

# The First Cadenza Signal Processing Challenge: Improving Music for Those With a Hearing Loss

Gerardo Roa Dabike<sup>1</sup>, Scott Bannister<sup>2</sup>, Jennifer Firth<sup>3</sup>, Simone Graetzer<sup>1</sup>,  
Rebecca Vos<sup>1</sup>, Michael A. Akeroyd<sup>3</sup>, Jon Barker<sup>4</sup>, Trevor J. Cox<sup>1</sup>, Bruno Fazenda<sup>1</sup>,  
Alinka Greasley<sup>2</sup> and William Whitmer<sup>3</sup>

<sup>1</sup>Acoustics Research Centre, University of Salford, UK

<sup>2</sup>School of Music, University of Leeds, UK

<sup>3</sup>Hearing Sciences, School of Medicine, University of Nottingham, UK

<sup>4</sup>Department of Computer Science, University of Sheffield, UK

## Abstract

The Cadenza project aims to improve the audio quality of music for those who have a hearing loss. This is being done through a series of signal processing challenges, to foster better and more inclusive technologies. In the first round, two common listening scenarios are considered: listening to music over headphones, and with a hearing aid in a car. The first scenario is cast as a demixing-remixing problem, where the music is decomposed into vocals, bass, drums and other components. These can then be intelligently remixed in a personalized way, to increase the audio quality for a person who has a hearing loss. In the second scenario, music is coming from car loudspeakers, and the music has to be enhanced to overcome the masking effect of the car noise. This is done by taking into account the music, the hearing ability of the listener, the hearing aid and the speed of the car. The audio quality of the submissions will be evaluated using the Hearing Aid Audio Quality Index (HAAQI) for objective assessment and by a panel of people with hearing loss for subjective evaluation.

## Keywords

hearing loss, hearing aids, inclusive music, music quality, machine learning, signal processing, challenge

## 1. Introduction

There are many causes of hearing loss, including congenital hearing loss, chronic middle ear infections, noise exposure and age-related hearing loss [1]. The World Health Organization estimates that over 1.5 billion people worldwide have hearing loss. This is projected to rise to 2.5 billion by 2050. In the UK, nearly 12 million people – 1 in 5 – have hearing loss, with more

---


*HCMIR23: 2nd Workshop on Human-Centric Music Information Research, November 10th, 2023, Milan, Italy*

✉ g.roadabike@salford.ac.uk (G. Roa Dabike); S.C.Bannister@leeds.ac.uk (S. Bannister);  
Jennifer.firth@nottingham.ac.uk (J. Firth); S.N.Graetzer@salford.ac.uk (S. Graetzer); r.r.vos1@salford.ac.uk (R. Vos);  
Michael.Akeroyd@nottingham.ac.uk (M. A. Akeroyd); j.p.barker@sheffield.ac.uk (J. Barker); t.j.cox@salford.ac.uk  
(T.J. Cox); B.M.Fazenda@salford.ac.uk (B. Fazenda); a.e.greasley@leeds.ac.uk (A. Greasley);  
bill.whitmer@nottingham.ac.uk (W. Whitmer)

ORCID 0000-0001-7839-8061 (G. Roa Dabike); 0000-0003-4905-0511 (S. Bannister); 0000-0002-7825-0945 (J. Firth);  
0000-0003-1446-5637 (S. Graetzer); 0000-0002-2629-6271 (R. Vos); 0000-0002-7182-9209 (M. A. Akeroyd);  
0000-0002-1684-5660 (J. Barker); 0000-0002-4075-7564 (T.J. Cox); 0000-0002-3912-0582 (B. Fazenda);  
0000-0001-6262-2655 (A. Greasley); 0000-0001-8618-6851 (W. Whitmer)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

than 40% of cases affecting people over 50 years old, and this figure rises to more than 70% for people over 70 years old [2].

Hearing loss can have a major impact on a person's quality of life, making it difficult to communicate, participate in social activities, and enjoy music. Despite this, only 40% of people who could benefit from hearing aids actually have them and use them often enough. This is partly because people perceive hearing aids as performing poorly [3, 4, 5, 6] or find little benefit in using them [7, 8]. Historically, hearing aids have focused on speech communication. However, music listening is also important as it benefits health and well-being. Music is a universal human phenomenon that exists in many contexts and has a powerful impact on our emotions [9]. Hearing loss can make it difficult to appreciate music. To take two examples, it can affect the ability of listeners to pick out the lyrics and melody lines, as well as to hear the high frequencies that give the music its richness and detail. As a result, music can sound dull, which can lead to disengagement from music.

