Clarity-2021 challenges : machine learning challenges for advancing hearing aid processing

Graetzer, SN, Barker, J, Cox, TJ, Akeroyd, M, Culling, JF, Naylor, G, Porter, E and Viveros Munoz, R

10.21437/Interspeech.2021-1574

INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Clarity-2021 challenges: Machine learning challenges for advancing hearing aid processing

Simone Graetzer[1], *Jon Barker*[2], *Trevor J. Cox*[1], *Michael Akeroyd*[3], *John F. Culling*[4], *Graham Naylor*[3], *Eszter Porter*[3], *Rhoddy Viveros Muñoz*[4]

[1] Acoustics Research Group, University of Salford, UK
[2] Department of Computer Science, University of Sheffield, UK
[3] School of Medicine, University of Nottingham, UK
[4] School of Psychology, Cardiff University, UK

claritychallengecontact@gmail.com

Abstract

In recent years, rapid advances in speech technology have been made possible by machine learning challenges such as CHiME, REVERB, Blizzard, and Hurricane. In the Clarity project, the machine learning approach is applied to the problem of hearing aid processing of speech-in-noise, where current technology in enhancing the speech signal for the hearing aid wearer is often ineffective. The scenario is a (simulated) cuboid-shaped living room in which there is a single listener, a single target speaker and a single interferer, which is either a competing talker or domestic noise. All sources are static, the target is always within $\pm 30^{\circ}$ azimuth of the listener and at the same elevation, and the interferer is an omnidirectional point source at the same elevation. The target speech comes from an open source 40-speaker British English speech database collected for this purpose. This paper provides a baseline description of the round one Clarity challenges for both enhancement (CEC1) and prediction (CPC1). To the authors' knowledge, these are the first machine learning challenges to consider the problem of hearing aid speech signal processing.

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

1. Introduction

By 2035, there will be 15 million people with hearing loss in the UK at an annual economic cost of 30 billion pounds [1, 2]. People with hearing loss are more susceptible to interference from background noise than unimpaired listeners. Yet speech in noise remains a major problem for hearing aid technology. Current hearing aids are often ineffective when the signal-to-noise ratio (SNR) is relatively low. Hearing aid wearers often complain that speech intelligibility is poor, and this is a common reason for lack of use [3]. Traditional devices tend to amplify the noise in addition to the target speech.

Over the last few decades, there have been major advances in machine learning applied to speech technology. For example, in automatic speech recognition (ASR), performance is unrecognisable when compared with what was possible ten years ago. In the CHiME, REVERB, Blizzard, and Hurricane challenges, researchers have made rapid progress by building on open source baseline software that is improved in each round [4, 5, 6, 7]. Advances can also be attributed to the availability of speech corpora recorded in various environments. Recent developments in machine learning applied to noise reduction and speech enhancement indicate that this is a promising approach for hearing aid speech signal processing. However machine learning challenges typically assume healthy hearing, and access to hearing-impaired listeners for rigorous algorithm evaluation is limited.

In the first round of the Clarity project challenges, we address the problem of speech-in-noise in everyday home environments. This paper is intended to be a reference for this round, in which the scenario is a simulated cuboid-shaped living room in which there is a single listener, a single target speaker and a single interferer, which is either a single competing talker or domestic noise. All sources are static. The target speech materials were collected specifically for the round. The software and datasets are publicly available [8]. This first round features

- A large target speech database of English sentences produced by 40 British English speakers;
- Simulated living rooms with a single static target speech source, interferer and listener, built using room impulse responses generated by the Real-time framework for the Auralization of interactive Virtual ENvironments (RAVEN, [9]) and head-related impulse responses (HRIRs) recorded for a number of humans and manikins [10];
- Baseline hearing aid software built on the open Master Hearing Aid (openMHA, [11]);
- Baseline hearing loss software based on the model developed by the Auditory Perception Group at the University of Cambridge (see, e.g., [12]);
- Baseline speech intelligibility software based on the Modified Binaural Short-Time Objective Intelligibility model or MBSTOI [13].

In the following, we introduce the round one scene generation and datasets in Section 2 and the baseline system in Section 3. We discuss the first Enhancement and Prediction challenges and the challenge instructions in Sections 4 and 5, respectively. We conclude in Section 6. More details can be found on the challenge website [1].

