Temporal Fine-Structure Coding and Lateralized Speech Perception in Normal-Hearing and Hearing-Impaired Listeners

Gusztáv Lőcsei[1], Julie H. Pedersen[2], Søren Laugesen[2], Sébastien Santurette[1], Torsten Dau[1], and Ewen N. MacDonald[1]

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, two-, four-, or eight-talker babble played reversed, or a nonreversed two-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 and interaural phase difference thresholds 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 benefit due to binaural unmasking differed only slightly between the groups. Neither the frequency discrimination threshold nor the interaural phase difference threshold 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.

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 (SI) 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 & Plomp, 1988).

A listener's head will acoustically shadow lateral incoming sound, resulting in ILDs, which allow for

[1]Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark
[2]Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

Corresponding author:
Gusztáv Lőcsei, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads 352, Kongens Lyngby 2800, Denmark.
Email: guloc@elektro.dtu.dk

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, Sandel, Teas, & Jeffress, 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 (IPDs) between the ears. ITDs play a dominant role in sound localization of relatively distant stimuli with a low-frequency content (Wightman & Kistler, 1992) and also contribute in the facilitation of speech understanding in spatial settings (Bronkhorst & Plomp, 1988). If the ITDs associated with target and masker streams are different, SI increases as compared with situations where the ITDs are the same (e.g., Carhart, Tillman, & Kenneth, 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 article, 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 SI in quiet that can be fully compensated for by appropriately amplifying the speech stimuli. The distortion component affects SI 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 suprathreshold stimuli (e.g., Dreschler & Plomp, 1985; Glasberg & Moore, 1989; Plomp, 1978). The sources of such deficits could include broadening of the auditory filters (e.g., Glasberg & Moore, 1989) or degraded temporal coding (e.g., Hopkins, Moore, & Stone, 2008; Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Papakonstantinou, Strelcyk, & Dau, 2011; Strelcyk & Dau, 2009).

The cochlea can be modeled 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 (Sek & Moore, 1995) and might play a role up to as high as 8 kHz (Ernst & 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 (Moore & Peters, 1992; Tyler, Wood, & Fernandes, 1983) or in low-rate frequency-modulation detection (Lacher-Fougère & Demany, 1998; Santurette & Dau, 2012; Strelcyk & Dau, 2009). 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 & Moore, 2011; King, Hopkins, & Plack, 2014; Ross, Fujioka, Tremblay, & Picton, 2007) or by binaural masking-level differences (Papakonstantinou et al., 2011; Strelcyk & Dau, 2009; Strouse, Ashmead, Ohde, & Grantham, 1998).

In an attempt to map difficulties in speech perception in noisy environments to suprathreshold processing deficits, multiple studies have investigated the relationship between deficits in 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 & Moore, 2008; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), reduced access to TFS has been associated with deteriorated speech perception when noise is present (Lorenzi et al., 2006; Lunner, Hietkamp, Andersen, Hopkins, & Moore, 2012; Papakonstantinou et al., 2011; Strelcyk & 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, Griffin, & Oxenham, 2012; Oxenham & Simonson, 2009; Strelcyk & 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, Carlile, Kopco, & van Shaik, 2011;

Lorenzi, Gatehouse, & Lever, 1999; Neher, Jensen, & Kragelund, 2011; Ruggles & Shinn-Cunningham, 2011), increasing SRTs, and reducing the amount of SRM (Best, Mason, & Kidd, 2011; Bronkhorst, 2000; Bronkhorst & Plomp, 1992; Neher et al., 2009, 2011; Peissig & 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., Bronkhorst, 2000; Bronkhorst & Plomp, 1992; Lorenzi et al., 1999; Neher et al., 2011).

Impairments in TFS coding have been associated with reduced SRTs in tasks involving spatial cues (Neher et al., 2011; Neher, Lunner, Hopkins, & Moore, 2012; Strelcyk & 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 with 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, SRTs in lateralized speech-shaped noise (SSN) and two-talker babble showed a significant correlation with measures of TFS processing, including 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 three-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 & Moore, 2010). However, these correlations became nonsignificant 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 & Plomp, 1988; Carhart et al., 1967; Culling, Hawley, & Litovsky, 2004; Glyde, Buchholz, Dillon, Cameron, & Hickson, 2013). Under the assumption that SRM in this experiment will be mainly governed by ITDs in the low-frequency domain (Bronkhorst & 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.

