

Speech intelligibility with various head-related transfer functions: A computational modelling approach

Axel Ahrens,^{1,a)} Maria Cuevas-Rodriguez,^{2,b)} and W. Owen Brimijoin^{3,c)}

¹Hearing Systems Section, Department of Health Technology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

²Universidad de Málaga, ETSI Telecomunicación, 29010, Málaga, Spain

³Facebook Reality Labs, Redmond, Washington 98052, USA

aahr@dtu.dk, mariacuevas@uma.es, owen.brimijoin@fb.com

Abstract: Speech intelligibility (SI) is known to be affected by the relative spatial position between target and interferers. The benefit of a spatial separation is, along with other factors, related to the head-related transfer function (HRTF). The HRTF is individually different and thus, the cues that affect SI might also be different. In the current study, an auditory model was employed to predict SI with various HRTFs and at different angles on the horizontal plane. The predicted SI threshold was found to be largely different across HRTFs. Thus, individual listeners might have different access to SI cues, dependent on their HRTF. © 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Martin Cooke]

<https://doi.org/10.1121/10.0003618>

Received: 12 October 2020 **Accepted:** 10 February 2021 **Published Online:** 2 March 2021

1. Introduction

Speech intelligibility has been shown to be improved in a scenario with spatially separated target and interferer in comparison to a scenario where they are co-located (Hirsh, 1950). This spatial advantage is commonly referred to as spatial release from masking (SRM). Two mechanisms have been shown to be the main contributors to this spatial benefit: a binaural unmasking (BU) component that arises due to different interaural time differences (ITDs) of target and interferer (Durlach, 1963; Licklider, 1948) and a signal-to-noise ratio (SNR) component (often referred to as better-ear SNR) that arises due to a direction dependent spectral filtering of the source signals from torso, head and pinna. This filter function is also referred to as the head-related transfer function (HRTF) or as head-related impulse response (HRIR) in the time domain. It should be noted that the spectral filtering gives rise to interaural level differences (ILDs), but the HRTF/HRIR also includes ITDs. Previous studies have investigated speech intelligibility while varying ITDs or ILDs and showed that larger differences in ITDs and ILDs lead to better speech intelligibility (Bronkhorst and Plomp, 1988). Culling *et al.* (2004) examined the role of ITD and ILD cues on speech intelligibility separately and in combination, finding better speech intelligibility performances when both cues were present (HRTF without manipulation). However, when interferers were presented in both hemispheres, the advantage caused by the ITD-only HRTFs was bigger than by the ILD-only HRTFs.

HRTFs have been shown to be highly individual. The individual differences have been shown to be due to anthropometric features such as the shape and the size of the head and the pinnae (Watanabe *et al.*, 2007). Listening with non-individual HRTFs has been shown to affect perception [see Xu *et al.* (2007) for a review]. For example, sound source localisation in the median plane has been shown to be reduced, while localisation on the horizontal plane has shown to not be affected by the HRTF (Møller *et al.*, 1996). However, the effect of HRTFs on speech intelligibility remains unclear. Jakien *et al.* (2017) measured speech intelligibility with individual HRTFs when reproducing the stimuli over loudspeakers as well as with artificial head HRTFs when using headphones. They found worse speech intelligibility with the dummy head HRTF than with individual HRTFs. However, they did not include head-movements in the headphone condition but did in the loudspeaker condition, rendering direct comparison problematic. Picinali *et al.* (2019) presented preliminary speech intelligibility results collected with different non-individual HRTFs in the same listeners. They found that some HRTFs lead to a better speech intelligibility than others. Furthermore, they noted that the “better HRTFs” were different for each listener. While these studies suggest an effect of the HRTF on speech intelligibility, no studies have examined what cues across HRTFs affect the differences in speech intelligibility.

^{a)} Author to whom correspondence should be addressed, ORCID: 0000-0001-8800-2497.

