILDs was decreased in the NH group when the contribution of ITDs in the high-frequency band was removed. In contrast, the HI listeners had similar BILDs in the TT_{bb} and TT_{lp} conditions.

A mixed-design ANOVA was conducted on the BILDs obtained in the TT_{bb} and TT_{lp} conditions, with *filtering* as within-subject and *listener group* as between-subject factors. The analysis revealed a significant main effect of *filtering* ($F(1, 13) = 5.647, p = 0.034$), *listener group* ($F(1, 13) = 14.95, p = 0.002$), and interaction between *filtering* and *listener group* ($F(1, 13) = 4.69, p = 0.0496$). For the NH listeners, there was a tendency towards greater BILDs in the TT_{bb} than in the TT_{lp} condition, even though the difference in group means was not significant (paired t-test, $t(5) = 2.44, p = 0.058$). The presence of ITDs in the high-passed band alone was not sufficient to produce BILDs in either of the listener groups, as group means were not significantly different from 0 in the TT_{hp} condition, as confirmed by two-tailed one-sample t-tests ($p > 0.05$).

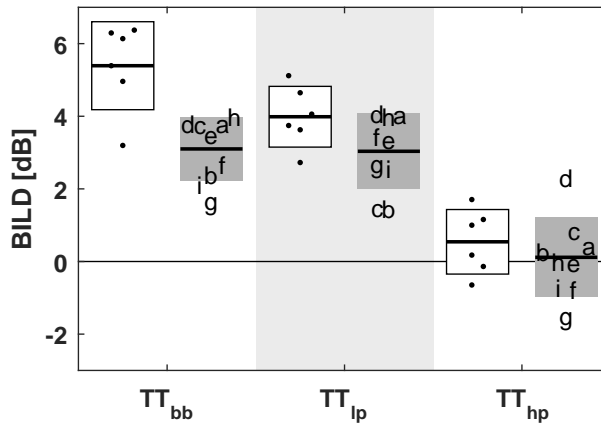


Figure 5.3: Binaural intelligibility level differences of the NH and HI listeners obtained in the speech intelligibility tests. Condition notations are the same as in Figure 5.2. The dark horizontal lines with the white/gray boxes show mean BILDs and ± 1 SD for the NH and HI listener groups, respectively.

5.4 Discussion

The results of the present study indicated that, for young NH listeners, frequencies above 1.25 kHz can also contribute to the BU of speech, which is consistent with the findings of Edmonds and Culling (2005). Furthermore, it was found that young NH listeners exhibited larger BILDs than older HI listeners when ITDs were imposed on the whole frequency range. Both listener groups benefitted from BU when the target and the maskers were separated by ITDs only below 1.25 kHz, and the magnitude of the BILDs was comparable in the two groups. When ITDs were imposed above, but not below, 1.25 kHz, no BILD was observed. The results suggest that, both in young NH and elderly HI listeners, BILDs are mainly facilitated in the low-frequency region of the stimuli. This finding is consistent with the conclusions of earlier reports investigating BU (e.g. Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005). The contributions of ITDs at high frequencies to BILDs seem to be negligible when presented in isolation. However, in contrast to the HI listeners, NH listeners could utilize high-frequency ITD information to some degree to aid speech understanding in the TT_{bb} condition. For the NH listeners, the difference in BILDs between the TT_{bb} and TT_{lp} did not reach significance, likely due to the relatively low number of listeners tested.

As the splitting frequency between the low- and high-frequency speech bands was set to 1.25 kHz, it is possible that the NH listeners with the highest IPD_{fr} thresholds had some limited access to TFS information in the high-frequency band. This could explain why the group differences in BILDs were greater in the TT_{bb} than in the TT_{lp} condition. Edmonds and Culling (2005) utilized a similar paradigm in the presence of brown noise or a single interfering talker, separating the low- and high-frequency bands at either 750 Hz or 1.5 kHz. Their results showed that, for young NH listeners, changing the splitting frequency did not affect BILDs elicited by the low-frequency band or by both bands. Since they tested young NH listeners, it is likely that the listeners' access to TFS cues was drastically reduced when the cut-off frequency was lowered from 1.5 kHz to 750 Hz; yet, the BILDs in these two lateralized conditions remained similar. Therefore, it is unlikely that the differences in BILDs between the NH and the HI groups were driven by the NH listeners' ability

to utilize TFS ITDs above 1.25 kHz.

