

# Contribution of low- and high-frequency bands to binaural unmasking in hearing-impaired listeners

GUSZTÁV LŐCSEI<sup>1</sup>, SÉBASTIEN SANTURETTE<sup>1,2</sup>, TORSTEN DAU<sup>1</sup>,  
AND EWEN N. MACDONALD<sup>1,\*</sup>

<sup>1</sup> *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

<sup>2</sup> *Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Rigshospitalet, Copenhagen, Denmark*

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

## INTRODUCTION

Studies investigating the binaural unmasking (BU) of speech in noise have typically found that binaural intelligibility level differences (BILDs) are determined by interaural timing differences (ITDs) in the low-frequency domain (e.g., Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005) suggesting that ITD cues related to temporal fine-structure (TFS) rather than the envelope (ENV) carry critical information for the BU of speech. While hearing impaired (HI) listeners often exhibit reduced TFS sensitivity, several studies (e.g., Neher *et al.*, 2012; Santurette and Dau, 2012; Lőcsei *et al.*, 2016) have shown no or only a moderate correlation between speech intelligibility scores in spatial settings and behavioural measures of TFS ITD sensitivity. A possible explanation for the weak relationship between TFS

---

\*Corresponding author: emcd@elektro.dtu.dk

ITD sensitivity and BILDs in various conditions can be that HI listeners utilize ITD cues in the ENV of high-frequency channels.

Edmonds and Culling (2005) investigated how ITDs in isolated frequency bands contributed to BILDs in young normal-hearing (NH) listeners. Their results indicated that ITDs in the frequency regions both below and above 1.5 kHz provided some masking release, but also that the full advantage was only achieved when ITDs were present over the full spectrum. Therefore, it appears that NH listeners exploit ENV ITDs at higher frequencies to aid speech perception.

In the current study, the contribution of TFS and ENV ITDs in different frequency regions to BILD was evaluated in young NH and older HI listeners. BILDs were measured in a speech-on-speech task. The target and the interferers were divided into two independent low- and high-frequency regions, and ITDs were imposed on the interferers in the low, high, or both frequency domains.

## METHODS

### Listeners

Six young NH (mean age: 24.2 years, standard deviation (SD): 2.2) and 9 older HI (mean age: 69.6 years, SD: 5.5) participated in the study. For each listener, air-conduction audiometric thresholds were measured at octave frequencies between 125 Hz and 8 kHz and between 750 Hz and 6 kHz. All NH listeners had thresholds below 25 dB HL at all measured frequencies. Most of the HI listeners had normal hearing below 1.5 kHz, but a mild-to-moderate hearing loss at frequencies above 1.5 kHz. In all listeners, the hearing thresholds between the ears differed by at most 15 dB at each tested audiometric frequency. The average hearing thresholds for the HI listeners are displayed in Table 1. All listeners provided written consent and received compensation for their efforts. All but one listener were tested over a single visit lasting between two and three hours. One NH listener was tested over two visits.

ID	Sex	Age	Audiometric thresholds averaged between the ears [dB HL]											Pure-tone averages		
			125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k	PTA <sub>low</sub>	PTA <sub>high</sub>	PTA <sub>oct</sub>
a	f	58	5	0	2.5	2.5	2.5	-5	7.5	17.5	30	35	17.5	0.5	21.5	8.5
b	f	66	12.5	7.5	7.5	10	10	5	10	10	25	25	40	8	22	12
c	f	75	22.5	15	10	7.5	5	5	15	15	22.5	17.5	50	8.5	24	13.5
d	m	68	20	12.5	10	7.5	7.5	5	5	22.5	32.5	37.5	67.5*	8.5	33	13.5
e	f	72	10	5	10	7.5	5	15	22.5*	40	30	45	72.5	8.5	42	14.5
f	f	67	17.5*	10	12.5	7.5	7.5	12.5*	35	50	55	52.5	50	10	48.5	24
g	f	72	22.5	12.5	20	15	20	22.5*	22.5*	27.5	27.5	30	52.5	18	32	20.5
h	m	76	15	12.5	20	20	15	25	25	45	57.5	52.5	67.5	18.5	49.5	26
i	f	72	17.5	25	22.5	20	20	22.5	27.5*	37.5*	40	52.5*	65	22	44.5	27

**Table 1:** Gender, age, and audiometric thresholds (air-conduction, averaged across both ears), and pure-tone averages of the HI listeners. Except for cases marked by asterisks, the differences in audiometric thresholds between left and right ears were less than or equal to 10 dB.

