

# Release from energetic and informational masking as revealed by listening effort and speech intelligibility

Jan Rennies<sup>1</sup>, Melanie Krüger<sup>2</sup>, Virginia Best<sup>3</sup>, Elin Roverud<sup>3</sup>, Gerald Kidd Jr<sup>3</sup>

<sup>1</sup> Fraunhofer IDMT, Hearing, Speech and Audio Technology & Cluster of Excellence Hearing4all, Oldenburg, Germany, E-Mail: jan.rennies@idmt.fraunhofer.de

<sup>2</sup> Hörzentrum Oldenburg GmbH, Oldenburg, Germany

<sup>3</sup> Department of Speech, Language and Hearing Sciences, Boston University, MA, USA

## Introduction

Normal-hearing listeners can exploit different unmasking cues to recognize target speech in complex speech-on-speech masking conditions [1]. To what degree hearing loss affects this capability is still unclear. Furthermore, research in this area mostly focuses on speech intelligibility (SI). This study additionally investigated the release from perceived listening effort (LE) that could be achieved by three unmasking cues (spatial separation, gender differences, and masker time reversal). Ideal time-frequency segregation was used to estimate the relative amounts of informational and energetic masking as well as the effort associated with source segregation [2, 3]. The goal was to directly compare the release from energetic and informational masking in listeners with and without hearing loss in terms of SI and LE.

## Methods

### Subjects

Eight listeners with normal hearing (NH) and (so far) three hearing-impaired listeners (HI,  $n=3$ ) with moderate sloping sensorineural hearing loss participated in this study. The HI listeners were aided in the experiment by linear NAL-RP amplification.

### Stimuli and conditions

Target speech always consisted of a fixed frontal male talker of the Oldenburg sentence test [4]. Two competing talkers were used as maskers. The maskers also uttered the same type of matrix sentences as the target talker. Eight different masking conditions were created by combining the same or different masker gender as the target (F / M), the same or different ( $\pm 90$  deg.) masker location as the target (see Figure 1), and intelligible or time-reversed maskers. All eight conditions were measured via headphones for natural and “glimpsed” speech. Glimpsed speech was created by eliminating all time-frequency tiles in which the energy of the combined maskers exceeded the target energy. This essentially rendered the masking talkers unintelligible (eliminating informational masking), while preserving the target “glimpsed” which can be assumed to be available to the listener after accounting for energetic masking [2, 3].

### Procedures

SI was measured using lists of 30 sentences, separated into two interleaved adaptive tracks to converge to the  $SRT_{20}$  and  $SRT_{80}$ , respectively. Psychometric functions were fitted to these data points to estimate (amongst others) the speech

recognition threshold  $SRT_{50}$ , i.e., the SNR at which 50% of the word were understood correctly.

LE was measured using adaptive categorical listening effort scaling (ACALES) [5] to derive psychometric functions from „no effort“ (1 effort scaling categorical unit, ESCU) to „extreme effort“ (13 ESCU) as a function of SNR.

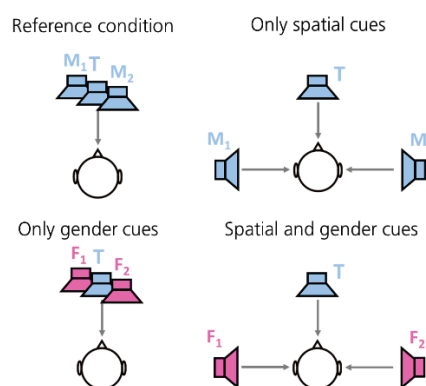
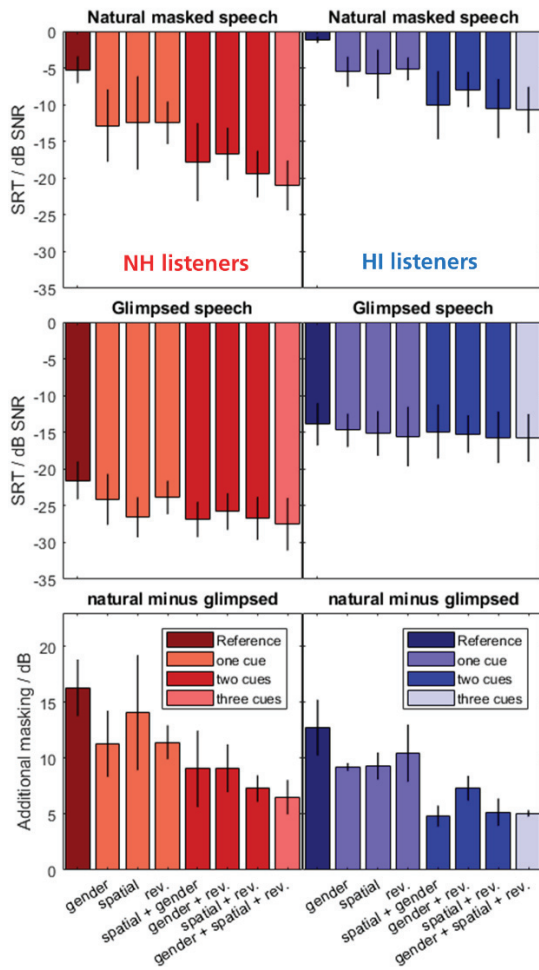


Figure 1: Illustration of gender and spatial unmasking cues conditions.

