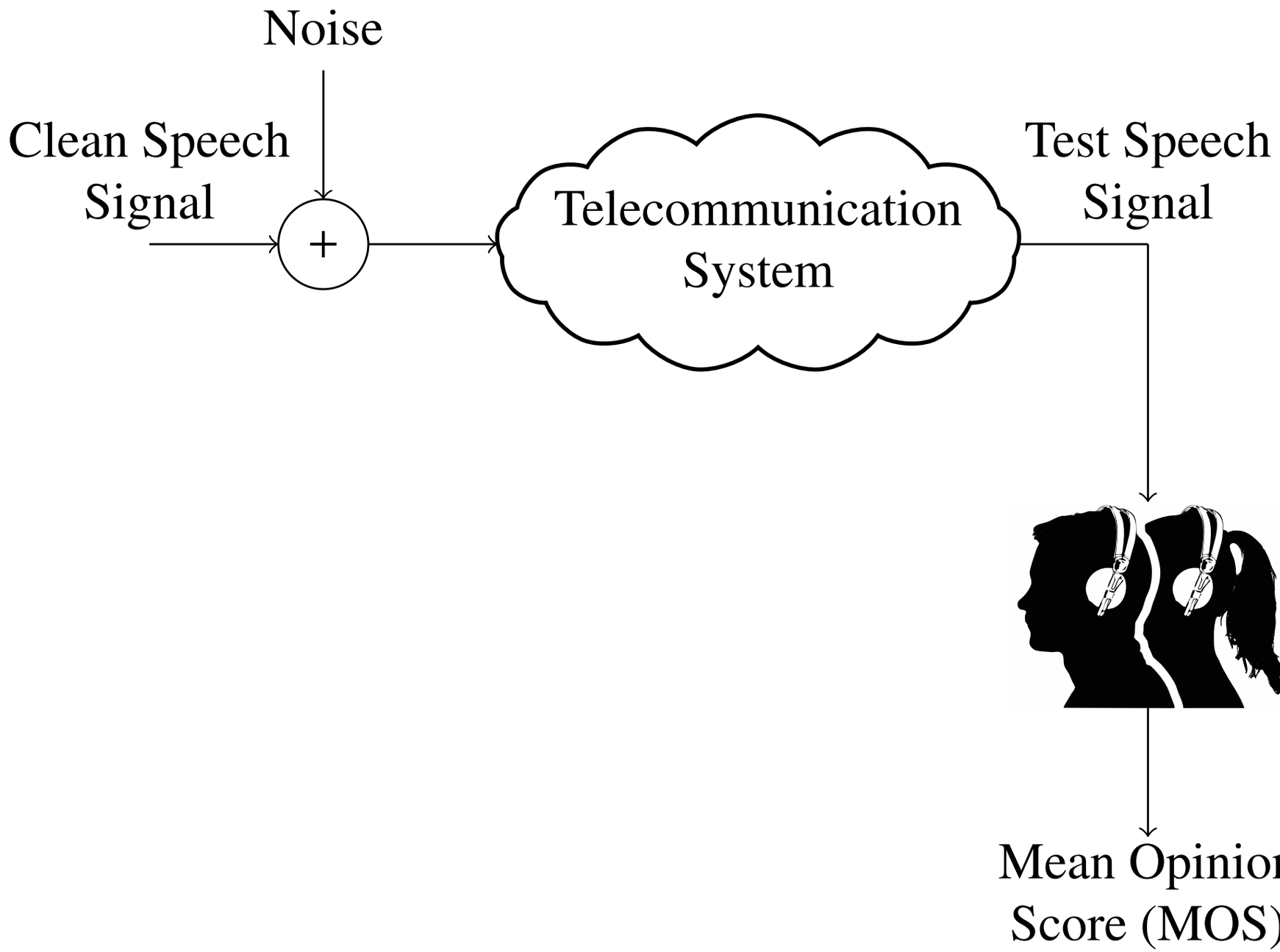
Perceptual Modeling Through an Auditory-Inspired Sparse Representation

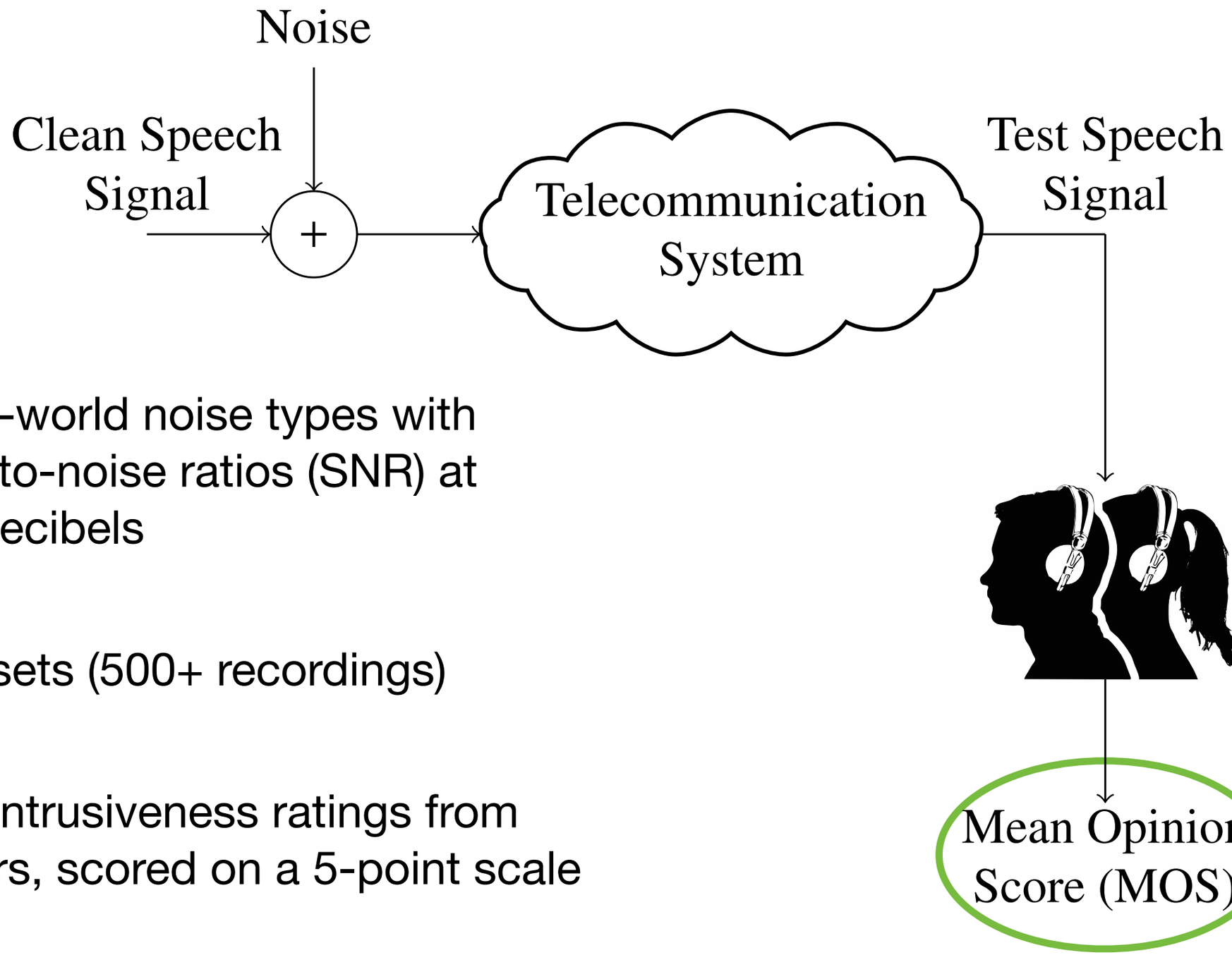
EPFL-Idiap-ETH Sparsity Workshop 2015

Raphael Ullmann[1,2] and Hervé Bourlard[1,2]
[1]Idiap Research Institute and [2]EPFL

