CEC3 plans and discussion

Trevor J. Cox, Jon Barker, Will Bailey, Simone Graetzer, Michael A. Akeroyd, John F. Culling, Graham Naylor

Universities of Salford, Sheffield, Nottingham, Cardiff, UK

https://claritychallenge.org/

https://claritychallenge.org/

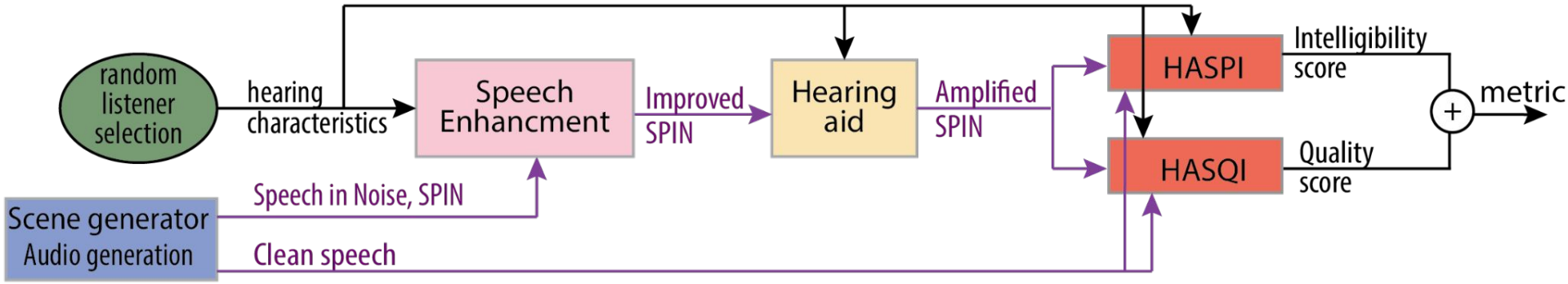
Running a series of signal processing challenges:

- *Clarity Prediction challenges* (CPC) to improve prediction of speech intelligibility
- *Clarity Enhancement challenges* (CEC) to improve processing by hearing aids
  - CEC3 to early 2024

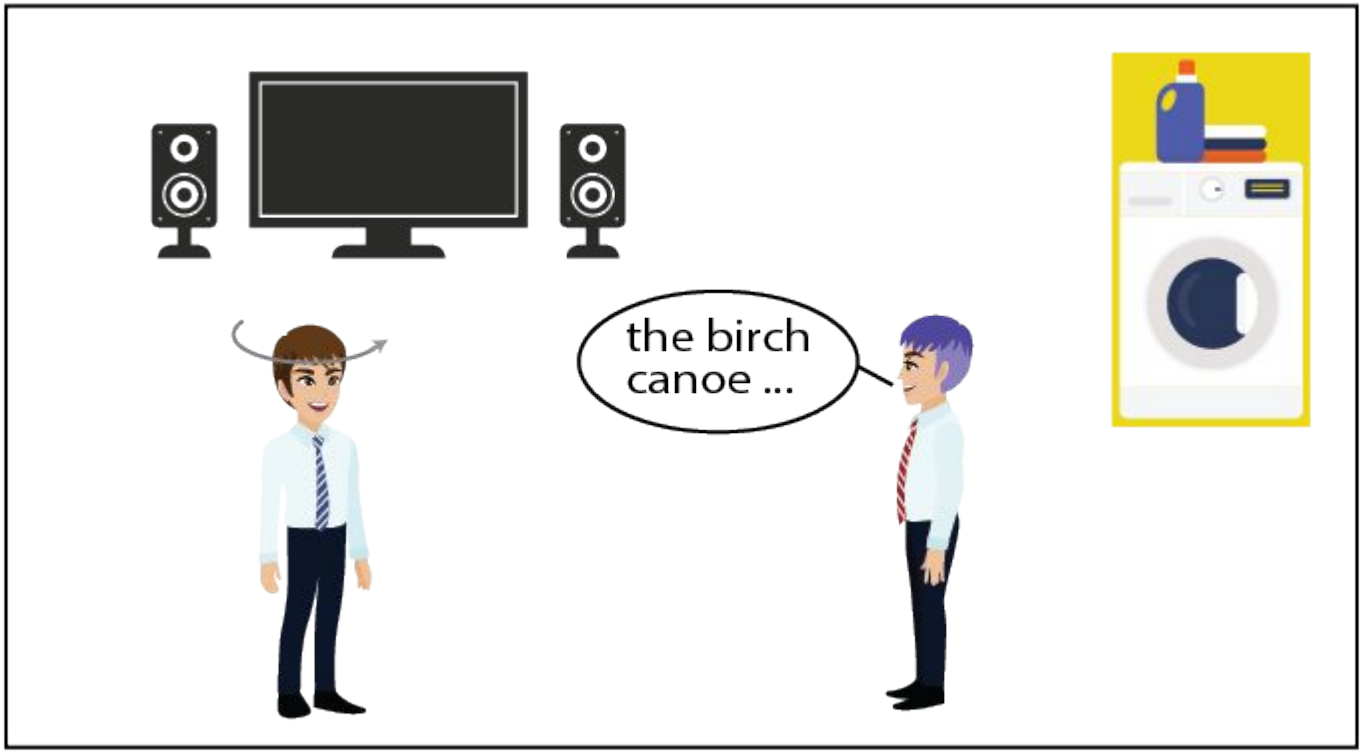
Companion project doing challenge series for music:

- https://cadenzachallenge.org/

Baseline (ICASSP 2023 challenge)



Scenario, CEC1, CEC2, ICASSP 2023



Review of previous challenges

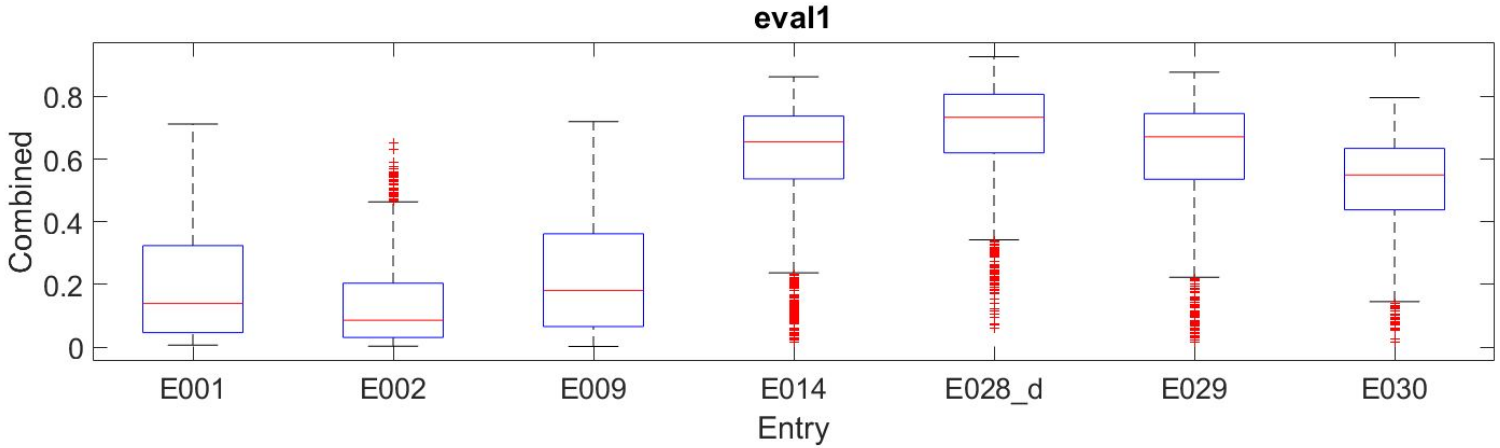
- CEC1
  - Static scenes, beam forming very effective
- CEC2
  - Much more complex scenes e.g. head rotation
  - Best system
    - Was complex: iterative neurally-informed beamforming informed by speaker-embeddings and carefully optimised amplification stage
    - Created highly intelligible utterances from scenes at SNRs as low as -12 dB
  - Datasets all simulated
- ICASSP 2023
  - Similar to CEC2 but with an additional measured data set for evaluation