^{b)} ORCID: 0000-0002-4698-5170.

^{c)} ORCID: 0000-0001-6124-3987.

In the present study, it was investigated whether HRTFs measured from various individuals have an effect on the predicted speech intelligibility with a computational binaural speech intelligibility model. For the analysis we used 96 HRTFs from two publicly available databases. Thus, it can be investigated if some listeners might have access to more advantageous cues for speech perception based on their HRTFs. Using an auditory model allows one to specifically investigate the differences in cue availability, while eliminating cognitive factors that might also play a role in understanding speech (Heinrich *et al.*, 2015). Furthermore, the application of an auditory model allows the analysis of a larger number of HRTFs than commonly possible in listener studies. Such a modelling approach can guide the selection of conditions in future listening experiments. Additionally, we investigated whether the BU component or the SNR component of the model have more influence on the simulated speech intelligibility with various HRTFs.

2. Methods

2.1 Binaural speech intelligibility model

We used the computational binaural speech intelligibility model from Jelfs *et al.* (2011) as implemented in the Auditory Modelling Toolbox (Soendergaard and Majdak, 2013). The model predicts the target-to-interferer ratio as the sum of a BU component and the SNR component at the ears. The BU component is based on binaural masking level difference predictions (Culling *et al.*, 2004) and is calculated from the ITDs of the target and the interferers and the interaural coherence of the interferer. The SNR component is calculated from the maximum of the long-term SNR of both ears within gammatone filters. The SNR component is also often referred to as better-ear SNR, but can also be nonzero for a symmetric interferer configuration, where spectral filtering leads to SNR advantages over a co-located configuration, that are the same in both ears. The two components are calculated in 43 gammatone filters and integrated over frequency using the speech intelligibility index (SII) weighting. The model takes binaural room impulse responses for the target and the interferers as input signals, but has also been verified with anechoic HRIRs (Jelfs *et al.*, 2011). Using the impulse responses directly without a convolution with speech material or noise resembles a situation with stationary noise interferers having the same spectrum as the target speech. This allows one to specifically investigate the effect of the HRTF on the model outcome. When multiple interferers are used, the HRIRs were concatenated as described in Jelfs *et al.* (2011).

2.2 HRTF databases

Two HRTF databases were used in this study, the LISTEN (Warusfel, 2003) and the CIPIC (Algazi *et al.*, 2001) database. The LISTEN database contains HRTFs measured from 187 directions for 51 subjects. The horizontal plane is fully sampled in 15° steps. Here, the processed version of the HRIRs was taken where the HRIRs were truncated to a length of 512 samples (sampling frequency 44.1 kHz) and diffuse field equalised. The CIPIC database consists of HRIRs from 45 subjects, measured from 125° directions. The horizontal plane is sampled at 80°, 65°, 55° on both hemispheres as well as in 5° steps between ±45°. Two of the subjects in the CIPIC database are artificial head measurements on a KEMAR (GRAS Sound & Vibration A/S, Holte, Denmark) with a small or a large pinna. The HRIRs are 200 samples long (sampling frequency 44.1 kHz) and were also diffuse-field equalised.

In the current study, only HRTFs measured on the horizontal plane were used. While both databases also contain HRTFs measured at elevation angles above and below the horizontal plane, here we focussed on the effects arising from different locations on the horizontal plane as speech sources are commonly found at approximately ear level.

2.3 Experimental setup

The target source was always simulated from the front, i.e., 0° azimuth, 0° elevation. Two conditions were investigated, a single and a two interferer condition. In the single interferer condition a large better-ear SNR is expected, while with symmetric interferers this effect is expected to be reduced. The interferers were simulated at various azimuths either on the right hand side in the single interferer condition or symmetrically on both sides in the two interferer condition. Here, we report the predicted SRM, i.e., the difference between the co-located and the separated interferer configurations.

