

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

GUSZTÁV LŐCSEI¹, SÉBASTIEN SANTURETTE^{1,2}, TORSTEN DAU¹,
AND EWEN N. MACDONALD^{1,*}

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

² *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

Spatial release from masking (SRM) elicited by interaural timing differences (ITDs) only can be almost normal for listeners with symmetrical hearing loss. This study investigated whether elderly hearing-impaired (HI) listeners still achieve similar SRMs as young normal-hearing (NH) listeners, when SRMs are elicited by small ITDs. Speech reception thresholds (SRTs) and SRM due to ITDs were measured over headphones for 10 young NH and 10 older HI listeners, who had normal or close-to-normal hearing below 1.5 kHz. Diotic target sentences were presented in diotic or dichotic speech-shaped noise or two-talker babble maskers. In the dichotic conditions, maskers were lateralized by delaying the masker waveforms in the left headphone channel. Multiple magnitudes of masker ITDs were tested in both noise conditions. Although deficits were observed in speech perception abilities in speech-shaped noise and two-talker babble in terms of SRTs, HI listeners could utilize ITDs to a similar degree as NH listeners to facilitate the binaural unmasking of speech. A slight difference was observed between the group means when target and maskers were separated from each other by large ITDs, but not when separated by small ITDs. Thus, HI listeners do not appear to require larger ITDs than NH listeners do in order to receive a benefit from binaural unmasking.

INTRODUCTION

If a target and maskers are separated in space, the intelligibility of the target typically improves, a phenomenon termed spatial release from masking (SRM). While SRM is mainly facilitated by better-ear listening, binaural unmasking (BU) can also play a role. Several studies have found normal or close-to-normal binaural intelligibility level difference (BILDs) in hearing-impaired (HI) listeners with symmetrical hearing loss (Bronkhorst and Plomp, 1989; Strelcyk and Dau, 2009; Lőcsei *et al.*, 2016). These results are surprising given that HI listeners usually exhibit degraded temporal fine

*Corresponding author: emcd@elektro.dtu.dk

structure (TFS) processing. However, most studies that investigate BILDs in normal-hearing (NH) and HI listeners use relatively large interaural time differences (ITDs).

In the present study, binaural intelligibility level differences (BILD) were measured for speech stimuli embedded in noise and separated by either large or small ITDs for a group of young NH listeners and older HI listeners in a series of headphone experiments. The hypothesis was that deficits in BU abilities in HI listeners, as measured by BILDs, should be more prominent when triggered by small ITDs than by large ITDs. In addition to BILDs, TFS interaural phase difference (IPD) thresholds were measured in pure-tone carriers over a range of frequencies. BILDs in the large and small ITD conditions were compared between the listener groups and the resulting IPD threshold profiles were contrasted with the size of BILDs in both cases.

METHODS

Participants

Ten young NH (20-27 years, mean: 23, standard deviation (SD): 2.31) and ten older HI (50-76 years, mean: 66.9, SD: 7.48) listeners participated in the study (see Table 1).

ID	Sex	Age	Audiometric thresholds averaged between the ears [dB HL]										
			125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k
a	m	60	7.5	0	5	5	0	0	7.5	25	32.5	27.5	35 ⁺
b	m	50	2.5	0	5	5	0	0	10	25	37.5	35	45
c	m	67	22.5	15	7.5	2.5	2.5	0	2.5	22.5	30	32.5*	65
d	f	65	7.5	2.5	5	10	10	7.5	7.5	12.5	22.5	32.5*	45
e	f	72	10	5	7.5	5	7.5	15	22.5	40	32.5	52.5	67.5
f	f	66	15	15	12.5	15	12.5	20	37.5*	50	45	50	65
g	f	72	22.5	12.5	15	15	17.5	17.5*	20 ⁺	25	27.5	35	52.5
h	f	69	5	5	12.5	17.5	22.5	32.5	40	47.5	52.5	55	60
i	m	76	15	12.5	20	20	15	25	25	45	57.5	52.5	67.5
j	f	72	15	25	22.5	20	20	27.5	32.5	32.5	40	45	65

Table 1: Gender, age, and audiometric thresholds (air-conduction, averaged across both ears) of the HI listeners. In some cases, differences in audiometric thresholds between left and right ears were as large as 15 dB (*) or 20 dB (+). In all other cases, these differences were less than or equal to 10 dB.

