



# The 1st Clarity Prediction Challenge: A machine learning challenge for hearing aid intelligibility prediction

Jon Barker<sup>1</sup>, Michael A. Akeroyd<sup>2</sup>, Trevor J. Cox<sup>3</sup>, John F. Culling<sup>4</sup>, Jennifer Firth<sup>2</sup>, Simone Graetzer<sup>3</sup>, Holly Griffiths<sup>2</sup>, Lara Harris<sup>3</sup>, Rhoddy Viveros-Munoz<sup>4</sup>, Graham Naylor<sup>2</sup>, Zuzanna Podwinska<sup>3</sup>, Eszter Porter<sup>2</sup>

<sup>1</sup> Department of Computer Science, University of Sheffield, UK

<sup>2</sup> School of Medicine, University of Nottingham, UK

<sup>3</sup> Acoustics Research Centre, University of Salford, UK

<sup>4</sup> School of Psychology, Cardiff University, UK

claritychallengecontact@gmail.com

## Abstract

This paper reports on the design and outcomes of the 1st Clarity Prediction Challenge (CPC1) for predicting the intelligibility of hearing aid processed signals heard by individuals with a hearing impairment. The challenge was designed to promote the development of new intelligibility measures suitable for use in developing hearing aid algorithms. Participants were supplied with listening test data comprising 7233 responses from 27 individuals. Data was split between training and test sets in a manner that fostered a machine learning approach and allowed both closed-set (known listeners) and open-set (unseen listener/unseen system) evaluation. The paper provides a description of the challenge design including the datasets, the hearing aid algorithms applied, the listeners and the perceptual tests. The challenge attracted submissions from 15 systems. The results are reviewed and the paper summarises, compares and contrasts approaches.

**Index Terms:** speech-in-noise, speech intelligibility, hearing aid, hearing loss, machine learning

## 1. Introduction

New approaches to hearing aid (HA) signal processing are emerging that use machine learning (ML) techniques (e.g., [1]). These approaches may have the potential to revolutionise hearing aid effectiveness for speech-in-noise listening, a situation for which current aids often provide little benefit. However, to fully exploit ML for HAs requires better objective speech intelligibility measures, i.e., to act as targets for optimisation and automatic evaluation. This paper presents the first Clarity Prediction Challenge (CPC1), the first open signal processing competition to directly address this need. Competitors were tasked with designing intelligibility predictors for listeners with a hearing impairment (HI). The challenge ran between November 2021 and April 2022 and below the challenge design is outlined and initial outcomes presented.

While speech intelligibility prediction is not new, the requirements for HAs go beyond what most traditional approaches provide. For example, many intelligibility measures have been developed for the telecommunications industry, evolving out of the speech transmission index [2]. Measures focus on modelling the impact of additive or convolutive channel noise (a key concern in the telecoms industry), but are insufficient for modelling the highly non-linear processing performed by HAs. More recent methods, notably short-time ob-

jective intelligibility (STOI) [3], have been designed using time-frequency weighting. These have typically been developed using young adults with so-called ‘normal hearing’, however, and nearly all lack approaches for including hearing loss.

CPC1 was designed with the development of individualised measures and hearing impairment at its core. Entrants were supplied with speech-in-noise signals that have been processed by a range of experimental HAs. This audio has also been evaluated by a panel of listeners with hearing impairment. CPC1 competitors were tasked to predict the intelligibility of a specific sentence as heard by a listener with quantified hearing characteristics (audiograms and other standard test results.) This provides a high degree of challenge for speech intelligibility algorithms which, to score well, need to produce good predictions over a wide range of signal and listener types.

While intelligibility prediction is a growing area, with much new work emerging (e.g., [4, 5, 6, 7, 8, 9, 10]), there are a lack of common datasets and tasks. So a further motivation for CPC1 has been to benchmark recent approaches. The challenge has attracted input from some of the major groups working in this area, and so represents a useful snapshot of the state-of-the-art.

The paper is set out as follows. Section 2 describes the materials including the signals and listener data from which the challenge is constructed. Section 3 describes the challenge tasks and rules, and briefly outlines the baseline system that was provided to entrants. The challenge submissions are reviewed in Section 4 with results presented in Section 5. The paper concludes in Section 6 with initial findings and future directions.

