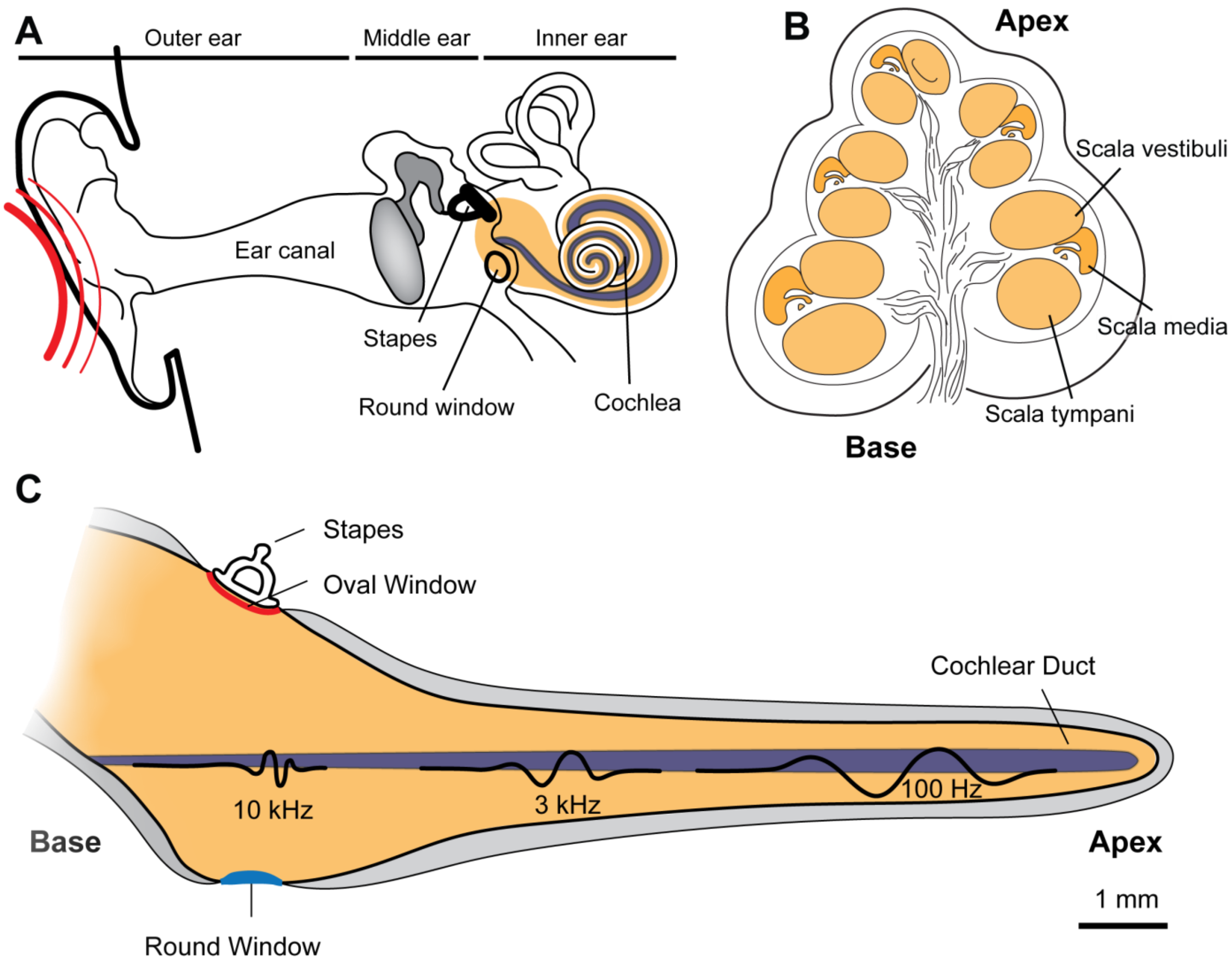
Hearing Hearing Loss

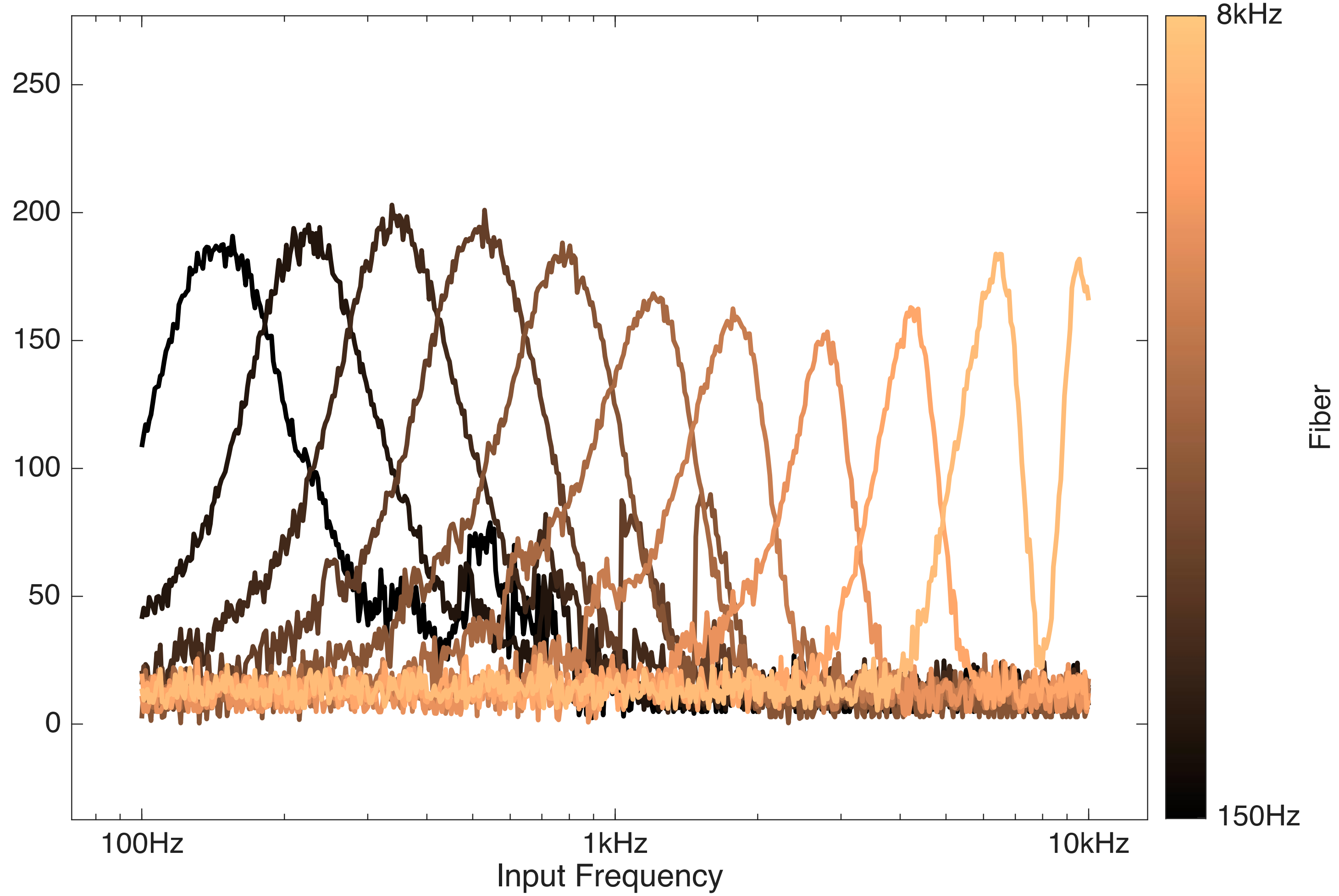