There are several spectro-temporal differences between speech and music that makes hearing aids optimised for speech perform poorly for music [10]. Although manufacturers have been developing special programs for music listening, the effectiveness has been mixed, with 68% of users reporting difficulty listening to music through their hearing aids [11]. There is a pressing need for better and more inclusive technology to enhance music accessibility to people with hearing loss.

The Cadenza project aims to improve the audio quality of music for people with hearing loss who wear hearing aids, using signal processing and machine learning challenges. These challenges are designed to bring together various research communities to make music more accessible to everyone, taking into consideration the diversity of listeners and making it more inclusive. In the first round (CAD1), we focused on two common scenarios for listening to music: over headphones and in a car in the presence of noise. Firstly, we introduce the general structure and design of CAD1 challenge. Sections 3 and 4 describe the specifics of Task 1 and Task 2. We conclude in Section 6. More details can be found on the challenge website<sup>1</sup>.

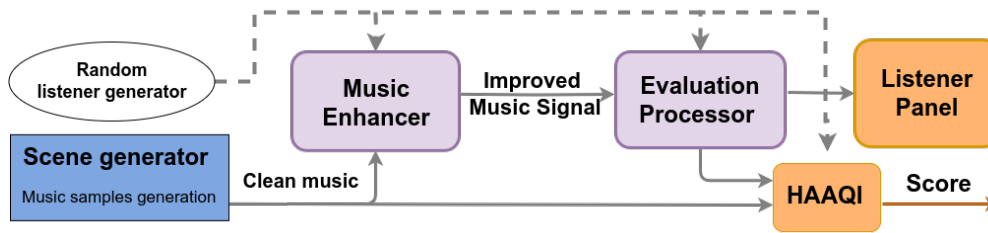
## 2. Overview of the Challenge Tasks

In the first scenario (Task 1), a person with hearing loss listens to music through headphones without using their hearing aids. In the second scenario (Task 2), the listener is inside a moving car, listening to the music that is coming from the car stereo in the presence of noise, while wearing their hearing aids. Entrants to the challenges are tasked with personalizing the music signals to improve the audio quality.

Figure 1 shows a diagram with the general structure of the challenges. Entrants must develop a *Music Enhancer* that takes in *clean music* and listener characteristics as input. The *Evaluation Processor* then takes the improved music signals, applying any acoustic conditions that are relevant to the task. Finally, the signals are evaluated using the Hearing Aid Audio Quality Index (HAAQI) [12] and a listener panel.

---

<sup>1</sup><https://www.cadenzchallenge.org>



**Figure 1:** Diagram of the structural design of the first challenge

## 2.1. Listener characterisation databases

Listeners are characterised by bilateral pure-tone audiograms. This gives the audible thresholds at standardised frequencies ([250, 500, 1000, 2000, 3000, 4000, 6000, 8000] Hz) as measured by an audiometer [13]. While a wider frequency range might have been useful for music, we were restricted to this range because these are the standard frequencies that have been tested in the available databases.

For the training (*train*) set, we used the 83 audiograms employed by the 2nd Clarity Enhancement Challenge [14] from the Clarity Project<sup>2</sup>. These correspond to anonymised examples of real audiograms drawn from the Scottish Section of Hearing Sciences at the University of Nottingham dataset.

For the development (*dev*) set, we selected 50 audiograms from [15]. We first filtered the audiograms to better-ear 4-frequency hearing loss between 20 and 75 dB. We then randomly chose the necessary number of audiograms to maintain the same distribution per band as in the original Clarity dataset. This *dev* set has an equal male-female distribution.

For the evaluation (*eval*) set, we recruited 52 bilateral hearing aid users, with symmetric or asymmetric hearing loss. The listeners were recruited via the University of Leeds. Hearing loss severity was mild for 15 listeners, moderate for 17 listeners, moderately severe for 18 listeners and severe for 2 listeners. In the *train*, *dev* and *eval* sets, hearing loss levels, at each frequency, were limited to 80 dB Hearing Level (HL).

## 2.2. Challenge Evaluations

Both scenarios will be subjected to two evaluation processes. The first is an objective evaluation using HAAQI. This is an intrusive metric in which the processed and reference signals are compared. In the evaluation, the HAAQI function is configured so that the reference signal has an amplification applied to it, so that all frequency bands contribute equally to its loudness. This amplification is the NAL-R hearing aid prescription [16]. This prescribes the gain to apply based on the individual's audiogram thresholds (in dB HL). This linear amplification improves audibility; there is no dynamic range compression.