Methods

Participants

Nineteen elderly HI listeners participated in the study (55–85 years, mean: 71.7, standard deviation (SD): 7.19). As the goal was to investigate suprathreshold factors affecting SI 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 PTAs 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 homogeneity 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 1.

The control group consisted of 10 young NH listeners (21–29 years, 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.

Temporal Processing and Cognitive Skills

To assess the robustness of monaural and binaural TFS coding, FDTs and 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 & Oxenham, 2005), FDT thresholds show only a weak relationship with frequency selectivity (Moore & Peters, 1992; Tyler et al., 1983), 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
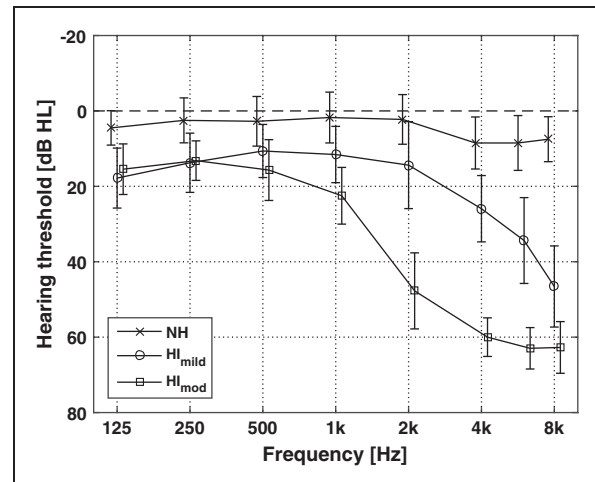

Figure 1. Audiometric data of the listener groups, averaged over both ears of the listeners. Horizontal bars denote ± 1 SD. The data of the NH and HI_{mod} listener groups are shifted on the x-axis for better readability.

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 three-interval three-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 two. 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 eight reversals. Thresholds were calculated as the geometrical mean of these reversal points. Overall, five runs were performed by each subject. The final threshold was calculated as the geometric mean of the thresholds in the last three runs. All stimuli were presented monaurally at 65 dB sound pressure level (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 & 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 one-up two-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 $\Delta\phi$. Initially, $\Delta\phi$ 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 five threshold estimation tests and the final threshold was calculated as the geometrical mean of the last three runs. The stimuli were presented at 30 dB sensation level.

The cognitive abilities of the listeners' were assessed using a Danish version of the reading span test (Daneman & Carpenter, 1980; Rönnberg, Arlinger, Lyxell, & Kinnefors, 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 three-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 s in each sentence and listeners had 1.75 s to give their response about the semantics of the sentence after the last word. After a block of three, four, five, or six 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 three 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 three sentences was included at the beginning of the test.

Speech Perception in Noise

SRTs were measured using target sentences uttered by a female talker from the Danish DAT corpus (Nielsen, Dau, & Neher, 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: SSN, reversed speech with two, four, or eight

streams of competing male talkers from the Grid corpus (Cooke, Barker, Cunningham, & Shao, 2006), or forward speech with single sentences uttered by the two 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 two, four, or eight competing talkers (R_2, R_4, R_8), or two 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 versus informational masking (e.g., Kidd, Mason, Deliwala, Woods, & Colburn, 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 spectrotemporal 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 s 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. Fifty nonoverlapping tokens of 5 s were selected from each of these talkers. When generating masker tokens, single random tokens were drawn from the pregenerated pool of tokens for each of the first two, four, or eight 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 toward 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 with 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 toward 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 have been 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 toward the target side was varied in two's (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 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 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, Stone, Füllgrabe, Glasberg, and Puria (2008), and also compensated for the loss of stimulus audibility. The elevated hearing thresholds of the HI listeners were compensated for by applying frequency-dependent 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 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 three lists before each visit. We tested the S_1, R_2, and R_4 conditions during the first and the R_8 and D_2 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.

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 honest significant difference 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.