2. Scene generation and datasets

The software is implemented in Python. It allows entrants to compare the performance of their system with the baseline system, and comprises a scene generation module, a hearing aid module, a hearing loss module and a speech intelligibility module.

[1] http://claritychallenge.org

http://dx.doi.org/10.21437/Interspeech.2021-1574

Table 1: *Round one Enhancement challenge (CEC1) timeline, where eval refers to the evaluation dataset.*

Stage	Date
$CEC1\ initial\ release$	$15-03-2021$
$CEC1\ eval\ release$	$01-06-2021$
$CEC1\ submission\ deadline$	$15-06-2021$
$Results\ announced/Interspeech$	$17-09-2021$

The scene generation software was used to create 10,000 unique scenes. The sound at the listener or receiver is generated first by convolving the source signals with the Binaural Room Impulse Responses (BRIRs), which are created in RAVEN and draw on HRIRs from the OlHead-HRTF database [10]. Target and interferer are mixed to obtain a specific speech-weighted better ear SNR at the reference microphone (front). The SNRs for the speech interferer range from 0 to 12 dB, while the SNRs for the noise interferer range from -6 to 6 dB. These ranges were chosen on the basis of pilot testing with 13 unaided hearing-impaired listeners. The reverberated speech and noise signals are then summed. The interferer always precedes the onset of the target by 2 s and follows the offset by 1 s.

2.1. Scenario, room geometry and materials

In the scenes, the listener is either sitting (with a height, *H*, of 1.2 m) or standing (H = 1.6 m), with the sound sources at the same elevation. These heights correspond to the centre of the listener's head and the target speaker's mouth. The target is always placed at a distance of $\geq$ 1 m and within $\pm 30^\circ$ azimuth inclusive (with a step of 7.5°) of the front of the listener and at an elevation of 0°. The target is always facing the listener. The interferer can be located in any position except within 1 m of the walls or the listener and is omnidirectional. Both target and interferer are point sources. The room is cuboid in shape.

The dimensions and reverberation times of the room were based on published statistics on British living rooms [14]. The reverberation time was low to moderate at 0.2 to 0.4 s on average between 125 Hz and 1 kHz. Rooms feature variations in surface absorption to represent doors, windows, curtains, rugs and furniture, combined with scattering coefficient of 0.1. The room dimensions, boundary materials, and the locations of the listener, target and interferer are randomised. The rooms have a length, L, set using a uniform probability distribution pseudo-random number generator with $3 \leq L(m) \leq 8$ and a height set using a Gaussian distribution with a mean of 2.7 m and standard deviation of 0.8 m. The area is set using a Gaussian distribution with mean 17.7 m^2 and standard deviation of 5.5 m^2.

One of the walls of the room is randomly selected for the location of the door. The door can be at any position with the constraint that it is at least 0.2 m from any corner of the room. A window is placed on one of the other three walls. The window can be at any position on the wall but is at H = 1.9 m and at 0.4 m from any corner. The curtains are simulated to the side of the window. For larger rooms, a second window and curtains are simulated following a similar method. A sofa is simulated at a random position as a layer on the wall and the floor. A rug is simulated at a random location on the floor.

The listener is positioned within the room using a uniform probability distribution for the x and y coordinates. There are constraints to ensure that the receiver is not within 1 m of the wall. The listener is positioned so as to be roughly facing the target talker (*i.e.*, within $\pm 30^\circ$ azimuth inclusive where $angle = 7.5n$ where n is an integer and $|n| \leq 4$). The target talker has a speech directivity pattern, while the interferer is a single point source radiating speech or non-speech noise omnidirectionally. Both target and interferer are randomly placed within the room with a uniform distribution except not within 1 m of the walls or receiver, and are at the same elevation as the receiver.

2.2. Head-related impulse responses

HRIRs were drawn from the OlHead-HRTF database [10]. Measurements made close to the ear drum (ED) and via a behind-the-ear (BTE) hearing aid form factor were used. The BTE model was equipped with three miniature microphones (front, middle and rear) with a distance between microphones of approximately 7.6 mm: front-mid 0.0076 m and mid-rear 0.0073 m. In the horizontal plane there is a uniform resolution of 7.5°.

