Lateralized speech perception in normal-hearing and hearing-impaired listeners and its relationship to temporal processing

GUSZTÁV LŐCSEI[1,*], JULIE HEFTING PEDERSEN[2], SØREN LAUGESEN[2], SÉBASTIEN SANTURETTE[1], TORSTEN DAU[1], AND EWEN N. MACDONALD[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark*

This study investigated the role of temporal fine structure (TFS) coding in spatially complex, lateralized listening tasks. Speech reception thresholds (SRTs) were measured in young normal-hearing (NH) and two groups of elderly hearing-impaired (HI) listeners in the presence of speech-shaped noise and different interfering talker conditions. The HI subjects had either a mild or moderate hearing loss above 1.5 kHz and reduced audibility was compensated for individually in the speech tests. The target and masker streams were presented as coming from the same or from the opposite side of the head by introducing 0.7-ms interaural time differences (ITD) between the ears. To assess the robustness of TFS coding, frequency discrimination thresholds (FDTs) and interaural phase difference thresholds (IPDTs) were measured at 250 Hz. While SRTs of the NH subjects were clearly better than those of the HI listeners, group differences in binaural benefit due to spatial separation of the maskers from the target remained small. Neither the FDT nor the IPDT tasks showed a clear correlation pattern with the SRTs or with the amount of binaural benefit, respectively. The results suggest that, although HI listeners with normal hearing in the low-frequency range might have elevated SRTs, the binaural benefit they experience due to spatial separation of competing sources can remain similar to that of NH listeners.

INTRODUCTION

Normal-hearing (NH) listeners are extremely skillful in following a particular talker in the presence of multiple interfering acoustic sources. Through the use of binaural cues, interaural level differences (ILDs) and interaural time differences (ITDs), listeners can segregate sources that are spatially separated. While NH listeners can exploit spatial cues to aid robust speech identification in cocktail-party scenarios, hearing loss has been shown to negatively affect spatial perception of speech (Neher *et al.*, 2011). Furthermore, speech intelligibility performance can vary substantially across individual hearing-impaired (HI) listeners with similar audiograms. One potential explanation for this are individual differences in temporal fine structure (TFS) coding (e.g., Strelcyk and Dau, 2009; Papakonstantinou *et al.*, 2011).

---

*Corresponding author: guloc@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

The purpose of this study was to investigate the relationship between monaural and binaural TFS coding and speech intelligibility in lateralized conditions. To assess the robustness of low-frequency TFS coding, we measured frequency discrimination thresholds (FDTs) and interaural phase difference detection thresholds (IPDTs) for pure tones at 250 Hz. In addition, a speech intelligibility experiment was conducted where the stimuli were presented over headphones and "spatialized" with frequency-independent ITD cues only. We hypothesized that listeners who have elevated pure tone IPD detection thresholds will have limited capabilities to exploit ITD disparities between target and masker streams, and thus have a reduced spatial release from masking (SRM) once maskers are separated from the target.

METHODS

Listeners

10 young NH (21-29 yrs, mean: 23 std: 3.01) and 19 older HI (55-85 yrs, mean: 71.7, std: 7.19) listeners participated in the study. Members of the NH group had audiometric thresholds lower than 20 dB HL at octave frequencies between 125 and 8000 Hz. 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. For each listener the difference in audiometric thresholds between the ears was at most 15 dB. The HI group was divided into two age-matched subgroups: those having pure-tone average thresholds (PTAs) less or equal to 40 dB HL above 1.5 kHz on average were classified as mildly impaired (HI_{mild}, 8 listeners) and the others were classified as moderately impaired (HI_{mod}, 11 listeners), respectively. This homogeneity of audiograms within groups was desirable in order to minimize audibility confounds at high frequencies once investigating the results of the speech intelligibility experiments.