Results

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}: $F(2, 26) = 24.34$, $p < .001$; PTA_{high}: $F(2, 26) = 213.98$, $p < .001$; PTA_{oct}: $F(2, 26) = 79.07$, $p < .001$) and post hoc analyses ($p < .05$ in all cases). A one-way ANOVA have been conducted on the HLTs at 250 Hz (HTL_{250}), at which temporal processing abilities were assessed. The effect of listener group was significant, $F(2, 26) = 11.09$, $p < .001$. Pairwise post hoc tests revealed that NH group significantly differed from both of the HI groups ($p = .001$), while the thresholds of the HI groups did not show any significant difference.
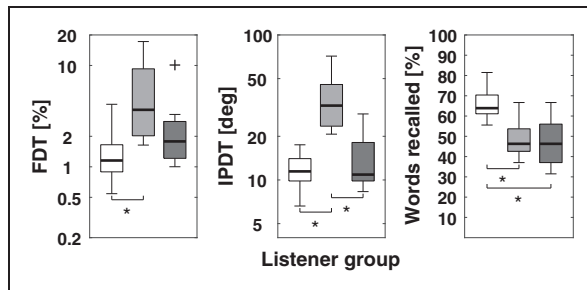
Figure 2. Box plots illustrating the FDT, IPDT, and Reading span test results. White, light gray, and dark gray boxes stand for the data of the NH, HI$_{mild}$, and HI$_{mod}$ groups, respectively. The thick black lines denote the medians and the boxes extend to the 25th and 75th percentiles. The thin 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.

Temporal Processing

The results from the FDT and IPDT experiments are displayed in Figure 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 (Lacher-Fougére & Demany, 2005; Strelcyk & Dau, 2009). 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, $p = .014$). 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 significant, $F(2, 21) = 5.67$, $p = .011$, and post hoc analyses revealed that the only significant difference in group means was the one between the NH and HI$_{mild}$ group ($p = .008$).

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 & 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, $F(2, 26) = 21.73$, $p < .001$. Post hoc analyses revealed that the differences in mean thresholds between the NH and HImild and between the HImild and HImod groups were both statistically significant (p < .001).

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, $F(2, 26) = 10.35$, $p < .001$. Post hoc tests confirmed that the NH listeners performed significantly better than both the HI$_{mild}$ ($p = .004$) and HI$_{mod}$ ($p = .001$) listener groups, while the difference between the two HI groups remained nonsignificant. These results are consistent with earlier studies showing an age-related decline of working memory (e.g., Schoof & Rosen, 2014).

Speech Perception in Noise

In Figure 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 with 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 1). The other spatial distributions were left out from the calculation, as those would bias the SRTs toward higher SNRs for those noise types that had a higher number of spatial distributions. Group differences were smallest in the steady-state masker
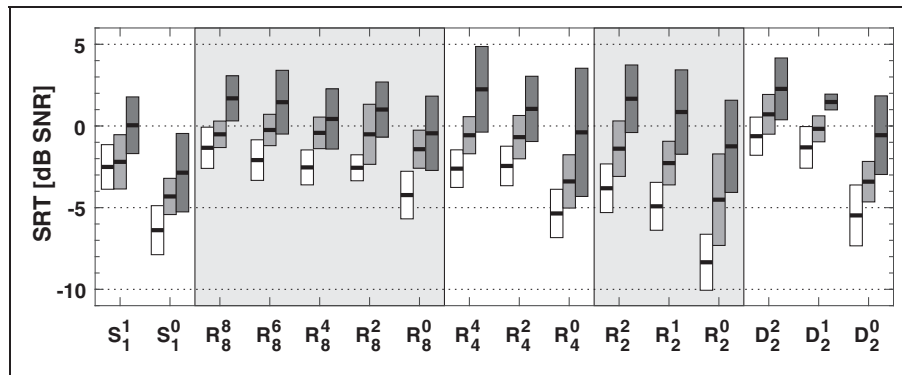
Figure 3. SRTs for NH (white), HI_mild (light gray), and HI_mod (dark gray) listeners. Horizontal black bars denote group means and the boxes represent ± 1 *SD*. The white or gray areas in the background denote condition groups with the same masker type. Condition group notations: S_x^y: speech shaped 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.

Table 1. SRTs of the listener groups averaged over the fully colocated and fully separated masker distributions for the five noise types.

	Average SRT (dB SNR)		
	Listener group		
Noise type	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

Note. SRT: speech reception thresholds; SNR: signal-to-noise ratio. Condition group notations as in Figure 3.

conditions (S_1) and gradually increased as more spectro-temporal fluctuations appear in the background noise. The NH 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 with the stationary S_1 conditions. While NH listeners performed better as the number of reversed interferers decreased from eight to four to two, 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 & Dau, 2012; Festen & Plomp, 1990; Strelcyk & 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.