### **Binaural temporal fine structure coding**

The listeners' sensitivity to binaural TFS information was assessed by measuring the upper frequency limit at which listeners were able to detect an interaural phase shift of  $180^\circ$  ( $IPD_{fr}$ ) using the same 3-interval 3-alternative forced-choice paradigm as Lőcsei *et al.* (2016). Reference and target intervals were presented at 30 dB sensation level (SL) and contained 4 tone bursts presented diotically or alternated between the diotic and dichotic presentation modes. The tone bursts were 300-ms long, gated with 50-ms raised-cosine ramps and separated by 100-ms silent gaps. The intervals were separated by 400-ms silent gaps. Six thresholds were evaluated for each listener, and the final threshold was calculated as the geometric mean of the last 3 thresholds.

### **Speech intelligibility tests**

Speech intelligibility was assessed using the open-set DAT corpus (Nielsen *et al.*, 2014). The "Dagmar" sentences were presented via headphones as target material against a two-talker masker (TT), which consisted of sentence pairs spoken by the two other talkers of the same corpus.

The target sentences were always presented diotically. In the reference condition, the maskers were colocated with the target ( $TT_{co}$ ). In the remaining conditions, the maskers were lateralized towards the right side by imposing a 0.68-ms timing delay in the left channel. This delay was either imposed on the full spectrum ( $TT_{bb}$  for "broad band"), the low spectral region ( $TT_{lp}$ , "low pass"), or the high spectral region ( $TT_{hp}$ , "high pass") of the maskers. In order to manipulate the ITD relations in low- and high-frequency bands independently, low-pass and high-pass filtered versions of the original maskers were created and the time delays were applied to the left-ear channel in the corresponding frequency regions. The resulting low- and high-frequency time signals were then added in each channel and presented to the listener. The cutoff frequencies of the low-pass and high-pass filters were set to 1173 Hz and 1332 Hz, respectively, corresponding to a 1 equivalent rectangular band (ERB) notch centered at 1.25 kHz between the low-pass filtered and high-pass filtered parts. Filter slopes were greater than 500 dB/oct in both cases in order to prevent any interactions between the two spectral regions.

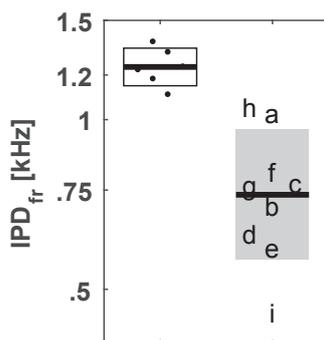
In each condition, sentence correct SRTs were measured by varying the level of the maskers in 2-dB steps. The initial signal-to-noise (SNR) ratio was set to 3 dB in the  $TT_{co}$  and to 0 dB in all the other conditions. When calculating the SRTs for each list, the presentation levels of the maskers from the 5th to the hypothetical 21st sentence were averaged and subtracted from the presentation level of the target. SRTs were measured over 3 lists in each condition and the final SRT value for each condition was calculated as the average of these, expressed in SNR. Overall, 12 lists were used in the testing phase, and 3 additional lists in the training phase, two of which were presented in the  $TT_{co}$  condition and one in the  $TT_{bb}$  condition.

The frequency range of the stimuli was restricted to between 100 Hz and 10 kHz.

A 512 order finite impulse response filter was used to compensate for the frequency response of the headphones, and to simulate the frequency response of the outer ear in a diffuse-field listening scenario. This filter also compensated for the loss of stimulus audibility based on the hearing thresholds of the individuals and the long-term average spectrum of the target speech. The audibility criterion used was 15 dB at and below 3 kHz, which was reduced to 12 dB, 8 dB and 0 dB at 4, 6 and at 8 kHz and above. The target sentences were presented at a nominal level of 65 dB sound pressure level (SPL) “free field”. The presentation levels were limited to at most 94 dBA. If the estimated level of a trial exceeded this level, it was scaled down in 2 dB steps to be below this upper limit.

## RESULTS

The results of the  $IPD_{fr}$  experiments are shown in Fig. 1. Horizontal black lines denote the group means and the white and gray boxes indicate  $\pm 1$  SD of the NH and HI listener groups, respectively. For analysis purposes, the thresholds were log-transformed. The difference in average  $IPD_{fr}$  thresholds between groups was statistically significant [ $t(13) = -4.65, p < 0.001$ ].