## 2. Materials

Entrants were challenged to predict the intelligibility of speech-in-noise signals processed by HAs for given listeners. The key factors outlined below are: the signals that were processed; the HA systems; the listener characteristics; and the intelligibility measurements.

### 2.1. The Signals

A 1,500 subset of the 10,000 speech-in-noise scenes generated for the first Clarity Enhancement Challenge (CEC1) [1] were used. Target utterances are 7 to 10 word studio-recorded sentences [11] and the interferers are recordings of common domestic noises such as washing machines. The scenes are simulated first by convolving the source signals with Binaural Room Impulse Responses, which are created in a geometric room

acoustic model [12] using the OIHead-HRTF database [13] (which has recordings for behind-the-ear HA devices with three microphones). As outlined in [1]: room dimensions and materials; and listener, target and noise locations, are randomised.

The level of the interferer signal was adjusted to obtain a specific speech-weighted better ear SNR at the front HA microphone. The SNRs range from -6 to 6 dB chosen on the basis of pilot testing with 13 unaided hearing-impaired listeners. The reverberated speech and noise signals were then summed. The interferer always preceded the onset of the target speech by 2 s and continues for 1 s after the target offset.

## 2.2. The Hearing Aid Systems

The signals have been processed with 10 different HA algorithms from the Clarity Enhancement Challenge (CEC1) [1]. In that challenge, entrants had been tasked with producing HA systems that maximise the intelligibility of speech-in-noise scenes for listeners with known audiograms. Algorithms were allowed to run offline with no constraint on computational cost, however, they had to use causal signal processing with a maximum allowed algorithmic latency of 5 ms.

The HA systems varied considerably, with different approaches to: single channel source separation; multichannel beamforming, and signal amplification. (Full details appear in the Proceedings of the Clarity-2021 Workshop [14]). The algorithms also varied considerably in effectiveness, both in terms of objective measures and listening test outcomes.

## 2.3. The Listeners and Listening Tests

A panel of hearing-impaired listeners was recruited each characterised by bilateral pure-tone audiograms measured at [250, 500, 1000, 2000, 3000, 4000, 6000, 8000] Hz. Only listeners who had a hearing loss of no larger than 80 dB HL in more than two bands were included. Exclusion criteria included: use of hearing intervention other than acoustic hearing aids; diagnosis of Meniere’s disease or hyperacusis, or of severe tinnitus. Hearing loss severity, defined as the average loss in dB HL between 2 and 8 kHz inclusive, was mild (15–35 dB) for 1 listener, moderate (35–56 dB) for 9 listeners and severe (>56 dB) for 17 listeners with a range of 35 dB to 76 dB.

The listening tests were conducted using the project’s Listen@Home software running on Lenovo 10e Chromebook tablets with Sennheiser PC-8 headsets. Participants used these in quiet rooms in their own homes. The software presented the signals in blocks (one HA algorithm per block) and listeners were asked to repeat what they heard. The voice recordings were first converted to text using the Google Cloud Speech-to-Text API <sup>1</sup> and then a team of transcribers validated the Speech-to-Text outputs and corrected errors. Ethical approval was obtained from Nottingham Audiology Services and NHS UK (IRAS Project ID: 276060).

27 listeners completed the tests, providing a total of 7233 responses. An intelligibility score was computed for each listener’s response to a sentence. The transcription of the listener’s response was aligned with the ground-truth text and the number of correctly identified words was counted. The percentage words correct was used as a proxy for sentence intelligibility.

The variable effectiveness of the systems, the different levels of hearing impairment, and SNR range of the source material, provided a wide spread in intelligibility scores and a challenging prediction task.

<sup>1</sup><https://cloud.google.com/speech-to-text>

## 3. Challenge Tasks and Baseline

Participants were asked to predict the intelligibility scores using either ‘intrusive’ or ‘non-intrusive’ systems. The distinction is explained in Figure 1. Intrusive systems, such as the challenge baseline, use representations of the noise-free reference speech signal as an additional input. This is compared to the hearing aid outputs and the distance is mapped onto an intelligibility score. Note, in ML approaches (such as many of the systems submitted to the challenge), separate feature extraction, distance measure and mapping functions may be combined into a complex non-linear model such as a deep neural network with many trained parameters. Non-intrusive systems attempt to estimate intelligibility directly from the HA output signal itself. Generally, this is a harder task and non-intrusive systems are expected to have worse scores.