2.3. Target speech database

The target speech materials were collected specifically for the first round: 10,000 unique sentences recorded by 40 British English speakers. These sentences were selected from the British National Corpus (XML edition, 2007, [15]) of (mainly) written text materials, including novels, pamphlets, etc., but excluding poetry. These sentences contain 7-10 words, all with a word frequency of at least one in the Kucera and Francis (1967) [16] database, and hand checked for acceptable grammar and vocabulary by the authors. The sentences were recorded in home studios (due to COVID-19) by forty voice actors from a radio production company, reading 250 sentences each. Semi-automated segmentation was performed using Google Speech-to-Text API with Python and unix shell scripts [17]. Segmented recordings were equalised in active speech level [18]. The database is publicly available [8].

2.4. Interferer databases

The types of interferers included in the databases were informed by a patient discussion group hosted by the University of Nottingham in March 2020. In half of the scenes, a speech interferer is used, while in the other half, the interferer is one of several types of domestic noise sources.

The speech interferer data, which come from the Open-source Multi-speaker Corpora of the English Accents in the British Isles [19], includes speakers with a range of UK and Ireland English accents. For each speaker, utterances are concatenated with a short period of silence in between (300 ms in addition to any silence at the beginning and ends of original recordings). The noise interferer data is a collection of samples, mainly from Freesound [20], under Creative Commons licences.

2.5. Listener characterisation databases

For the training (*train*) and development (*dev*) datasets, listeners are characterised only by bilateral pure-tone audiograms at the following frequencies: [250, 500, 1000, 2000, 3000, 4000, 6000, 8000] Hz. These audiograms were obtained for people who are **not** members of the listener panel; hence there is listener independence between the evaluation (*eval*) and non-eval datasets. As the baseline hearing loss module may not produce sensible results for hearing losses greater than 80 dB Hearing Level (HL), only listeners who had a hearing loss no

greater than 80 dB HL in more than two bands were included ($N = 83$). Any losses greater than 80 dB HL were limited to 80 dB HL. Hearing loss severity, defined according to the Cambridge hearing loss model as the average loss in dB HL between 2 and 8 kHz inclusive, was mild for four listeners (defined as $15 > HL(dB) > 35$), moderate for 26 (defined as $35 > HL(dB) > 56$), and severe for 53 ($HL > 56$ dB).

For the initial, MBSTOI evaluation, additional audiograms were obtained from the same source, and the same rules were applied. For the listening test evaluation, the audiograms are those obtained for the members of the listener panel, which comprises 50 bilateral hearing aid users with symmetric or asymmetric hearing loss. They have an averaged sensorineural hearing loss between 25 and about 60 dB in the better ear. Exclusion criteria included the following: use of any hearing intervention other than acoustic hearing aids, use of a programmable ventriculo-peritoneal (PVP) shunt, diagnosis of Meniere's disease or hyperacusis, and diagnosis of severe tinnitus. Ethical approval was obtained from Nottingham Audiology Services and the National Health Service UK for collection and use of these data (IRAS Project ID: 276060).

2.6. Challenge datasets

Scenes were pseudo-randomly allocated to one of three datasets. The train and dev datasets are 6000 and 2500 scenes in size, respectively. Therefore, the models are trained on 6000 target speech utterances, where half of the scenes include a speech interferer, and half, a non-speech interferer. The eval dataset is 1500 scenes in size. Target speakers are allocated to datasets such that there is an equal representation of female and male talkers. Train and dev datasets included extensive metadata including the ID of the target and interferer, the azimuths of both sources relative to the listener, the positions of the listener and sources, and the offset used to identify the start of the segment of interferer used in the scene. The metadata also included the assignment of scenes to listeners, where for the train and dev sets, each scene is processed for three listeners. Both train and dev datasets were provided when the challenges were launched to be used during system development. Additionally, scene generation code was provided to generate the datasets, and to augment them if desired (however any submission can only be based on the predefined datasets).

The eval dataset was held back for evaluation of the developed systems and was provided shortly before the challenge submission deadline. For the enhancement challenge, participants were required to run their systems on the mixed hearing aid input signals and submit the system outputs to the organisers for evaluation.

3. Baseline system

The baseline system - hearing aid, hearing loss and speech intelligibility modules - is provided for optional use by competition entrants.