Speech tests

SRTs were measured using target sentences uttered by a female talker from the Danish DAT corpus (Nielsen *et al.*, 2014). We used the "Dagmar" sentences as targets in the presence of the following interferers: speech shaped noise (SSN), reversed speech with 2, 4, or 8 competing male talkers from the Grid corpus (Cooke *et al.*, 2006) and forward speech with single sentences uttered by the 2 other female talkers from the DAT corpus. Target and masker stimuli were presented as coming from a lateral direction by introducing 0.7-ms ITDs between the ears for each of the streams. Two spatial configurations were used in each masker conditions: target and maskers leading on the same side (later referred to as co-located conditions) and target and maskers leading on opposite sides (separated condition). The side of the target was randomized from trial to 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 run on an interleaved manner. The notations S_1, C_2, C_4, C_8, and D_2 are used to denote the set of conditions where SSN, reversed speech of 2, 4, or 8 competing talkers, or 2 interferers from the DAT corpus are used as maskers, respectively. When referring to a specific spatial condition within each of these sets, the "co" and "sep" indicators will be used as superscripts (e.g., S_1^{co} refers to the condition with the SSN masker, where target and masker are presented as coming from the same side).

The maskers in the S_1, C_2, C_4, and C_8 conditions were spectrally shaped to have the same long-term average spectrum of 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 C_2, C_4, and C_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-to-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.

The stimuli were presented over headphones. To simulate free-field presentation, the target sentences were first set to a nominal level of 65 dB SPL "free field", mixed with the maskers, and finally amplified by adding open ear gain components (Moore *et al.*, 2008). The elevated hearing thresholds of the HI subjects were compensated for by applying frequency dependent linear gains based on their audiograms and the long-term average spectrum of the target speech (Neher *et al.*, 2011; Nielsen *et al.*, 2014). The audibility criterion was set to 15 dB at and below 3 kHz which was reduced to 4 dB at 8 kHz by logarithmic interpolation at the intermediate frequencies. SRTs corresponding to the 50% sentence correct values were tracked by adapting the masker level in 2-dB steps. SRTs were estimated based on the performance over one list in each condition. The speech tests were performed in two sessions and subjects were trained on 3 lists before each visit. We tested the S_1, C_2, and C_4 conditions during the first and the C_8 and D_2 conditions during the second visit. Within each visit, the presentation order of the conditions was balanced using a latin-square design. List numbers used for the target sentences were balanced between conditions with the same technique.

Temporal processing

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, respectively.

The FDT test was similar to that of Papakonstantinou *et al.* (2011). A 3-interval 3-alternative forced-choice (3I-3AFC) paradigm was applied in combination with a multiplicative one-up two-down tracking rule. Listeners had to indicate the target tone that had a higher frequency than the two references, which were presented at 250 Hz. The initial difference between target and reference was set to 25%, and the initial step-size to 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 geometrical 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. 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 pick the target stimulus containing an interaural phase shift of $\Delta\phi$ degrees in a 2I-2AFC task 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 ramped with a 20-ms long Hann window and separated by 100-ms silent intervals. For the reference stimuli, each of the four tones had 0° interaural phase. 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 was 3.375 and was decreased to 2.25 and 1.5 after the first and second reversals. 8 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 SL.

RESULTS

The SRTs for the NH (white), HI_{mild} (light grey), and HI_{mod} (dark grey) groups are shown in Fig. 1. In the box plots, the thick black lines denote the medians and the boxes extend to the 25^{th} and 75^{th} percentiles. The thin lines extend to the most extreme data points within 1.5 interquartile ranges from the 25^{th} and 75^{th} percentiles. A repeated-measures ANOVA was performed on the SRTs with masker type and spatial distribution as within-subject factors and listener group as between-subject factor. The degrees of freedom were adjusted with Greenhouse-Geisser correction where the assumption of sphericity was violated.