3. Results

3.1 Single interferer configuration

Figure 1 shows the model predictions with the various HRTFs from the LISTEN (left column) and the CIPIC (right column) database with respect to the azimuth interferer angles (gray lines). The top row shows the predicted SRM, the middle row the predicted SNR component and the bottom row the predicted BU component. The median SRM (black line) across the HRTFs increases with increasing angular separation between the target and the interferer up to about 75° and 65° for the LISTEN and the CIPIC database, respectively. Beyond this angle the median SRM decreases slightly. However, for some HRTFs the predicted SRM continues to increase, for some it remains constant, and for some it decreases. A similar trend can be seen for the predicted SNR advantage. The BU advantage increases up to an angle of 45°.

The variance of the spatial benefit across HRTFs was found to increase with the separation angle for the LISTEN database but remain relatively constant for the CIPIC database. The largest difference in predicted SRM was

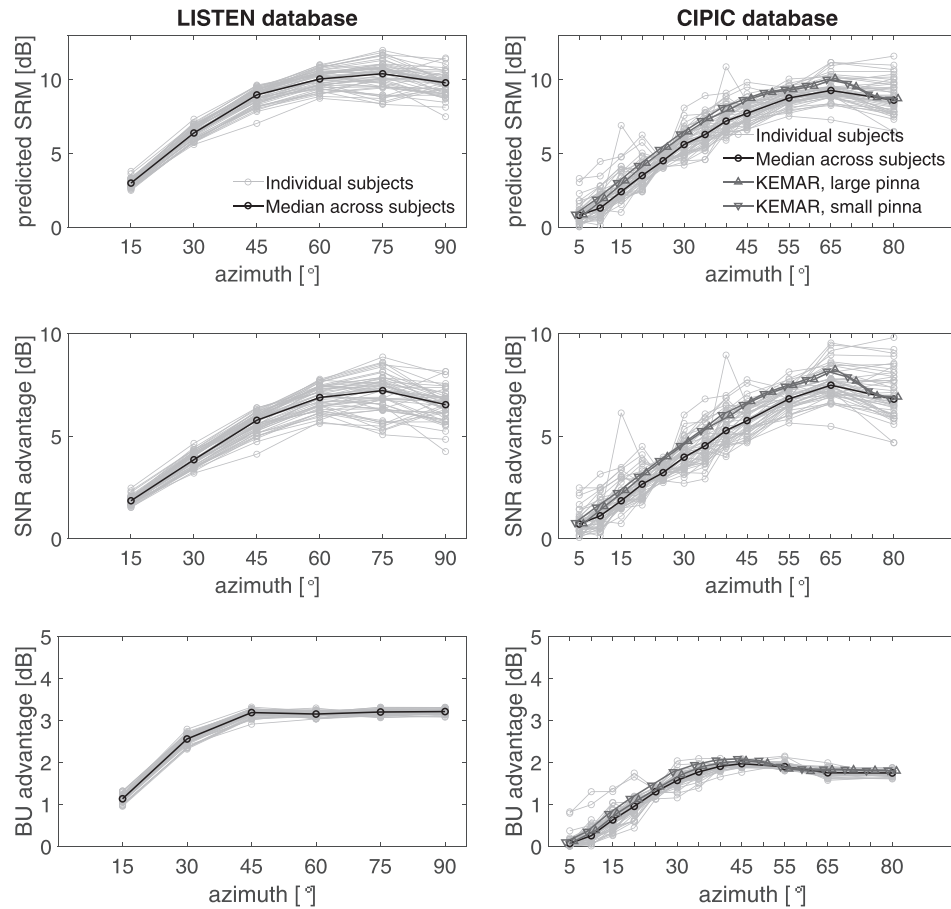


Fig. 1. The predicted SRM (top), SNR advantage (middle), and BU advantage (bottom) with respect to the azimuth angle of a single interferer with a target from the front (0°). The gray curves depict individual subjects' HRTFs from the LISTEN (left) and the CIPIC (right) database. The black curves depict the median calculated over the HRTFs. For the CIPIC database also predictions with artificial head HRTFs are shown (triangles).

