Using a theoretical model of auditory processing improves spectrally-degraded speech perception

Evan C. Smith and Lori L. Holt

Carnegie Mellon University

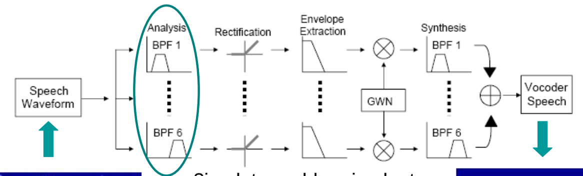
Department of Psychology & Center for the Neural Basis of Cognition

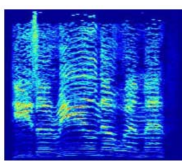
evan@cncb.cmu.edu & lholt@andrew.cmu.edu

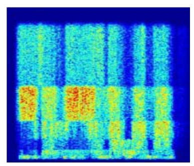
Abstract

Smith & Lewicki (2005, 2006) showed that auditory neurons efficiently code for natural sounds, maximizing information rate while minimizing coding cost, and also that speech acoustics are optimally adapted to the mammalian auditory code. The present work applies efficient coding theory to the problem of speech perception in individuals using cochlear implants (CI). We present a machine-learning method for CI filterbank design based on the efficient-coding hypothesis. Further, we describe a pair of experiments which evaluate this approach using noise-excited vocoder speech. Participants' recognition of continuous speech and isolated syllables is significantly more accurate for speech filtered through the theoretically-motivated, efficient-coding filterbank relative to the standard cochleotopic filterbank, particularly for speech transients. These findings offer insight in CI design and provide behavioral evidence for efficient coding in human perception.

Noise-excited Vocoder Speech







Simulate cochlear implant hearing by removing spectral resolution of speech

The **filterbank** determines the content of remaining channels

Can we learn an "optimal" filterbank?

Filterbanks based on standard processing principles:

1) Cochleotopic frequency tiling "pseudo-logarithmic" motivated by the cochlea

2) Linear frequency tiling as a baseline for comparison

VS.

Filterbank based on maximizing coding efficiency

Learn a set of kernel functions to optimize efficiency under the following model:

$$x(t) = \sum_{m=1}^{M} \sum_{i=1}^{n_m} s_i^m \phi_m (t - \tau_i^m) + \varepsilon(t)$$

Convert functions to filterbank

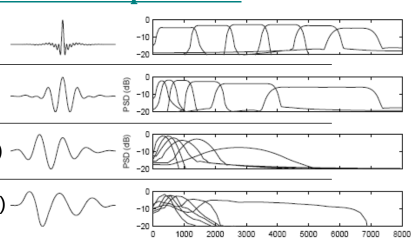
Filterbanks for Experiment

Linear frequency

Cochleotopic frequency

Efficient (gammatone fit)

Efficient (spline smoothed)



Experiments

Experiment 1

When participants hear spectrally-degraded spoken sentences, will they recognize more words when the channels are based on the efficient principle or when they are based on a cochleotopic frequency map?

Experiment 2

When participants hear spectrally-degraded non-lexical items (e.g., /aba/ or /ti/), will they still show improved performance using the "efficient" filterbank? And, fro where is the increased accuracy coming (how does efficiency improve performance)?

Methods

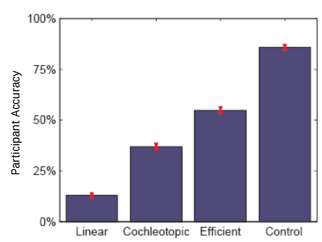
Experiment 1:

- 16 normal-hearing participants
- 168 spoken sentences
- Conditions:
  - Linear-f vocoder speech
  - Cochleotopic-f vocoder speech
  - Efficient-f vocoder speech
    - Gammatone fit
  - Control (normal) speech
- Dictation
- Word-by-word accuracy

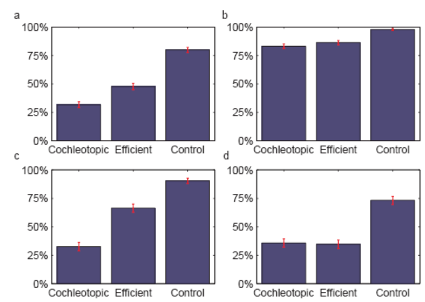
Experiment 2:

- 15 normal-hearing participants
- 186 non-lexical items
- Conditions:
  - Cochleotopic-f vocoder speech
  - Efficient-f vocoder speech
    - Spline-smoothed
  - Control (normal) speech
- Repetition + Dictation
- Phone-by-phone accuracy

Results

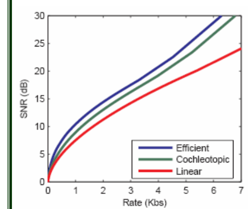


Experiment 1: The mean percentage of words correctly identified is plotted for each filtering condition. Recognition of words in sentences was, on average, 50% more accurate for vocoder speech using an "efficient" filterbank versus a cochleotopic filterbank. All error bars are 95% confidence interval of the mean.



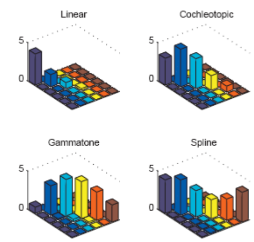
Experiment 2: Accuracy for non-lexical items: (**a**) whole items, (**b**) vowels, (**c**) non-fricative consonants (e.g., /b/ /d/ /n/ /r/), and (**d**) fricatives (e.g., /sh/). Found that the overall gain from better coding of non-fricative consonants. This matches previous findings that most linguistic information comes from timing.

Confirming Coding Efficiency



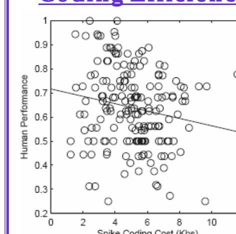
Rate-distortion curves computed using each set of filters confirms that the learned filters are more efficient than the cochleotopic set.

Maximizing Channel Capacity



Log covariance matrices of filter outputs. Efficient filterbank optimizes tradeoff between maximizing within-channel variance and minimizing between-channel redundancy

Coding Efficiency vs. Accuracy



Coding efficiency predicts 10 % of variance in participant accuracy for spectrally-degraded conditions in experiment 2.