In most of the tested conditions, NH listeners performed the best, followed by the HI_{mild} group and then by the HI_{mod} listeners. This was supported by the significant main effect of listener group [$F(2,26)=24.171$, $p<0.001$]. Differences between groups were smallest in the S_1 conditions and greatest in the C_2 conditions. NH listeners yield the lowest SRTs in the C_2 conditions, while HI listeners performed best in the S_1 conditions. Despite the inherent spectro-temporal fluctuations in the C_8 backgrounds, all groups had elevated thresholds as compared to the stationary S_1 conditions. While the NH listeners performed better as the number of reversed interferers decreased from 8 to 4 to 2, the HI listeners performed similarly in all of these conditions. Consistent with these observations, SRTs differed significantly between the various masker types [$F(3.11,80.95)=28.02$, $p<0.001$], and the interaction between masker type and listener group was also significant [$F(6.23,80.96)=5.03$, $p<0.001$].

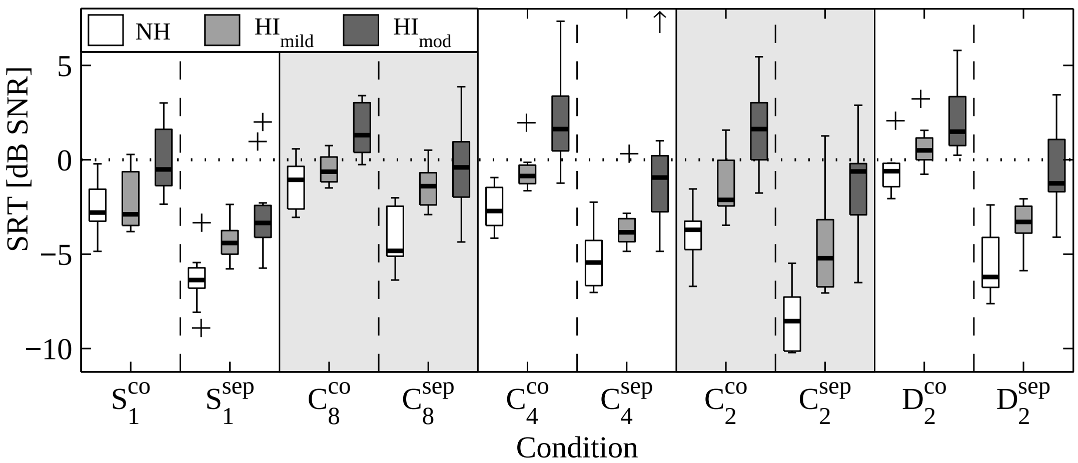

Fig. 1: SRTs across conditions (white: NH, light grey: HI_{mild}, dark grey: HI_{mod}). Shaded areas denote condition groups with the same type of background noise, while the dashed lines separate the spatial distributions within the noise groups (left: co-located, right: separated).

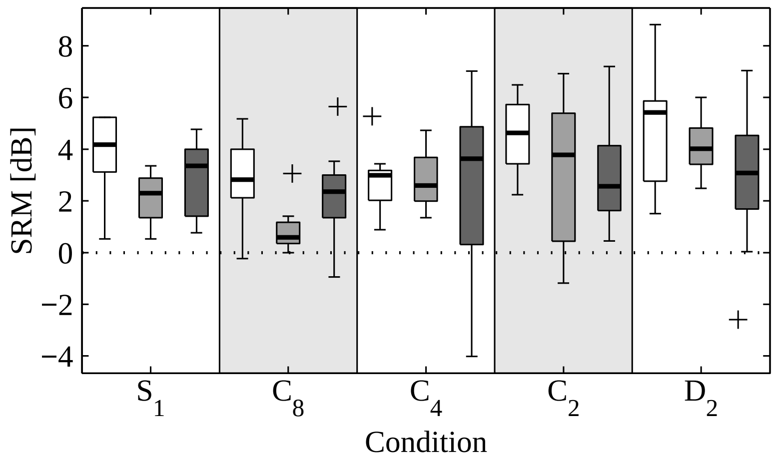

Fig. 2: SRMs across conditions (white: NH, light grey: HI_{mild}, dark grey: HI_{mod}). SRM was calculated as the difference in SRTs between the co-located and separated spatial configurations within noise condition groups.