Prediction of Perceived Noise Intrusiveness

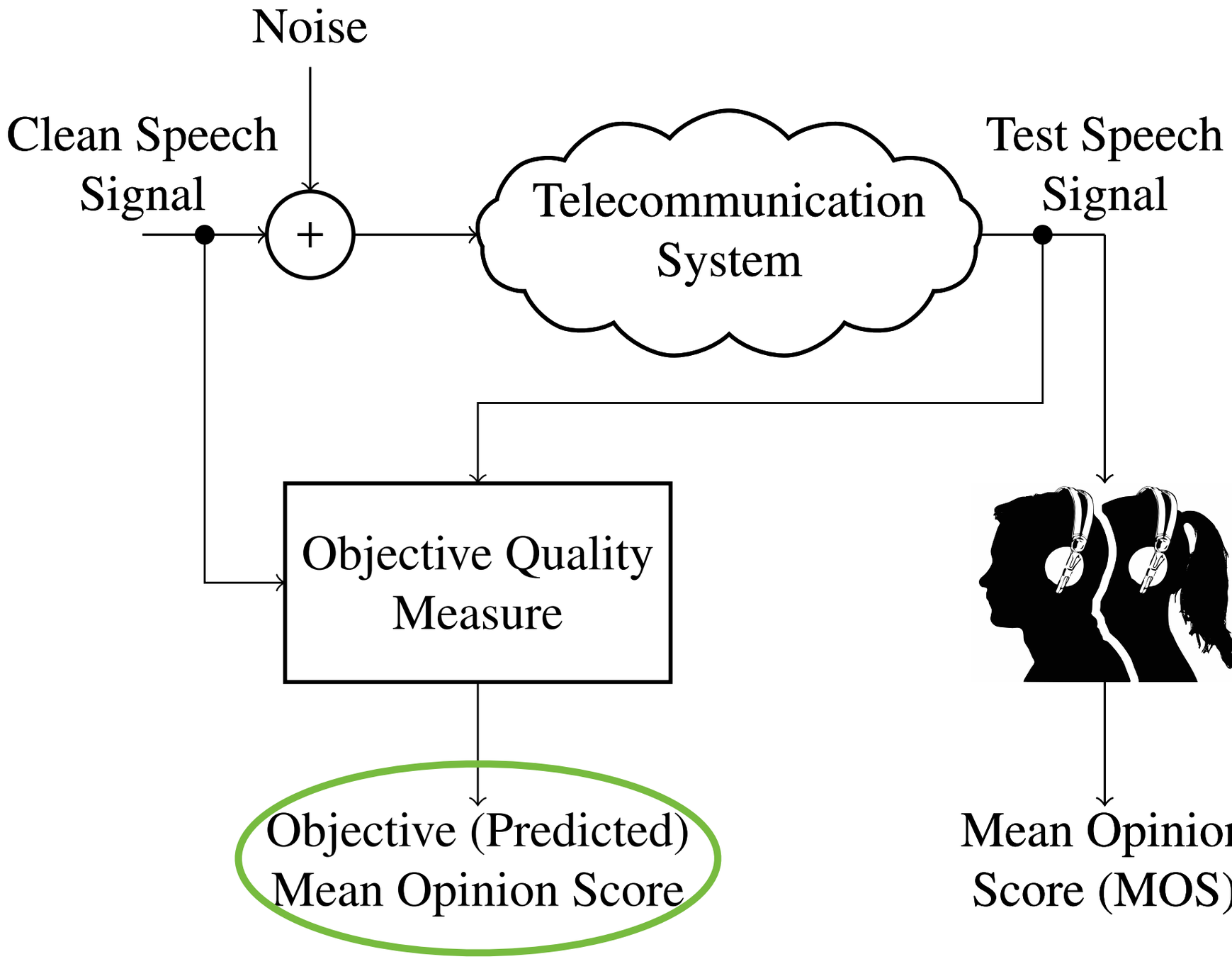


Prediction of Perceived Noise Intrusiveness



- 10 real-world noise types with signal-to-noise ratios (SNR) at 3–40 decibels
- 3 datasets (500+ recordings)
- Noise intrusiveness ratings from listeners, scored on a 5-point scale

Prediction of Perceived Noise Intrusiveness



Why Sparsity?

- Traditional approach
  - Combine acoustic features (noise level, variance, spectral composition)

- Our study: Focus on **low-level sensory coding principles**
  - Efficient Coding Hypothesis:
    “(...) our perceptions are caused by the activity of a rather small number of neurons selected from a very large population (...)” — [Barlow, 1972]
  - Redundancy reduction to help make sense of sensory inputs
    [Olshausen & Field, 1996]

- Barlow H. B. (1972) Single units and sensation: A neuron doctrine for perceptual psychology? Perception.
- Olshausen B. A., Field D. J. (1996) Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Neural Computation.

Efficient Auditory Coding — Model

- Generative waveform model [Lewicki & Sejnowski,1999]:

$$\hat{x}(t) = \sum_{m=1}^{M} \sum_{i=1}^{I_m} \alpha_m^i \phi_m \left(t - \tau_m^i\right)$$

- Shiftable kernels $\{\phi_1(t), \ldots, \phi_m(t), \ldots, \phi_M(t)\}$, can have different lengths

- Use Matching Pursuit to approximate $\hat{\mathbf{x}} = \Phi\boldsymbol{\alpha}$, includes translation of kernels

- May think of each kernel instance as a population of spiking auditory neurons ➔“Spike Coding”

- Lewicki M. S., Sejnowski T. J. (1999) Coding time-varying signals using sparse, shift-invariant representations. Adv. NIPS 11.

Efficient Auditory Coding — Dictionary

- How to choose the dictionary $\mathbf{\Phi} = \{\phi_1(t), \ldots, \phi_m(t), \ldots, \phi_M(t)\}$?