Binaural fine structure processing

In the measurements assessing sensitivity to binaural TFS information, the task of the listeners was to detect IPDs of pulsating pure-tones at different frequencies. Thresholds were estimated using a 3-interval 3-alternative forced-choice paradigm. Each interval contained a sequence of four 200-ms pure tones presented at the same frequency, separated by 100-ms silent gaps. The gaps between presentation intervals were 400 ms long. In the reference intervals, all of the tones were presented diotically. In the target interval, the first and third tones were presented with zero IPD, and the

second and fourth tones with a starting phase of $-\frac{\Delta\phi}{2}$ and $\frac{\Delta\phi}{2}$ in the left and right channels of the headphones, respectively, yielding a total IPD of $\Delta\phi$.

For each listener, the frequency range at which an IPD of 180° could be detected (IPD_{fr}) was measured first. Thereafter, IPD thresholds at fixed frequencies ranging from 250 Hz up to IPD_{fr} were measured in 250-Hz steps (IPD_{lf} experiments). Presentation levels were set to 30 dB sensation level (SL).

Speech perception in noise

Speech intelligibility was evaluated using the DAT corpus (Nielsen *et al.*, 2014), both in speech-shaped noise (SSN) and in an interfering two-talker background (TT). In the SSN condition, the “Dagmar” sentences were used as target material, and the long-term average spectrum of the noise was matched to that of the “Dagmar” sentences. To avoid repeating any lists within the experiment, the “Asta” sentences were used in the TT condition. In these cases, sentences spoken by the two other talkers were applied as maskers. No spectral matching was applied between target and maskers in the TT conditions. The SSN tokens were semi-randomly chosen from a pool of fifty 5-second noise samples, which were then truncated to match with the length of the target sentence. The TT maskers started at the same time as the target but could end earlier or later than the target. The target sentences were always presented diotically while the maskers were delivered in one of the following lateralization settings: (1) diotic presentation, colocated with the target (SSN_{co} and TT_{co}), (2) lateralized to the side through large ITDs of 0.68 ms (SSN_{lrg} and TT_{lrg}), or (3) lateralized to the side through small ITDs of 0.27 ms (SSN_{sm}). SRTs were measured adaptively using 20 sentence lists. In the TT_{sm} condition, instead of measuring SRTs at a fixed ITD, the 50% sentence-correct point was tracked as a function of ITD at a fixed SNR. The SNR for this condition was set to 3 dB lower than each individuals’ SRT in the TT_{co} condition. Thus, TT_{sm} tracks the ITD needed to achieve a BILD of 3 dB.

All stimuli were delivered via headphones. Audibility of the stimuli was restored by applying individualized linear gains based on the individual listeners’ audiogram and on the long-term average spectrum of the “Dagmar” sentences. The audibility criterion was set to be 15 dB above the individual hearing thresholds for one-third octave bands between 110 Hz and 3 kHz, which was reduced to 12, 8, and 0 dB at 4, 6, and 8 kHz. Then, the target stimulus was scaled to a nominal level of 65 dB SPL when measured at the eardrums of a HATS and mixed with the scaled maskers. The individualized gains were applied to this mixture amplifying both target and maskers. These filters also compensated for the headphone frequency response. Presentation levels were limited to 94 dBA and if the estimated overall presentation level of a stimulus exceeded this, it was downscaled in 2-dB steps.