In Fig. 2 the spatial release from masking (SRM), calculated as the difference in SRTs between spatially separated and co-located target-masker conditions, is shown for the different noise conditions. All of the listener groups benefited from spatially separating the maskers. The benefit varied between 1 and 5 dB, depending on noise condition and listener group. Except for the C_4 condition group, the NH listeners benefited the most from spatial separation. On average, the NH group yielded 3.78 dB

SRM, while the HI_{mild} and HI_{mod} groups yielded 2.66 and 2.62 dB, respectively. The main effect of spatial distribution was significant [$F(1,26)=311.41$, $p<0.001$]. On average, the greatest SRM was obtained in D_2 and the smallest in C_8. Both interactions between spatial distribution and masker type [$F(3.08,79.96)=4.57$, $p=0.005$] and between spatial distribution and listener group [$F(2,26)=4.91$, $p=0.016$] were significant.

The FDT and IPDT results are presented in Fig. 3. For both measures, the NH group performed significantly better than the HI group (HI_{mild} and HI_{mod} collapsed). The mean thresholds were 1.22 and 2.87 percent in the FDT (two-tailed t-test: $p=0.014$), and 11.47 and 19.5 degrees in the IPDT experiment (two-tailed t-test: $p=0.0186$), respectively.

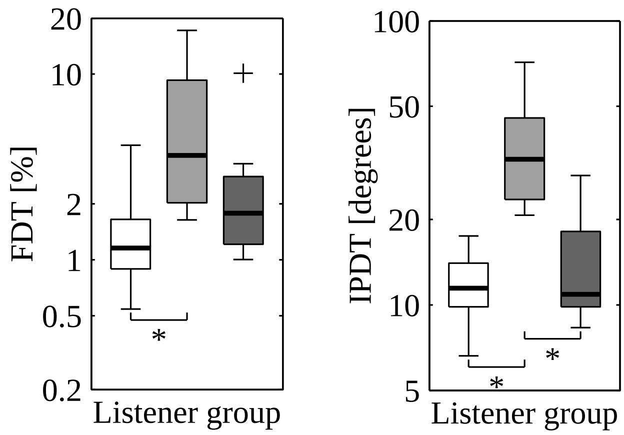

Fig. 3: FDT (left) and IPDT (right) results for NH (white), HI_{mild} (light grey) and HI_{mod} (dark grey). Asterisks denote statistically significant differences in means.

Figure 4 shows scatter plots of SRTs averaged across all conditions (SRT_{avg}) vs. PTAs at octave frequencies from 0.5 to 4 kHz (left panel), average SRTs across co-located conditions (SRT_{co}) vs. FDT (middle panel), and the average SRM benefit of all conditions (SRM_{all}) vs. IPDT (right panel). The correlations between FDT and SRT_{co} and between IPDT and SRM_{all} were not significant. The correlation between PTA and SRT_{avg} was significant when the HI_{mild} and HI_{mod} groups were pooled together ($r=0.55$, $p=0.015$). The slope of the regression line was 0.11, showing that, on average, a 9-dB increment in PTA yielded about 1-dB increment in SRT. This correlation was not significant when the HI_{mild} and HI_{mod} groups were considered separately.

DISCUSSION

The aim of the current study was to investigate the relationship between TFS processing and speech perception in lateralized conditions. We hypothesized that reduced FDTs and IPDTs would be associated with elevated SRTs in conditions where target and maskers are co-located or with reduced SRMs, respectively. Individualized linear gains were applied to all speech stimuli to reduce possible confounds due to stimulus inaudibility.

The HI listeners showed elevated SRTs as compared to the NH population, and the differences were largest in fluctuating masker condtions. No correlation was found between SRTs in the co-located conditions and FDTs. This contradicted our hypothesis and some previous results (Papakonstantinou *et al.*, 2011). Instead, SRTs were positively correlated with audiometric thresholds. It is still likely 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. One such reason could be the broadening of auditory filters at the higher sound pressure levels of the stimuli persented to the HI subjects (Studebaker *et al.*, 1999).