➔Learn a dictionary from natural environmental noises [Smith & Lewicki, 2006]

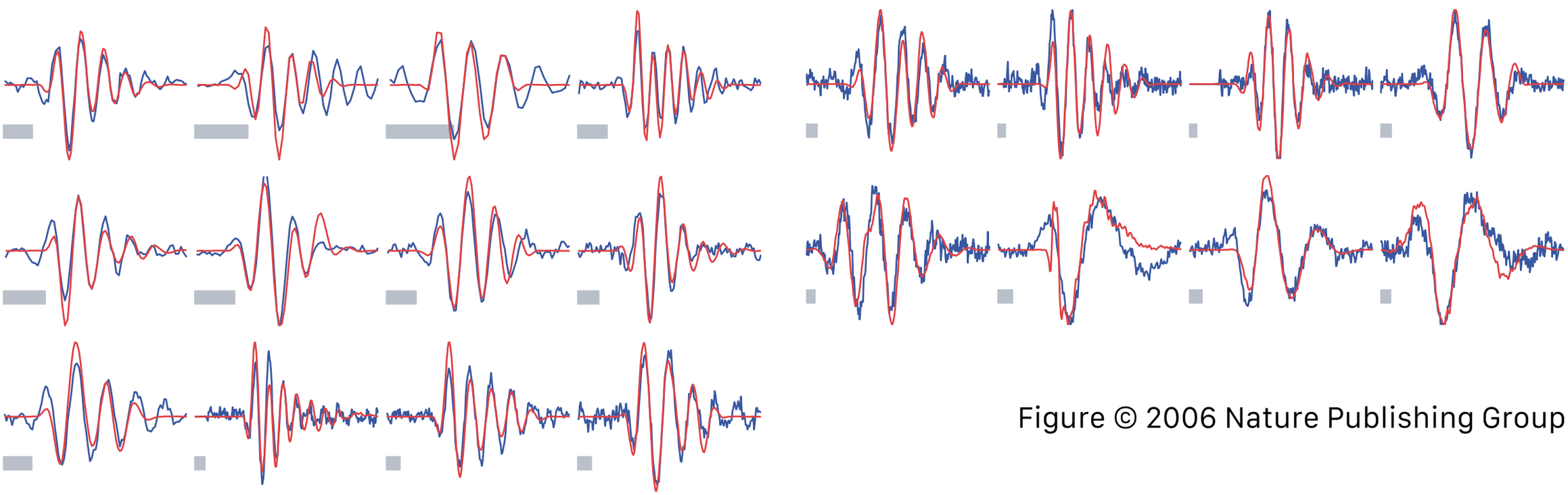
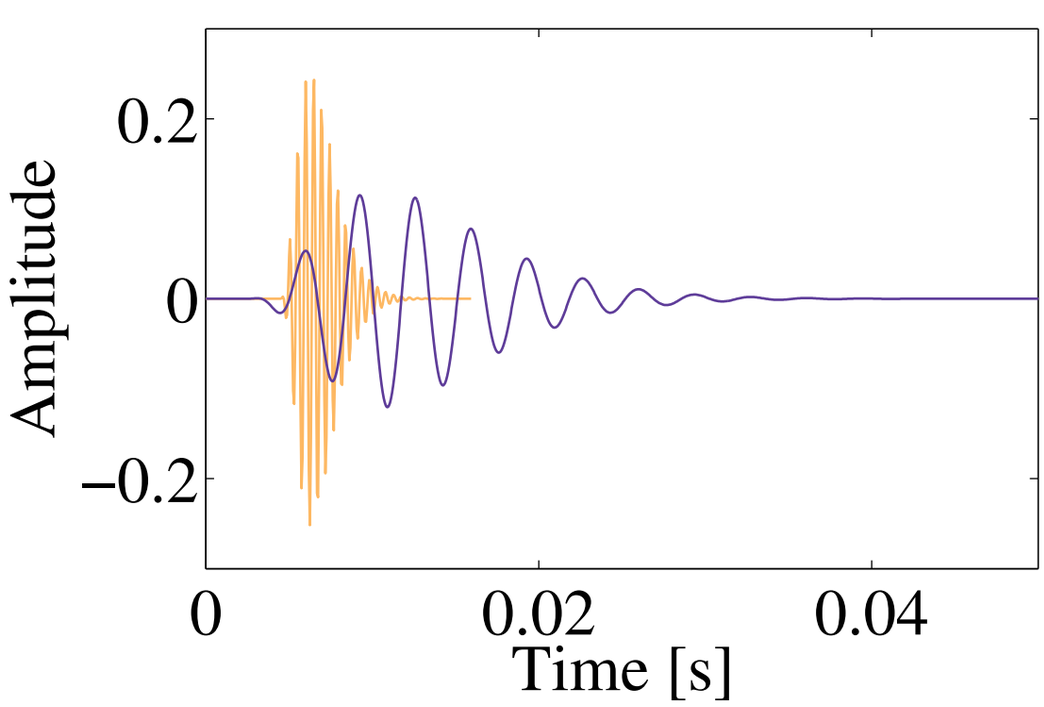


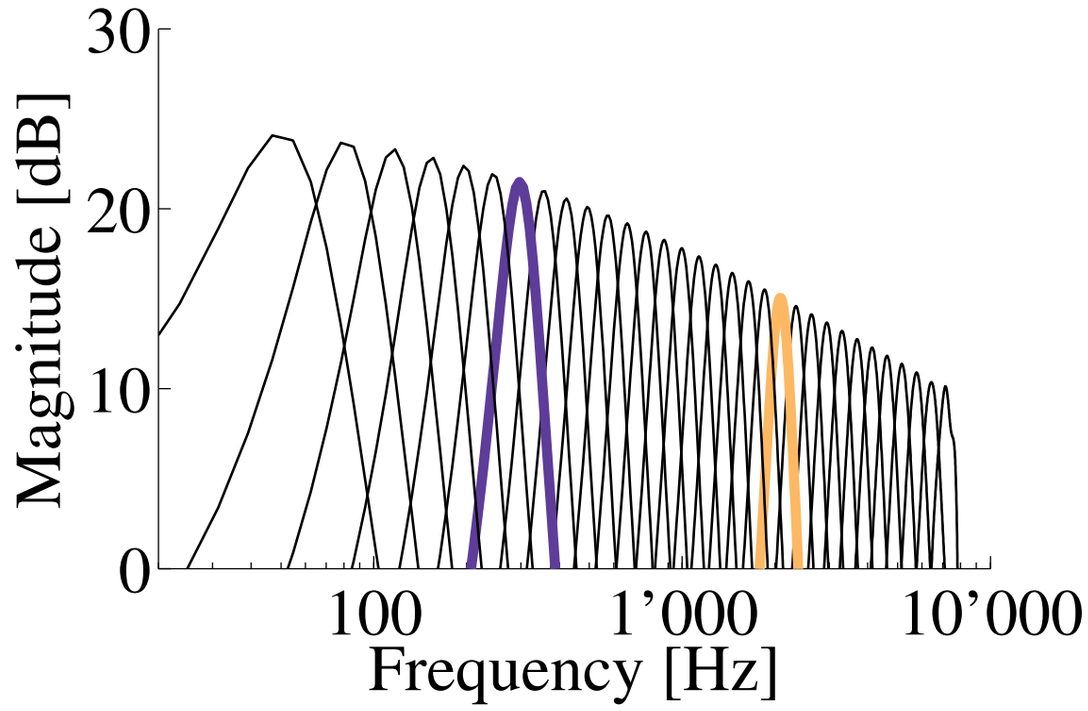
Figure © 2006 Nature Publishing Group

• Smith E. C., Lewicki M. S. (2006) Efficient Auditory Coding. Nature.