RESULTS

The IPD thresholds measured at fixed frequencies are shown in Fig. 1 for both the NH (dots) and HI (letters) listeners. The solid horizontal black lines denote the group

means and the corresponding boxes represent ± 1 SD. Significant differences were confirmed between the log-transformed group means for IPD_{250} [$t(18) = -2.79, p = 0.012$]. Note, however, that in the IPD_{fr} tests at frequencies at or above 750 Hz, thresholds could not be measured for all of the HI listeners, biasing the group means towards lower values than the true group average. This is also clearly reflected in Fig. 2, which shows the results of the IPD_{fr} experiment. Differences in group means were significant for the IPD_{fr} [$t(18) = 5.67, p < 0.001$] thresholds, and also for the ITD_{min} thresholds [$t(10.16) = -3.234, p < 0.009$].

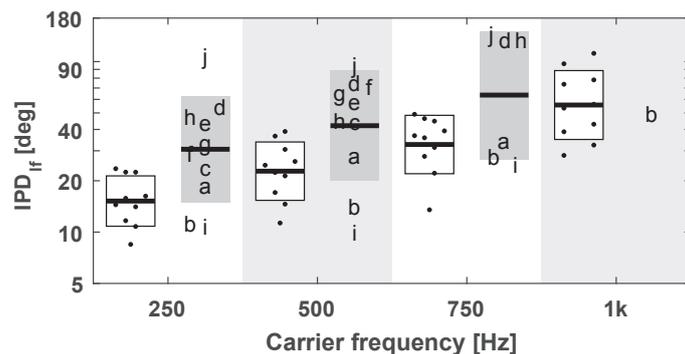


Fig. 1: IPD_{fr} thresholds for the NH (dots) and HI (letters) listener groups. Black horizontal lines mark group means and the boxes denote ± 1 SD of the corresponding groups. The shading of the background is according to the conditions with different carrier frequencies.

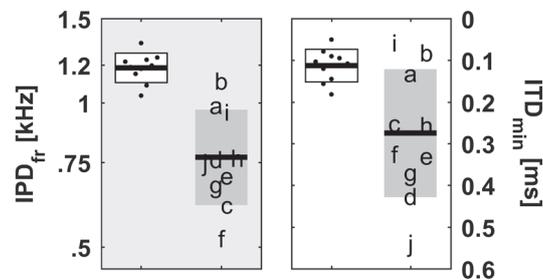


Fig. 2: Maximum frequency for detection of 180° IPD and minimum ITD thresholds (ITD_{min}) of the NH (dots) and HI (letters) listeners. Black horizontal lines mark group means and the boxes denote ± 1 SD of the corresponding groups. Note that the y-axis in the right panel is reversed, so that data points located further towards the top of each panel represent better performance.

Figure 3 shows the SRTs for the NH and the HI listeners obtained in the fixed-ITD conditions. A mixed-design ANOVA was conducted on the SRT data for the SSN_{co} , SSN_{lrg} , TT_{co} and TT_{lrg} conditions. The model contained the SRTs as the dependent

variable, and used noise type (SSN or TT) and lateralization (co or lrg) as within-subject factors and listener group (NH or HI) as between-subject factors. All main effects and two-way interactions were significant.

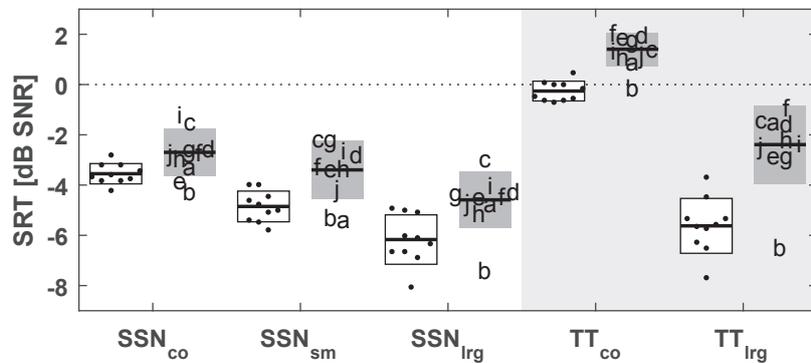


Fig. 3: SRTs in SSN and two-talker babble (TT) for NH (dots) and HI (letters) listeners. Solid black horizontal lines indicate group means and the boxes denote ± 1 SD. The background shadings indicate condition groups using the same type of background noise. In each condition the target was presented diotically. The different test conditions are denoted on the x-axis. Subscripts indicate the ITD configuration of the masker: **co**: diotic presentation, colocated with the masker; **sm**: masker lateralized with a small ITD (0.27 ms); **lrg**: masker lateralized with a large ITD (0.68 ms).