HI listeners experienced less SRM than NH, but the difference was small. Thus, HI listeners retain some benefit from large ITDs differences between target and maskers. While the HI group performed worse in the IPDT experiments, the IPDTs and SRM scores were not correlated. One reason for this could be the relatively large ITDs used to trigger different spatial positions. These time differences were clearly detectable to almost all of our HI subjects at 250 Hz. 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.

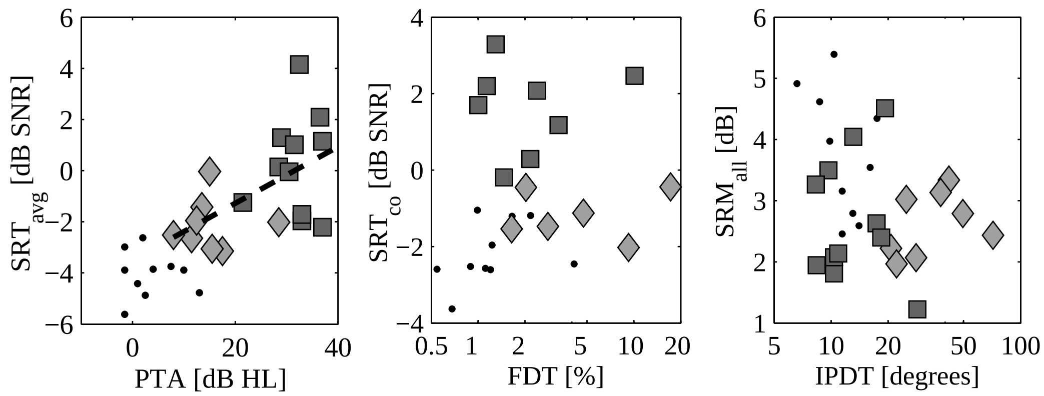

Fig. 4: Scatter plots between audiometric thresholds or TFS coding and speech reception performance (dots: NH, diamonds: HI_{mild}, squares: HI_{mod}). The dashed regression line was fitted to the data of the HI group (HI_{mild} and HI_{mod} collapsed). See text for further details.

The 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. 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 have been observed in earlier studies as well (Hopkins and Moore, 2011; Papakonstantinou *et al.*, 2011). Given the relatively small number of subjects in each group, it might be that this distribution of the data was just a casual result of partitioning the HI group into two subgroups.

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 been 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 auditiory filters at higher presentation levels), as the audibility of the target stimuli has been individually compensated for. The amount of SRM was smaller for HI than for NH listeners, but only in the order of 1 dB. Low-frequency IPDTs did not correlate with SRM. SRMs in an experimental paradigm applying smaller ITDs to separate target from maskers would be more limited by IPDTs at low frequencies and may thus be a more sensitive measure to investigate the effect of binaural TFS processing on spatial speech perception.

REFERENCES

Cooke, M., Barker, J., Cunningham, S., and Shao, X. (**2006**). “An audio-visual corpus for speech perception and automatic speech recognition,” J. Acoust. Soc. Am., **120**, 2421-2424.

Hopkins, K. and Moore, B.C.J. (**2010**). “Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies,” Int. J. Audiol., **49**, 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,” J. Acoust. Soc. Am., **130**, 334-349.

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,” Ear Hearing, **29**, 907-922.

Nielsen, J., Dau, T., and Neher, T. (**2014**). “A Danish open-set speech corpus for competing-speech studies,” J. Acoust. Soc. Am., **135**, 407-420.

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?” J. Acoust. Soc. Am., **130**, 1542-1558.

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

Strelcyk, O. and Dau, T. (**2009**). “Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing,” J. Acoust. Soc. Am., **125**, 3328-3345.

Studebaker, G.A., Scherbecoe, R.L., and McDaniel, D.M., and Gwaltney, C. (**1999**). “Monosyllabic word recognition at higher-than-normal speech and noise levels,” J. Acoust. Soc. Am., **105**, 2431-2444.