DOI: 10.61782/fa.2023.0876

THE CLARITY & CADENZA CHALLENGES

Michael A Akeroyd[1*] **Will Bailey**[2] **Scott Bannister**[3]
Jennifer Firth[3] **Simone Graetzer**[4] **Gerardo Roa Dabike**[4]
Zehai Tu[2] **Rebecca Vos**[4] **Jon Barker**[2]
Trevor J. Cox[4] **John F. Culling**[5] **Bruno Fazenda**[4]
Alinka Greasley[3] **Graham Naylor**[1] **William Whitmer**[1]

[1] Hearing Sciences, University of Nottingham, Nottingham UK
[2] Dept. of Computer Science, University of Sheffield, Sheffield, UK
[3] School of Music, University of Leeds, Leeds, UK
[4] Acoustical Engineering, University of Salford, Salford, UK
[5] School of Psychology, Cardiff University, UK

ABSTRACT

The Clarity (Speech in noise) and Cadenza (music) projects are two large, complementary research projects that are exploiting the latest in machine learning to create improved listening experiences for those with a hearing loss. In both, we are running a series of open competitions, for which entrants are challenged to improve and personalise the audio for listeners with a hearing loss. This challenge methodology fosters a new research community devoted to making music and speech more accessible, as well as creating open-source tools and databases to facilitate future investigations. The challenges pose a variety of dilemmas to the competitors: for instance, while a hearing aid must manipulate live speech with low latency and limited computing power, recorded music from consumer devices can be pre-processed with non-causal techniques using cloud computing. In this presentation we will update the latest news on the third Clarity challenge and the first Cadenza challenge and report on the open-access computational tools and rating scales we have developed.

Keywords: *hearing loss, speech in noise, music.*

[*]***Corresponding author***: *mszmaa@nottingham.ac.uk*

1. INTRODUCTION

At least 1.5 billion people live with hearing loss, and this number is set to increase as the global population ages. Many of them would benefit from a hearing aid, yet only a fraction of those have them, and further many of them do not use their devices often enough. A major reason for this low uptake is the perception that hearing aids do not perform well enough.

The projects are aiming to stimulate progress in signal processing. Clarity focusses on speech in noise via hearing aids, whereas Cadenza focusses on music processing via consumer devices and hearing aids. Both projects are organizing open challenges, in which we supply a scenario, software tools, datasets (both training and evaluation) and baseline systems (including computational models of hearing aids and hearing loss). Entrants are challenged to develop their own systems that can do better on the evaluation tests than our baseline and other competitors. For speech in noise, this means improving the percent-correct of words identified, for music, this means the perceived audio quality of the music. Systems are evaluated using both computational measures and panels of hearing-impaired listeners.

In this paper we briefly summarize the high-level designs of the scenarios in both projects.

2. THE CLARITY CHALLENGES

The Clarity project is the earlier of the two. We have run three enhancement challenges ("CEC1" [1], "CEC2" [2], "ICASSP 2023" [3]), the latter two more complex than the first, and a speech intelligibility prediction challenge ("CPC1" [4]) following on from the first enhancement challenge.

The scenario in CEC1 was a small room, with low to moderate reverberation, in which a listener is listening to a sentence from a target talker. The sentence was one of over 10000 recordings from the new, open-access 7-10 word Clarity utterances [5]. An interferer sound was playing simultaneously; this was either a competing talker or a continuous noise source (e.g., a washing machine). All the signals were processed by binaural room impulse responses to give the signals that would be received by hearing-aid microphones. There was no target or listener motion. The target sound started 2 seconds after the start of the interferer, so it was clear and unambiguously identifiable for the real listening tests. It also gave the hearing aid algorithms some time to adjust to the background noise. The SNR was set to broadly cover the full psychometric function.

The scenario in CEC2 was made harder in three main respects. First, there could be two or three interferer sounds running throughout the audio, being a stream of competing speech, continuous domestic noise (e.g., a washing machine), or music. Second, the listener started looking away from the target's direction and rotated their head early on in the sentence. Third, there was also some variability in target speaker onset time.

The ICASSP 2023 challenge built on CEC2. The significant innovation here was to include an evaluation set based on new measurements, to assess the generalisability of the machine learning algorithms beyond simulation.

CEC2 and ICASSP 2023 were evaluated using HASPI [6]. In addition, CEC2 used a panel of listeners with hearing loss.

