Speech Intelligibility Prediction for Hearing-Impaired Listeners with the bBSIM-STI Model

Saskia Röttges[1,4], *Jana Roßbach*[2,4], *Christopher F. Hauth*[1,4], *Thomas Biberger*[1,4], *Bernd T. Meyer*[2,4], *Rainer Huber*[3,4], *Jan Rennies*[3,4], *Thomas Brand*[1,4]

[1]Medizinische Physik, Carl von Ossietzky University, Oldenburg, Germany
[2]Communication Acoustics, Carl von Ossietzky University, Oldenburg, Germany
[3]Fraunhofer IDMT, Hearing, Speech and Audio Technology, Oldenburg, Germany
[4]Cluster of Excellence Hearing4all, Germany

saskia.roettges@uni-oldenburg.de, jana.rossbach@uni-oldenburg.de,
christopher.hauth@uni-oldenburg.de, thomas.biberger@uni-oldenburg.de,
rainer.huber@idmt.fraunhofer.de, jan.rennies@idmt.fraunhofer.de,
bernd.meyer@uni-oldenburg.de, thomas.brand@uni-oldenburg.de

Abstract

As part of the first Clarity Prediction Challenge (CPC1), this study predicted the speech intelligibility of hearing-impaired subjects. A hybrid model was used for the predictions, using a blind equalization-cancellation (EC) model as binaural front-end and a non-blind version of the speech transmission index (STI) using a correlation method as back-end. The model presented here (bBSIM-STI) is very similar to the basic CPC1 non-blind model. The data to be predicted were split into two data sets. The closed data set includes the information of all listeners and enhancement algorithms. The open data set includes just a subset of the listeners and the enhancement algorithms. bBSIM-STI (open set: RMSE = 28.09 % and ρ = 0.63, closed set: RMSE = 27.52 % and $\rho = 0.66$) produces lower RMSEs and higher correlations than the baseline model (open set: RMSE = 36.52 % and ρ = 0.53, closed set: RMSE = 28.52 % and ρ = 0.62) for both the open and the closed set.

Index Terms: speech recognition, hearing impairment, binaural hearing, Speech Transmission Index

1. Introduction

Several studies (e.g., [1, 2, 3]) demonstrated that normal-hearing (NH) listeners can benefit from spatially separated target speaker and masker sources compared to situations, where the target is spatially co-located with the masker. Such a benefit is generally reduced in reverberant situations compared to anechoic situations as binaural cues (interaural level differences, ILDs, and interaural time differences, ITDs) are impaired by reverberation, which also reduces temporal modulation of the target signal and reduces the chance to listen into the dips of fluctuating masker signals.

From previous studies (e.g., [2, 4]) it is known that hearing-impaired (HI) listeners are less capable of taking advantage when the target is spatially separated from the masker than NH listeners, which might be explained by an impaired representation of binaural cues and a reduced audibility of the target signal in the better ear. Since the acoustic environments considered in the first Clarity Enhancement Challenge (CEC1) [5] typically comprise reverberation and spatially separated target and masker sources, a modelling approach consisting of a binaural processor, which simulates binaural interaction and better-ear processing under consideration of hearing-impairment, followed by an analysis of the temporal envelope seems to be reasonable choice for this task.

Thus, this contribution (entry ID E019) to the CPC1 [6] is based on the latest version of the blind Binaural Speech Intelligibility Model (BSIM20) [7] and the correlation-based version of the Speech Transmission Index (STI) [8]. Former versions of BSIM [9, 10] did not work blindly (i.e., they required separated speech and noise signals) and applied the Speech Intelligibility Index (SII) [11] as back-end. In this contribution we use the blind front-end of BSIM20 which is called bBSIM in the following. bBSIM produces equal results as the non-blind version but requires no auxiliary information about the target speech and the masking noise, so that it can be combined with arbitrary back-ends predicting speech recognition scores (see, e.g., [12, 13]).