Perceptual Model — Dictionary

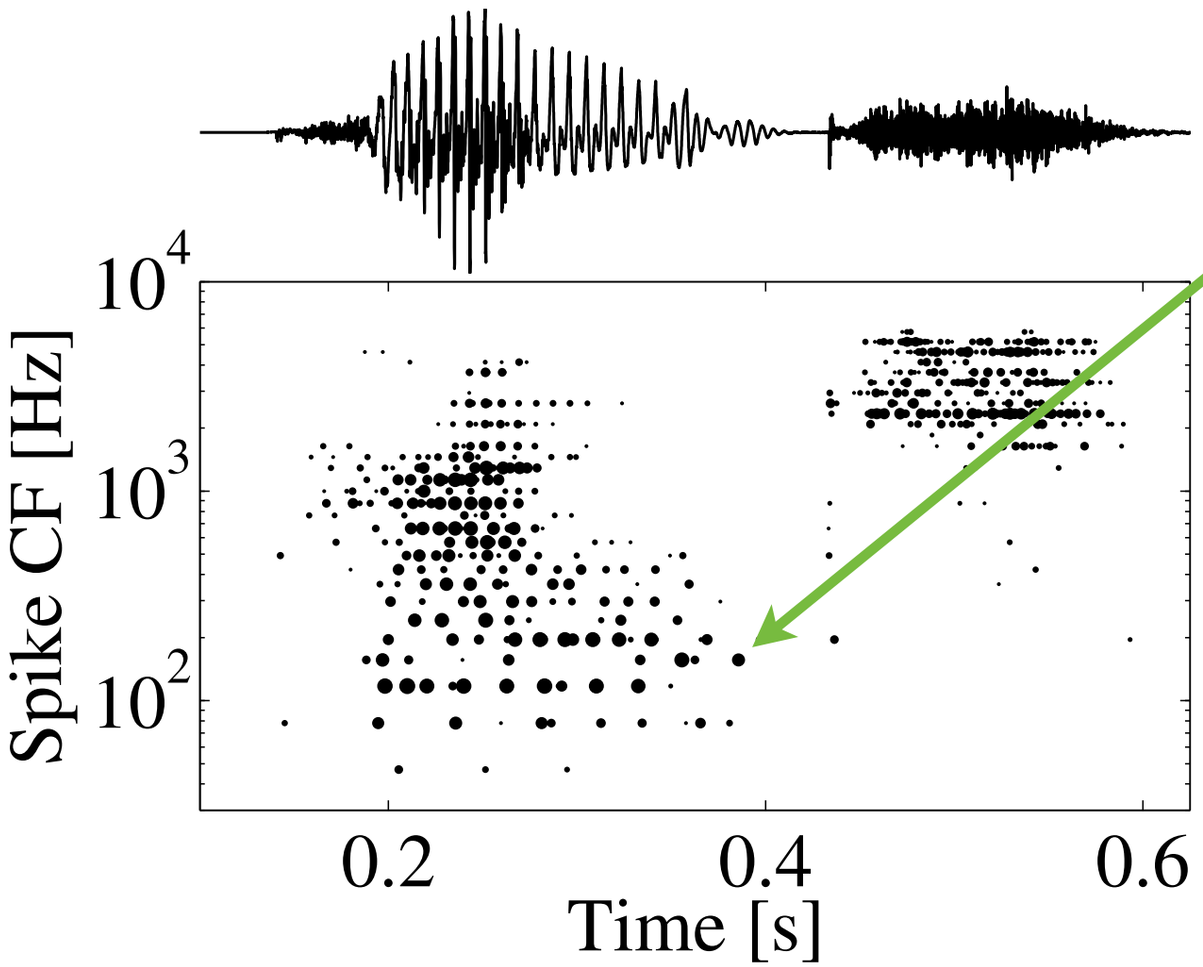
- Use a dictionary of analytically defined auditory filter shapes (“gammatones”)
- We use 32 gammatones sampled at 16 kHz, generated with Slaney’s toolbox





- Slaney M. (1998) Auditory Toolbox — Version 2. Technical Report. Interval Research Corp.

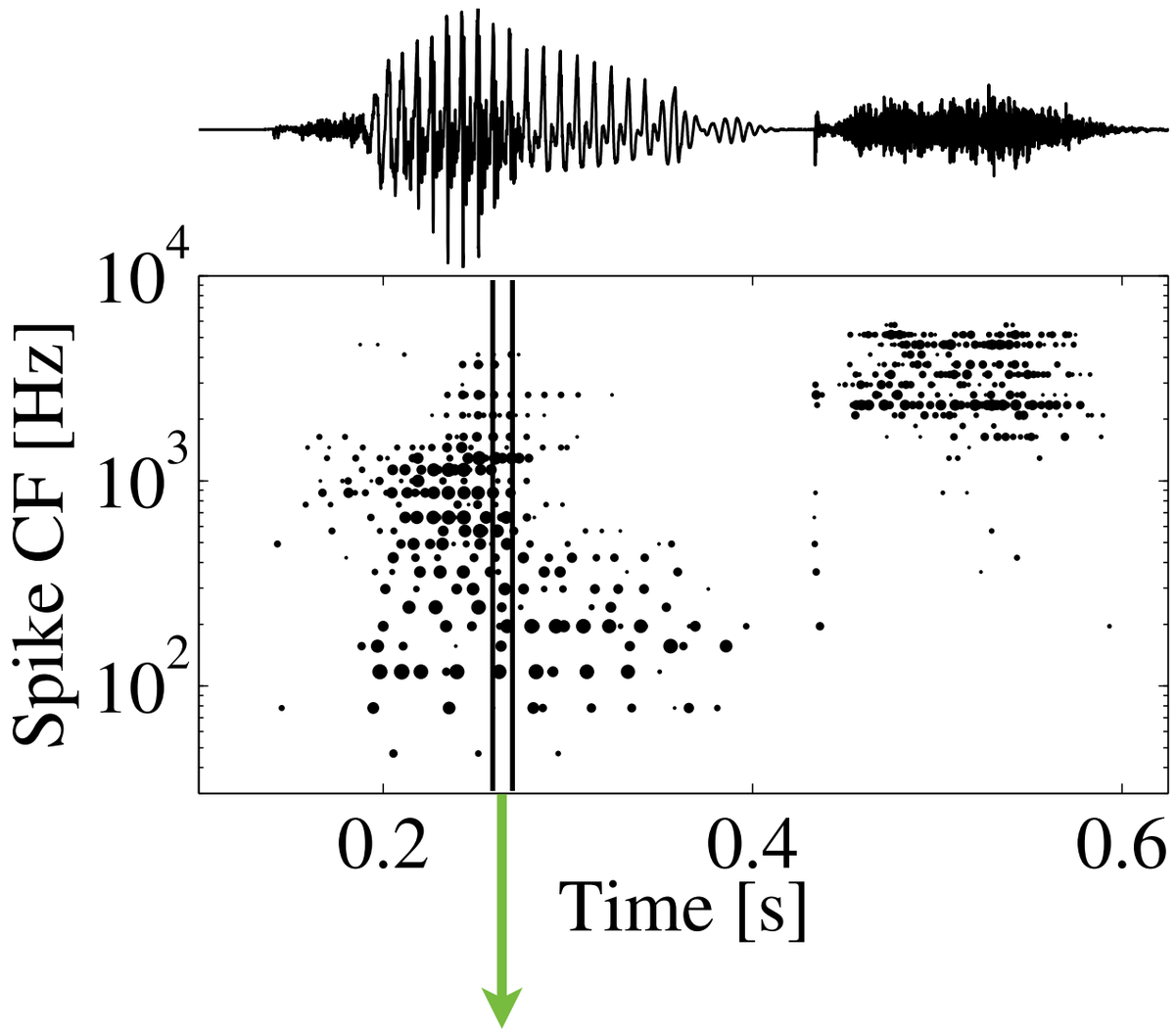
Perceptual Model — Noise Signal Analysis