There are several possibilities why the HI listeners, compared to the NH group, showed greater deficits in BILDs in the TT_{bb} than in the TT_{lp} condition. First, aging has been associated with a general reduction in temporal coding abilities, degrading TFS and ENV processing simultaneously (He et al., 2008; King et al., 2014). In terms of ENV processing, aging has also been shown to affect performance both in monaural tasks, like gap detection or amplitude modulation detection (e.g. Strouse et al., 1998; He et al., 2008), and in binaural tasks like interaural phase discrimination (King et al., 2014). Therefore, it is possible that, besides their impoverished binaural TFS coding ability, the older HI listeners were less sensitive to ENV ITDs than the young NH listeners, rendering the relatively small contribution of ITDs at high-frequencies ineffective. In contrast, the reduced binaural TFS coding abilities might still have allowed for a reasonable amount of binaural information to facilitate BILDs both in the TT_{bb} and TT_{lp} conditions. As sensitivity to binaural temporal cues at higher frequencies was not measured in the current study, it is unclear whether the older HI listeners indeed had a reduced sensitivity to ENV ITDs. Second, the reduced sensation level at which the HI listeners received the stimuli could also have affected the contribution of ENV ITDs in the BILDs. Even though elevated hearing thresholds are not necessarily related to greater-than-normal ENV ITD detection thresholds when stimuli are presented at a fixed sensation level (King et al., 2014), thresholds tend to worsen with decreasing SL even for NH listeners (see e.g. Lacher-Fougère and Demany, 2005). In the current study, stimulus audibility was controlled by compensating for elevated hearing thresholds. Nevertheless, the HI listeners generally received the speech stimuli at lower sensation levels than the NH listeners, especially at higher frequencies where the audibility criterion was gradually reduced. Thus, it is possible that, for the HI listeners, stimulus audibility was not sufficient to contribute to BILDs. Finally, a combination of both reduced temporal processing abilities and reduced stimulus audibility is also possible. In any case, the data demonstrate that, in contrast to their NH peers, the HI listeners could not utilize ITD cues above 1.25 kHz to facilitate BILDs.

The reason why the sensitivity to TFS IPDs using pure tones was only a weak

predictor of the amount of BU of speech in the previous chapters thus remains unresolved. Three possibilities seem reasonable to consider. First, the evaluation of TFS coding abilities assessed at multiple frequencies may be necessary to successfully predict the amount of BILDs in speech-in-noise tasks, as the BU of speech is facilitated over a broad frequency range, possibly operating independently within peripheral channels (Culling and Summerfield, 1995; Edmonds and Culling, 2005). Second, even in the frequency range where TFS ITDs are accessible, listeners are also sensitive to ENV ITDs or onset delays (Zurek, 1993). Even though TFS ITDs clearly dominate ENV ITDs at these frequencies, ENV ITDs might still affect the degree of perceived laterality in complex stimuli (see e.g. Bernstein and Trahiotis, 1985). Therefore, it cannot be excluded that ENV ITDs affected the performance in the SI tasks in some way. Finally, it is possible that other measures of binaural TFS coding abilities might be more indicative of BU effects, and thus represent better predictors of the BILDs individual listeners can reach in speech-in-noise tasks. Binaural masking level differences of pure tones in noise, or lateralization tasks where the extent of laterality is assessed instead of lateralization detection thresholds might represent possible candidates.

5.5 Summary and conclusion

The present study showed that BILDs were similar for a group of young NH and older HI listeners when elicited by ITDs below 1.25 kHz, but slightly lower for the HI group when triggered by ITDs over the full frequency range of the stimuli. When ITDs were imposed above 1.25 kHz only, no BILDs were found in any of the groups. Overall, the results suggest that, while the young NH listeners might have utilized both TFS ITDs at low frequencies and ENV ITDs at high frequencies to facilitate BU, older HI listeners relied exclusively on ITDs at the low frequencies. Therefore, the weak link between the binaural measures of TFS coding and the BILDs observed in the HI data from the previous chapters cannot be attributed to utilizing ENV ITDs at the higher frequencies to facilitate BU. It still remains possible that BILDs were affected by the sensitivity to ENV ITDs in the low frequency region, or that IPD

detection thresholds with pure-tones at single frequencies are not directly related to the processes underlying BU.

Acknowledgments

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

6

Overall discussion

This thesis presented a series of experiments studying the relationship between the auditory coding of supra-threshold stimuli and spatial speech perception. In particular, the relationship between lateralized speech perception and TFS coding in the normal and impaired auditory system was investigated. SRTs and BILDs were measured in a variety of background noise conditions in young NH and elderly HI listener groups. Also, the sensitivity to monaural and binaural TFS information in non-speech stimuli was assessed using frequency discrimination thresholds (FDTs) and IPD discrimination thresholds (IPDTs). The role of TFS cues on binaural unmasking (BU) was assessed by comparing the results of the speech intelligibility tests to the measures of TFS coding. Furthermore, it was explored how ITDs in the low- and high-frequency parts of the stimuli facilitate BU in NH and HI listeners.

6.1 Summary of main findings

Chapter 2 presented a study where SI and lateralization performance was measured for young NH listeners. Speech was presented in lateralized noise conditions with 2, 4, 8 or 12 interferers. The results of the speech tests indicated that, when all maskers were separated from the target side, BILDs increased as the number of maskers increased from 2 to 8. However, spatially separating the last 2 or 4 maskers from the target side yielded similar BILDs independent of the overall number of maskers in the scene. The findings demonstrated the robustness of BU to background noise: for a fixed number of maskers, the BILDs were independent of the underlying noise floor in the actual condition. Also, it was found that speech lateralization thresholds were good predictors of the SI performance of each listener.