The use of bBSIM helps to understand how relevant the binaural information in the CPC1 is for speech understanding. In this contribution, we use the correlation-based STI as back-end, as it is takes reverberation effects into account and produced the best predictions during the training phase of CPC1 compared to other back-ends we tried. This back-end is not blind as it requires the clean target speech separately and thus the combination of bBSIM and STI is a hybrid model. Note that in this contribution no machine learning is applied but two classic approaches from psychoacoustics are combined that are very easy to compute. In this respect this contribution is very close to the baseline model of CPC1 which used a very similar binaural front-end [14] combined with a back-end that also analyses the modulations of the signal [15]. Hence, this contribution can be seen as an alternative baseline model that shows how far we (the authors) were able to get without machine learning and training to the test data.

2. Method

2.1. Data basis

The provided data basis of the CPC1 was used, which contains audio signals [16], characteristics of the HI listeners, and the speech intelligibility scores from listening test (correct response rates given for single sentences) [5]. The listeners' task in the listening test was to repeat the words that were understood in the presented test signal, which varied in the acoustic scene and also in speech enhancement algorithm of the CEC1 [5]. The different acoustic scenes were generated by convolving the audio signals with binaural room impulse responses (BRIRs). The acoustic scenes always included one target speaker from a set of 40 speakers, uttering a 7- to 10-word sentence. The target speech was masked by continuous noise as interferer. All stimuli were played in a small room with low to moderate reverberation. Each acoustic scene consisted of a unique target utterance and a unique interferer segment, which were mixed together. These signals and the audiograms of the HI listeners were processed by the hearing aid algorithms of the CEC1 [5] and subsequently used for the listening tests. The data basis consisted of two parts, an open set and a closed set, each consisting of training and evaluation data. The closed set contained information about all 27 listeners and 10 speech enhancement algorithms, for the test and evaluation data, while the training data of the open data set contained a subset of 22 listeners and 9 enhancement algorithms. Not all provided information from the database was used, but only the information mentioned here.

2.2. Baseline model

The baseline model used in this challenge is a composite of a hearing loss model [14] and a speech intelligibility model [15]. Decreased audibility, reduced dynamic range, and the loss of temporal and frequency resolution is simulated by the hearing loss model. The model uses the output of the hearing aid processor and the audiograms of the listener as input. The speech intelligibility model is a binaural, modified version of the short time objective intelligibility model (STOI) [17] and calculates the correlation of the speech envelopes of the clean and the degraded speech. A minimum root-mean-square error (RMSE) sigmoid fitting is used to map the MBSTOI values to speech recognition in percent. The fitting parameters were estimated only from the training data of the respective data set.

2.3. bBSIM

The bBSIM proposed in [7] is used as binaural front-end. It receives the mixed target speech and interferer signals at the left and the right ear as input. The stimuli provided in the challenge were preprocessed by removing the first 2 seconds and the last 1 second that were known to contain only noise. We additionally applied a simple rms-based voice activity detection to remove remaining silent frames.