Anes Macić

Computer Audition F2023

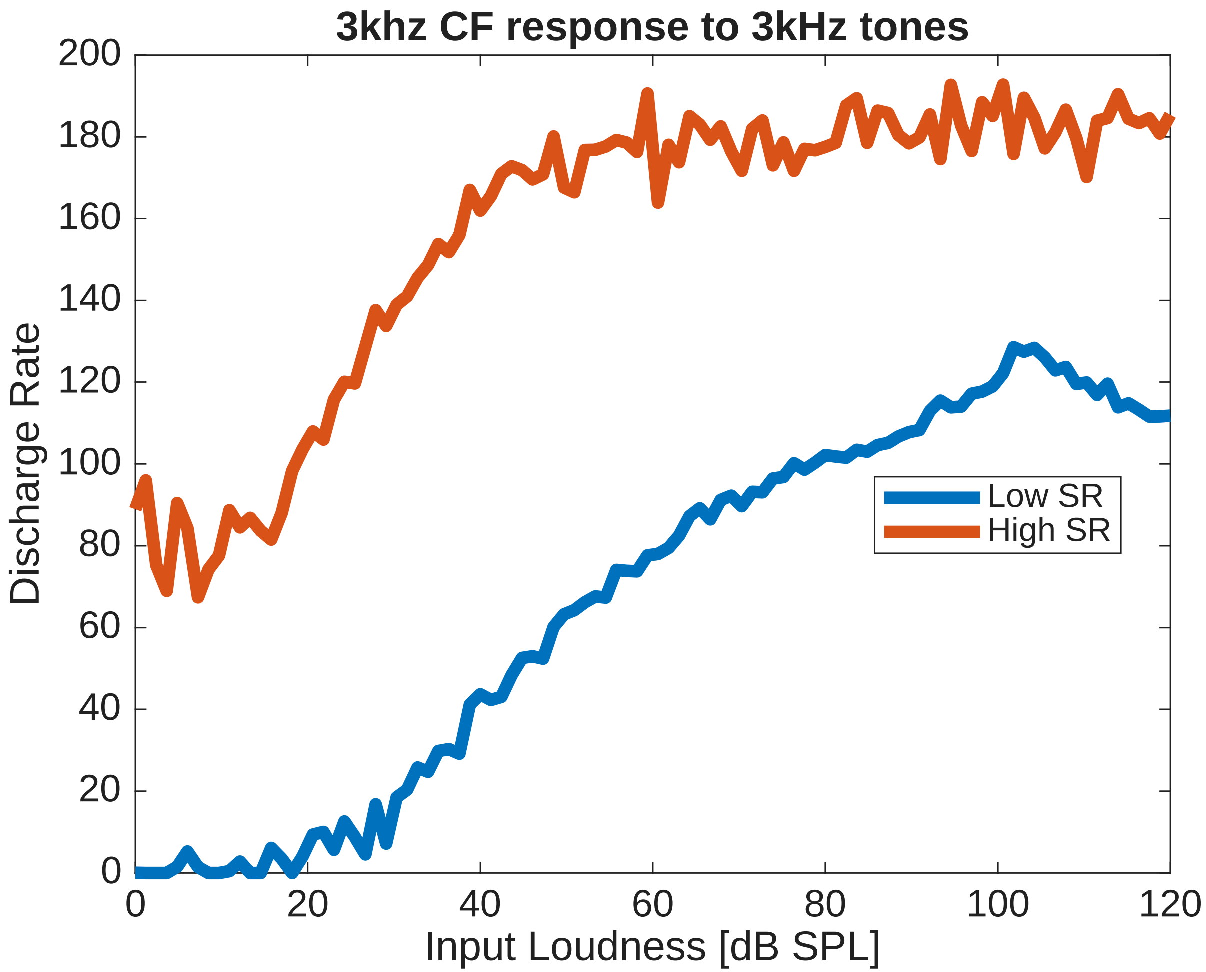