The measures characterizing BU of speech are plotted in Fig. 4. In the left panel, the BILDs due to masker lateralization are plotted, which were calculated as the difference in SRTs between the co and the sm or lrg conditions. The right panel indicates the results obtained in the TT_{sm} condition, which indicates the ITD needed to achieve a BILD of 3 dB. In general, the NH listeners showed a slightly better performance than the HI listeners in all conditions. For BILDs at fixed ITDs, a statistically significant interaction between lateralization and listener group [$F(1, 18) = 8.81, p = 0.008$] was observed in the ANOVA model. Most listeners benefitted from masker lateralization in all of the tested conditions. While BILDs were small in the SSN_{sm} condition, they increased as the ITD magnitudes of the maskers increased from 0.27 to 0.68 ms. The benefit was greatest in the TT_{lrg} condition, where it reached 5.4 dB and 3.8 dB for the NH and HI listeners, respectively. It appears that NH listeners exhibited greater BILDs in the conditions with fixed ITDs and a 3-dB BILD at smaller ITDs than the HI listeners. However, independent t -tests on the BILD data indicated that the group differences were only statistically significant in the TT_{lrg} condition [$t(18) = 3.03, p = 0.007$].

Pearson's correlations were calculated between each of the four measures of BU and the ITD_{min} or IPD_{fr} results within the group of HI listeners. None of the correlations were significant.

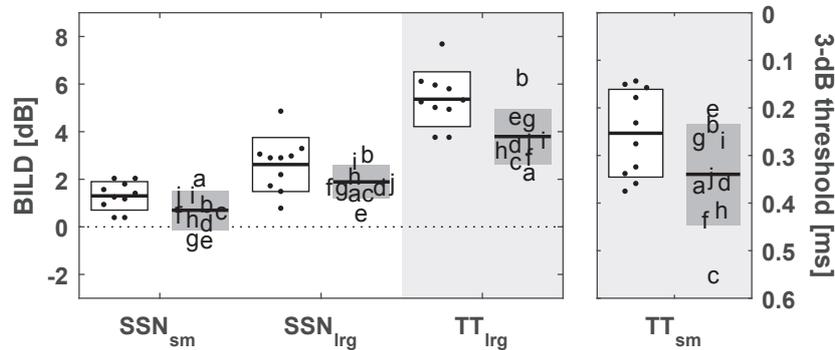


Fig. 4: BILDs at fixed ITDs in SSN and two-talker babble (TT) and the ITD threshold needed to yield a fixed 3-dB BILD in the TT noise (i.e., a 3-dB decrease in SRTs as compared to the TT_{co} condition). Solid horizontal black lines and the boxes around denote group means and ± 1 SD for the NH (dots) and HI individuals (letters). Background shadings mark condition groups with the same noise type. Note that the first 3 conditions to the left are expressed in dB, while the last condition in ms. Condition notations are the same as in Fig. 3.

DISCUSSION

In the present study, HI listeners exhibited poorer thresholds compared to NH listeners for the binaural measures of TFS processing IPD_{fr} , IPD_{lf} and ITD_{min} . These results are consistent with previous studies (Ross *et al.*, 2007; Hopkins and Moore, 2011; Neher *et al.*, 2011; King *et al.*, 2014).