To test the statistical significance of the abovementioned observations, mixed ANOVAs were performed on the SRTs and BILDs in the fully colocated and fully separated conditions. Figure 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, and D_2) and *lateralization* as within-subject and *listener group* (NH, HI_mild, and HI_mod) as between-subject factors showed a main effect of *lateralization*, $F(1, 26) = 311.41$, $p < .001$, *noise type*, $F(3.11, 80.95) = 28.02$, $p < .001$, and *listener group*, $F(2, 26) = 24.171$, $p < .001$. The interaction was significant between *noise type* and *listener group*, $F(6.23, 80.9) = 5.03$, $p < .001$. Bonferroni-corrected paired *t*-tests within listener groups showed that the SRTs were significantly greater in the R_8 than in the S_1 condition for all of the listener groups ($p < .0125$ in each case), indicating that substituting the SSN noise masker with the reversed eight-talker babble resulted in increased masking, despite the inherent spectro-temporal fluctuations of the masker. In contrast, differences between the S_1 and R_2 noise condition were lower than zero for the NH ($p = .006$), not significantly different from zero for the HI_mild ($p = .6$), and greater than zero for the HI_mild listeners ($p = .002$). This supports the observation that NH listeners' performance improves as the number of interfering streams decreases from eight to two, 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*, $F(2, 26) = 4.91$, $p = .016$, and *lateralization* and *noise type*, $F(3.08, 6.15) = 4.57$, $p = .005$, a second mixed ANOVA was conducted on
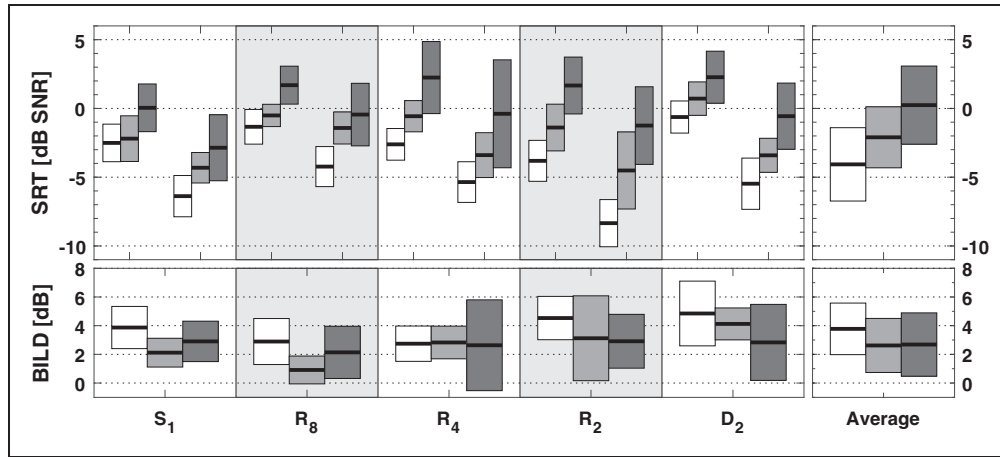
Figure 4. Top panel: SRTs in the fully colocated and fully separated target-masker distributions (left and right part of each block, respectively, repeated from Figure 3). Bottom panel: BILDs calculated as the difference between the colocated and separated SRTs. The horizontal black lines with the white, light gray, and dark gray boxes stand for the data of the NH, HI_{mild}, and HI_{mod} groups (mean and SD). The white or gray areas in the background denote condition groups with the same masker type. Condition group notations: S1: speech shaped noise masker; R_8, R_4, R_2: reversed speech masker consisting of eight, four, or two competing talkers; D_2: forward speech masker consisting of two competing talkers.

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 group*, $F(2, 26) = 4.91$, $p = .016$, and *noise type*, $F(3.08, 79.96) = 4.57$, $p = .005$, and no significant interaction. Post hoc pairwise comparisons with Bonferroni correction revealed that the BILD in the R_8 condition was significantly lower than the BILD in the S_1 and D_2 conditions ($p < .005$ in each case). Compared with the NH listeners, the average BILD was lower by about 1 dB for both the HI_{mild} and HI_{mod} listeners ($p < .017$).

In Figure 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 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 ($p > .005$).