Inner Ear - Mechanotransduction of Pressure Waveform to a "Neurogram"



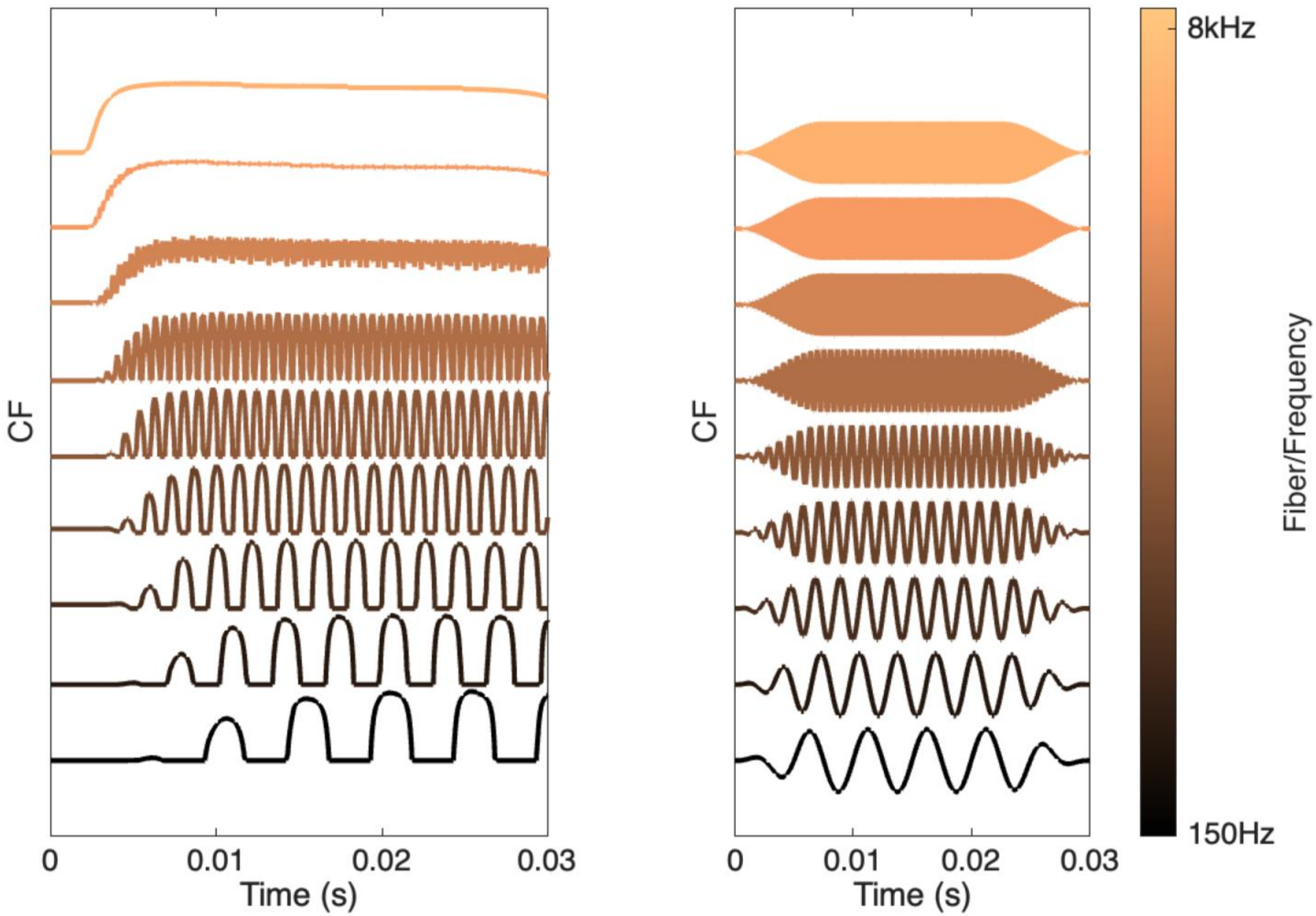
Auditory Nerve Properties: Frequency Sensitivity