The second evaluation consists of a listener panel of 52 listeners (the *eval* listeners) who will rate the audio quality of the music samples. The panel will use a number of scales: *clarity*, *harshness*, *distortion*, *frequency balance*, *overall audio quality*, and *liking*. Overall audio quality

<sup>2</sup><https://www.claritychallenge.org>

captures whether the audio quality is poor or good, and liking is how much the listener liked the specific piece they just listened to. These dimensions have been developed for this purpose, through a sensory evaluation study [17].

Overall audio quality captures whether the audio quality is poor / good, and liking is how much the participant liked the specific piece they just listened to

### 3. Design of Task 1

This is presented as a demixing-remixing problem. The demixing stage follows the same design as previous music separation challenges [18, 19]. The goal is to decompose stereo music into vocal, drums, bass, and other (VDBO). However, unlike past demixing challenges, we use HAAQI for the evaluation instead of the signal-to-distortion ratio. In the remixing stage, the separated VDBO components allow for personalised remixing for each listener. For example, for some music, the vocals could be amplified to improve the audibility of the lyrics. In Task 1, for Figure 1, the Evaluation Processor only focuses on computing HAAQI.

We use the MUSDB18-HQ music dataset [20]. This contains 150 tracks where 86 are for *train*, 14 for *dev* and 50 for *eval*. For training data augmentation, entrants are allowed to use the BACH10 [21], FMA-Small [22] and MedleyDB versions 1 [23] and 2 [24].

The *eval* set includes 49 tracks, after removing one track where the lyrics might cause offence. For the objective evaluation, 30-second segments of music were selected at random. For the listening panel, 15-second segments were selected at random, ensuring that no explicit language was present.

### 4. Design of Task 2

The goal is to process music emitted by a car stereo, while accounting for the presence of simulated car noise. However, this is not a denoising problem, as participants do not have access to the exact noise signal. The evaluator processor, for Figure 1, adds the car acoustic conditions (head-related impulse responses (HRIRs) and simulated car noise) and applies the fixed hearing aid processing algorithms to the enhanced signals before computing the HAAQI score.

The music datasets for task 2 are based on the FMA-Small [22] and the MTG Jamendo datasets [25]. FMA-Small contains 30-second music segments from eight genres; however, we only included the genres ‘Hip-Hop’, ‘Instrumental’, ‘International’, ‘Pop’ and ‘Rock’ as they are the most likely to be found in a car listening environment for our target listeners. We also included samples from the ‘classical’ and ‘orchestral’ genres from the MTG-Jamendo dataset, as people with hearing loss are more likely to be older and listeners to classical music [26].

This process resulted in 7000 30-second tracks distributed as follows: 5600 *train*, 700 *dev*, and 700 *eval*. However, as the *eval* samples will be listened to by the listener panel, the *eval* set was reduced to 70 samples by randomly selecting 10 samples per genre. HAAQI will be computed over the whole 30-second segments but, only 15 seconds will be scored by the panel.

We simulate the different components of the car noise based on a car speed and gear as follows: (i) a mono complex tone corresponding to the engine noise, (ii) two mono signals

**Table 1**

Baseline HAAQI scores for Task 1 and Task 2 of the First Cadenza Challenge.

Task	System	mean HAAQI	std HAAQI
Task 1	Baseline 1 - Demucs	0.2548	0.0403
	Baseline 2 - OpenUnmix	0.2253	0.0290
Task 2	Baseline	0.1256	0.0321

generated by filtering a white noise by a lowpass filter with a 6-dB/octave slope corresponding to aerodynamic and rolling noise from the left and right side of the car.

HRIRs were drawn from the eBrIRD - ELOSPHERES database [27]. Each scene contains 74 measurements from  $-90^\circ$  to  $90^\circ$  azimuth in  $2.5^\circ$  steps. Anechoic HRIRs are used for the car noise, and in-car HRIRs are used for the enhanced signals.

## 5. Baseline Details and Results

Two baseline systems were proposed for Task 1. The first baseline, ‘Baseline 1 - Demucs’, utilizes the out-of-the-box hybrid-demucs source separation model [28] to estimate the VDBO components of the music. This hybrid architecture leverages both time-domain and spectrogram-based approaches. The second baseline, ‘Baseline 2 - OpenUnmix’, employs the OpenUnmix source separation model [29]. OpenUnmix is a purely spectrogram-based approach and served as the baseline for the SiSEC 2018 challenge [18].