The first stage of bBSIM simulates the frequency selectivity of the human auditory system by splitting the input signals (left and right ear signals) into 30 Equivalent Rectangular Bandwidth-(ERB-)spaced frequency bands [18] using a gammatone filterbank [19] with center frequencies from 150 Hz to 8000 Hz. Based on the individual pure tone audiograms, two internal threshold-simulating noises are added to the left and right input signals to simulate the hearing loss. The left and right threshold simulating noises are generated as uncorrelated signals, so that the equalization-cancellation (EC, see below) stage of bBSIM cannot cancel them out. For frequencies up to 1500 Hz, binaural processing is realized as a blind EC [7] mechanism, where the differences in ITDs and ILDs between target and interfering signal can be used to improve the signal-to-noise ratio. For frequencies above 1500 Hz, the better ear is selected blindly. In the equalization step the two ear channels of each gammatone filter channel are equalized in level and phase. Then, the cancellation step is applied, which uses two different strategies: 1) a minimization of the output power and 2) a maximization of the output power. While the first strategy can be assumed to be the better strategy at negative SNRs, because it attenuates the interfering signal, the second strategy can be assumed to be better at positive SNRs, because the power of the target signal is increased. To choose the best of both strategies in each frequency channel, the speech-to-reverberation modulation energy ratio (SRMR) [20] is used. SRMR describes the ratio between speech-like and non speech-like amplitude modulations by calculating a ratio between the energy in modulation frequency channels below 16 Hz and above 16 Hz. The SRMR is calculated for both strategies and both ear channels and, subsequently, the EC channel and the ear channel with the higher SRMR are combined to produce a single channel signal with enhanced SNR. Due to its simple calculation SRMR can be applied independently to each ERB channel. In theory, an EC mechanism allows for a complete cancellation of interfering sounds, which would produce unrealistically high speech intelligibility at very low SNRs. To avoid this, the imperfections of human binaural processing have to be accounted for. In the current model version, this is achieved by introducing uncertainties of binaural processing when adjusting the delay and gain factors in the EC stage. These were chosen to fluctuate around the optimal factor, which can be described as a jitter. Since this jitter is directly imposed on the signal, Monte Carlo simulations (MCSs) must be performed, which is quite time consuming. In this study, we used a constant mistuning of the equalization parameters ("fixed jitter") determined by the jitter's standard deviation [7] and can thus dispense with the time-consuming MCSs.

2.4. Speech Transmission Index

The back-end of the model employed in this study is a specific version of the STI [21], which receives bBSIM's output signals of the clean target speech and degraded speech as input. The calculation of the separate target and interfering signals is possible as bBSIM's processing is linear with respect to the signals, so that speech and noise can be processed separately using the EC parameters determined by the blind model (see [7] for details). The STI analyzes the modulation transfer function by comparing the envelopes of the input signals to calculate the modulation transmission index for each frequency band. Here, the normalized covariance method [8] is applied: The covariance between the envelopes of the target speech and the degraded speech is calculated and then normalized with the individual variances of the target speech and the degraded speech. The weighted average of the transfer index of all frequency bands gives the STI and is very similar to the later proposed short-time objective intelligibility (STOI) measure [17].

2.5. Mapping from STI to speech recognition

In this challenge speech recognition in percent correct has to be predicted. The employed STI back-end produces index values ranging from 0 to 1 and, therefore, the back-end values are mapped to speech recognition using

$$f(x) = \frac{1}{1 + exp(4 \cdot s_{50} \cdot (L_{50} - x))}, \tag{1}$$

where s_{50} denotes the slope at the midpoint of the intelligibility function and L_{50} denoting the level of this midpoint, which is equal to the speech recognition threshold (SRT) at which 50 % of the words are correctly understood. For both data sets, the psychometric functions are fitted only to the training data. For the open data set, L_{50} and s_{50} have been chosen to fit best to all points of the training data, and are subsequently used to map the STI value of the open data set. For the closed data set the mapping is slightly different as it is done for each listener individually. This means that for each of the 27 listeners the optimal mapping parameters are used and the corresponding listener ID is used to map the STI index values of the closed data set.

3. Results

Figure 1 shows scatter plots of measured vs. predicted intelligibility scores for each tested sentence for the open test set (left panel) and the closed test set (right panel). A wide scatter can be observed indicating that the prediction accuracy is limited. One reason for this wide scatter is certainly that the predictions were done sentence by sentence which implies a large measurement uncertainty as discussed below.