found to be 4 dB for a separation angle of 90° and 5.9 dB for a separation angle of 15° , for the LISTEN and the CIPIC database, respectively. The large difference in the CIPIC database at 15° might be driven by a single outlier. The second largest difference was found at 80° (5.2 dB). To avoid outliers in the data, a percentile analysis was performed. The range between the 5th and the 95th percentile was found to be 3.2 dB (at 75°) for the LISTEN and 3.6 dB (at 80°) for the CIPIC database.

To analyse the contribution of the SNR component and the BU component to the variance of the predicted SRM across HRTFs, the R^2 of a linear regression model was investigated. Therefore, models were fitted either between the SRM and the BU or between the SRM and the SNR. This analysis was performed for the angle with a largest variance only. The correlation analysis between the predicted SRM and the SNR or the BU component of the model revealed that the SNR component alone can account for most variance across HRTFs. The SNR component accounts for 99.7% of the variance in the CIPIC database and 99.5% in the LISTEN database. The BU component accounts for 9.6% in the CIPIC and 1.2% in the LISTEN database.

The predicted SRM was found to be about 2 dB larger for the HRTFs in the LISTEN database than for the ones in the CIPIC database. This difference is mainly driven by the BU component of the model while the SNR component is similar across the two databases.

The CIPIC database also contains two HRTF measurements of an artificial head (KEMAR) with two different pinna sizes. The predicted SRM with the KEMAR HRTFs (dark gray lines) was found larger than the median of the subjects' HRTFs. This difference was driven by the SNR component of the model, while the BU component was found similar between subjects and artificial head. The difference between the two pinna sizes was found negligible.

3.2 Symmetric interferer configuration

Figure 2 shows the model predictions as in Fig. 1 but for symmetric interferers. The median predicted SRM (top panels, black lines) increases with separation angle but less than with a single interferer. The reason is a significantly smaller SNR