Kernel instances $\phi_{j(k)}(t)$ are localized in time and frequency

Kernel instances are called atoms or **“spikes”**

Perceptual Model — Noise Signal Analysis

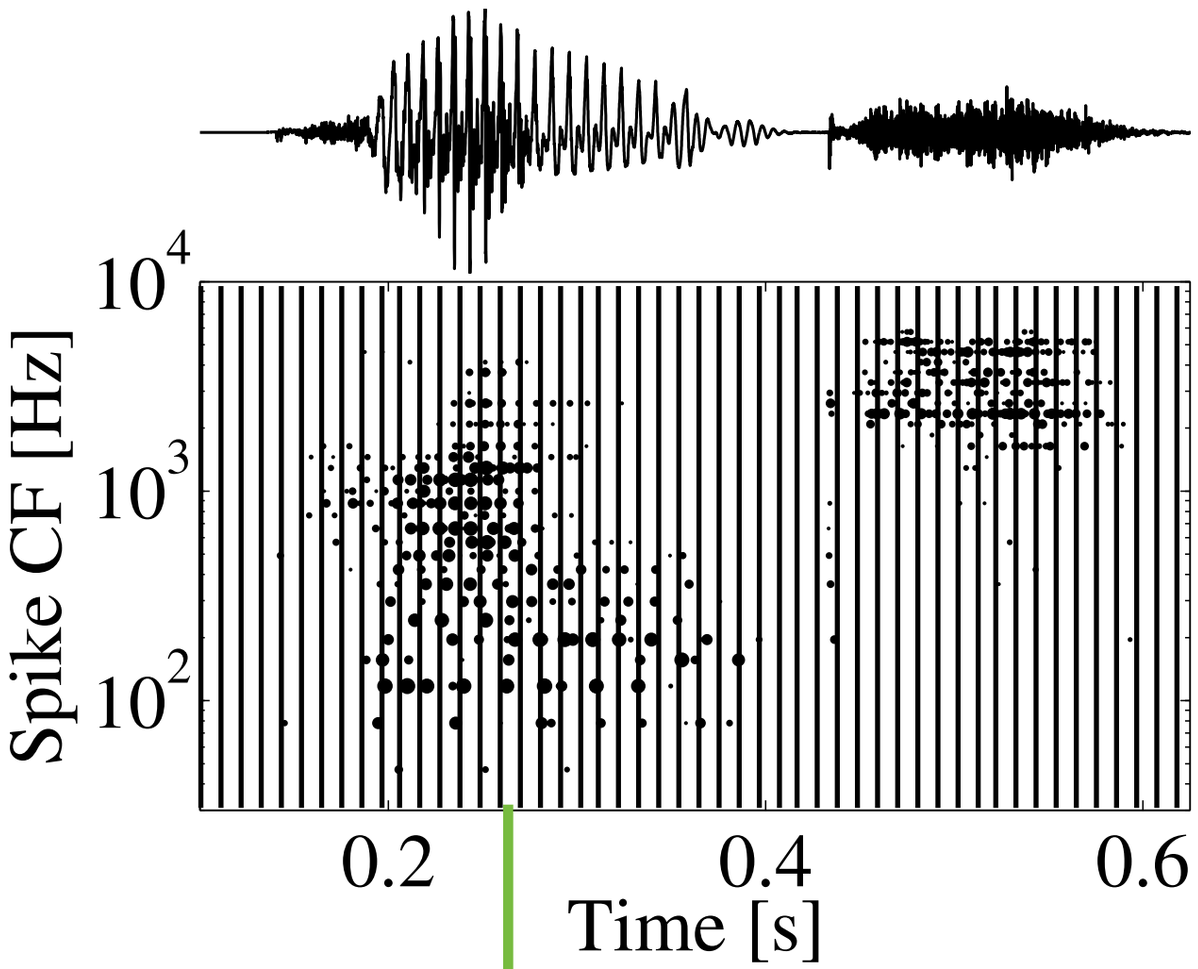


Compute number of spikes (i.e., $\ell 0$ norm)

Kernel instances $\phi_{j(k)}(t)$ are localized in time and frequency

Kernel instances are called atoms or **“spikes”**

Perceptual Model — Noise Signal Analysis



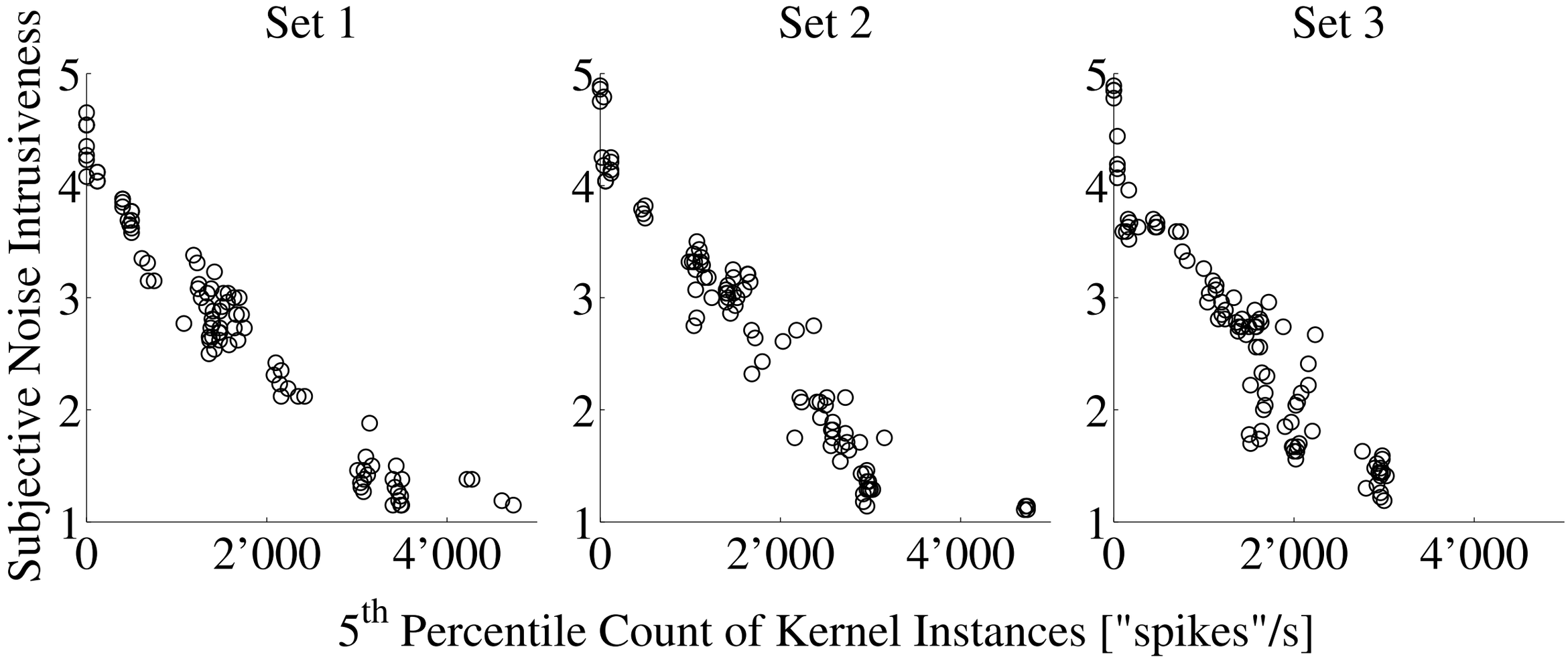
Get $\ell 0$ norm over time
Take the 5th percentile