In the speech experiments, both groups exhibited lower average SRTs in SSN than in TT noise. Listeners in both groups showed a clear benefit when the maskers were lateralized to the side, indicating the presence of an active BU mechanism. The amount of BILDs differed slightly between the two groups, and this difference was only statistically significant in the TT_{lrg} condition. Therefore, the results obtained in the TT conditions do not support the hypothesis that the HI listeners' processing deficits in BU are more pronounced when triggered by small rather than by large ITDs. Rather, the deficits in the BU of speech manifested themselves mainly by reducing the overall benefit HI listeners could achieve when target and maskers were separated by large ITDs. Nonetheless, the SRTs obtained in the SSN_{lrg} and TT_{lrg} conditions suggest the possibility that BILDs in the TT_{lrg} condition were, at least partly, affected by monaural deficits in temporal processing. The SRTs in the TT conditions were different from those in the SSN conditions as the two maskers differ in the amount of modulation and informational masking. While the NH listeners yielded similar SRTs in the TT_{lrg} and SSN_{lrg} conditions, the HI listeners had about 2-dB higher SRTs in the TT_{lrg} condition than in the SSN_{lrg} condition. However, informational masking is substantially reduced when target and maskers are spatially separated (Arbogast *et*

al., 2002). Therefore, the performance in the TT_{lrg} condition can be assumed to be limited by factors other than informational masking (c.f. Best *et al.*, 2002). Several studies have shown that HI listeners are more susceptible to modulation masking than NH listeners, which manifests itself in less-than-normal fluctuating-masker benefit when modulations are imposed on a stationary masker (Festen and Plomp, 1990; Strelcyk and Dau, 2009). Therefore, it is possible that, compared to the NH listeners, the HI listeners would have elevated thresholds in the TT_{lrg} condition due to their susceptibility to modulation masking, even if they had intact binaural processing abilities. The extent to which such monaural factors might have contributed to the reduced BILDs in the current study is nonetheless difficult to evaluate, as it is likely that both informational and modulation masking are involved in the TT_{co} and TT_{lrg} conditions.

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

CONCLUSIONS

HI listeners showed a reduction in binaural TFS coding abilities compared to NH listeners, as reflected in a reduction of the IPD_{fr} and an increase of the ITD_{min} thresholds. Although deficits were observed in speech perception abilities in SSN and two-talker babble in terms of SRTs, HI listeners could utilize ITDs to a similar degree as NH listeners to facilitate the binaural unmasking of speech. A slight difference was observed between the group means when target and maskers were separated from each other by large ITDs, but not when separated by small ITDs. Therefore, HI listeners did not experience greater difficulties in terms of reduced BILDs when spatial differences between target and maskers were induced by small ITDs.

ACKNOWLEDGEMENTS

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

REFERENCES

- Arbogast, T.L., Mason, C.R., and Kidd, G. (2002). "The effect of spatial separation on informational and energetic masking of speech," *J. Acoust. Soc. Am.*, **112**, 2086-2098. doi: 10.1121/1.1510141
- Best, V., Thompson, E.R., Mason, C.R., and Kidd, G. (2013). "An energetic limit on spatial release from masking," *J. Assoc. Res. Otolaryngol.*, **14**, 603-610. doi: 10.1007/s10162-013-0392-1
- Bronkhorst, A.W., and Plomp, R. (1989). "Binaural speech intelligibility in noise for hearing-impaired listeners," *J. Acoust. Soc. Am.*, **86**, 1374-1383. doi: 10.1121/1.398697
- Festen, J.M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.*, **88**, 1725-1736. doi: 10.1121/1.400247
- 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.
- King, A., Hopkins, K., and Plack, C.J. (2014). "The effects of age and hearing loss on interaural phase difference discrimination," *J. Acoust. Soc. Am.*, **135**, 342-351. doi: 10.1121/1.3585848
- Lócssei, G., Pedersen, J.H., Laugesen, S., Santurette, S., Dau, T., and MacDonald, E.N. (2016). "Temporal fine-structure coding and lateralized speech perception in normal-hearing and hearing-impaired listeners," *Trends Hear.*, **20**, 1-15. doi: 10.1177/2331216516660962
- 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.
- 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. doi: 10.1121/1.4835935
- Ross, B., Tremblay, K.L., and Picton, T.W. (2007). "Physiological detection of interaural phase differences," *J. Acoust. Soc. Am.*, **121**, 1017-1027. doi: 10.1121/1.2404915
- Strelec, 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. doi: 10.1121/1.3097469