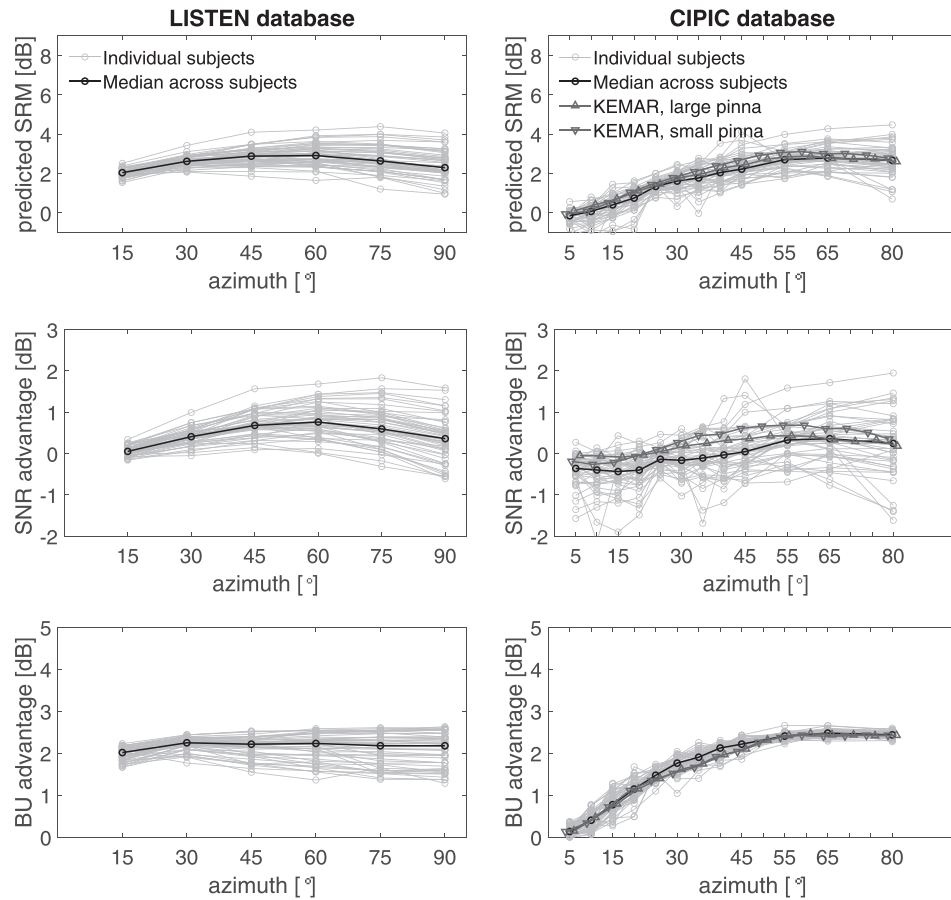


Fig. 2. The predicted SRM (top), SNR advantage (middle), and BU advantage (bottom) with respect to the azimuth angle of two symmetric interferers with a target from the front (0°). The gray curves depict individual subjects' HRTFs from the LISTEN (left) and the CIPIC (right) database. The black curves depict the median calculated over the HRTFs. For the CIPIC database also predictions with artificial head HRTFs are shown (triangles).

component (middle panel) due to a reduced better-ear listening advantage. The SNR advantage is close to zero and for some conditions negative. The BU advantage is similar as with a single interferer.

The largest inter-HRTF difference was found 3.2 dB (at 75° azimuth) and 3.8 dB (at 80° azimuth) in the LISTEN and CIPIC database, respectively. The range between the 5th and the 95th percentile was found 2.5 dB at 90° azimuth for the LISTEN and 2.8 dB at 80° azimuth for the CIPIC database.

A correlation analysis using R^2 as in Sec. 3.1 showed that with symmetric interferers, the SNR advantage alone accounts for 70.4% (at 90° azimuth) and 99.1% (at 80° azimuth) of the variance in the predicted SRM for the LISTEN and the CIPIC database, respectively. The BU component alone can account for 46.4% and 1.2% of the variance, for LISTEN and CIPIC, respectively.

Comparing the two HRTF databases shows a different trend than with a single interferer. For the 15° separation angle between target and interferers, the median predicted SRM is about 1 dB higher with HRTFs from the LISTEN database than from the CIPIC database. The difference is driven by the BU advantage. The SNR advantage in both databases is close to 0 dB. The predicted SRM using the KEMAR HRTFs was comparable to the SRM of the subjects' HRTFs in the CIPIC database.

4. Discussion and conclusion

In the present study we investigated the influence of various head-related transfer functions (HRTFs) on the spatial release from masking (SRM). We used a computational auditory model and simulated the SRM in various conditions with HRTFs from two publicly available databases. It was demonstrated that the predicted SRM of a binaural speech intelligibility model is dependent on the HRTF. The variance of the predicted spatial benefit across HRTFs was found larger for wider separation angles between target and interferers than for smaller separation angles. The differences across the HRTFs could mainly be explained by the signal-to-noise ratio (SNR) at the ears while the binaural unmasking (BU)

advantage showed less variations. The predicted speech intelligibility with the HRTFs of an artificial head was better than the subjects' HRTFs medians in the condition with a single interferer but very similar in the presence of two symmetric interferers. With the artificial head the BU component was closer matched to the subjects' HRTFs than the SNR component.