Table 1 shows the calculated RMSEs with standard errors (SE) and the correlation (ρ) between the predicted and the measured intelligibility scores for the bBSIM-STI and the baseline model for the open and the closed data set. bBSIM-STI achieves a lower RMSE than the baseline model (MBSTOI) in both the open set and the closed set. This difference is particularly pronounced for the open data set (about 7%), and is smaller (about 1%) for the closed data set. In terms of correlation, bBIM-STI achieves a slightly higher correlation than the baseline model in the open and the closed set. In contrast to the baseline model, there is not much difference between the RMSEs of the two data sets, which also applies to the correlation.

Table 1: *Root mean squared error (RMSE) of the predictions with standard errors (SE) and correlations for the closed and open data set using the bBSIM-STI and the baseline model MBSTOI.*

Model	**data**	**RMSE**	**SE**	ρ
bBSIM-STI	open set	28.09 %	1.12 %	0.63
	closed set	27.52 %	0.56 %	0.66
MBSTOI	open set	36.52 %	1.35 %	0.53
	closed set	28.52 %	0.58 %	0.62

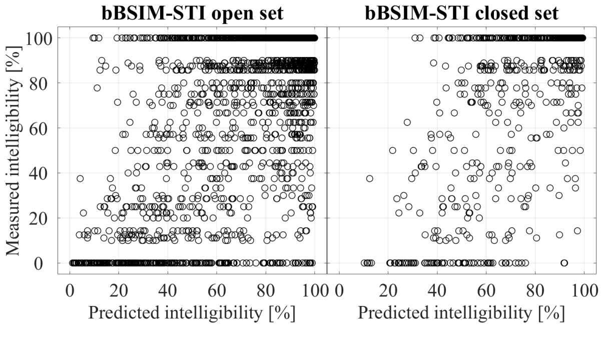

Figure 1: *Scatter plots visualizing the relation between measured and predicted intelligibility for the open set (left panel) and closed set (right panel) for the bBSIM-STI.*

Figure 2 shows histograms of the prediction errors (difference between the predicted and measured recognition scores per sentence) in per cent for bBSIM-STI (green) and MBSTOI (blue). The left panel shows the results for the open set and the right panel shows the results for the closed set. All histograms have their main maximum close to 0 %, with a prominent peak close to 0 % that clearly deviates from a Gaussian distribution. This peak can be explained by the limitation of the recognition scores to values between 0 and 100 %: when the measured score is 0 or 100 % the prediction model has the chance to hit this value exactly, because there is no remaining variance in the measured data (see Figure 1). For the open set a prominent side maximum in the histogram of MBSTOI in the range from -95 to -75 % can be seen representing a systematic bias in the predictions. All of these outliers, except for three, occurred for the same speech enhancement algorithm (E018). Obviously this speech enhancement algorithm [22] produced signals that irritated MBSTOI more than the bBSIM-STI model. The averaged RMSEs over all listeners and scenes for the different speech enhancement algorithms for the bBSIM-STI model ranges from 21.0 % to 34.5 %, where the combination with the E018 algorithm yielded in a high RMSE (34.4 %) compared to the other algorithms.

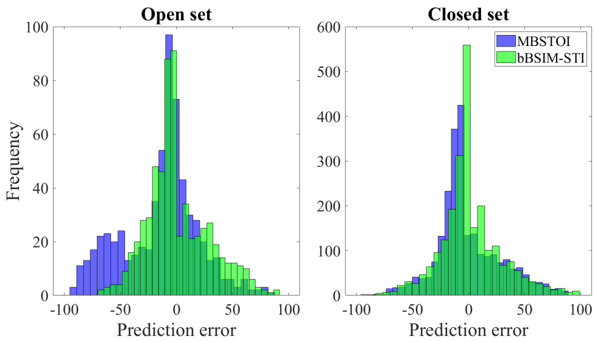

Figure 2: *Histogram visualizing the prediction error (difference between predicted and measured intelligibility) for bBSIM-STI (green) and MBSTOI (blue) for the open set (left panel) and closed set (right panel).*