Chapter 3 investigated the relationship between monaural and binaural TFS processing abilities in young NH and two groups of older HI listeners with normal hearing below 1.5 kHz but a mild or moderate hearing loss (HI_{mild} or HI_{mod}) at frequencies above 1.5 kHz. SRTs and BILDs were assessed in SSN, reversed babble with 8, 4 or 2 interferers and two-talker babble (TT) noise. Monaural and binaural TFS coding abilities were evaluated at 250 Hz by measuring FDTs and IPDTs. Regarding temporal processing, the HI groups performed worse than the NH group. BILDs were only slightly (1 dB) smaller than normal for both the HI_{mild} and HI_{mod} listeners. Temporal processing abilities were not correlated with the SI performance of the listeners, when considering either SRTs or BILDs, which was opposite to the hypothesis of the study and contrasted with some earlier studies (Strelcyk and Dau, 2009; Neher et al., 2012). It was concluded that performance in the SI tests might have been affected on an attentional level, as the presentation side of the target was randomized from trial-to-trial when estimating SRTs. It was also hypothesized that triggering BILDs by small ITDs may reveal greater between-group differences, as HI listeners may have limited access to such cues, which, in turn, may affect BU.

Although not mentioned in Chapter 3, the results contrasted with the findings of Chapter 2 in the sense that BILDs for a fixed number of colocated maskers *did* depend on the overall number of maskers in the scene. BILDs for separating the last 2 maskers from the target side were calculated in the reversed-speech conditions for each listener group. As the overall number of interferers dropped from 8 to 4 to 2, BILDs increased by about 2 dB on average. A mixed ANOVA conducted on the BILDs with *noise type* (R_8 , R_4 and R_2) as within-subject and *listener group* (NH, HI_{mild} and HI_{mod}) as between-subject factors revealed a significant main effect of noise type ($F(2, 52) = 9.66$, $p < 0.001$), but no main effect of *listener group* and no interaction. Similar results were found for separating the last 4 maskers in the 4- and 8-masker conditions. The findings of the two studies further contrasted in the sense that while BILDs increased with an increasing number of maskers in Chapter 2, they decreased in Chapter 3. Since the same masker materials were used in both experiments, the most likely explanation is that changes in the target material interacted with the pattern of BILDs.

Chapter 4 addressed this hypothesis by investigating how BILDs are affected by the amount of masker lateralization in young NH and older HI listeners. Also, the effect of TFS sensitivity in the two lateralized conditions was evaluated by comparing the obtained BILDs to measures of binaural TFS coding (IPD_{fr} and ITD_{min}). SRTs were measured both in SSN and in TT noise. To minimize the effect of attentional factors, the target was kept in the center of the head when estimating SRTs. Listeners were subjected to the same test material to further increase sensitivity to individual differences. Consistent with the findings of Chapter 3 and with earlier reports, the HI listeners showed a degraded binaural sensitivity to TFS both in terms of IPD_{fr} and ITD_{min} . Group differences in BILDs were generally small and the hypothesis that group differences in BILDs would be greatest when target and maskers were separated by small ITDs was not supported. Overall, BILDs were on average only slightly lower for the HI than for the NH listeners. Regarding TFS coding and BILDs in the HI group, IPD_{fr} thresholds were better predictors of BILDs than ITD_{min} thresholds, suggesting an important role of the frequency range over which TFS ITD were accessible to the listeners. Importantly, the correlations between BILDs and IPD_{fr} were of similar magnitude when the target and maskers were separated by small and large ITDs, indicating that reduced TFS sensitivity affected BU to a similar degree, independent of the magnitude of lateralization. It was concluded that, even though sensitivity to binaural TFS information correlates moderately with the listeners' ability to benefit from BU, IPD detection thresholds of pure tones at single frequencies can only explain a limited amount of the variance present in the BILD data. It was argued that the relatively weak link between these measures may reflect the limited potential of the IPD tests to characterize the most relevant aspects of the processing mechanisms leveraging the BU of speech, or that HI listeners may have partly relied on binaural cues in the envelope.

Chapter 5 investigated BILDs in TT noise in similar groups of NH and HI listeners as in Chapter 4. In order to assess how much ITDs in the low- and high-frequency domain contribute to BILDs, both the target and the maskers were split into two frequency regions at a splitting frequency of 1.25 kHz. Maskers were lateralized to the side in the low, in the high, or in both frequency regions. The experiments revealed that, for both listener groups, BILDs were mainly facilitated by

ITDs in the low-frequency parts of the stimuli (consistent with Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005) and that ITDs imposed on the high-frequencies only were insufficient for providing BU. BILDs were similar between the groups when elicited by ITDs in the low-frequencies only, but HI listeners had smaller BILDs than NH listeners when ITDs were imposed over the full frequency spectrum. The findings indicated that, while the NH listeners were also able to benefit to some degree from ITDs in the high-frequencies, older HI listeners relied solely on the low-frequency stimulus content to bring about BU. The HI listeners' reduced sensitivity may have originated from e.g. a general age-related reduction in temporal coding abilities (affecting both TFS and ENV coding), and/or from the reduced sensation level of the high-frequency stimulus content due to their high-frequency HL.