3.1. Baseline hearing aid module

The intention was to simulate a basic hearing aid without various aspects of signal processing that are common in high-end hearing aids, but tend to be implemented in proprietary forms so cannot be replicated exactly. Most modern devices include both multiband dynamic compression and some sort of directional microphone processing. These devices are typically monaural, with no bilateral wireless link.

The baseline fitting algorithm, the Camfit compressive algorithm [21], is used to calculate compression ratios per frequency band where bands have centre frequencies as follows: [177, 297, 500, 841, 1414, 2378, 4000, 6727] Hz. A one-to-one input-output ratio is used below the compression thresholds in each band. The compression thresholds are determined according to the levels in each frequency band of speech with a standard long-term average spectrum and an overall level of 45 dB SPL, *i.e.*, speech that is just audible for a normal hearing listener, where speech produced with a normal vocal effort has an overall level of 60 $L_{A,eq}$ at 1 m [22, 23].

The baseline hearing aid involves a configuration of the openMHA system for a simple Behind The Ear (BTE) model with front and rear microphones with distances of 0.0149 m between them (determined by the HRIRs). This configuration of openMHA includes multiband dynamic compression via the *dc* plugin and directional processing - non-adaptive differential processing - via the *adm* plugin with no additional noise reduction or bilateral link. The aim of the compression component is to compensate for the listener's hearing loss (raised auditory thresholds) and to fit the output level into the listener's dynamic range. The aim of the directional processing is to improve the signal-to-noise ratio and, in particular, to attenuate sources in the rear hemisphere of the listener (in this case, using a hypercardioid polar pattern). The plugin combines the signal from two omnidirectional microphones on each hearing aid: in this case the front and rear microphones [24]. In the first round, the baseline does not simulate the direct path.

3.2. Baseline hearing loss module

The baseline hearing loss module is a Python translation of the MATLAB model developed by Moore, Stone and other members of the Auditory Perception Group at the University of Cambridge (see, e.g., [12]). It uses a gammatone filterbank model of the auditory system and simulates four main aspects of hearing loss: decreased audibility (the raising of auditory thresholds), reduced dynamic range (loudness recruitment), and the loss of temporal and frequency resolution (frequency selectivity). For people with a hearing impairment, the auditory thresholds are increased while the loudness discomfort threshold remains at the same level; hence the dynamic range is reduced. The loss of frequency selectivity reduces the ability to discriminate between sounds at different frequencies, and is due to loss of cochlea hair cell sensitivity. Signals are attenuated in each frequency band according to the listener's audiogram to simulate the raising of auditory thresholds. The loudness recruitment filterbank comprises 28 filters with two times broadening. Frequency smearing is performed according to the severity of the hearing loss as defined in section 2.5. The higher the degree of smearing, the higher the level of noise between signal components.

3.3. Baseline speech intelligibility module

The baseline speech intelligibility module is MBSTOI [13], which is a binaural speech intelligibility metric based on the Short-Time Objective Intelligibility metric (STOI) [25], and is translated from MATLAB into Python. Both MBSTOI and STOI are invasive metrics that require both the clean speech reference and the processed or degraded signal. STOI-based approaches are suitable for non-linear processing, such as clipping, and do not require access to the noise separately. To date, most of the published literature applying MBSTOI to the hearing aid context concerns beamforming approaches to noise reduction. The findings indicate that binaural versions of STOI

perform well in evaluating the effects of these processes on intelligibility with or without additive noise and reverberation (e.g., [26]).

MBSTOI downsamples the input signals to 10 kHz, and analyses the signals with a short-time Discrete Fourier Transform (386 ms), with parameter values as in the case of STOI. The DFT coefficients from the left ear and the right ear are combined using an Equalisation-Cancellation (EC) stage. This stage models the binaural advantage obtained by having two ears in situations where there is spatial separation between target and interferer. Independent noise sources, referred to as jitter, are added to the grid search over interaural time and level differences to align any interferer or distortion components to limit performance to be consistent with human performance. The combined DFT coefficients are used to compute power envelopes in one-third octave bands. These envelopes are arranged in vectors or regions of $N = 30$ samples. Intermediate correlation coefficients are then calculated per band and time frame (over the N sample region). In the better ear stage, intermediate correlation coefficients are calculated similarly, but per ear. For each band and frame/region, the maximal intermediate correlation coefficients are chosen from the EC and better ear stage outputs. The final MBSTOI measure, d, is obtained by averaging the intermediate correlation coefficients across time and frequency. See [13] for more detail.