Kernel instances $\phi_{j(k)}(t)$ are localized in time and frequency

Kernel instances are called atoms or **“spikes”**

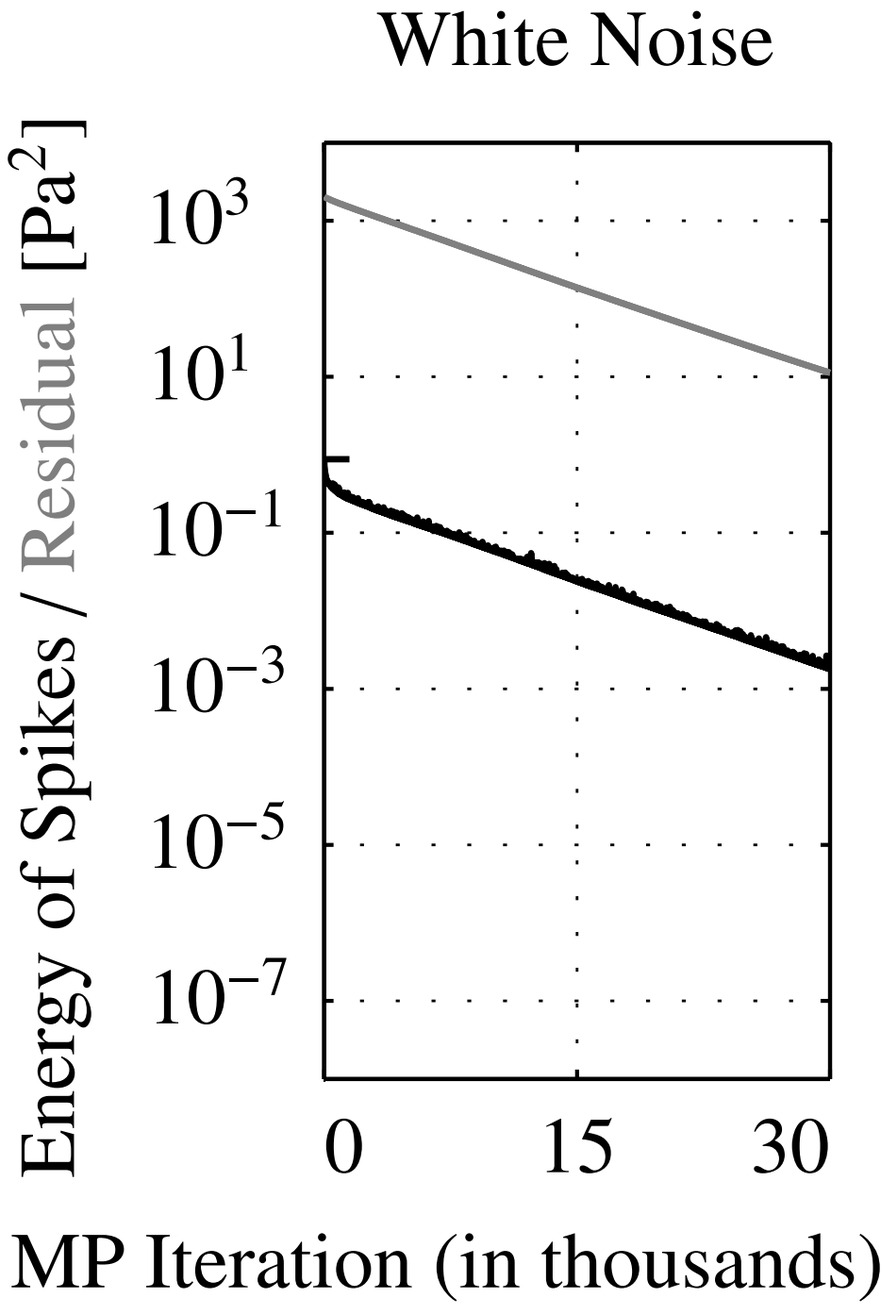
Perceptual Model — Evaluation

5th percentile of “spikes” over time highly correlates with subjective scores of noise intrusiveness



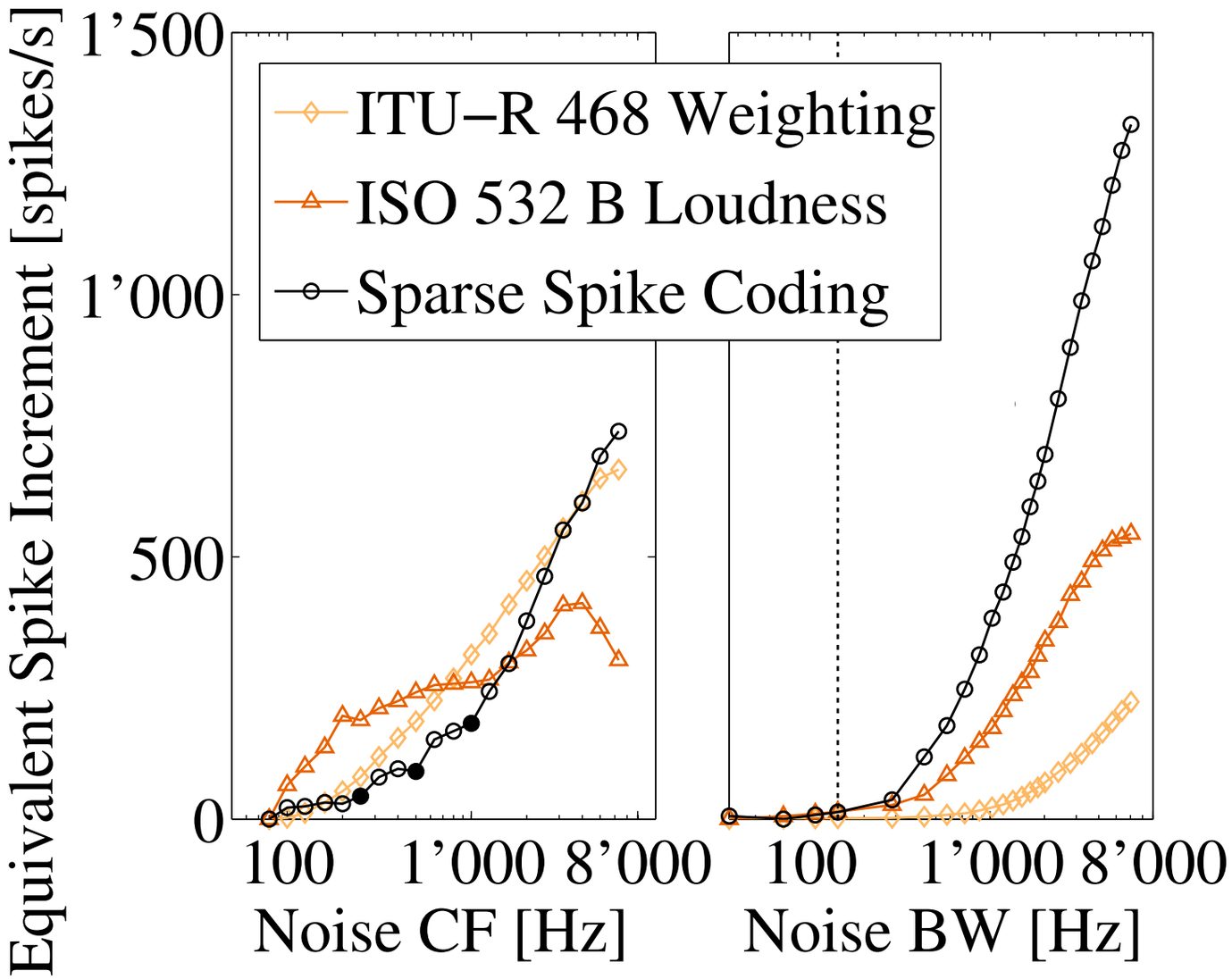
Why Does It Work? — Because of Greedy Pursuit