6.2 Implications for lateralized speech perception

The current study revealed a systematic difference in the amount of BU between young NH and elderly HI listeners. The moderate correlation between BILDs and binaural TFS coding abilities (Chapter 4), and the older HI listeners' inability to use the high-frequency content of the stimuli to facilitate the BU of speech (Chapter 5) suggest that the HI listeners were largely relying on binaural TFS cues. As the difference in BILDs between the NH and HI listeners remained relatively small, it appears that the HI listeners sufficiently retained their sensitivity to binaural temporal information at low-frequencies to aid the BU of speech, at least in the current experimental setups. Interaural temporal cues at low frequencies might be especially useful for facilitating SRM in listening scenarios where interaural level differences are reduced. This might be the case, e.g., in the presence of multiple interferers positioned in both hemifields of the listeners (e.g. Hawley et al., 2004; Culling et al., 2004). The current findings underline the importance of intact interaural timing information at low frequencies. Providing such information might be crucial for wearers of hearing assistive devices, as they may be able exploit these cues to facilitate BU and therefore to aid SI.

In Chapter 4, binaural TFS coding abilities showed a moderate correlation with BILDs, which was independent of the magnitude of ITDs applied to spatially separate the maskers from the target. BILDs elicited by small ITDs did not show a greater correlation with IPD detection thresholds (IPDTs) than when elicited by large ITDs, even though the performance in the small-ITD condition was expected to relate more to the listeners' absolute sensitivity to detect subtle timing differences between the ears. The findings of Chapter 4 also suggested that a reduction in the frequency range at which listeners were sensitive to binaural TFS information was a better predictor of BILDs than best ITD sensitivity within a narrow frequency range. Even though binaural TFS information might be essential both for the detection of IPDs and the facilitation of BILDs, the moderate correlation between the two tasks might originate from the involvement of partly different processing mechanisms. This might include processing differences driven by the fundamentally different acoustic structure of the stimuli (e.g. narrow-band vs broadband stimuli), but also a greater involvement of cognitive processes in the SI tests.

6.3 Limitations and perspectives

The current study tested elderly HI listeners differing from the control group both in terms of their age and HL. As both aging and HL can negatively affect TFS processing and also SI performance, it was not possible to directly assess the independent contributions of these underlying factors. Also, the current study only tested listeners with symmetrical hearing loss and normal or close-to-normal hearing at low frequencies. Sensorineural hearing loss in the low-frequency domain and aging (without any hearing loss) might both lead to reduced TFS coding abilities, which might potentially originate from different physiological changes in the auditory pathway (Moore, 2014). Studying such subgroups of listeners might reveal how different underlying processing deficits affect TFS coding abilities and their relationship to speech perception.

A major question remaining for future research is how well the findings of this thesis generalize to more realistic listening scenarios. In the current headphone

experiments, identical target and/or masker waveforms were delayed between the ears of the listeners. In contrast, acoustic waveforms can differ greatly between the ears in realistic situations, due to e.g. interaural decorrelation caused by reverberation, leading to more ambiguous interaural cues (Licklider, 1948; Lavandier and Culling, 2008). It is possible that, despite having close-to-normal BILDs in headphone experiments, elderly and HI listeners would suffer from greater difficulties in scenarios where the interaural cues are degraded. In fact, in a recent study, Ruggles et al. (2012) demonstrated that, compared to young listeners, middle-aged listeners tend to rely to a greater degree on TFS ITDs than on ENV ITDs in spatial listening tasks. They also showed that reverberation is particularly detrimental for the interaural coherence of the TFS at low-frequencies. This might expose middle-aged and elderly listeners to greater difficulties in reverberant environments. Therefore, in future studies, it would be important to test how reduced temporal coding abilities affect the listeners' ability to cope with interaural stimulus perturbations in terms of BU and spatial attention.

In Chapter 4, it was argued that measuring BILDs elicited by large ITDs may provide clear interaural cues even for listeners with reduced sensitivity to ITDs. Applying small ITDs, and therefore small perceptual separation between the target and the maskers, was assumed to reveal differences between listener groups with normal and degraded TFS processing abilities. However, no such effect was observed. A limitation of this approach was that BILDs were relatively small even in the conditions with the large ITDs. The smaller separation between target and maskers resulted in even smaller BILDs, such that any group differences might be difficult to detect. A possibility to remedy this issue was mentioned in Chapter 4; using identical speakers for the target and the maskers would likely raise the amount of BILDs due to an increased release from informational masking (Kidd et al., 2008). Similar BILDs might not necessarily imply the absence of performance differences between listener groups. Even if the SRT benefit due to the spatial separation of the target and maskers is similar between listener groups, it is possible that the listeners with degraded temporal processing abilities require greater mental effort to sustain close-to-normal BILDs. Degradations in the peripheral encoding of acoustic stimuli, e.g., due to a hearing impairment, will also

negatively affect the perceptual features associated with the stimuli (such as e.g. spatial location). Such perceptual cues can be used to segregate competing sound sources (Shinn-Cunningham and Best, 2008), therefore aiding speech perception. Effortful listening due to signal degradation poses more cognitive demands on the listener (Pichora-Fuller and Singh, 2006). Therefore, another way to assess how spatial separation affects speech perception could be to monitor how cognitive involvement changes in different spatial settings.