## Results

### Speech intelligibility

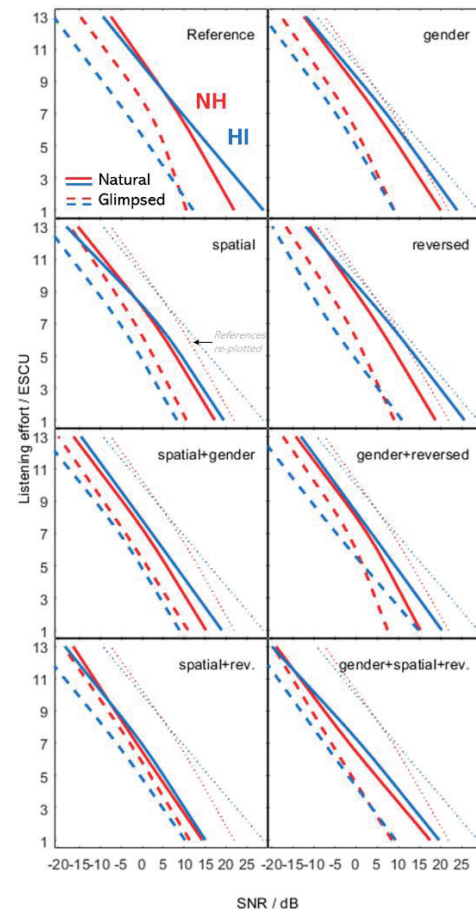
For NH listeners (Fig. 2, left panel / red bars) and HI listeners (right panel / blue bars), the reference condition without unmasking cue was the most difficult one as expected.  $SRT_{50}$ s in all conditions with at least one unmasking cue were considerably lower, and combining more than one cue was beneficial (top left panel). The general pattern was similar for HI subjects (top right panel). However, the  $SRT_{50}$  was generally much higher (by up to 10 dB), and the available unmasking cues produced a smaller benefit than for NH subjects. For both groups ‘glimpsed’ speech produced much lower and more similar  $SRT_{50}$ s across conditions, indicating a significant amount of informational masking in the natural speech conditions, which was largest in the reference condition (mid panels). The differences between  $SRT_{50}$  for natural and glimpsed speech are shown in the bottom panels. They represent “additional masking”, which can be considered an estimate of the informational masking component in the natural speech mix [2,3]. “Additional masking” was largest in the reference condition for both subject groups and tended to decrease with increasing number of unmasking cues. “Additional masking” was slightly smaller for HI than for NH subjects.



**Figure 2:** SI data expressed as mean SRT<sub>50</sub> for natural speech (top) and glimpsed speech (middle). The difference indicates “additional masking” (bottom).

### Listening effort

Psychometric functions (parametrically averaged across listeners) are shown for NH (red) and HI subjects (blue) in Figure 3. Solid and dashed lines represent natural and glimpsed speech, respectively. Different panels show different conditions. For comparison, the data of the reference condition are re-plotted in each panel as dotted lines. For both groups, the reference condition was the most difficult one at a given SNR (top left panel), i.e., available unmasking cues reduced perceived listening effort (compare solid and dotted lines in each panel). The unmasking benefit was similar for both groups and tended to be larger for spatial separation than for gender differences or masker time-reversal. In other words, HI listeners tended to benefit similarly from unmasking cues as NH listeners for natural speech. For glimpsed speech, however, considerable differences between both groups were found: glimpsing produced a much stronger reduction of perceived listening effort in HI than in NH listeners (see distance between dashed and solid lines). This suggests that HI listeners experienced a much larger effort associated with segregating the target talker from the maskers. In other words, they had a much larger benefit from an algorithm that conducted the target-from-masker-segregation “for them”.



**Figure 3:** LE data shown as psychometric functions. Each panel represents one masking conditions. Solid and dashed lines represent natural and glimpsed speech, respectively.

### Discussion & Conclusions

The NH data collected in this study essentially replicated results of [1] for the German language, although the unmasking effect due to the different cues was somewhat smaller. The (aided) HI listeners measured so far also showed a benefit from the tested unmasking cues, but the SRT<sub>50</sub> decrease was considerably smaller, which is consistent with [2]. A particular disadvantage for HI listeners was that they could not benefit as much from combined unmasking cues or from glimpsing in terms of SRT<sub>50</sub>. In contrast, the LE data suggested a comparable benefit of unmasking cues for natural speech, and a much larger benefit of glimpsing in HI listeners. This may indicate that HI listeners experienced significantly greater effort associated with segregating a target talker from a talker mix. Algorithms that perform this segregation “for the listeners” could thus be very beneficial.

### Acknowledgements

This study was funded by the Klaus Tschira Stiftung gGmbH and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 A1.

## References

- [1] Rennies, J., Best, V., Roverud, E. & Kidd Jr, G.(2019). “Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort,” *Trends Hear.* 23, 1-21.
- [2] Kidd Jr, G., Mason, C.R., Best, V., Roverud, E., Swaminathan, J., Jennings, T., Clayton, K. & Colburn, H.S. (2019). “Determining the energetic and informational components of speech-on-speech masking in listeners with sensorineural hearing loss,” *J. Acoust. Soc. Am.* 145, 440-457.
- [3] Brungart, D.S., Chang, P.S., Simpson, B.D., and Wang, D. (2006). “Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation,” *J. Acoust. Soc. Am.* 120, 4007-4018.
- [4] Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests [Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test]. *Zeitschrift für Audiologie*, 38, 86-95.
- [5] Krueger, M., Schulte, M., Brand, T., and Holube, I. (2017). “Development of an adaptive scaling method for subjective listening effort,” *J. Acoust. Soc. Am.* 141, 4680-4693.