4. Discussion

When experimenting with the provided stimuli and ground truth data of CPC1, we observed that the model's binaural processing did not generate relevant spatial or binaural unmasking. This indicates that the signals used in CPC1 do not provide usable binaural information. It is possible that the realistic scenes including at least moderate amounts of reverberation had limited potential for eliciting effects of spatial unmasking. This could be tested by applying binaural prediction models to the unprocessed signals. A further reason for the missing binaural benefit might be that the applied signal enhancement algorithms destroyed most of the binaural information. Whatever the reason, the binaural front-end employed in the current model did not deliver substantial advantages for CPC1. However, since it did predict binaural unmasking well in earlier studies, we expect it may provide useful information in future versions of the prediction challenges if conditions and/or algorithms with stronger binaural cues are included.

The present CPC1 also provides some insights into incorporating hearing loss simulation into prediction models. In the current modeling approach, hearing-impairment is simulated by (stimulus-independent) additive noise according to the listener's audiogram of the left and right ears within the bBSIM. Such additive noise has consequences on the audibility of the target signal, but also on the equalization-cancellation mechanism mimicking binaural interaction. Surprisingly, consequences of hearing-impairment simulations had only minor influence on the prediction performance and for that reason we did not use the pure tone audiogram at all in our second submission (E022). This finding might mirror the fact that the listeners adjusted the overall level themselves and that, consequently, audibility did not play an important role in these measurements, and/or that supra-threshold hearing deficits are not well described by the pure tone audiogram. Recently, [23] successfully predicted SRTs for speech in stationary and fluctuating noise maskers for unaided HI listeners, by simulating consequences of an impaired audibility and suprathreshold hearing deficits, whereas the latter one was simulated by stimulus-dependent additive noise. Conversely, for SRT predictions of a noisy speech signal processed by binaural noise-reduction algorithms, [24] applied the same modeling framework as used in [23], but without taking suprathreshold hearing deficits into account. Although, they achieved a reasonable prediction performance, their results imply that including aspects of suprathreshold hearing deficits would be desirable and probably improve the accuracy of predictions. In the context of binaural hearing, [25] successfully simulated consequences of hearing deficits for listeners having no more than a slight hearing loss by including aspects of audibility and suprathreshold hearing deficits realized by stimulus-independent and stimulus-dependent additive noise. [25] hypothesized that at higher overall levels, as they typically occur in conversation scenarios, suprathreshold hearing deficits are probably dominant. Those findings indicate that extending our suggested modeling approach with a processing stage that incorporates consequences of suprathreshold hearing deficits may potentially improve the prediction performance for HI listeners.

Because the data of this challenge provided no information about suprathreshold hearing loss, we tried to take an individual component into account, which describes each respective listener. This was achieved by fitting the STI–to–intelligibility mapping individually for each listener. Note that this individual mapping not necessarily describes the consequence of suprathreshold hearing deficits, as there are also other reasons for individual differences, like cognitive processing. This individual mapping was done for the closed data set only and improved the prediction accuracy of our model. For the open data set we did not perform this individual mapping, as this was not possible for the unknown listeners. Instead, for the open data set, we used a general mapping for all listeners. Surprisingly, our model performed nearly as accurately for the open data set as for the closed data set. This indicates that the individual mapping was not that relevant for the listeners of the open data set of CPC1.

For the interpretation of the prediction accuracy it has to be taken into account that the human recognition data is binomially distributed and that, consequently, the standard error of each measured recognition score can roughly be estimated by