The results from the current study are in general agreement with previous speech intelligibility studies. Multiple studies investigated speech intelligibility with a target talker from the front and a single interferer at multiple separation angles on the horizontal plane (Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997). The studies generally found the same trend [see Bronkhorst (2000) for a summary], in that the largest SRM occurs when the interferer is between 90° to 120°. However, the differences between an interferer at 60° and at 90° and 120° was small in comparison to the difference between 0° and 60°. These findings are in agreement with the predictions shown in Figure 1. Marrone *et al.* (2008) measured speech intelligibility with a target at 0° and two symmetrically separated interferers and found a plateauing release from masking for separation angles between 45° and 90°, which is in line with the predicted intelligibility results found in the current study (see Fig. 2). Marrone *et al.* (2008) used speech interferers, while in this paper noise interferers were assumed, however, Cosentino *et al.* (2014) found similar results with noise interferers. The reduction in SRM and particularly in better-ear SNR advantage has previously been shown when interferers are placed in both hemispheres [e.g., Hawley *et al.* (2004)].

Here, we entirely relied on predictions from the computational auditory model from Jelfs *et al.* (2011) instead of listener studies. The reason was to only investigate the cues that are available to the auditory system, while avoiding within-listener variations as well as cognitive effects. The risk with a pure modelling approach is that the model does not reflect listener performance. However, Jelfs *et al.* (2011) evaluated the auditory model with various listener studies and showed that the model predictions match the speech intelligibility data well. The most relevant dataset for the current study that they used to evaluate their model was from Bronkhorst and Plomp (1988), where speech intelligibility was measured with a single noise source positioned at various azimuth angles. The results of the model and the measurements both showed an increase in SRM with increasing separation angle. However, the model predicted a lower SRM at 90° than at 60°, while the measured SRM did not show this drop. This drop in SRM at 90° can also be seen in Fig. 1 of the current study. Such a drop in intelligibility has been shown previously (Peissig and Kollmeier, 1997). Jelfs *et al.* (2011) argued that the reason for the difference between the model and the data are the use of different HRTFs. This explanation is in line with the findings in the current study. For some HRTFs the 90° drop can be seen, while for others the SRM remains constant or increases.

To our knowledge only one previous study investigated speech intelligibility with human listeners measured with a larger number of HRTFs. Picinali *et al.* (2019) found small but significant effects on speech intelligibility across seven HRTFs from the LISTEN database when using a target talker at 0° and two interfering speech sources at ±90°. They reported from preliminary data that different listeners performed best with different HRTFs. That finding is not in line with the findings of the current study. Our results suggest that one HRTF is associated with a higher speech intelligibility than the other HRTFs. However, many more factors influence the performance in speech intelligibility tests. For example, it might be that we perform best with an HRTF that is closest to our own HRTF (Orduña-Bustamante *et al.*, 2018).

The contribution of the SNR component to the variance in predicted speech intelligibility across HRTFs was more pronounced in the condition with a single interferer than with symmetric interferers. The likely reason is that with a single interferer, the ear contralateral to the interferer receives a more favourable SNR than the ipsilateral ear. This effect is commonly referred to as better-ear SNR listening. When the better-ear SNR is large, any variance in HRTFs is likely to contribute relatively more to the SNR component of the model. When the interferers are symmetric, on the other hand, the SNR component is comprised of only differences in the spectra across the ears. We argue that this distinction underlies the observed difference in predicted intelligibility across the symmetric versus asymmetric conditions. However, in both interferer conditions (with symmetric and asymmetric interferers) the variance across HRTFs was mainly driven by the SNR component. Thus, the HRTFs seem to affect the SNR more than the binaural unmasking component. Cubick *et al.* (2018) measured speech intelligibility in normal-hearing listeners with and without hearing aids. They showed that hearing aids vary the HRTF as well as reduce the spatial benefit. Employing a similar model as in the current study, they showed that mainly the better-ear SNR could account for the findings, in line with the findings of the current study.