In the current thesis, binaural TFS sensitivity was assessed by measuring IPDTs in low-frequency pure tones, which correlated only moderately with BILDs. As BILDs are facilitated over a broad frequency range (Levitt and Rabiner, 1967), estimates of TFS sensitivity over multiple frequencies might serve as better predictors for BILDs. Furthermore, IPDTs assess the smallest timing differences in the TFS of the stimuli listeners can detect, but might not be representative for the extent of laterality such timing differences can provide, which might be a more important feature when considering BILDs. Binaural masking level differences of tones in noise could also be used as estimators of binaural TFS coding abilities at different frequencies (Hall et al., 1984; Moore, 2014). BMLDs are especially appealing as they assess directly how much SNR increment the auditory system can provide when the signal and the noise are presented in different binaural modes.

Bibliography

- Arbogast, T. L., Mason, C. R., and Kidd, G. (2002). “The effect of spatial separation on informational and energetic masking of speech”. In: *The Journal of the Acoustical Society of America* 112.5, pp. 2086–2098.
- Bernstein, L. R. and Trahiotis, C. (1985). “Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities”. In: *The Journal of the Acoustical Society of America* 77.5, pp. 1868–1880.
- Bernstein, L. R. and Trahiotis, C. (2009). “How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli.” In: *The Journal of the Acoustical Society of America* 125.5, pp. 3234–3242.
- Best, V., Carlile, S., Kopco, N., and Schaik, A. van (2011). “Localization in speech mixtures by listeners with hearing loss.” In: *The Journal of the Acoustical Society of America* 129.5, E210–E215.
- Best, V., Thompson, E. R., Mason, C. R., and Kidd, G. (2013). “An energetic limit on spatial release from masking.” In: *Journal of the Association for Research in Otolaryngology* 14.4, pp. 603–610.
- Bianchi, F., Fereczkowski, M., Zaar, J., Santurette, S., and Dau, T. (2015). “Effects of cochlear compression and frequency selectivity on pitch discrimination of complex tones with unresolved harmonics”. In: *Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*. Ed. by Santurette, S., Dau, T., Dalsgaard, J. C., Tranebjærg, L., and Andersen, T. Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 51–58.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

- Bronkhorst, A. W. and Plomp, R. (1988). "The effect of head-induced interaural time and level differences on speech intelligibility in noise". In: *The Journal of the Acoustical Society of America* 83.4, pp. 1508–1516.
- Bronkhorst, A. W. and Plomp, R. (1989). "Binaural speech intelligibility in noise for hearing-impaired listeners." In: *The Journal of the Acoustical Society of America* 86.4, pp. 1374–1383.
- Bronkhorst, A. W. and Plomp, R. (1992). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing." In: *The Journal of the Acoustical Society of America* 92.6, pp. 3132–3139.
- Bronkhorst, A. W. (2000). "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions". In: *Acta Acustica united with Acustica* 86.1, pp. 117–128.
- Brughera, A., Dunai, L., and Hartmann, W. M. (2013). "Human interaural time difference thresholds for sine tones: the high-frequency limit." In: *The Journal of the Acoustical Society of America* 133.5, pp. 2839–2855.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers". In: *The Journal of the Acoustical Society of America* 109.3, pp. 1101–1109.
- Carhart, R., Tillman, T. W., and Kenneth, R. J. (1967). "Release of Masking for Speech through Interaural Time Delay". In: *The Journal of the Acoustical Society of America* 42.1, pp. 124–138.
- Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears". In: *Journal of the Acoustical Society of America* 25.5, pp. 975–979.
- Christiansen, C. and Dau, T. (2012). "Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise." In: *The Journal of the Acoustical Society of America* 132.3, pp. 1655–1666.
- Cooke, M., Barker, J., Cunningham, S., and Shao, X. (2006). "An audio-visual corpus for speech perception and automatic speech recognition." In: *The Journal of the Acoustical Society of America* 120.5, pp. 2421–2424.
- Culling, J. F. and Summerfield, Q. (1995). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay". In: *The Journal of the Acoustical Society of America* 98.2, pp. 785–797.

- Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (2004). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources". In: *The Journal of the Acoustical Society of America* 116.2, pp. 1057–1065.
- Daneman, M. and Carpenter, P. A. (1980). "Individual differences in working memory and reading". In: *Journal of Verbal Learning and Verbal Behavior* 19, pp. 450–466.
- Dreschler, W. A. and Plomp, R. (1985). "Relation between psychophysical data and speech perception for hearing-impaired subjects. II." In: *The Journal of the Acoustical Society of America* 78.4, pp. 1261–1270.
- Edmonds, B. A. and Culling, J. F. (2005). "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay". In: *The Journal of the Acoustical Society of America* 117.5, pp. 3069–3078.
- Ernst, S. M. A. and Moore, B. C. J. (2012). "The role of time and place cues in the detection of frequency modulation by hearing-impaired listeners". In: *The Journal of the Acoustical Society of America* 131.6, pp. 4722–4731.
- Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (1957). "Localization of High-Frequency Tones". In: *The Journal of the Acoustical Society of America* 29.9, pp. 988–991.
- Festen, J. M. and Plomp, R. (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.
- Freyman, R. L., Griffin, A. M., and Oxenham, A. J. (2012). "Intelligibility of whispered speech in stationary and modulated noise maskers." In: *The Journal of the Acoustical Society of America* 132.4, pp. 2514–2523.
- George, E. L. J., Festen, J. M., Goverts, S. T., and Theo Goverts, S. (2012). "Effects of reverberation and masker fluctuations on binaural unmasking of speech." In: *The Journal of the Acoustical Society of America* 132.3, pp. 1581–1591.
- Glasberg, B. R. and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments." In: *The Journal of the Acoustical Society of America* 79.4, pp. 1020–1033.

- Glasberg, B. R. and Moore, B. (1989). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech." In: *Scandinavian Audiology. Supplementum* 32, pp. 1–25.
- Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., and Hickson, L. (2013). "The importance of interaural time differences and level differences in spatial release from masking." In: *The Journal of the Acoustical Society of America* 134.2, EL147–EL152.
- Goverts, S. T. and Houtgast, T. (2010). "The binaural intelligibility level difference in hearing-impaired listeners: the role of supra-threshold deficits." In: *The Journal of the Acoustical Society of America* 127.5, pp. 3073–3084.
- Goverts, S. T., Delreux, M., Festen, J. M., and Houtgast, T. (2007). "The influence of masker type on the binaural intelligibility level difference". In: *Hearing—From Sensory Processing to Perception*. Springer, pp. 551–558.
- Hall, J. W., Tyler, R. S., and Fernandes, M. A. (1984). "Factors Influencing the Masking Level Difference in Cochlear Hearing-Impaired and Normal-Hearing Listeners". In: *Journal of Speech and Hearing Research* 27, pp. 145–154.
- Hartmann, W. M., Dunai, L., and Qu, T. (2013). "Interaural Time Difference Thresholds as a Function of Frequency". In: *Basic Aspects of Hearing: Physiology and Perception*. Ed. by Moore, B. C. J., Patterson, D. R., Winter, M. I., Carlyon, P. R., and Gockel, E. H. New York, NY: Springer New York, pp. 239–246.
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer". In: *The Journal of the Acoustical Society of America* 115.2, pp. 833–843.
- He, N.-j., Mills, J. H., Ahlstrom, J. B., and Dubno, J. R. (2008). "Age-related differences in the temporal modulation transfer function with pure-tone carriers". In: *The Journal of the Acoustical Society of America* 124.6, pp. 3841–3849.
- Henning, G. B. and Ashton, J. (1981). "The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay". In: *Hearing Research* 4.2, pp. 185–194.

- Henry, K. S., Kale, S., and Heinz, M. G. (2014). "Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers." In: *Frontiers in systems neuroscience* 8.20, pp. 1–10.
- Hopkins, K. and Moore, B. C. J. (2010). "Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies." In: *International journal of audiology* 49, pp. 940–946.
- Hopkins, K. and Moore, B. C. J. (2011). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise." In: *The Journal of the Acoustical Society of America* 130.1, pp. 334–349.
- Hopkins, K., Moore, B. C., and Stone, M. A. (2008). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech". In: *The Journal of the Acoustical Society of America* 123.2, pp. 1140–1153.
- Houtgast, T. (1989). "Frequency selectivity in amplitude-modulation detection". In: *The Journal of the Acoustical Society of America* 85.4, pp. 1676–1680.
- Jerger, J., Brown, D., and Smith, S. (1984). "Effect of Peripheral Hearing Loss on the Masking Level Difference". In: *Archives of Otolaryngology* 110.5, pp. 290–296.
- Kaernbach, C. (1991). "Simple adaptive testing with the weighted up-down method." In: *Perception & psychophysics* 49.3, pp. 227–229.
- Kidd, G., Mason, C. R., Best, V., and Marrone, N. (2010). "Stimulus factors influencing spatial release from speech-on-speech masking." In: *The Journal of the Acoustical Society of America* 128.4, pp. 1965–1978.
- Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). "Reducing informational masking by sound segregation". In: *The Journal of the Acoustical Society of America* 95.6, pp. 3475–3480.
- Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking". In: *Auditory perception of sound sources*. Ed. by Yost, W. A., Popper, A. N., and Fay, R. R. Springer US, pp. 143–189.
- King, A., Hopkins, K., and Plack, C. J. (2014). "The effects of age and hearing loss on interaural phase difference discrimination." In: *The Journal of the Acoustical Society of America* 135.1, pp. 342–351.