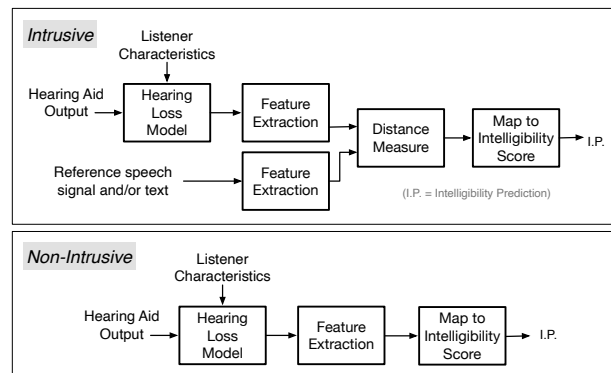


Figure 1: Examples of intrusive vs. non-intrusive models.

The listener data was partitioned into training and test sets. The training set provided full knowledge of every aspect of the data such as the simulation parameters, HA inputs/outputs and listener responses. The test data was released immediate prior to the submission date with just the HA outputs and the listener audiograms. The clean target signals and text was provided for those building intrusive systems.

The data was partitioned in two ways:

- Track 1 (closed set). Same listeners and HA systems in the training set (4812 responses) and test (2421 responses).
- Track 2 (open set). 22 of the 27 listeners and 9 of the 10 systems were in the training set (3545 responses), and the test data had responses from either 5 unseen listeners (432 responses) or the unseen system (249 responses).

In both tracks, the target sentences appearing in the training set are disjoint from those in the test set.

Entrants were provided with a baseline software system which they could choose to adapt or ignore. This baseline had previously been developed as an objective measure for scoring HA algorithms in CEC1 [1]. In brief, the system was a combination of the MSBG hearing loss model [15] and the intrusive MBSTOI intelligibility measure [16]. The MSBG was a Python translation of the MATLAB model developed by the Auditory Perception Group, University of Cambridge. It takes the noisy hearing aid output and the listener’s audiogram, and produces a degraded signal that mimics various aspects of hearing loss. The degraded signal is then compared to the noise-free reference signal using MBSTOI – a binaural version of STOI. Finally, a logistic function maps MBSTOI scores onto the per-

centage words correct, c.f., [3]. The parameters of the logistic were learnt from the training data using a least-mean-square error fit. Note, although billed as a ‘baseline’, this is a sophisticated approach that is expected to produce scores that are not easy to beat. Raw baseline predictions are shown in Figure 2.

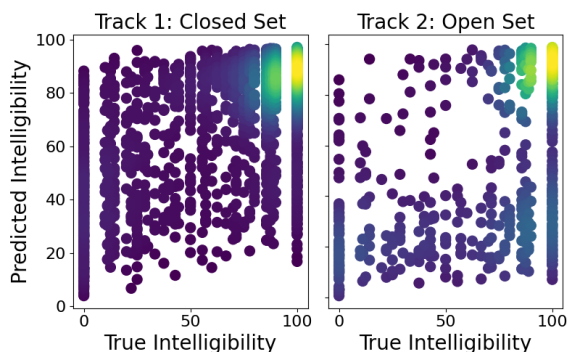


Figure 2: Baseline predictions for Track 1 and Track 2.

## 4. Submissions

A total of 15 systems were submitted originating from 9 separate teams, with most evaluating their systems for both the open and closed sets. Systems have been categorised as intrusive or non-intrusive with a roughly even split between the two types (see column ‘Intr’ of Table 1).

### 4.1. Modelling Hearing Loss

Hearing loss (HL) was modelled in a variety of ways. Many systems used the MSBG model from the baseline (E16, E19, E23, E29, E31, E32, E33, E36). Of these E19 adapted the model by adding noise at the audiogram thresholds to ensure inaudible signal components could not contribute to intelligibility. System E36 employed MSBG indirectly by using the baseline system predictions as one of several inputs to a more complex deep-learning fusion system. In contrast, System E02 and E38 modelled hearing loss using an auditory peripheral model that simulated spike activity in the auditory nerves [17]; their model could be tuned for a given audiogram. Systems E06 and E34 were built on top of the hearing-aid speech perception index (HASPI), which has an internal HL model [18].