The underlying reason for the differences in intelligibility across HRTFs is acoustic in nature, and thus driven by differences in the anatomy of the body, head, and ears. Nevertheless, we cannot rule out differences that are due to the measurements of the HRTFs. A total of 96 HRTFs from two HRTF databases were used in the current study, and the results across the two databases were similar, reducing the likelihood that the observed effects were due to measurement uncertainties. The databases are likely to cover a diverse range of people, however, it remains unclear if conclusions can be drawn for a worldwide population. The artificial head, which is supposed to reflect an average across the population, for example, showed results slightly above the median predicted intelligibility in the single interferer condition. Thus, the artificial head might not reflect an average head, the databases do not cover an average of the population, or both. However, with symmetric interferers the difference diminishes.

For applications like virtual reality, it is often not possible to use individual HRTFs, thus non-individual HRTFs from artificial heads or databases need to be employed. When choosing a non-individual HRTF for a task that involves

speech intelligibility, e.g., a cocktail party situation, it is important to keep in mind which HRTF to choose, as individual characteristics of each HRTF seem to have an impact on the speech intelligibility, as shown in this study. In the case of the CIPIC database, it has been shown that the artificial head HRTF is a good average for most human HRTFs. Thus, using an artificial head HRTF seems to be a viable option when no individual HRTFs are available. However, some individual HRTFs show a better predicted speech intelligibility in comparison to the artificial head HRTF. Thus, speech intelligibility could be maximised by choosing an “ideal” HRTF.

Throughout this study it has been shown that there is an effect of HRTFs on modelled speech intelligibility, however, it remains unclear if listeners are affected by the differences in cue availability or if the auditory system is trained to a specific set of cues. Furthermore, other factors are known to affect speech intelligibility. For example, the type of the interferer. Using the current modelling approach assumes stationary noise interferers with the same spectrum as the target speech. When using speech instead of noise interferers, the speech intelligibility thresholds tend to be higher [e.g., Carhart *et al.* (1969)] and the SRM is often larger [e.g., Freyman *et al.* (1999)]. If speech interferers affect the speech intelligibility with various HRTFs differently than the modelled noise interferers, could be investigated in listener studies. Another factor affecting speech intelligibility is room reverberation [e.g., Plomp (1976)]. Reverberation has been shown to mask spectral differences in impulse responses (Ahrens *et al.*, 2020; Oreinos and Buchholz, 2015), thus, differences across HRTFs might also be masked. Furthermore, individual perception differences (e.g., due to cognitive factors or threshold differences), head-motion, adaptation, and learning of HRTFs might introduce larger effects than the variations across HRTFs and thus, could mask differences in speech perception as shown in this study. Nevertheless, the modelling approach in this study showed possible differences in cue availability for speech intelligibility with various HRTFs. Thus, future auditory models and the selection of HRTFs for applications such as virtual reality should incorporate the effect of different HRTFs and should be aware of the variance in cue availability.

Acknowledgments

We would like to thank Pavel Zahorik and Sebastià Vicenç Amengual Garí for the discussion that led to this manuscript. A.A. was supported by the Centre for Applied Hearing Research (CAHR) and by a grant from Facebook Reality Labs.

