Vocoders and Objective measures: How much to trust for designing new sound coding strategies in cochlear implants?

Behnam Molaee-Ardekani ; Yue Zhang ; Rafael Attili Chiea ; Manuel Segovia-Martinez

Oticon Medical (France)

Background

The perception in **cochlear implant (CI)** patients is significantly reduced in challenging environments like multi-talkers and music. New coding strategies may improve this limitation.

However, validation of the performance of a new strategy and the underlying hypothesis behind it is costly and time consuming through clinical validations.

Therefore, the use of **vocoders (VOCs)** and **mathematical objective measures (MOMs)**, such as eSTOI and NCM among others, have been considered more than before.

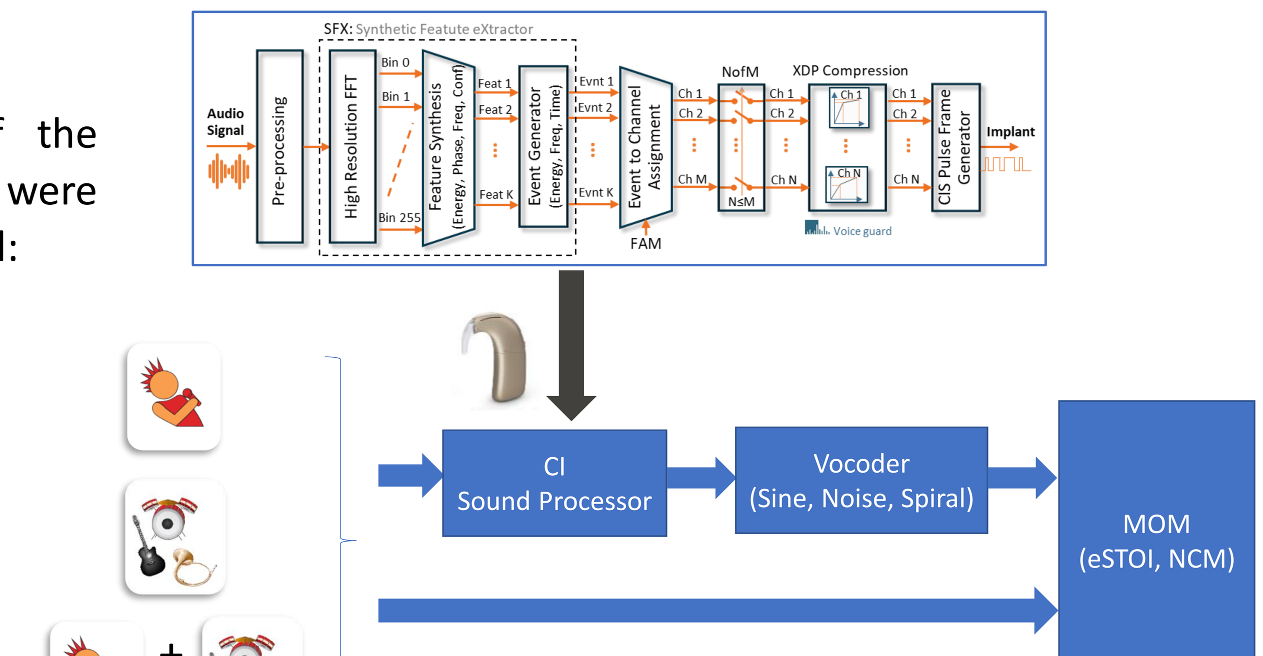
Abstract

We (Oticon medical) as a CI manufacturer have recently used vocoders and objective measures (such as eSTOI and NCM) to assess the performance of a newly designed coding strategy (see SFE strategy in posters **#1634** and **#1556**).

We think that there are some fundamental limitations in this approach that should be discussed with the CI field experts for hopefully finding better and more practical solutions.

Methodology

- We designed a new CIS-based coding strategy (posters **#1634** and **#1556**) and sent its output pulses to three different VOCs:
  - Sine (Dorman et al 1997, Luo et al 2007)
  - Noise (Shannon et al 1995, Ananthakrishnan et al 2017)
  - Spiral (Grange et al 2017)
- We estimated the speech intelligibility of the vocoded sound where the input sounds were extracted from different database and included:
  - Speech
  - Music
  - POP music including lyrics
- We used two widely accepted MOMs (Van Kuyk et al 2018) (see poster **#1636**) :
  - NCM (normalized covariance measure), (Koch et al 1992)
  - eSTOI (extended short-time objective intelligibility measure), (Jensen et al 2016)