Systems E22 and E30 were unusual in that they contained no explicit HL model. E30 used standard better-ear STOI between the HA output and the reference, but then introduced listener individuality by combining STOI (and other features) with the listener ID in a final regression stage (possible only in the closed-set track). E22 applied an existing non-intrusive model directly to the HA outputs and modelled listener variability in the closed-set track by fitting different intelligibility mapping functions for each listener.

### 4.2. Binaural modelling

All entrants had to handle the binaural nature of the challenge. Listening tests were based on the outputs of binaural HA algorithms responding to spatialised scenes, often with spaced target and interferer locations. Further, the listeners may have had different degrees of hearing loss in each ear. This aspect was challenging because binaural cues in the audio might have been degraded by the HA processing. Unsurprisingly, no entrant

used monaural processing by averaging signals or using only one HA output.

The most common strategy, employed in the baseline and by entrants E02, E06, E19, E22, E29, E32, E38, mimicked the ‘better ear’ processing that humans listeners employ. In this strategy, intelligibility scores are made independently for each ear and the overall intelligibility is formed from the better of the two scores. At one extreme this can be performed for the entire signal (e.g., E30). Alternatively, better ear decisions can be made based on short temporal windows and then integrated over time (e.g., E32). The static nature of the scenes probably favoured the former, simpler approach.

Alternatively (or in addition), left and right ears outputs can be combined at the signal level - mimicking human binaural unmasking. The equalization-cancellation (EC) model of binaural processing [19] uses cross-correlation to detect and cancel the effects of noise interference. System E19 used EC processing for frequencies up to 1,500 Hz and better ear processing above 1,500 Hz. The baseline system used an EC approach in combination with better ear processing. E02 and E38 used spiking neural models and a form of EC processing via spike alignment. Most systems chose not to employ traditional binaural unmasking, however. In those that did (e.g., the baseline, E19), the binaural information was found to be too weak to be useful and intelligibility scores were dominated by the better-ear effect.

Other entrants performed binaural processing via data-driven information fusion at various levels. For example, E16 and E33 used a linear layer to fuse frame-level intelligibility decisions. At the other extreme, E23 and E35 concatenated windows from the left and right HA signals, from which latent features were extracted considering maximum mutual information.

### 4.3. Intrusive systems

Intrusive systems traditionally operate by comparing the corrupted signal with that of a clean reference signal using a perceptual-motivated distance measure. Problems arise with HA processing because algorithms can introduce non-linear distortion to the signal to *increase* intelligibility (e.g., frequency shifting). So the appropriate distance measure is not clear, especially if the characteristics of the HA processor are unknown. Nevertheless, techniques like MBSTOI used in baseline, which measure distance using within-band envelope correlations, can produce reasonable scores over a range of operating conditions.

For the systems submitted, E19 used the Binaural Speech Intelligibility model (BSIM) [20] for representing the signals and performed the comparison using the Speech Transmission Index [2] (similar to STOI). E02 used a neural spiking representation and measured mutual information for the comparison. E30 had multiple intrusive and non-intrusive components with the intrusive measure being based on STOI.

Other systems used more abstract representations and/or non-linear regressions to compare the signals. E31 used a single Convolutional Neural Network (CNN) to jointly extract features from the within channel envelopes of the reference and degraded signals. Whereas E32 passed both signals separately through an end-to-end speech recognition system and compared the latent representations formed at various stages in the model. E31 and E36 also used speech recognition modelling to extract deep representations of the signal. Whereas, E30 took the simpler approach of using the prompt text directly (also considered intrusive) to judge the predictability of the sentence, as listeners would find high probability sentences more intelligible.

Table 1: Evaluation of 15 submitted systems plus baseline for RMS prediction error (RMSE) and ground-truth vs prediction correlation (Corr). Results are shown for closed set (Track 1) and open set (Track 2). ‘Intr’ Yes indicates an intrusive system. ‘Prior’ is a system blindly always guessing the mean of the training data intelligibility.

