# Auditory kernels informed by cochlear processing

Mihai Bratu
Supervisors: Jorge Martinez, Dimme de Groot

References:
[1] Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. Nature, 439 (7079),978–982.
[2] Barlow, H. B., et al. (1961). Possible principles underlying the transformation of sensory messages. Sensory communication, 1 (01), 217–233.
[3] Carney, L. H. and Yin, T. C. (1988). Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model. Journal of Neurophysiology, 60 (5), 1653–1677.
[4] Thoret, E., Ystad, S., & Kronland-Martinet, R. (2023). Hearing as adaptive cascaded envelope interpolation. Communications Biology, 6 (1), 671.

## 1. Introduction

- Auditory kernels [1] are one popular approach of sparsely representing a speech signal. They are adapted to speech and based on the Efficient Coding Hypothesis [2]. They were additionally motivated by correlation to cat revcor filters [3].
- A recent model of cochlear processing, CEI [4], breaks a signal into multiple modes. It successfully predicted multiple features observed in hearing.

## 2. Research Question

To what extent do kernels adapted to the modes preserve features of speech signals vs. kernels directly adapted to the speech signal?

## 3. Methods

- Kernels adapted to the TIMIT train set (English speakers), 20Hz-7kHz, using Matching Pursuit.
- Testing on the TIMIT test set.
- Data split into modes using CEI [3].
- Vowel activations on a multilingual dataset containing 95 languages.

## 4. Results

- Recombining the signal using activations of kernels on modes performs worse than activations on the signal (Fig. 1).
- Correlation with cat revcor filters [3] is not as prominent for kernels adapted to modes as it was shown in [1] for kernels adapted to the original signal (Fig. 2).
- Clustering performance of kernel activations of vowels slightly improves for the kernels adapted to modes, when compared to the kernels adapted to the original signal, though poor for both (Table 1).

| Dictionary | BCubed recall for 8 vowels |
|---|---|
| $D_{TIMIT}$ | 0.244 ± 0.005 |
| combined $D^{[i]}_{mode}$ | 0.262 ± 0.004 |

Table 1

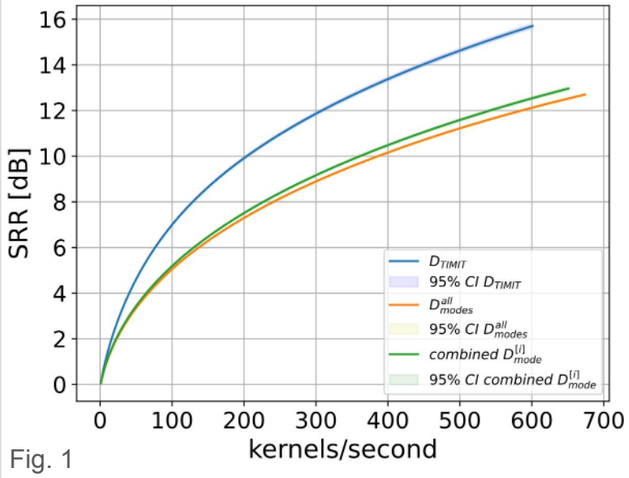


Fig. 1

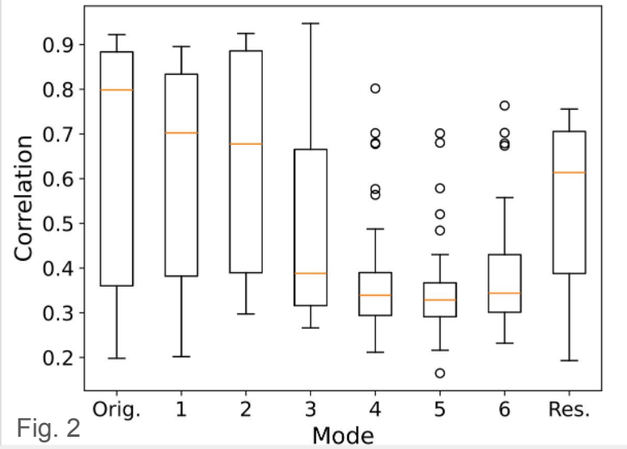


Fig. 2

## 5. Conclusion

- Recovering the original signal from the modes encoded using kernels performs worse and is less efficient than the encoding of the original signal.
- The kernels adapted to modes appear to preserve defining features of vowels, though both scored poorly.

## 6. Future work

- Frequency range can be modified to encompass higher frequencies or a narrower band.
- The encoding of the defining features of vowels in the modes of CEI should be investigated.
- Encoding of consonants should be also investigated for both kernels and CEI.