For Task 2, a single baseline was proposed, which applies a level constraint to the mixture at the hearing aid microphones to prevent clipping caused by the NAL-R hearing aid amplification.

Table 1 shows the baseline HAAQI scores for Task 1 and Task 2. For Task 1, the results correspond to the mean and standard deviation (std) of all eight left and right VDBO signals. In Task 2, the scores represent the mean and std of the scores per genre.

## 6. Conclusion and Future Work

The Cadenza project aims to improve the audio quality of music for those who have a hearing loss. It consists of two common music listening scenarios; Task 1, listening to music over headphones and, in Task 2, listening to music in a car. While Task 1 presents a demixing-remixing problem, Task 2 presents a near-end music enhancement problem. Both scenarios will be evaluated using HAAQI for objective evaluation and a listener panel for subjective evaluation in autumn/winter 2023. The project team will run an online workshop in December 2023 where entrants will outline their approaches, and the results of the evaluations will be presented. Currently, we are running an ICASSP 2024 Grand Challenge and the CAD2 challenge will launch in 2024.

## Acknowledgments

The Cadenza project is supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number: EP/W019434/1].

We thank our partners: BBC, Google, Logitech, RNID, Sonova, Universität Oldenburg.

## References

- [1] World Health Organization (WHO), Deafness and hearing loss, 2023. URL: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>, accessed: 2023-07-05.
- [2] Royal National Institute for the Deaf (RNID), Prevalence of deafness and hearing loss, 2020. URL: <https://rnid.org.uk/about-us/research-and-policy/facts-and-figures/prevalence-of-deafness-and-hearing-loss/>, accessed: 2023-07-05.
- [3] S. Kochkin, Consumers Rate Improvements Sought in Hearing Instruments, *Hearing Review* 9 (2002) 18–22.
- [4] S. Kochkin, MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing, *The Hearing Journal* 63 (2010) 19–20.
- [5] L. V. Knudsen, M. Öberg, C. Nielsen, G. Naylor, S. E. Kramer, Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A review of the literature, *Trends in Amplification* 14 (2010) 127–154. doi:10.1177/1084713810385712.
- [6] C. Meyer, L. Hickson, What factors influence help-seeking for hearing impairment and hearing aid adoption in older adults?, *International journal of audiology* 51 (2012) 66–74. doi:10.3109/14992027.2011.611178.
- [7] Z. Polat, T. Eyupler, The hearing aid use and its efficiency in geriatric population, *Mega J Case Rep* 1 (2022).
- [8] D. Tsimpida, E. Kontopantelis, D. Ashcroft, M. Panagioti, Comparison of self-reported measures of hearing with an objective audiometric measure in adults in the english longitudinal study of ageing, *JAMA network open* 3 (2020) e2015009–e2015009. doi:10.1001/jamanetworkopen.2020.15009.
- [9] S. A. Mehr, M. Singh, D. Knox, D. M. Ketter, D. Pickens-Jones, S. Atwood, C. Lucas, N. Jacoby, A. A. Egner, E. J. Hopkins, R. M. Howard, J. K. Hartshorne, M. V. Jennings, J. Simson, C. M. Bainbridge, S. Pinker, T. J. O’Donnell, M. M. Krasnow, L. Glowacki, Universality and diversity in human song, *Science* 366 (2019) eaax0868. doi:10.1126/science.aax0868.
- [10] M. Chasin, F. A. Russo, Hearing aids and music, *Trends in Amplification* 8 (2004) 35–47. doi:10.1177/108471380400800202.
- [11] A. Greasley, H. Crook, R. Fulford, Music listening and hearing aids: perspectives from audiologists and their patients, *International Journal of Audiology* 59 (2020) 694–706. doi:10.1080/14992027.2020.1762126.
- [12] J. M. Kates, K. H. Arehart, The Hearing-Aid Audio Quality Index (HAAQI), *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24 (2016) 354–365. doi:10.1109/TASLP.2015.2507858.
- [13] British Society of Audiology, Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncom-



fortable loudness levels, 2018. URL: <https://www.thebsa.org.uk/resources/pure-tone-air-bone-conduction-threshold-audiometry-without-masking/>, accessed: 2023-09-20.