Entrant	Intr.	Track 1 (closed)		Track 2 (open)	
		RMSE ↓	Corr ↑	RMSE ↓	Corr ↑
E30 [22]	Yes	<b>22.5 ± 0.5</b>	0.79	–	–
E32 [23]	Yes	23.1 ± 0.5	0.77	<b>23.5 ± 0.9</b>	0.76
E29 [24]	No	23.3 ± 0.5	0.77	24.6 ± 1.0	0.73
E36 [25]	Yes	24.0 ± 0.5	0.76	29.2 ± 1.2	0.60
E33 [26]	No	24.1 ± 0.5	0.75	28.9 ± 1.1	0.65
E16 [26]	No	24.7 ± 0.5	0.74	30.7 ± 1.2	0.59
E22 [27]	No	25.9 ± 0.5	0.70	32.1 ± 1.2	0.54
E19 [28]	Yes	27.5 ± 0.6	0.66	28.1 ± 1.1	0.63
Base. [1]	Yes	28.5 ± 0.6	0.62	36.5 ± 1.4	0.53
E06 [29]	No	32.0 ± 0.7	0.50	–	–
E34 [29]	No	33.4 ± 0.7	0.43	–	–
E35 [30]	No	35.4 ± 0.7	0.25	35.7 ± 1.4	0.22
Prior	No	36.4 ± 0.7	–	36.2 ± 1.4	–
E31 [31]	Yes	37.2 ± 0.7	0.41	28.3 ± 1.1	0.67
E23 [32]	No	41.5 ± 0.7	0.07	43.7 ± 1.5	0.05
E02 [33]	Yes	–	–	35.2 ± 1.4	0.38
E38 [33]	Yes	–	–	49.7 ± 1.5	0.30

#### 4.4. Non-intrusive systems

The challenge also attracted diverse non-intrusive approaches. Some directly trained systems that tried to learn the speech intelligibility (SI) scores from the degraded signals directly. Whereas others built statistical models of intelligible speech (e.g., training a speech recogniser) and then looked at confidence measures when the system was presented with degraded speech. For example, E29 used confidence measures extracted from an end-to-end speech recognition system, which was first trained on a large speech dataset and then adapted using the MSBG processed signals from the CPC1 training set. A similar approach using automatic speech recognition (ASR) confidence measures was used in E22. E35 used contrastive predictive coding and vector quantization to extract features from a deep learning based SI predictor trained on the CPC1 training set [21]. Whereas, as an example of the direct approach, E16 attempted to learn the SI score from the hearing loss model outputs using a CNN-LSTM architecture with an attention layer. Unsupervised learning was used on MSBG outputs to form a suitable latent representation of the hearing impaired signals. E06 and E34 attempted to build a network trained to predict HASPI scores from the degraded signal only.

## 5. Results

Results are summarised in Table 1 showing: the RMS prediction error with one standard error, and also the correlation between predicted and ground-truth scores. Systems are ranked by the Track 1 RMSE performance (best first). The table also includes the RMSE score achieved when simply guessing the mean of the training data intelligibility for all signals (‘Prior’).

The very best systems had an RMSE of 22.5 (closed) and 23.5 (open) and correlations of 0.79 and 0.76, respectively. This was significantly better than the baseline system. It was also no-

table that the best approaches are quite close in performance despite the diversity of the approaches taken. It was expected from the outset that it would not be possible for systems to achieve very low RMSE scores: there will be some aspects of the listener responses that are simply not predictable from the information provided, e.g., momentary distractions or loss of concentration leading to spurious responses etc.

As expected nearly all systems scored higher on the closed-set challenge than on the open-set, other than system E31. The baseline system scored 28.5 (closed) and 36.5 (open). The performance drop was larger for most of the submitted systems, but unlike some entries, the baseline system had an identical configuration in both tracks. For the closed-set task, the baseline was quite competitive with 8 submissions performing better but 5 performing worse, whereas in the open-track 11 out of 13 systems outperformed the baseline. Note, surprisingly two submissions in each track have performed poorer than the prior system, which guessed the training-set mean speech intelligibility score for every case. One of these was a new approach using spike activity that did not use individual listener characteristics, and so still has potential for improvement [33].