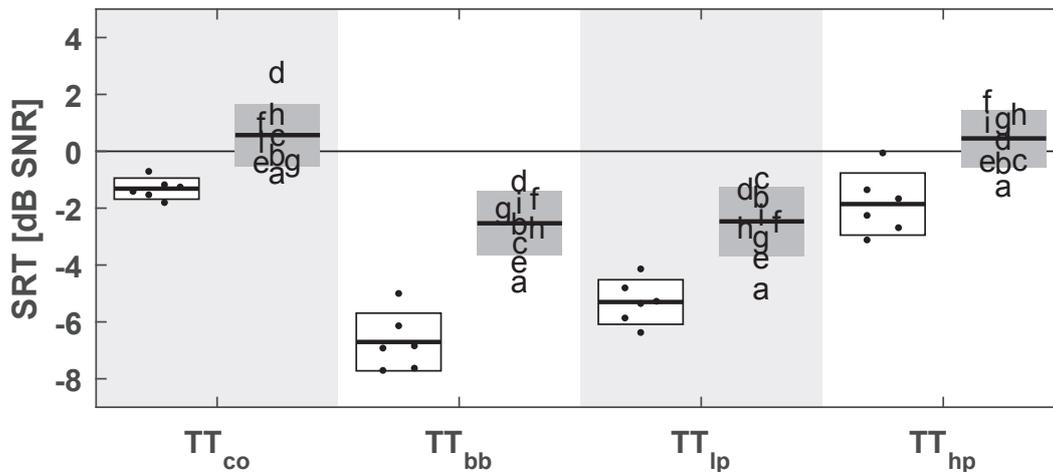


**Fig. 1:** The upper frequency limit for detecting a  $180^\circ$  interaural phase difference in the TFS of a pure tone. The dark horizontal lines with the white and gray boxes stand for the mean and  $\pm 1$  SD of the NH and HI listener groups, respectively. Dots and letters denote individual thresholds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

The lowest and highest thresholds in the HI group were 456 Hz and 1056 Hz, showing a large spread of how the individual listeners performed. Within the HI group, neither age nor  $PTA_{low}$  was correlated with the  $IPD_{fr}$  thresholds. The HI listeners showed similar binaural TFS processing abilities as the HI listeners tested in Lőcsei *et al.* (2016).

Figure 2 shows the group means and standard deviations of the SRTs for the two listener groups (NH: white boxes, HI: gray boxes) in the SI experiments. SRTs in the  $TT_{bb}$  and  $TT_{lp}$  conditions were lower than in the  $TT_{co}$  condition, indicating a masking release when ITDs were imposed at least on the low-frequency part of the maskers. In contrast, SRTs were similar in the  $TT_{co}$  and  $TT_{hp}$  conditions for both listener groups. For the NH listeners, SRTs were slightly higher in the  $TT_{lp}$  condition than in the  $TT_{bb}$  condition. In contrast, HI listeners performed similarly in the two conditions.

The BILDs were calculated as the difference between SRTs in the  $TT_{co}$  and all the other conditions. The group means and standard deviations are shown in Fig. 3. Group



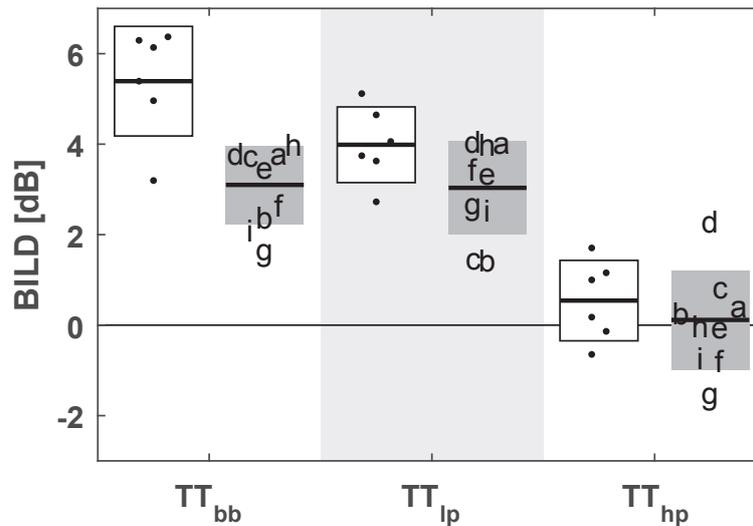
**Fig. 2:** Speech reception thresholds of the NH and HI listeners in the speech intelligibility tests. The target was always presented diotically. The maskers were presented either diotically (TT<sub>co</sub>), or lateralized to the right side in the low or high frequency domains (TT<sub>lp</sub> or TT<sub>hp</sub>), or over the full frequency domain (TT<sub>bb</sub>). The dark horizontal lines and the white and gray boxes indicate for the mean and  $\pm 1$  SD of the NH and HI listener groups, respectively. Dots and letters denote individual thresholds within the corresponding groups.

differences in BILDs were highest in the TT<sub>bb</sub> condition. Group differences were less pronounced in the TT<sub>lp</sub> condition. HI listeners exhibited similar BILDs in the TT<sub>bb</sub> and TT<sub>lp</sub> conditions.