Predicting SI

As mentioned earlier, 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
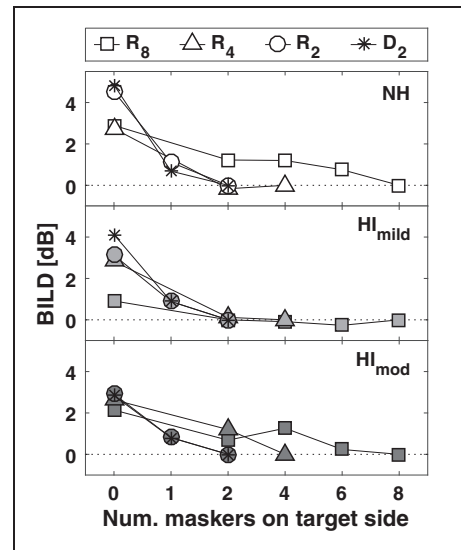

Figure 5. BILDs as a function of number of interfering talkers colocated with the target side in the different interferer conditions. The top, middle, and bottom panels illustrate the results of the NH, HI_{mild}, and HI_{mod} listener groups. Condition group notations: R_8, R_4, R_2: reversed speech masker of eight, four, or two competing talkers; D_2: forward speech masker of two competing talkers.

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 nonsignificant. The correlation between age and PTA_{oct} was also nonsignificant. 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 6 presents performance measures from the SI tests compared with 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 colocated 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, $r(17) = .55$, $p = .015$. 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.
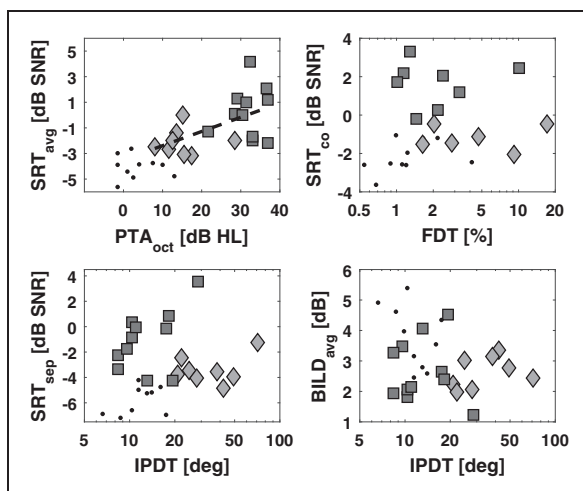

Figure 6. Scatter plots between predictor measures and speech perception results for the NH (black dots), HI_{mild} (light gray diamonds), and HI_{mod} (dark gray squares) listeners. Top left: SRT_{avg} as a function of PTA_{oct}. The dashed line stands for the fitted regression line on the data of the 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.

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 (Lunner et al., 2012; Strelcyk & Dau, 2009), for example, 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 colocated noise 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 puretone 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 SRM by reducing the magnitude of the BILDs. The top right panel of Figure 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, $r(17) = -.64$, $p = .014$. 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 reran the analysis by controlling for PTA_{oct}. The correlation between IPDT and SRT_{sep} became nonsignificant, 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.

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 colocated 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 interindividual 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 & 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 HLTs 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, $t(17) = 2.5$, $p = .023$. 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, three 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}, $t(17) = -1.49$, $p = .165$, and no significant correlation between number of visits and the FDT or IPDT results ($p > .1$ 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 & 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 R_8 condition, yielding higher SRTs than the conditions with the S_1 masker in each of the listener groups. This difference, however, cannot be explained based on energetic masking in the classical sense, as the R_8 maskers have a sparser spectro-temporal structure than the S_1 maskers. It is more likely that these differences arise from susceptibility to modulation masking (Houtgast, 1989; Takahashi & Bacon, 1992). The differences in average SRTs between the R_8 and R_2 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 R_2 masker condition. Nonetheless, the results clearly show that HI listeners had difficulties with understanding speech in modulated noise compared with the NH group, which is consistent with earlier reports (Christiansen & Dau, 2012; Festen & Plomp, 1990; Strelcyk & 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, 2000; Bronkhorst & Plomp, 1989; Santurette & 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 colocated 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, for example, its formant structure. It appears, therefore, that performance in the SI 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 & Simonson, 2009; Strelcyk & Dau, 2009), but it rather facilitates speech understanding in noise by providing cues for the perceptual segregation of target from the maskers (Lunner et al., 2012; Strelcyk & Dau, 2009). This conjecture nonetheless needs further investigation, especially that there are indications that TFS coding deficits can be associated with degraded SI 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 reduces 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 & Culling, 2005). Since the upper frequency limit of sensitivity to TFS ITDs reduces with progressing age (e.g., Ross et al., 2007), it is possible that BILDs listeners can obtain in a particular listening scenario is 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, for example, the broadening of the auditory filters and reduced cochlear compression (Bianchi, Fereczkowski, Zaar, Santurette, & Dau, 2016; Henry, Kale, & Heinz, 2014). If the high-frequency hearing loss of the HI listeners in the current study was coupled with broader auditory filters and reduced compression, for example, 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 & 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 & 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. When maskers are high on 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, Mason, & Kidd, 2008; 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 predefined 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 SPLs, which has been shown to affect SRTs (Studebaker, Scherbecoe, McDaniel, & Gwaltney, 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.