While the challenge quantified hearing abilities, this data appears to have been less useful for predicting speech intelligibility than expected. Given that the listeners had control over the volume level on the tablet and the HA processors already had amplification stages, this probably meant that audibility was not crucial to the listening task, and hence the audiogram was less useful than anticipated. Only a small number of entrants used the other hearing measures (a suprathreshold metric and results from hearing and HA use questionnaires). Nevertheless, the best system in Track 1 did use the listener ID, which would act as a proxy to hearing acuity. In addition, it captured other non-acoustic factors from the listening tests (e.g., how quickly subjects gave up when the task got difficult).

The ranking of the intrusive systems was on average higher than that of the non-intrusive systems, however, the best non-intrusive system was able to come very close to the top performance, i.e., 23.3 vs. 22.5 (closed) and 24.6 vs. 23.5 (open). This was surprising given the large amount of extra information that was provided by access to the reference speech signal.

## 6. Conclusions

This paper outlined the first ever open-challenge for speech intelligibility prediction for signals processed by hearing aids. The best competitors made significant improvements on the baseline model. For the closed track, the best system used a CNN-LSTM with a wide range of input data in addition to the audio from the HA. The best open track model used latent and other representations from a DNN-ASR as inputs to an SI model. The best intrusive methods performed only slightly better than non-intrusive methods. This bodes well for future Clarity Enhancement Challenges to improve HA processing, because accurate non-intrusive models allow optimization of more non-linear ML approaches. The next Prediction Challenge (CPC2) in 2023 will include new data that will reduce the overfitting of models to a single database of listening tests.

## 7. Acknowledgements

Clarity is funded by UKRI (EP/S031448/1, EP/S031308/1, EP/S031324/1 and EP/S030298/1). We thank Amazon, the Hearing Industry Research Consortium and the Royal National Institute for the Deaf (RNID) for their support.