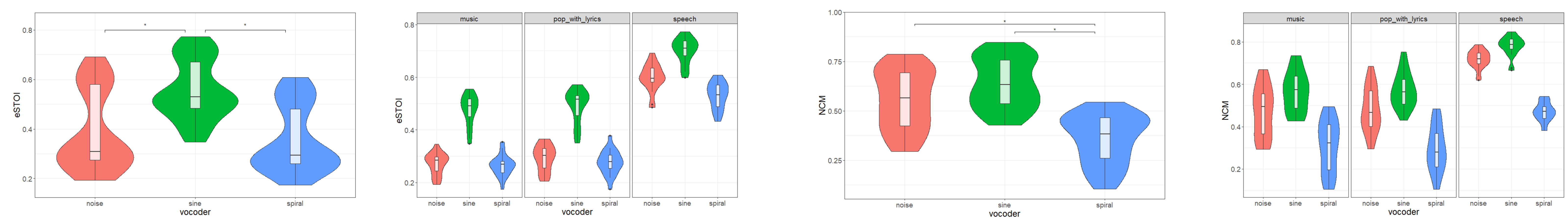


Vocoders info:

- Sine, Noise, Spiral: Fc at default Crystalis FAM
- Noise: Filter Order 3 equivalent to a current spread of 16 dB/Octave
- Spiral: Current Spread 16 dB/Octave, Carriers 80 [20,8000] Hz

Results

Results varied depending on the different VOCs, music stimuli and MOMs



- There was significantly higher STOI scores for Sine vocoded speech compared to using Noise (β=1.17, SE=0.11, p<0.05) and Spiral vocoders (β=1.01, SE=0.11, p<0.05) and only significantly lower NCM scores for Spiral vocoder compared to using Sine (β=0.64, SE=0.11, p<0.05) and Noise vocoders (β=0.53, SE=0.11, p<0.05).
- The difference also depended on the acoustic inputs: speech and music with lyrics showed similar order of difference but music inputs showed that sine vocoded speech having the highest STOI (β=0.55, SE=0.09, p<0.05; β=0.47, SE=0.03, p<0.05) and NCM scores (β=0.14, SE=0.03, p<0.05; β=0.39, SE=0.01, p<0.05).

Discussions

Different interpretations:

- Our results indicate that one may get **quite different interpretations** when comparing different CI strategies using this methodology, **depending on the chosen VOC or MOM** and their combination.
- We are wondering if using the VOC-MOM method might even **mislead the researchers** towards wrong design paths.

Two levels of uncertainty:

- The main issue is that the potential benefits of a CI strategy should be verified by a compound method that has **two levels of uncertainty (i.e., VOC and MOM)** that need to be validated individually.

Uncertainty at the VOC level:

VOC evaluations by CI patients with single-sided deafness have shown that the preference of VOC is very subjective. However, each VOC's functionality is based on various assumptions that differ from one another. As a result, VOCs cannot guarantee that they accurately present all aspects of the benefits and improvements of newly developed coding strategies in their vocoded sounds. Although designing a strategy-specific vocoder is beneficial, it must undergo validation first. Thus, different VOCs must undergo statistical examination (e.g., sensitivity measurement, test-retest reliability, dropout testing).

Uncertainty at the MOM level:

The second level of ambiguity is generated at the MOM level. Current MOMs assume similarity to unprocessed speech audio wave as a measure of the information transfer in the CI device. This is not so reliable to quantify the differences between coding strategies. Accentuating certain aspects of the original signals could be helpful to CI patient, and different physical features have different perceptual importance. Therefore, the CI community need to work on the definition of what is a better information transfer for CI.

Mismatching between acute and VOC-MOM:

- **Mismatched** outcomes between an **acute testing** clinical protocol and the **VOC-MOM method** is expected due to acclimatization of the strategy.

  MOMs assume an ideal outcome, which will not appear in acute clinical testing. A predicted acute testing outcome that includes margin of error in acclimatization and evolution in time would be useful for estimating the realistic strategy effectiveness.

References

Ananthakrishnan et al, Ear Hear. 2017; 38(5): e256–e267. doi:10.1097/AUD.0000000000000432
Dorman et al, *J Acoust Soc Am.* 1997;102:2403–2411.
Grange et al, J. Acoust. Soc. Amer., vol. 142, no. 5, pp. EL484–EL489, Nov. 2017.
Jensen and Taal, IEEE/ACM Trans. Audio, Speech, Language Process., vol. 24, no. 11, pp. 2009–2022, 2016.
Koch R, "Auditory sound analysis for the prediction and improvement of speech intelligibility," Ph.D. dissertation, University of Goettingen, Goettingen, 1992.
Shannon et al, Speech Recognition with Primarily Temporal Cues Sh0oo, vol. 270, October 1995.
Van Kuyk et al, IEEE/ACM Trans. Audio, Speech, Language Process., doi: 10.1109/TASLP.2018.2856374