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 (Papkonstantinou 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.

References

Bernstein, L. R., & Trahiotis, C. (2009). How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *The Journal of the Acoustical Society of America*, 125, 3234–3242.

Best, V., Carlile, S., Kopco, N., & van Schaik, A. (2011). Localization in speech mixtures by listeners with hearing loss. *The Journal of the Acoustical Society of America*, 129, EL210–EL215.

Best, V., Mason, C. R., & Kidd, G. (2011). Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers. *The Journal of the Acoustical Society of America*, 129, 1616–1625.

Bianchi, F., Fereczkowski, M., Zaar, J., Santurette, S., & Dau, T. (2016). Effects of cochlear compression and frequency selectivity on pitch discrimination of complex tones with unresolved harmonics. In S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranjebjærg, & T. Andersen (Eds.), *Individual hearing loss—Characterization, modelling, compensation strategies. 5th international symposium on auditory and audiological research (ISAAR 2015)* (pp. 51–58). Copenhagen, Denmark: The Danavox Jubilee Foundation.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. Cambridge, MA: The MIT Press.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica*, 86, 117–128.

Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83, 1508–1516.

Bronkhorst, A. W., & Plomp, R. (1989). Binaural speech intelligibility in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 86, 1374–1383.

Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *The Journal of the Acoustical Society of America*, 92, 3132–3139.

Carhart, R., Tillman, T. W., & Kenneth, R. J. (1967). Release of masking for speech through interaural time delay. *The Journal of the Acoustical Society of America*, 42, 124–138.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25, 975–979.

Christiansen, C., & Dau, T. (2012). Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. *The Journal of the Acoustical Society of America*, 132, 1655–1666.

Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, 120, 2421–2424.

Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America*, 116, 1057–1065.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, 19, 450–466.

Dreschler, W. A., & Plomp, R. (1985). Relations between psychophysical data and speech perception for hearing-impaired subjects. II. *The Journal of the Acoustical Society of America*, 78, 1261–1270.

Edmonds, B. A., & Culling, J. F. (2005). The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay. *The Journal of the Acoustical Society of America*, 117, 3069–3078.

Ernst, S. M. A., & Moore, B. C. J. (2012). The role of time and place cues in the detection of frequency modulation by hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 131, 4722–4731.

Feddersen, W. E., Sandel, T. T., Teas, D. C., & Jeffress, L. A. (1957). Localization of high-frequency tones. *The Journal of the Acoustical Society of America*, 29, 988–991.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88, 1725–1736.

Freyman, R. L., Griffin, A. M., & Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *The Journal of the Acoustical Society of America*, 132, 2514–2523.

Glasberg, B. R., & Moore, B. C. J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology Supplementum*, 32, 1–25.

Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., & Hickson, L. (2013). The importance of interaural time differences and level differences in spatial release from masking. *The Journal of the Acoustical Society of America*, 134, EL147–EL152.

Henry, K. S., Kale, S., & Heinz, M. G. (2014). Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Frontiers in Systems Neuroscience*, 8(20), 1–20.

Hopkins, K., & Moore, B. C. J. (2010). Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies. *International Journal of Audiology*, 49, 940–946.

Hopkins, K., & 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. *The Journal of the Acoustical Society of America*, 130, 334–349.

Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *The Journal of the Acoustical Society of America*, 123, 1140–1153.

Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *The Journal of the Acoustical Society of America*, 85, 1676–1680.

Kidd, G. Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *The Journal of the Acoustical Society of America*, 95, 3475–3480.

King, A., Hopkins, K., & Plack, C. J. (2014). The effects of age and hearing loss on interaural phase difference discrimination. *The Journal of the Acoustical Society of America*, 135, 342–351.