$$\sigma_p \approx \sqrt{\frac{p(1-p)}{j}}, \tag{2}$$

with p denoting the recognition score of a sentence (with values from 0 to 1) and j denoting the statistically independently recognized parts per sentence. The sentences of this challenge had 7 to 10 words. Note that due to sentence context the number of statistically independently recognized parts per sentence is smaller than the actual number of words. According to [26] the number of statistical independently perceived parts of a sentence can be estimated as $j = \log(p_w)/\log(p_p)$, with p_p denoting the average intelligibility (proportion of correctly perceived words) and p_w denoting the average proportion of sentences, for which all words have been repeated correctly. The average j in the open set data of CLC1 is 2.31, which is a typical value for short meaningful sentences. This gives an average standard error of the p estimate of approximately 33% for $p = 50\%$ and of 20% for $p = 90\%$ according to Equation 2. In other words, even much better models than ours can hardly achieve RMSE values close to 20 % or below. In order to compare the prediction accuracy of the different models that participated in this challenge, it might be helpful to analyze the average prediction accuracy across all sentences that have been presented to a listener using a given speech enhancement algorithm.

5. Conclusion

We contributed a hybrid model consisting of the blind Binaural Speech Intelligibility Model (bBSIM) and the Speech Transmission Index (STI) in a correlation-based version. This bBSIM-STI model is very similar to the baseline model MBSTOI and produces a slightly lower RMSE and a slightly higher correlation than the baseline model for the open set and the closed set. The main difference between bBSIM-STI and MBSTOI in the open set can be attributed to a systematic prediction error of MBSTOI for a single speech enhancement algorithm. Further conclusions are:

- The improved prediction accuracy is certainly not caused by the binaural front-end as bBSIM produces virtually the same predictions as the binaural front-end of the baseline model as both front-ends are based on [10].
- The improved prediction accuracy is probably due to small differences in the back-ends. In our back-end an SNR is derived from the correlation values, which is then limited to -15 to 15 dB which reduces the frequency of outliers in the predictions. Apart from this limitation and somewhat longer time frames for the short-term analysis, our back-end is virtually identical to the back-end of the baseline model.
- As the binaural front-end blindly predicts spatial and binaural release from masking (based on the mixed speech and noise signal and without knowledge of the clean speech) it can be combined with arbitrary prediction back-ends and we would be happy if it would be used by other groups in future rounds of the CPC.
- For the signals of CPC1 the bBSIM did not produce a relevant spatial release from masking which indicates that there were no usable binaural cues in the output of the speech enhancement algorithms tested in this challenge. It is unclear if this is a consequence of the listening conditions or of the tested algorithms.

6. Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project ID 352015383 – SFB 1330 A1/A2, Project ID 325439187 – Multilinguale modellbasierte rehabilitative Audiologie and Germany's Excellence Strategy – EXC 2177/1 – Project ID 390895286.

7. References

[1] R. Plomp, "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," *Acustica*, vol. 34, no. 4, pp. 200–211, 1976.

[2] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, 2006.

[3] T. Biberger and S. Ewert, "The effect of room acoustical parameters on speech reception thresholds and spatial release from masking," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2188–2200, 2019.

[4] V. Best, C. R. Mason, J. Swaminathan, E. Roverud, and G. Kidd Jr., "Use of a glimpsing model to understand the performance of listeners with and without hearing loss in spatialized speech mixtures," *The Journal of the Acoustical Society of America*, vol. 141, no. 1, pp. 81–91, 2017.

[5] S. Graetzer, J. Barker, T. J. Cox, M. Akeroyd, J. F. Culling, G. Naylor, E. Porter, and R. Viveros Muñoz, "Clarity-2021 challenges: Machine learning challenges for advancing hearing aid processing," *in Proceeding of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2021, Brno, Czech Republic, 2021.*

[6] J. Barker, M. Akeroyd, T. J. Cox, J. F. Culling, J. Firth, S. Graetzer, H. Griffiths, L. Harris, G. Naylor, Z. Podwinska, E. Porter, and R. Viveros Muñoz, "The 1st Clarity Prediction Challenge: A machine learning challenge for hearing aid intelligibility prediction." *submitted to Interspeech 2022*.

[7] C. F. Hauth, S. C. Berning, B. Kollmeier, and T. Brand, "Modelling binaural unmasking of speech using a blind binaural processing stage," *Trends in Hearing*, vol. 24, 2020.