- Lacher-Fougère, S. and Demany, L. (1998). "Modulation detection by normal and hearing-impaired listeners". In: *Audiology* 37.2, pp. 109–121.
- Lacher-Fougère, S. and Demany, L. (2005). "Consequences of cochlear damage for the detection of interaural phase differences." In: *The Journal of the Acoustical Society of America* 118.4, pp. 2519–2526.
- Lavandier, M. and Culling, J. F. (2008). "Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer". In: *The Journal of the Acoustical Society of America* 123.4, pp. 2237–2248.
- Levitt, H. and Rabiner, L. R. (1967). "Predicting binaural gain in intelligibility and release from masking for speech." In: *The Journal of the Acoustical Society of America* 42.3, pp. 820–829.
- Licklider, J. (1948). "The influence of interaural phase relations upon the masking of speech by white noise". In: *The Journal of the Acoustical Society of America* 20.2, pp. 150–159.
- Lócsi, G., Santurette, S., Dau, T., and MacDonald, E. N. (2014). "Influence of acoustic complexity on spatial release from masking and lateralization". In: *Proceedings of Forum Acusticum 2014*. European Acoustics Association - EAA.
- Lócsi, G., Pedersen, J. H., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E. N. (2015). "Lateralized speech perception in normal-hearing and hearing-impaired listeners and its relationship to temporal processing". In: *Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*. Ed. by Santurette, S., Dau, T., Dalsgaard, J. C., Tranebjærg, L., and Andersen, T. Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 389–396.
- Lorenzi, C. and Moore, B. C. J. (2008). "Role of temporal envelope and fine structure cues in speech perception: A review". In: *Auditory Signal Processing in Hearing-Impaired Listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007)*. Ed. by Dau, T., Buchholz, J. M., Harte, J. M., and Christiansen, T. U. Centertryk A/S., Denmark, pp. 263–272.
- Lorenzi, C., Gatehouse, S., and Lever, C. (1999). "Sound localization in noise in hearing-impaired listeners." In: *The Journal of the Acoustical Society of America* 105.3, pp. 3454–3463.

- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.49, pp. 18866–18869.
- Lunner, T., Hietkamp, R. K., Andersen, M. R., Hopkins, K., and Moore, B. C. J. (2012). "Effect of Speech Material on the Benefit of Temporal Fine Structure Information in Speech for Young Normal-Hearing and Older Hearing-Impaired Participants". In: *Ear and Hearing* 33.3, pp. 377–388.
- Macpherson, E. A. and Middlebrooks, J. C. (2002). "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited". In: *The Journal of the Acoustical Society of America* 111.5, pp. 2219–2236.
- Makary, C. A., Shin, J., Kujawa, S. G., Liberman, M. C., and Merchant, S. N. (2011). "Age-related primary cochlear neuronal degeneration in human temporal bones". In: *Journal of the Association for Research in Otolaryngology* 12.6, pp. 711–717.
- Marrone, N., Mason, C. R., and Kidd, G. (2008a). "Evaluating the benefit of hearing aids in solving the cocktail party problem." In: *Trends in Amplification* 12.4, pp. 300–315.
- Marrone, N., Mason, C. R., and Kidd, G. (2008b). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms." In: *The Journal of the Acoustical Society of America* 124.5, pp. 3064–3075.
- Marrone, N., Mason, C. R., and Kidd, G. (2008c). "Tuning in the spatial dimension: evidence from a masked speech identification task." In: *The Journal of the Acoustical Society of America* 124.2, pp. 1146–1158.
- Moore, B. C. J. (2008). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people". In: *Journal of the Association for Research in Otolaryngology* 9, pp. 399–406.
- Moore, B. C. J. (2014). *Auditory processing of temporal fine structure: effects of age and hearing loss*. World Scientific.

- Moore, B. C. J. and Peters, R. W. (1992). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity." In: *The Journal of the Acoustical Society of America* 91.5, pp. 2881–2893.
- Moore, B. C. J., Stone, M. A., Füllgrabe, C., Glasberg, B. R., and Puria, S. (2008). "Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss." In: *Ear and Hearing* 29.6, pp. 907–922.
- Moore, B. C. J., Glasberg, B. R., Stoev, M., Füllgrabe, C., and Hopkins, K. (2012). "The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L)". In: *The Journal of the Acoustical Society of America* 131.2, EL1003–EL1006.
- Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and Schaik, A van (2009). "Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition." In: *International Journal of Audiology* 48, pp. 758–774.
- Neher, T., Jensen, N. S., and Kragelund, L. (2011). "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?" In: *The Journal of the Acoustical Society of America* 130.3, pp. 1542–1558.
- Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. J. (2012). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L)." In: *The Journal of the Acoustical Society of America* 131.4, pp. 2561–2564.
- Nielsen, J., Dau, T., and Neher, T. (2014). "A Danish open-set speech corpus for competing-speech studies." In: *The Journal of the Acoustical Society of America* 135.1, pp. 407–420.
- Oxenham, A. J. and Simonson, A. M. (2009). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference." In: *The Journal of the Acoustical Society of America* 125.1, pp. 457–468.
- Oxenham, A. J. and Bacon, S. P. (2003). "Cochlear compression: perceptual measures and implications for normal and impaired hearing." In: *Ear and Hearing* 24.5, pp. 352–366.