A mixed-design ANOVA was conducted on the BILDs obtained in the TT<sub>bb</sub> and TT<sub>lp</sub> conditions, with filtering as within-subject and listener group as between-subject factors. The analysis revealed a significant main effect of filtering [ $F(1, 13) = 5.647$ ,  $p = 0.034$ ], listener group [ $F(1, 13) = 14.95$ ,  $p = 0.002$ ], and interaction between filtering and listener group [ $F(1, 13) = 4.69$ ,  $p = 0.0496$ ]. For the NH listeners, there was a trend towards larger BILDs in the TT<sub>bb</sub> than in the TT<sub>lp</sub> condition [paired  $t$ -test,  $t(5) = 2.44$ ,  $p = 0.058$ ]. For both groups, BILDs in the TT<sub>hp</sub> were not significantly different from 0 (one-sample  $t$ -tests,  $p > 0.05$ ).

## DISCUSSION

The results of the present study indicate that, for young NH listeners, frequencies above 1.25 kHz can contribute to the BU of speech, which is consistent with the findings of Edmonds and Culling (2005). Furthermore, it was found that young NH listeners exhibited larger BILDs than older HI listeners when ITDs were imposed on the whole frequency range. Both listener groups exhibited BU when the target and the maskers were separated by ITDs only below 1.25 kHz, and the magnitude of the



**Fig. 3:** Binaural intelligibility level differences of the NH and HI listeners obtained in the speech intelligibility tests. Condition notations are the same as in Fig. 2. The dark horizontal lines with the white/gray boxes show mean BILDs and  $\pm 1$  SD for the NH and HI listener groups, respectively.

BILDs was comparable in the two groups. When ITDs were imposed above, but not below, 1.25 kHz, no BILD was observed. The results suggest that, both in young NH and older HI listeners, BILDs are mainly facilitated in the low-frequency region of the stimuli. This finding is consistent with the conclusions of earlier reports investigating BU (e.g., Levitt and Rabiner, 1967; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005). The contributions of ITDs at high frequencies to BILDs seem to be negligible when presented in isolation. However, in contrast to the HI listeners, NH listeners could utilize high-frequency ITD information to some degree to aid speech understanding in the  $TT_{bb}$  condition. For the NH listeners, the differences in BILDs between the  $TT_{bb}$  and  $TT_{lp}$  were not statistically significant, likely due to the relatively low number of listeners tested.

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

remained similar. Therefore, it is unlikely that the differences in BILDs between the NH and the HI groups were driven by the NH listeners' ability to utilize TFS ITDs above 1.25 kHz.

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

## CONCLUSIONS

BILDs were found to be similar for a group of young NH and older HI listeners when elicited by ITDs below 1.25 kHz, despite the fact that the latter group showed a clear reduction in binaural TFS coding abilities. BILDs were slightly lower for the HI group when triggered by ITDs over the full frequency range of the stimuli. When ITDs were imposed above 1.25 kHz only, no BILDs were found in any of the groups. Overall, the results suggest that, while the young NH listeners might have utilized both TFS ITDs at low frequencies and ENV ITDs at high frequencies to facilitate BU, older HI listeners relied exclusively on ITDs at the low frequencies. It still remains possible that BILDs were affected by the sensitivity to ENV ITDs in the low frequency region.

## ACKNOWLEDGMENTS

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

## REFERENCES

- Bronkhorst, A.W., and Plomp, R. (1988). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.*, **83**, 1508-1516. doi: 10.1121/1.395906
- Edmonds, B.A., and Culling, J.F. (2005). "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay," *J. Acoust. Soc. Am.*, **117**, 3069-3078. doi: 10.1121/1.1880752
- He, N.-j., Mills, J.H., Ahlstrom, J.B., and Dubno, J.R. (2008). "Age-related differences in the temporal modulation transfer function with pure-tone carriers," *J. Acoust. Soc. Am.*, **124**, 3841-3849. doi: 10.1121/1.2998779
- 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: 10.1121/1.4838995
- Lacher-Fougère, S., and Demany, L. (2005). "Consequences of cochlear damage for the detection of interaural phase differences," *J. Acoust. Soc. Am.*, **118**, 2519-2526. doi: 10.1121/1.2032747
- Levitt, H., and Rabiner, L.R. (1967). "Predicting binaural gain in intelligibility and release from masking for speech," *J. Acoust. Soc. Am.*, **42**, 820-829. doi: 10.1121/1.1910654
- 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., Lunner, T., Hopkins, K., and Moore, B.C.J. (2012). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L)," *J. Acoust. Soc. Am.*, **131**, 2561-2564. doi: 10.1121/1.3689850
- 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
- Santurette, S., and Dau, T. (2012). "Relating binaural pitch perception to the individual listener's auditory profile," *J. Acoust. Soc. Am.*, **131**, 2968-2986. doi: 10.1121/1.3689554
- Strouse, A., Ashmead, D.H., Ohde, R.N., and Grantham, D.W. (1998). "Temporal processing in the aging auditory system," *J. Acoust. Soc. Am.*, **104**, 2385-2399. doi: 10.1121/1.423748