4. Enhancement challenge

In the enhancement challenge (CEC1), the task is to replace the baseline hearing aid algorithm with an algorithm that improves the speech intelligibility of the mixture signals for the listeners relative to baseline. To ensure that the number of entries to be evaluated by the listener panel is not too large, an initial ranking on the basis of MBSTOI scores over the eval dataset will be performed to identify the most promising candidates. Ultimately, entries will be ranked according to measured intelligibility scores from the listener panel over the eval dataset.

4.1. Challenge rules

A set of challenge "rules" were provided to participants. The rules were designed to keep systems close to the application scenario, to avoid unintended overfitting, and to make systems directly comparable. Algorithms were required to be causal; the output at time *t* must not use any information from input samples more than 5 ms into the future (*i.e.*, any information from input samples $> t$ + 5 ms). Entrants were required to submit their processed signals in addition to a technical document describing the system/model and any external data and pre-existing tools, software and models used. All parameters were to be tuned using only the provided train and dev datasets, and systems were to be run (ideally only once) on the eval dataset using parameter settings identified on the basis of dev dataset results. The rules are intended to allow some freedom to build systems that would be implementable in a hearing aid with the necessary computing power, even if these systems are not currently feasible.

4.2. Baseline system results

Baseline system performance as MBSTOI d by SNR (dB) for the dev set is shown in Figure 1, where the interconnected points indicate the means and the points with low opacity indicate the variability in the data according to, for example, speech material, listener hearing loss, distances between the listener and the target and interferer, room volume, and separation angle. The mean and median d are 0.41. For the most part, there is a positive monotonic relationship between MBSTOI and SNR, as anticipated. For the two interferer types, there is a weak to moderate positive correlation between MBSTOI and SNR (speech: τ = 0.35, $p < 0.001$; noise: τ = 0.49, $p < 0.001$) and a weak negative correlation between MBSTOI and target-listener distance (τ = -0.13, $p < 0.001$). That is, as the SNR becomes more favourable, or the distance between the target speaker and listener reduces, MBSTOI increases, as would be anticipated.

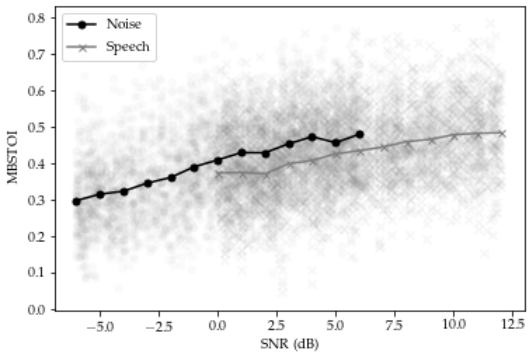

Figure 1: *MBSTOI by SNR (dB) per interferer type.*

5. Prediction challenge

In the prediction challenge (CPC1), the task is to replace the hearing loss and/or speech intelligibility models in the pipeline. The listening tests for CEC1 will provide the data for CPC1; *i.e.*, the outcomes of CEC1 will facilitate the improvement of prediction models. The proposed CPC1 launch date is October 2021. The baseline models are those described in sections 3.2 and 3.3. The models submitted may be a single speech intelligibility model that can account for the behaviour of both healthy hearing and hearing-impaired listeners, or two separate models for hearing loss and speech intelligibility. Ranking of entries will be determined by prediction accuracy over the eval dataset.

6. Conclusions

The first round of the Clarity challenges aims to produce improved hearing aid algorithms for speech signal processing. The scenario is a (simulated) cuboid-shaped living room in which there is a single listener and two static sources: a single target speaker, and a single interferer, where the interferer is a competing talker or domestic noise. The first enhancement challenge, CEC1, involves developing hearing aid models that improve on the baseline, while the first prediction challenge, CPC1, involves improving on the baseline speech intelligibility prediction model(s). The submitted CEC1 systems and results will be presented at the Clarity-2021 workshop in September, 2021.