- Decrease of spike energies (black line) depends on signal type
  - White noise is a kind of “worst case”, i.e., it does not correlate well with any kernel in the dictionary
- Logarithmic changes in sound energy produce *linear* changes in spike counts
- Greedy decomposition captures high-energy sounds first



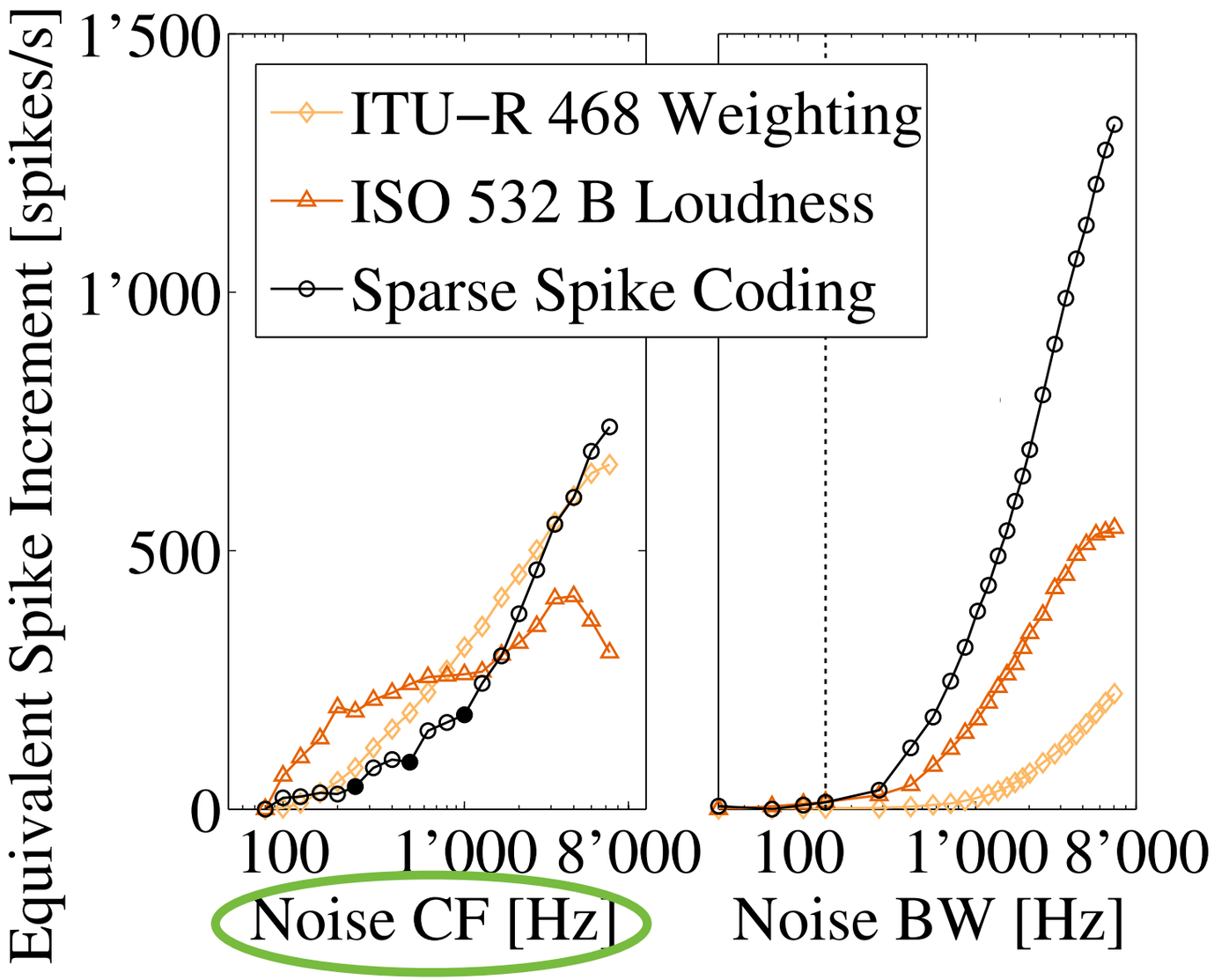
Why Does It Work? — Because of the Dictionary