As each listener responded individually, there is also considerable interest in designing intelligibility predictors for individual listeners. This was the purpose of the first prediction challenge, CPC1, which used the data from CEC1 (the corresponding 2nd prediction challenge using the data from CEC2 is currently ongoing). In the prediction challenges, we provide the audio produced by a variety of simulated hearing aids for speech-in-noise, the corresponding clean reference signals (the original speech), the listener characteristics such as their audiograms, and their measured speech intelligibility scores from CEC1 and CEC2. Systems were scored by their RMS prediction error against the original listener data.

3. THE CADENZA CHALLENGES

The Cadenza project generalizes the challenge methodology to the audio quality of music for someone with a hearing loss. There are two scenarios, "in the car" and "headphones".

The car scenario is a listener, wearing their hearing aids, sitting in a car driving at a (known) steady speed, while listening to recorded music played over the car stereo. Entrants are challenged to process the music played from the stereo to improve the audio quality allowing for the presence of car noise. This is an enhancement problem, parallel to CEC1. Importantly, this is not an active noise control task, as entrants don't know the exact noise waveforms of the car.

The headphone scenario imagines someone with a hearing loss who is listening to prerecorded music via headphones, and not their hearing aids. The intention is for entrants to demix stereo tracks into a vocal, drums, bass & other ("VDBO") representation then individually remix for a given listener to give improved audio quality. We will use HAAQI [7] to score the demixing objectively. The listening panel will rate the remixes for a set of quality metrics that are being developed by a focus group.

4. COMPUTATIONAL COST & CAUSALITY

Importantly, the scenarios allow different constraints on computational cost and causality across the two challenges.

In the Cadenza scenarios there are no computational constraints, as all the signals are prerecorded. In the real world, we imagine that whatever processing is done to the signals can be done in advance, or by the car's stereo.

In Clarity, however, the situation is different. A critical limitation of any real hearing aid is that it must be a real-time device that can deliver its signals within just a few milliseconds. But all the systems entered in the Clarity challenges are computational simulations of hearing aids. The first potential constraint, real-time, is not applied on the assumption that algorithms can be speeded up in the future if they prove useful. Therefore there is no constraint on computational cost.

But the second constraint is still needed. All the models must still be causal, in the sense that the output from the hearing aid at any time cannot use any information from more than 5 ms into the future. This is a requirement because a greater delay causes problems with lip reading and also causes unacceptable perceptual effects for hearing aids with open fittings

5. ACKNOWLEDGMENTS

Clarity is funded by UKRI (EP/S031448/1, EP/S031308/1, EP/S031324/1 and EP/S030298/1). Cadenza is funded by UKRI (EP/W019434/1). We thank Amazon, BBC R&D, the Hearing Industry Research Consortium, Google Inc., Logitech UK Ltd, Sonova AG, Carl von Ossietzky University Oldenburg and the Royal National Institute for the Deaf (RNID) for their support. For further details on both projects see https://claritychallenge.org and https://cadenzachallenge.org/

6. REFERENCES

[1] J. Barker, M.A. Akeroyd, T.J. Cox, J.F. Culling, J. Firth, S. Graetzer, H. Griffiths, L. Harris, G. Naylor, Z. Podwinska and E. Porter. "The 1st clarity prediction challenge: A machine learning challenge for hearing aid intelligibility prediction," in Proc. 23nd INTERSPEECH 2022-10821, 2022.

[2] T.J. Cox, J. Barker, W. Bailey, S. Graetzer, M.A. Akeroyd, J.F. Culling, G. Naylor. "Overview of the 2023 ICASSP SP Clarity Challenge: Speech Enhancement," in Proc. ICASSP 2023. 6621, 2023.

[3] T.J. Cox, M.A. Akeroyd, J. Barker, J.F. Culling, J. Firth, S. Graetzer, H. Griffiths, L. Harris, R. Viveros Munoz, G. Naylor and Z. Podwinska. "Predicting speech intelligibility for people with a hearing loss: The clarity challenges," in Proc. INTER-NOISE 265. 4599-4606, 2023.

[4] S. Graetzer, M.A. Akeroyd, J. Barker, T.J. Cox, J.F. Culling, G. Naylor, E. Porter, R. Viveros-Muñoz. "Dataset of British English speech recordings for psychoacoustics and speech processing research: The clarity speech corpus," Data Brief, 41, pp. 107951, 2022.

[5] J.M. Kates and K.H. Arehart. "The hearing-aid speech perception index (HASPI) version 2," Speech Communication, 131, pp. 35-46, 2021.

[6] J.M. Kates and K.H. Arehart. "The hearing-aid audio quality index (HAAQI)," IEEE/ACM transactions on audio, speech, and language processing, 24, pp. 354-365, 2015.