References and links

- Ahrens, A., Marschall, M., and Dau, T. (2020). “The effect of spatial energy spread on sound image size and speech intelligibility,” *J. Acoust. Soc. Am.* **147**(3), 1368–1378.
- Algazi, V., Duda, R., Thompson, D., and Avendano, C. (2001). “The CIPIC HRTF database,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, October, IEEE, pp. 99–102.
- Bronkhorst, A. W. (2000). “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” *Acta Acust. Acust.* **86**(1), 117–128. <https://www.ingentaconnect.com/content/dav/aaua/2000/00000086/00000001/art00016>.
- 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**(4), 1508–1516.
- Carhart, R., Tillman, T. W., and Greetis, E. S. (1969). “Perceptual masking in multiple sound backgrounds,” *J. Acoust. Soc. Am.* **45**(3), 694–703.
- Cosentino, S., Marquardt, T., McAlpine, D., Culling, J. F., and Falk, T. H. (2014). “A model that predicts the binaural advantage to speech intelligibility from the mixed target and interferer signals,” *J. Acoust. Soc. Am.* **135**(2), 796–807.
- Cubick, J., Buchholz, J. M., Best, V., Lavandier, M., and Dau, T. (2018). “Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners,” *J. Acoust. Soc. Am.* **144**(5), 2896–2905.
- Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (2004). “The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources,” *J. Acoust. Soc. Am.* **116**(2), 1057–1065.
- Durlach, N. I. (1963). “Equalization and cancellation theory of binaural masking-level differences,” *J. Acoust. Soc. Am.* **35**(8), 1206–1218.
- Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). “The role of perceived spatial separation in the unmasking of speech,” *J. Acoust. Soc. Am.* **106**(6), 3578–3588.
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). “The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer,” *J. Acoust. Soc. Am.* **115**(2), 833–843.
- Heinrich, A., Henshaw, H., and Ferguson, M. A. (2015). “The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests,” *Front. Psych.* **6**, 782.
- Hirsh, I. J. (1950). “The relation between localisation and intelligibility,” *J. Acoust. Soc. Am.* **22**(2), 196–200.
- Jakien, K. M., Kempel, S. D., Stansell, M. M., and Gallun, F. J. (2017). “Validating a rapid, automated test of spatial release from masking,” *Am. J. Audiology* **26**(4), 507–518.
- Jelfs, S., Culling, J. F., and Lavandier, M. (2011). “Revision and validation of a binaural model for speech intelligibility in noise,” *Hear. Res.* **275**(1-2), 96–104.
- Licklider, J. C. R. (1948). “The influence of interaural phase relations upon the masking of speech by white noise,” *J. Acoust. Soc. Am.* **20**(2), 150–159.
- Marrone, N., Mason, C. R., and Kidd, G. (2008). “Tuning in the spatial dimension: Evidence from a masked speech identification task,” *J. Acoust. Soc. Am.* **124**(2), 1146–1158.
- Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (1996). “Binaural technique: Do we need individual recordings?,” *AES: J. Audio Eng. Soc.* **44**(6), 451–469, available at <https://www.aes.org/e-lib/browse.cfm?elib=7897>.

- Orduña-Bustamante, F., Padilla-Ortiz, A., and Torres-Gallegos, E. A. (2018). "Binaural speech intelligibility through personal and non-personal hrtf via headphones, with added artificial noise and reverberation," *Speech Commun.* **105**, 53–61.
- Oreinos, C., and Buchholz, J. M. (2015). "Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones," *J. Acoust. Soc. Am.* **137**(6), 3447–3465.
- Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.* **101**(3), 1660–1670.
- Picinali, L., Cuevas-Rodriguez, M., Gonzalez-Toledo, D., and Reyes-Lecuona, A. (2019). "Speech-in-noise performances in virtual cocktail party using different non-individual Head Related Transfer Functions," in *Proceedings of the 23rd International Congress on Acoustics*, Vol. 1, pp. 2158–2159.
- Plomp, R. (1976). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," *Acta Acust. Acust.* **34**(4), 200–211. <https://www.ingentaconnect.com/content/dav/auaa/1976/00000034/00000004/art00004>.
- Soendergaard, P., and Majdak, P. (2013). "The auditory modelling toolbox," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), pp. 33–56.
- Warusfel, O. (2003). "IRCAM LISTEN HRTF database," <http://recherche.ircam.fr/equipements/salles/listen/> (Last viewed February 22, 2021).
- Watanabe, K., Ozawa, K., Iwaya, Y., Suzuki, Y., and Aso, K. (2007). "Estimation of interaural level difference based on anthropometry and its effect on sound localisation," *J. Acoust. Soc. Am.* **122**(5), 2832–2841.
- Xu, S., Li, Z., and Salvendy, G. (2007). "Individualization of head-related transfer function for three-dimensional virtual auditory display: A review," in *Virtual Reality* (Springer, Berlin), pp. 397–407.