Lacher-Fougère, S., & Demany, L. (1998). Modulation detection by normal and hearing-impaired listeners. *Audiology*, 37, 109–121.

Lacher-Fougère, S., & Demany, L. (2005). Consequences of cochlear damage for the detection of interaural phase differences. *The Journal of the Acoustical Society of America*, 118, 2519–2526.

Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Sound localization in noise in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 105, 3454–3463.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 18866–18869.

Lorenzi, C., & Moore, B. C. J. (2008). Role of temporal envelope and fine structure cues in speech perception: A review. In T. Dau, J. M. Buchholz, J. M. Harte, & T. U. Christiansen (Eds.), *Auditory signal processing in hearing-impaired listeners. 1st international symposium on auditory and audiological research (ISAAR 2007)* (pp. 263–272). Holbaek, Denmark: Centertryk A/S.

Lunner, T., Hietkamp, R. K., Andersen, M. R., Hopkins, K., & 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. *Ear and Hearing*, 33, 377–388.

Marrone, N., Mason, C. R., & Kidd, G. Jr. (2008). Tuning in the spatial dimension: Evidence from a masked speech identification task. *The Journal of the Acoustical Society of America*, 124, 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. *Journal of the Association for Research in Otolaryngology*, 9, 399–406.

Moore, B. C., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America*, 91, 2881–2893.

Moore, B. C. J., Stone, M. A., Füllgrabe, C., Glasberg, B. R., & 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. *Ear and Hearing*, *29*, 907–922.

Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., . . . van Schaik, A. (2009). Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition. *International Journal of Audiology*, *48*, 758–774.

Neher, T., Jensen, N. S., & Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *The Journal of the Acoustical Society of America*, *130*, 1542–1558.

Neher, T., Lunner, T., Hopkins, K., & Moore, B. C. J. (2012). Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L). *The Journal of the Acoustical Society of America*, *131*, 2561–2564.

Nielsen, J., Dau, T., & Neher, T. (2014). A Danish open-set speech corpus for competing-speech studies. *The Journal of the Acoustical Society of America*, *135*, 407–420.

Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *The Journal of the Acoustical Society of America*, *125*, 457–468.

Papakonstantinou, A., Strelcyk, O., & Dau, T. (2011). Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise. *Hearing Research*, *280*, 30–37.

Peissig, J., & Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of the Acoustical Society of America*, *101*, 1660–1670.

Plack, C. J., & Oxenham, A. J. (2005). The psychophysics of pitch. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (24 ed., pp. 7–55). (Springer Handbook of Auditory Research). New York, NY: Springer.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, *63*, 533–549.

Rönnberg, J., Arlinger, S., Lyxell, B., & Kinnefors, C. (1989). Visual evoked potentials: Relations to adult speechreading and cognitive function. *Journal of Speech and Hearing Research*, *32*, 725–735.

Ross, B., Fujioka, T., Tremblay, K., & Picton, T. W. (2007). Aging in binaural hearing begins in mid-life: Evidence from cortical auditory-evoked responses to changes in interaural phase. *The Journal of Neuroscience*, *27*, 11172–11178.

Ruggles, D., & Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners. *Journal of the Association for Research in Otolaryngology*, *12*, 395–405.

Santurette, S., & Dau, T. (2012). Relating binaural pitch perception to the individual listener's auditory profile. *The Journal of the Acoustical Society of America*, *131*, 2968–2986.

Schoof, T., & Rosen, S. (2014). The role of age-related auditory and cognitive declines in understanding speech in noise. *Frontiers in Aging Neuroscience*, *6*(307), 1–14.

Sek, A., & Moore, B. C. (1995). Frequency discrimination as a function of frequency, measured in several ways. *The Journal of the Acoustical Society of America*, *97*, 2479–2486.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Strelcyk, O., & Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, *125*, 3328–3345.

Strouse, A., Ashmead, D. H., Ohde, R. N., & Grantham, D. W. (1998). Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America*, *104*, 2385–2399.

Studebaker, G. A., Scherbecoe, R. L., McDaniel, D. M., & Gwaltney, C. A. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *The Journal of the Acoustical Society of America*, *105*, 2431–2444.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G. Jr., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, *5*, 11628.

Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech, Language, and Hearing Research*, *35*, 1410–1421.

Tyler, R. S., Wood, E. J., & Fernandes, M. (1983). Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *74*, 1190–1199.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, *91*, 1648–1661.