Some tests with narrowband noises



Why Does It Work? — Because of the Dictionary

Some tests with narrowband noises

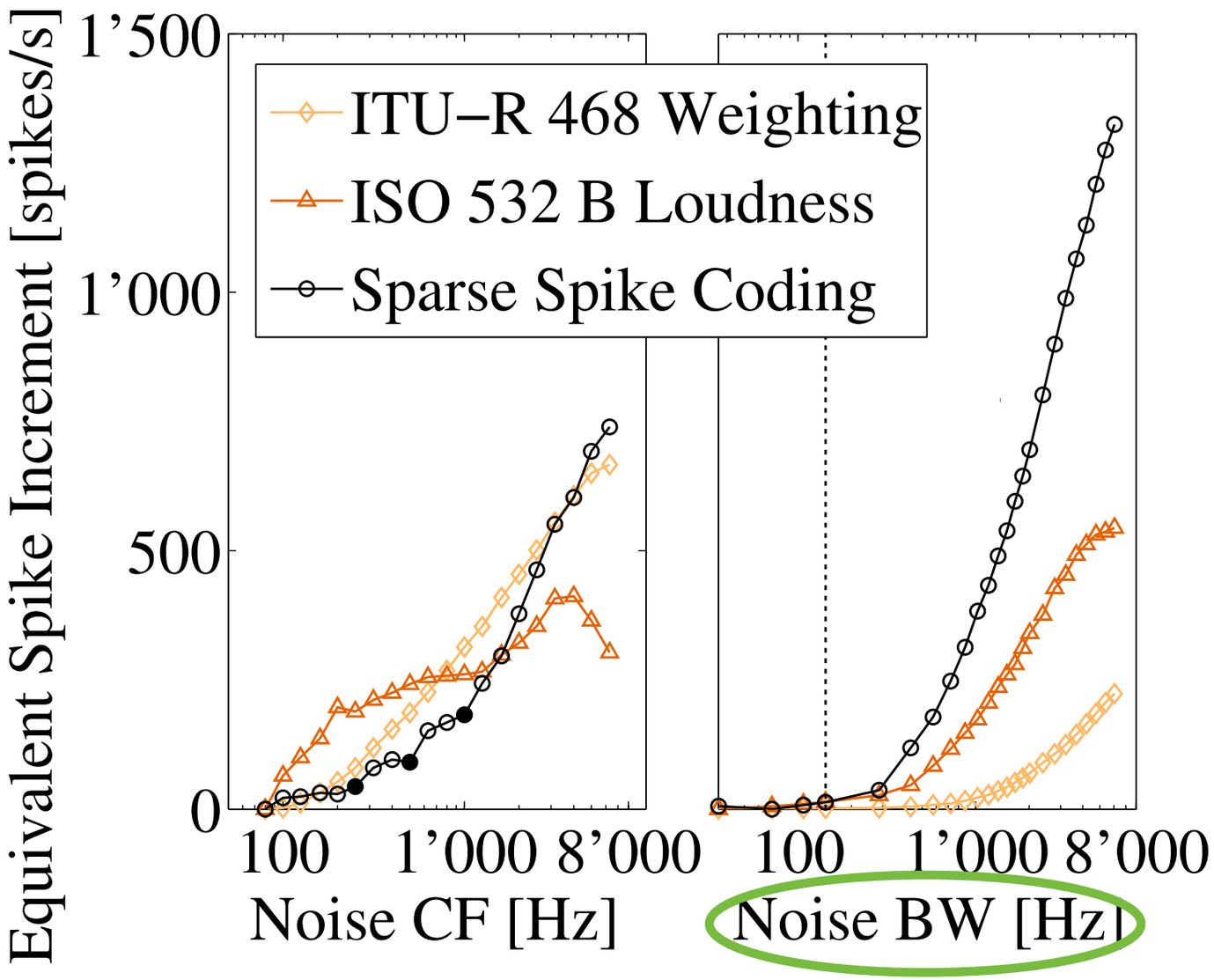


- ERB-wide noise at varying center frequencies (CF)

➔Spike count similar to noise weighting curves

Why Does It Work? — Because of the Dictionary

Some tests with narrowband noises



- ERB-wide noise at varying center frequencies (CF)

➔Spike count similar to noise weighting curves

- Fixed center frequency, increasing noise bandwidth

➔Spike count increases above auditory bandwidth (dotted line)

Results — Comparison to Other Measures

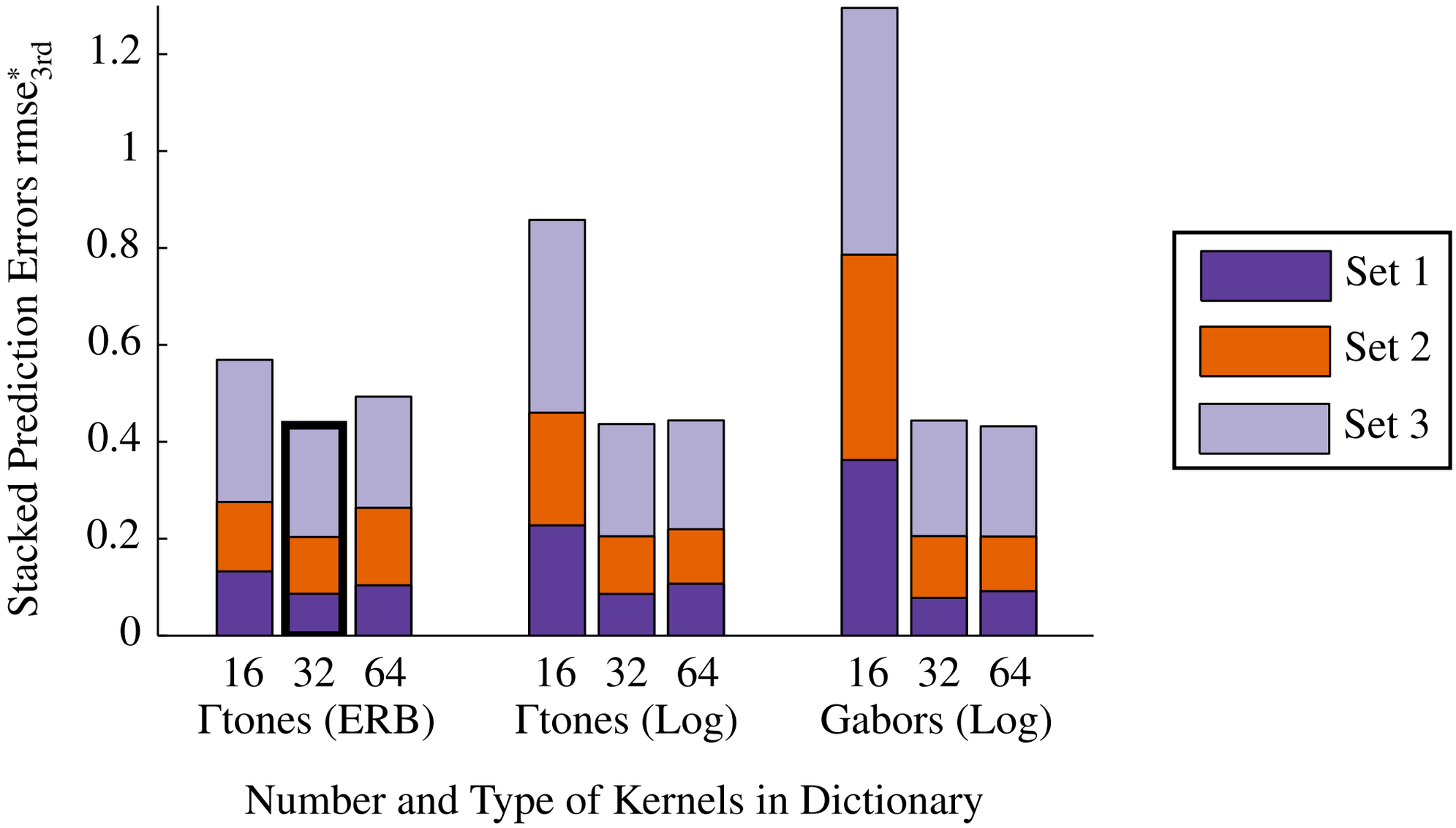
- Comparison to widely used acoustic indicators
  - Noise level in decibels with “A” frequency weighting, denoted “dB(A)”
  - Loudness (a psychoacoustic model of perceived sound intensity)

➔Significantly lower prediction error ($p < 0.01$) on 2 datasets

Measure	Prediction Error (lower values are better)		
	Set 1	Set 2	Set 3
Weighted Level [dB(A) SPL]	0.230	0.277	0.234
Mean Loudness [sone]	0.257	0.206	**0.197**
5th Percentile Loudness [sone]	0.191	0.234	0.270
5th Percentile Density [spikes/s]	**0.087****	**0.117****	0.231

Results — (In)sensitivity to Parameters

- Robust to changes in dictionary design



Conclusion

- We are doing audio processing, not speech processing
- Number of “spikes” reflects the level and type of noise
- Sparsity of noise over time highly correlates with perceived intrusiveness
- Efficient coding hypothesis offers a different interpretation of intrusiveness:
  - Complexity of the input stream to the auditory system
  - Activations of nerve spike populations in response to noise

Thank You for Your Attention

Thanks to

- Laboratory of Electromagnetics and Acoustics (LEMA), EPFL
- SwissQual AG
- Dr. Marc Ferras and Dr. Mathew Magimai-Doss for useful discussions