- Papakonstantinou, A., Strelcyk, O., and Dau, T. (2011). "Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise." In: *Hearing research* 280.1-2, pp. 30–37.
- Peissig, J. and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners." In: *The Journal of the Acoustical Society of America* 101.3, pp. 1660–1670.
- Pichora-Fuller, M. K. and Singh, G. (2006). "Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation". In: *Trends in Amplification* 10.1, pp. 29–59.
- Plack, C. J. and Oxenham, A. J. (2005). "The Psychophysics of Pitch". In: *Pitch: Neural Coding and Perception*. Ed. by Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. Vol. 24. Chapter 2. New York: Springer, pp. 7–55.
- 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.
- Rönnerberg, J., Arlinger, S., Lyxell, B., and Kinnefors, C. (1989). "Visual Evoked Potentials: Relation to Adult Speechreading and Cognitive Function". In: *Journal of Speech, Language, and Hearing Research* 32.4, pp. 725–735.
- Ross, B., Fujioka, T., Tremblay, K. L., and Picton, T. W. (2007a). "Aging in Binaural Hearing Begins in Mid-Life: Evidence from Cortical Auditory-Evoked Responses to Changes in Interaural Phase". In: *Journal of Neuroscience* 27.42, pp. 11172–11178.
- Ross, B., Tremblay, K. L., and Picton, T. W. (2007b). "Physiological detection of interaural phase differences." In: *The Journal of the Acoustical Society of America* 121.2, pp. 1017–1027.
- Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). "Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.37, pp. 15516–15521.
- Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2012). "Why middle-aged listeners have trouble hearing in everyday settings." In: *Current Biology* 22.15, pp. 1417–1422.

- Santurette, S. and Dau, T. (2012). "Relating binaural pitch perception to the individual listener's auditory profile." In: *The Journal of the Acoustical Society of America* 131.4, pp. 2968–2986.
- Schoof, T. and Rosen, S. (2014). "The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners". In: *Frontiers in Aging Neuroscience* 6.307, pp. 1–14.
- Seçk, A. and Moore, B. C. J. (1995). "Frequency discrimination as a function of frequency, measured in several ways". In: *The Journal of the Acoustical Society of America* 97.4, pp. 2479–2486.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues." In: *Science* 270.5234, pp. 303–304.
- Shinn-Cunningham, B. G. and Best, V. (2008). "Selective attention in normal and impaired hearing". In: *Trends in Amplification* 12.4, pp. 283–299.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). "Chimaeric sounds reveal dichotomies in auditory perception." In: *Nature* 416.6876, pp. 87–90.
- Staffel, J. G., Hall, W., Grose, J. H., and Pillsbury, C. (1990). " N_0S_0 and N_0S_{pi} detection as a function of masker bandwidth in normal-hearing and cochlear-impaired listeners". In: *The Journal of the Acoustical Society of America* 87.4, pp. 1720–1727.
- Strelcyk, O. and Dau, T. (2009). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing." In: *The Journal of the Acoustical Society of America* 125.5, pp. 3328–3345.
- Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). "Temporal processing in the aging auditory system." In: *The Journal of the Acoustical Society of America* 104.4, pp. 2385–2399.
- Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (1999). "Monosyllabic word recognition at higher-than-normal speech and noise levels." In: *The Journal of the Acoustical Society of America* 105.4, pp. 2431–2444.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., and Patel, A. D. (2015). "Musical training, individual differences and the cocktail party problem." In: *Scientific reports* 5.11628, pp. 1–11.

- Takahashi, G. A. and Bacon, S. P. (1992). "Modulation Detection , Modulation Masking , and Speech Understanding in Noise in the Elderly". In: *Journal of Speech and Hearing Research* 35, pp. 1410–1421.
- Tyler, R. S., Wood, E. J., and Fernandes, M. (1983). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners". In: *The Journal of the Acoustical Society of America* 74.4, pp. 1190–1199.
- Wagener, K., Jøsvassen, J., and Ardenkjær, R. (2003). "Design, optimization and evaluation of a Danish sentence test in noise". In: *International Journal of Audiology* 42.1, pp. 10–17.
- Wightman, F. L. and Kistler, D. J. (1992). "The dominant role of low-frequency interaural time differences in sound localization." In: *The Journal of the Acoustical Society of America* 91.3, pp. 1648–1661.
- Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (2005). "Speech recognition with amplitude and frequency modulations." In: *Proceedings of the National Academy of Sciences of the United States of America* 102.7, pp. 2293–2298.
- Zurek, P. M. (1993). "A note on onset effects in binaural hearing." In: *The Journal of the Acoustical Society of America* 93, pp. 1200–1201.
- Zwislocki, J. and Feldman, R. S. (1956). "Just Noticeable Differences in Dichotic Phase". In: *The Journal of the Acoustical Society of America* 28.5, pp. 860–864.

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 Feręczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.
- Vol. 21:** *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.
- Vol. 22:** *Federica Bianchi*, Complex-tone pitch representations in the human auditory system, 2016.
- Vol. 23:** *Johannes Zaar*, Measures and computational models of microscopic speech perception, 2016.

The end.

To be continued...