- [14] M. A. Akeroyd, W. Bailey, J. Barker, T. J. Cox, J. F. Culling, S. Graetzer, G. Naylor, Z. Podwińska, Z. Tu, The 2nd Clarity Enhancement Challenge for Hearing Aid Speech Intelligibility Enhancement: Overview and Outcomes, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, pp. 1–5. doi:10.1109/ICASSP49357.2023.10094918.
- [15] P. von Gablenz, E. Hoffmann, I. Holube, Prevalence of hearing loss in Northern and Southern Germany, HNO 65 (2017) 130–135. doi:10.1007/s00106-016-0318-4.
- [16] R. Peters, B. Moore, B. Glasberg, M. Stone, Comparison of the NAL(R) and Cambridge formulae for the fitting of linear hearing aids, British Journal of Audiology 34 (2000) 21–36. doi:10.3109/03005364000000115.
- [17] S. Bannister, A. Greasley, T. Cox, M. Akeroyd, J. Barker, B. Fazenda, J. Firth, S. Graetzer, G. Roa Dabike, R. Vos, W. Whitmer, Muddy, Muddled, or Muffled? Understanding the Perception of Audio Quality in Music by Hearing Aid Users, in: 17th International Conference on Music Perception and Cognition (ICMPC 2023), August 2023.
- [18] F.-R. Stöter, A. Liutkus, N. Ito, The 2018 signal separation evaluation campaign, in: Latent Variable Analysis and Signal Separation, 2018, pp. 293–305. doi:10.1007/978-3-319-93764-9\_28.
- [19] Y. Mitsufuji, G. Fabbro, S. Uhlich, F.-R. Stöter, A. Défossez, M. Kim, W. Choi, C.-Y. Yu, K.-W. Cheuk, Music Demixing Challenge 2021, Frontiers in Signal Processing 1 (2022) 808395. doi:10.3389/frsip.2021.808395.
- [20] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, R. Bittner, MUSDB18-HQ - an uncompressed version of musdb18, 2019. doi:10.5281/zenodo.3338373.
- [21] Z. Duan, B. Pardo, C. Zhang, Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions, IEEE Transactions on Audio, Speech, and Language Processing 18 (2010) 2121–2133. doi:10.1109/TASL.2010.2042119.
- [22] M. Defferrard, K. Benzi, P. Vandergheynst, X. Bresson, FMA: A dataset for music analysis, in: 18th International Society for Music Information Retrieval Conference (ISMIR), 2017. URL: <https://arxiv.org/abs/1612.01840>.
- [23] R. M. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, J. P. Bello, MedleyDB: A multitrack dataset for annotation-intensive MIR research., in: ISMIR, 2014. URL: [https://rachelbittner.weebly.com/uploads/3/2/1/8/32182799/bittner\\_ismir\\_2014.pdf](https://rachelbittner.weebly.com/uploads/3/2/1/8/32182799/bittner_ismir_2014.pdf).
- [24] R. M. Bittner, J. Wilkins, H. Yip, J. P. Bello, MedleyDB 2.0: New data and a system for sustainable data collection, ISMIR Late Breaking and Demo Papers (2016) 36. URL: <https://wp.nyu.edu/ismir2016/wp-content/uploads/sites/2294/2016/08/bittner-medleydb.pdf>.
- [25] D. Bogdanov, M. Won, P. Tovstogan, A. Porter, X. Serra, The MTG-Jamendo dataset for automatic music tagging, in: Machine Learning for Music Discovery Workshop, International Conference on Machine Learning (ICML 2019), Long Beach, CA, United States, 2019. URL: <http://hdl.handle.net/10230/42015>.
- [26] A. Bonneville-Roussy, P. J. Rentfrow, M. K. Xu, J. Potter, Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood, Journal of Personality and Social Psychology 105 (2013) 703–717. doi:10.1037/a0033770.

- [27] G. Hilkhuisen, eBrIRD - ELOSPHERES binaural room impulse response database, 2021. URL: <https://www.phon.ucl.ac.uk/resource/ebrird/>, accessed: 2023-03-14.
- [28] A. Défossez, Hybrid spectrogram and waveform source separation, arXiv preprint arXiv:2111.03600 (2021). URL: <https://arxiv.org/abs/2111.03600v3>.
- [29] F.-R. Stöter, S. Uhlich, A. Liutkus, Y. Mitsufuji, Open-unmix - a reference implementation for music source separation, Journal of Open Source Software (2019). doi:10.21105/joss.01667.