## 8. References

- [1] S. Graetzer, J. Barker, T. J. Cox, M. Akeroyd, J. F. Culling, G. Naylor, E. Porter, and R. V. Muñoz, “Clarity-2021 challenges: Machine learning challenges for advancing hearing aid processing,” in *Proceedings of Interspeech 2021*, Aug. 2021, pp. 686–690.
- [2] T. Houtgast and H. Steeneken, “A physical method for measuring speech-transmission quality,” *The Journal of the Acoustical Society of America*, vol. 67, pp. 318–326, 1980.
- [3] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time–frequency weighted noisy speech,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.
- [4] K. Yamamoto, T. Irino, T. Matsui, S. Araki, K. Kinoshita, and T. Nakatani, “Predicting speech intelligibility using a gammachirp envelope distortion index based on the signal-to-distortion ratio,” in *INTERSPEECH*, 2017, pp. 2949–2953.
- [5] C. Sørensen *et al.*, “Non-intrusive intelligibility prediction using a codebook-based approach,” in *EUSIPCO*. IEEE, 2017, pp. 216–220.
- [6] C. Spille *et al.*, “Predicting speech intelligibility with deep neural networks,” *Computer Speech & Language*, vol. 48, pp. 51–66, 2018.
- [7] K. Arai, S. Araki, A. Ogawa, K. Kinoshita, T. Nakatani, K. Yamamoto, and T. Irino, “Predicting speech intelligibility of enhanced speech using phone accuracy of DNN-based asr system,” in *Interspeech*, 2019, pp. 4275–4279.
- [8] M. Karbasi, S. Bleeck, and D. Kolossa, “Non-intrusive speech intelligibility prediction using automatic speech recognition derived measures,” *arXiv preprint arXiv:2010.08574*, 2020.
- [9] R. E. Zezario *et al.*, “STOI-Net: A deep learning based non-intrusive speech intelligibility assessment model,” in *APSIPA ASC*. IEEE, 2020, pp. 482–486.
- [10] A. M. C. Martínez, C. Spille, J. Roßbach, B. Kollmeier, and B. T. Meyer, “Prediction of speech intelligibility with DNN-based performance measures,” *Computer Speech & Language*, p. 101329, 2021.
- [11] S. Graetzer, M. A. A. amd Jon Barker, T. J. Cox, J. F. Culling, G. Naylor, and E. P. amd Rhoddy Viveros-Muñoz, “Dataset of British English speech recordings for psychoacoustics and speech processing research: The Clarity speech corpus,” *Data in Brief*, vol. 41, no. 107951, Apr. 2022.
- [12] D. Schröder and M. Vorländer, “RAVEN: A real-time framework for the auralization of interactive virtual environments,” in *Forum Acusticum*, Denmark: Aalborg, 2021, pp. 1541–1546.
- [13] F. Denk, S. M. Ernst, J. Heeren, S. D. Ewert, and B. Kollmeier, “The Oldenburg Hearing Device (OIHead) HRTF Database,” University of Oldenburg, Tech. Rep., 2018.
- [14] *Proc. ISCA Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2021)*, Virtual Workshop, Sep. 2021. [Online]. Available: <https://claritychallenge.github.io/clarity2021-workshop/>
- [15] 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, pp. 603–615, 1997.
- [16] A. H. Andersen, J. M. de Haan, Z. H. Tan, and J. Jensen, “Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions,” *Speech Communication*, vol. 102, pp. 1–13, 2018.
- [17] I. C. Bruce, Y. Erfani, and M. S. Zilany, “A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites,” *Hearing Research*, vol. 360, pp. 40–54, 2018.
- [18] J. M. Kates and K. H. Arehart, “The hearing-aid speech perception index (HASPI),” *Speech Communication*, vol. 65, pp. 75–93, 2014.
- [19] N. I. Durlach, “Equalization and cancellation theory,” in *Foundations of Modern Auditory Theory*, J. V. Tobias, Ed., 1972, vol. 2, pp. 371–463.
- [20] 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.
- [21] A. F. McKinney and B. Cauchi, “Non-intrusive binaural speech intelligibility prediction from discrete latent representations,” *Signal Processing Letters*, vol. to appear, 2022.
- [22] M. Huckvale and G. Hilkhuysen, “ELO-SPHERES intelligibility prediction model for the Clarity Prediction Challenge 2022,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.
- [23] Z. Tu, N. Ma, and J. Barker, “Exploiting hidden representations from a DNN-based speech recogniser for speech intelligibility prediction in hearing-impaired listeners,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.
- [24] —, “Unsupervised uncertainty measures of automatic speech recognition for non-intrusive speech intelligibility prediction,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.
- [25] N. Kamo, K. Arai, A. Ogawa, S. Araki, T. Nakatani, K. Kinoshita, M. Delcroix, T. Ochiai, and T. Irino, “Conformer-based fusion of text, audio, and listener characteristics for predicting speech intelligibility of hearing aid users,” in *Proceedings of the 2nd Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2022)*, Online, Jun. 2022.
- [26] R. E. Zezario, F. Chen, C.-S. Fuh, H.-M. Wang, and Y. Tsao, “Mbi-net: A non-intrusive multi-branched speech intelligibility prediction model for hearing aids,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.
- [27] J. Roßbach, R. Huber, S. Roßtges, C. F. Hauth, T. Biberger, T. Brand, B. T. Meyer, and J. RENNIES, “Speech intelligibility prediction for hearing-impaired listeners with the LEAP model,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.
- [28] S. Roßtges, J. R. C. F. Hauth, T. Biberger, B. T. Meyer, R. Huber, J. RENNIES, and T. Brand, “Speech intelligibility prediction using the bBSIM-STI model - technical report contribution E019,” in *Proceedings of the 2nd Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2022)*, Online, Jun. 2022.
- [29] G. Close, S. Hollands, S. Goetze, and T. Hain, “Non-intrusive speech intelligibility metric prediction for hearing impaired individuals,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.
- [30] A. F. McKinney and B. Cauchi, “Non-intrusive prediction of speech intelligibility for the first Clarity Prediction Challenge (CPC1),” in *Proceedings of the 2nd Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2022)*, Online, Jun. 2022.
- [31] C. O. Mawalim, B. A. Titalim, and M. Unoki, “CPC1 E031 system description,” in *Proceedings of the 2nd Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2022)*, Online, Jun. 2022.
- [32] A. F. McKinney and B. Cauchi, “Non-intrusive prediction of speech intelligibility for the first Clarity Prediction Challenge (CPC1),” in *Proceedings of the 2nd Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2022)*, Online, Jun. 2022.
- [33] F. Alvarez and W. Nogueira, “Predicting speech intelligibility using the spike activity mutual information index,” in *Proceedings of Interspeech 2022*, Incheon, South Korea, Sep. 2022.