Measured Eval2 dataset

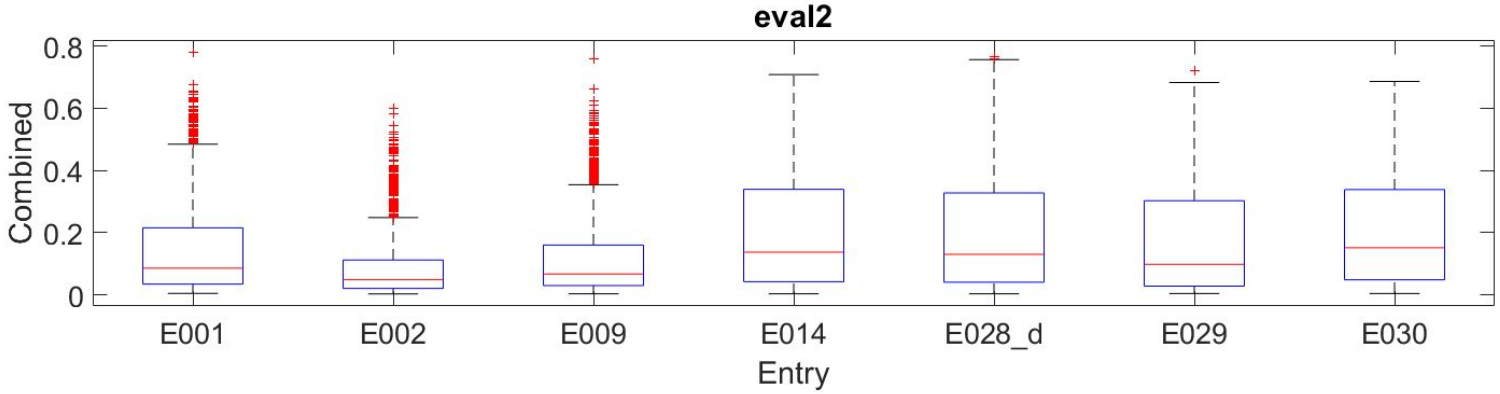
- A Listening room with many sources and listener positions.
- New target sentences and 10 live actors.
- Close cardioid mic & 1st-order Ambisonic mic at listener position.
- Speech recorded in noise-free conditions.
- Noise, music and speech interferers played from a loudspeaker.
- Post-processing:
  - Head rotations.
  - HRTFs → hearing-aid mic signals
  - Speech and noise mixed.

Results: ICASSP 2023

Simulated



Measured + processed



Results continued

- Simulated Eval 1
  - 5 teams had entries that improved on the baseline, with 2 producing worse scores.
  - HASPI and HASQI values were highly correlated.
  - Across the successful systems, the improvement in the HASQI quality scores were about
  - half the improvement in HASPI intelligibility scores.
- More ecologically valid Eval 2
  - 4 teams beat the baseline, the improvement was much smaller than simulated Eval 1.

Eval 1 vs Eval 2

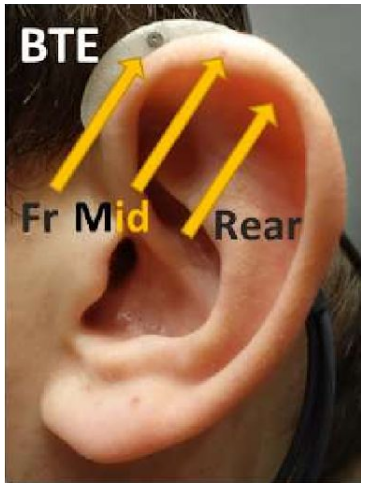
	Eval 1: simulated	**Eval 2: Measured + processed**
Actor behaviour	Close mic	Talking to distant mic
Directivity of interferers	Omnidirectional	Loudspeaker
Interferers	Could be more than one of each type (music, speech, noise)	At most, one of each type (music, speech, noise)
Listener-target distance	Exponential probability	Rectangular probability
Room impulse response	Geometric model	Real
Ambisonic order	6th	1st
Mic noise	Negligible	On Ambisonic mic
Known unknown	DNN using data in unexpected way	

CEC3 track 1

Research question: How to create additional data that allows machine learning algorithms to be more robust to more ecologically-valid situations?

Recording done on:

- 6th order ambisonic microphone
- Hearing aid formers



CEC3 track 2

Research question: How well can machine learning enhancement algorithms work with more dynamic scenes e.g. moving, intermittent interferers?

Scenario: listening to a talker while out and about in real dynamic environments, e.g. on a busy street.

- Scenes constructed from Eigenmike recordings of real backgrounds:
  - Railway, roadside, cafe, atria, stadia, construction noise
  - Places with high levels and intermittent/moving noise sources
- Close target talker mixed into scene based on studio recordings and Eigenmike-recorded impulse responses.

Discussion

- What do you think about the proposed tasks?
- How could they be improved?
- How do we increase engagement with Clarity challenges?