From spoken speech to sung speech - Cadenza Lyric Challenge

Cadenza Project Team
Universities of Salford, Nottingham,
Leeds and Sheffield

Clarity-2025, 22nd August 2025, Delft, The Netherlands

The Cadenza project is

- defining what music personalised for an individual with Hearing Loss (HL) should sound like.
- exploiting the latest advances in machine learning to create improved listening experiences.

The next challenge is on lyric intelligibility, which differs from spoken speech. Factors that affect lyric intelligibility include

- Vocal style and articulation.
- Song genre.
- Mixing and production techniques.
- Listening hearing acuity.

ICASSP SPGC Cadenza Challenge 2025: Predicting Lyric Intelligibility

Motivation

- Understanding the lyrics in music is key for music enjoyment [1].
- People with HL can have difficulties in hearing lyrics clearly and effortlessly [2].
- In speech technology, metrics to evaluate intelligibility automatically have driven improvements in speech enhancement.

Source: Sabena Costa, Pixabay.

1. Fine, P. A. and Ginsborg, J., 2014. Making myself understood: perceived factors affecting the intelligibility of sung text. Frontiers in Psychology, 5, 809.
2. Greasley, A., Crook, H. and Fulford, R., 2020. Music listening and hearing aids: perspectives from audiologists and their patients. International Journal of Audiology, 59(9), pp.694-706.

- Dataset: a novel dataset of a total of 11,100 sung excerpts
  - Cadenza Lyrics Intelligibility Challenge (CLIP) dataset.
  - English sung excerpts extracted from the FMA (Free Music Archive [1]) dataset.
  - Samples presented *as is* and with *mild* and *moderate HL* using a HL simulator.
  - Intelligibility corresponds to the ratio of correctly transcribed words to the total number of words.

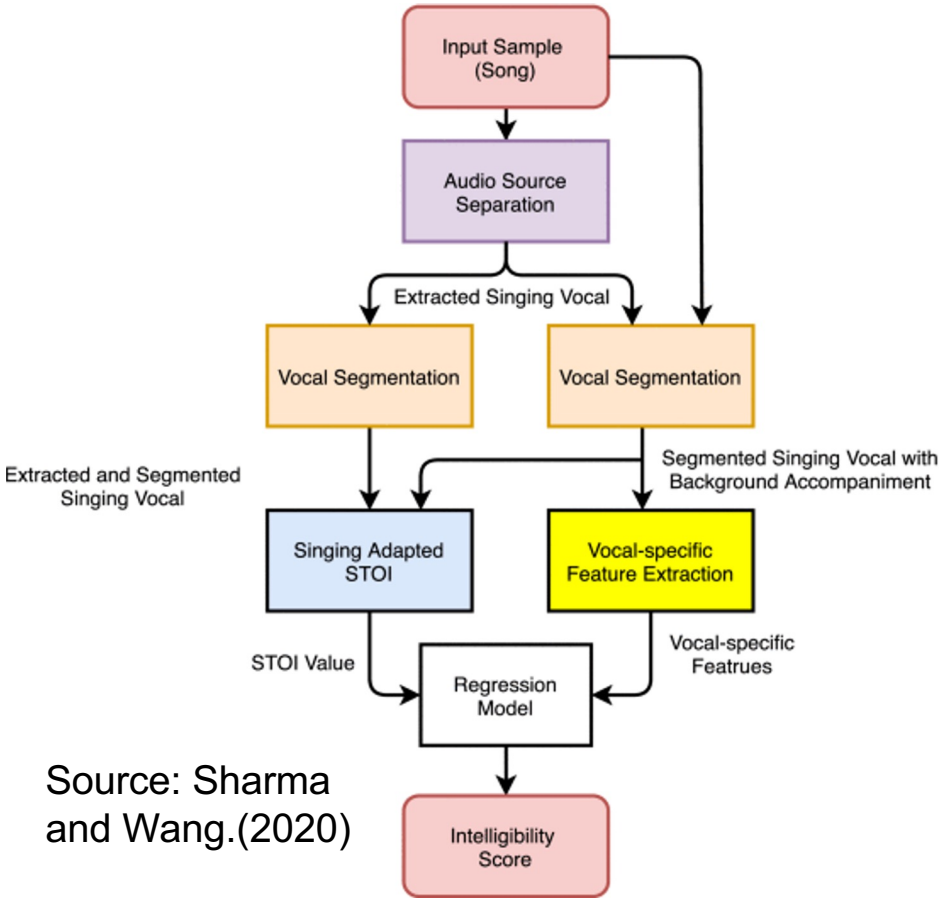
1. Defferrard, Michaël, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson., 2017, FMA: A DATASET FOR MUSIC ANALYSIS, International Society for Music Information Retrieval.

- Dataset: a novel dataset of a total of 11,100 sung excerpts
  - Ground truth generated using native British English speakers without limiting the number of times each sample is listened to.
  - Sets are disjoint by artist.

No Loss	Mild HL	Moderate HL	Transcript
			do you know i am truly alone
			banana run looks like i’m pulling a banana run

Software and Baseline

- Software provided as part of the PyClarity module
- Baseline based on Singing Adapted STOI (Sharma and Wang, 2020)
  - Implementation ported to Python
  - Retrained on CLIP dataset



Source: Sharma and Wang.(2020)

B. Sharma and Y. Wang, "Automatic Evaluation of Song Intelligibility Using Singing Adapted STOI and Vocal-Specific Features," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020.

Find out more before our 1st September 2025 challenge launch!

Website: https://cadenzachallenge.org/

Google Groups: https://groups.google.com/g/cadenza-challenge

Zenodo Community: https://zenodo.org/communities/cadenzachallenge/