7. Acknowledgements

This research is funded by the UK's Engineering and Physical Sciences Council under grants EP/S031448/1, EP/S031308/1, EP/S031324/1 and EP/S030298/1. We are grateful to Amazon, the Hearing Industry Research Consortium, the Royal National Institute for the Deaf (RNID), and Honda for their support.

8. References

[1] Royal National Institute for the Deaf (RNID), "Facts and figures," https://rnid.org.uk/about-us/research-and-policy/facts-and-figures/, 2021, accessed: 2021-03-01.

[2] S. Archbold, B. Lamb, C. O'Neill, and J. Atkins, "The Real Cost of Adult Hearing Loss: reducing its impact by increasing access to the latest hearing technologies," The Ear Foundation, Tech. Rep., 2014.

[3] S. Kochkin, "Marketrak vi: Consumers rate improvements sought in hearing instruments," *Hearing Review*, vol. 9, pp. 18–22, 2002.

[4] S. Watanabe, M. Mandel, J. Barker, E. Vincent *et al.*, "CHiME-6 challenge: Tackling multispeaker speech recognition for unsegmented recordings," 2020. [Online]. Available: https://arxiv.org/abs/2004.09249

[5] K. Kinoshita, M. Delcroix, S. Gannot *et al.*, "A summary of the reverb challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, vol. 1, pp. 1–19, 2016.

[6] K. Prahallad, H. A. Murthy, S. King, A. W. Black, K. Tokuda *et al.*, "The Blizzard challenge 2014," in *Proc. Blizzard Challenge workshop*, 2014.

[7] M. Cooke, C. Mayo, and C. Valentini-Botinhão, "Intelligibility-enhancing speech modifications: the hurricane challenge," in *Proceedings of Interspeech 2013*, 2013, pp. 3552–3556.

[8] J. Barker, S. Graetzer, and T. Cox, "Software to support the 1st Clarity Enhancement Challenge [software and data collection]," https://doi.org/10.5281/zenodo.4593856, 2021, accessed: 2021-06-03.

[9] 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.

[10] F. Denk, S. M. Ernst, J. Heeren, S. D. Ewert, and B. Kollmeier, "The Oldenburg Hearing Device (OlHeaD) HRTF Database," University of Oldenburg, Tech. Rep., 2018.

[11] H. Kayser, T. Herzke, P. Maanen, C. Pavlovic, and V. Hohmann, "Open Master Hearing Aid (openMHA): An integrated platform for hearing aid research," *The Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2879–2879, 2019.

[12] 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.

[13] 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.

[14] M. A. Burgess and W. A. Utley, "Reverberation times in British living rooms," *Applied Acoustics*, vol. 18, no. 5, pp. 369–380, 1985.

[15] BNC Consortium, "British National Corpus, version 3 (BNC XML Edition)," http://www.natcorp.ox.ac.uk/, 2007, distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Accessed: 2021-03-01.

[16] H. Kucera and W. Francis, *Computational analysis of present day American English*. Providence, RI: Brown University Press, 1967.

[17] Google, "Google Speech-to-Text Application Programming Interface," https://cloud.google.com/speech-to-text/, accessed: 2021-03-01.

[18] ITU-T P.56, *Objective measurement of active speech level*. ITU-T Recommendation P.56, 1993.

[19] 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

[20] F. Font, G. Roma, and X. Serra, "Freesound technical demo," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013.

[21] B. C. J. Moore, J. I. Alcántara, M. Stone, and B. R. Glasberg, "Use of a loudness model for hearing aid fitting: II. Hearing aids with multi-channel compression," *British Journal of Audiology*, vol. 33, no. 3, pp. 157–170, 1999.

[22] D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, C. Ludvigsen *et al.*, "An international comparison of long-term average speech spectra," *The Journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 2108–2120, 1994.

[23] BS EN ISO 9921, *Ergonomics—Assessment of Speech Communication*. International Organization for Standardization, Geneva, 2003.

[24] G. W. Elko and A. T. N. Pong, "A simple adaptive first-order differential microphone," in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE*, 1995, pp. 169–172.

[25] 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.

[26] A. H. Andersen, J. M. D. Haan, Z. H. Tan, and J. Jensen, "A binaural short time objective intelligibility measure for noisy and enhanced speech," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.