[8] I. Holube and B. Kollmeier, "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1703–1716, 1996.

[9] T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2801–2810, 2002.

[10] R. Beutelmann, T. Brand, and B. Kollmeier, "Revision, extension, and evaluation of a binaural speech intelligibility model," *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2479–2497, 2010.

[11] ANSI, "ANSI S3.5-1997, American national standard methods for calculation of the speech intelligibility index," *Am. Natl. Stand. Institute, New York*, 1997.

[12] D. Hülsmeier, C. F. Hauth, S. Röttges, K. P., J. Roßbach, M. R. Schädler, B. T. Meyer, A. Warzybok, and T. Brand, "Towards Non-Intrusive Prediction of Speech Recognition Thresholds in Binaural Conditions, in Proc. Conference on Speech Community (ITG)," 2021.

[13] S. Röttges, C. Hauth, J. Rennies, and T. Brand, "Using a blind EC mechanism for modelling the interaction between binaural and temporal speech processing," *Acta Acustica united with Acustica*, Accepted by Acta Acustica united with Acustica 2022.

[14] Y. Nejime and B. C. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *The Journal of the Acoustical Society of America*, vol. 102, no. 1, p. 603–615, 1997.

[15] A. Andersen, J. M. de Haan, Z. Tan, and J. J., "Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions," *Speech Communication*, vol. 102, pp. 1–13, 2018. [Online]. Available: https://doi.org/10.1016/j.specom.2018.06.001

[16] I. Demirsahin, O. Kjartansson, A. Gutkin, and C. Rivera, "Open-source Multi-speaker Corpora of the English Accents in the British Isles," in *Proceedings of The 12th Language Resources and Evaluation Conference (LREC)*. Marseille, France: European Language Resources Association (ELRA), May 2020, pp. 6532–6541. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.804

[17] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time – Frequency Weighted Noisy Speech," *IEEE Transaction on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

[18] B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 750–753, 1983.

[19] V. Hohmann, "Frequency analysis and synthesis using a Gammatone filterbank," *Acta Acustica united with Acustica*, vol. 88, no. 3, p. 433–442, 2002.

[20] J. F. Santos, M. Senoussaoui, and T. H. Falk, "An improved non-intrusive intelligibility metric for noisy and reverberant speech," *2014 14th International Workshop on Acoustic Signal Enhancement, IWAENC 2014*, pp. 55–59, 2014.

[21] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *The Journal of the Acoustical Society of America*, vol. 318, no. 1980, 1979.

[22] X. Chen, Y., W. Xiao, M. Weng, T. Wu, S. Shang, Q. Meng, and N. Zheng, "Clarity-2021 challenges: A Cascaded Speech Enhancement for Hearing Aids in Noisy-Reverberant Conditions," 2021.

[23] B. Kollmeier, M. R. Schädler, A. Warzybok, B. T. Meyer, and T. Brand, "Sentence Recognition Prediction for Hearing-impaired Listeners in Stationary and Fluctuation Noise With FADE: Empowering the Attenuation and Distortion Concept by Plomp With a Quantitative Processing Model," *Trends in Hearing*, vol. 20, pp. 1–17, 2016.

[24] M. R. Schädler, D. Hülsmeier, A. Warzybok, and B. Kollmeier, "Individual Aided Speech-Recognition Performance and Predictions of Benefit for Listeners With Impaired Hearing Employing FADE," *Trends in Hearing*, vol. 24, pp. 1–22, 2020.

[25] L. Bernstein and C. Trahiotis, "A crew of listeners with no more than "slight" hearing loss who exhibit binaural deficits also exhibit higher levels of stimulus-independent internal noise," *The Journal of the Acoustical Society of America*, vol. 147, no. 5, pp. 3188–3196, 2020.

[26] A. Boothroyd and S. Nittrouer, "Mathematical treatment of context effects in phoneme and word recognition," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, p. 101–114, 1988.