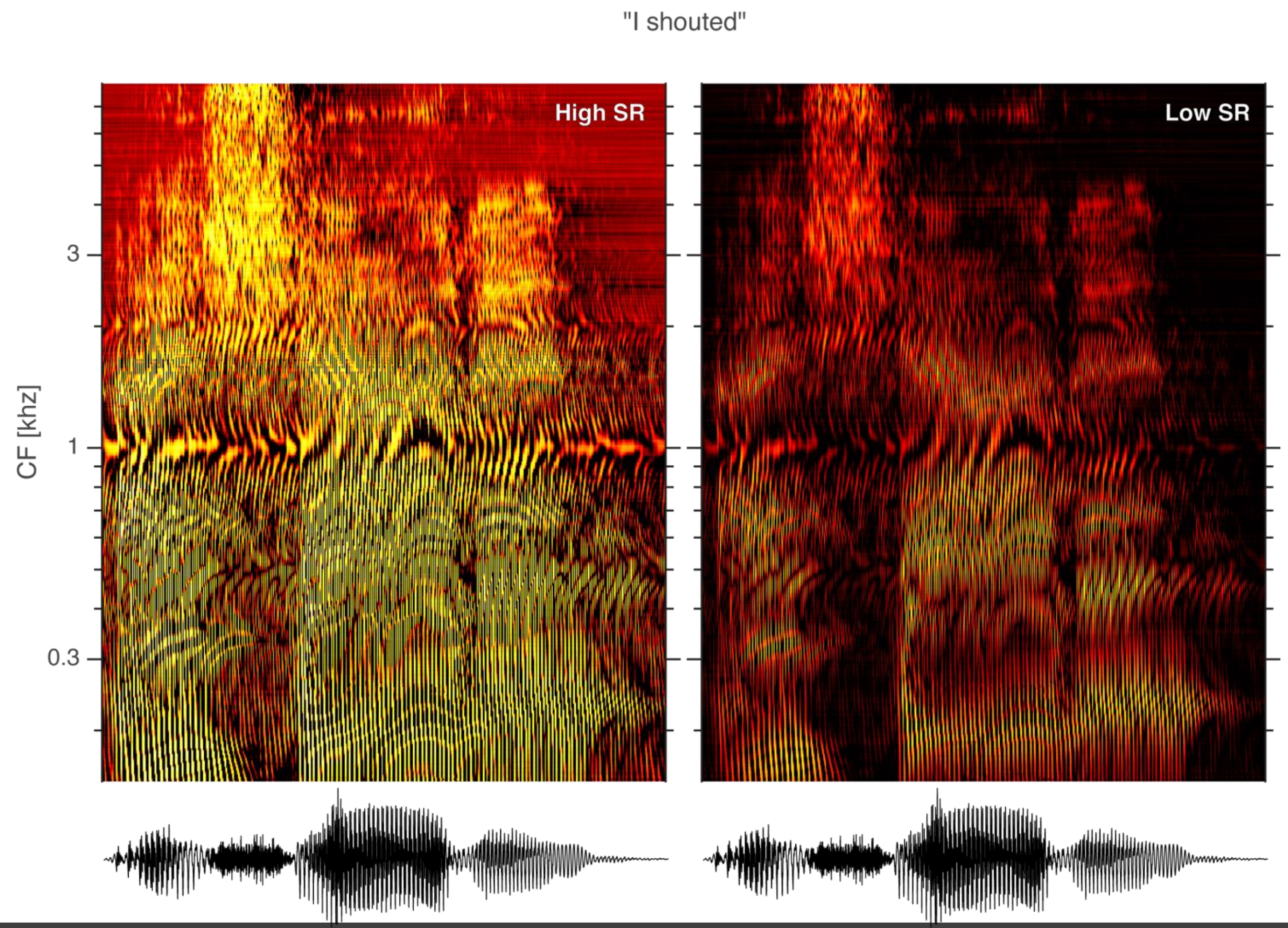
Auditory Nerve Properties: Dynamic Range



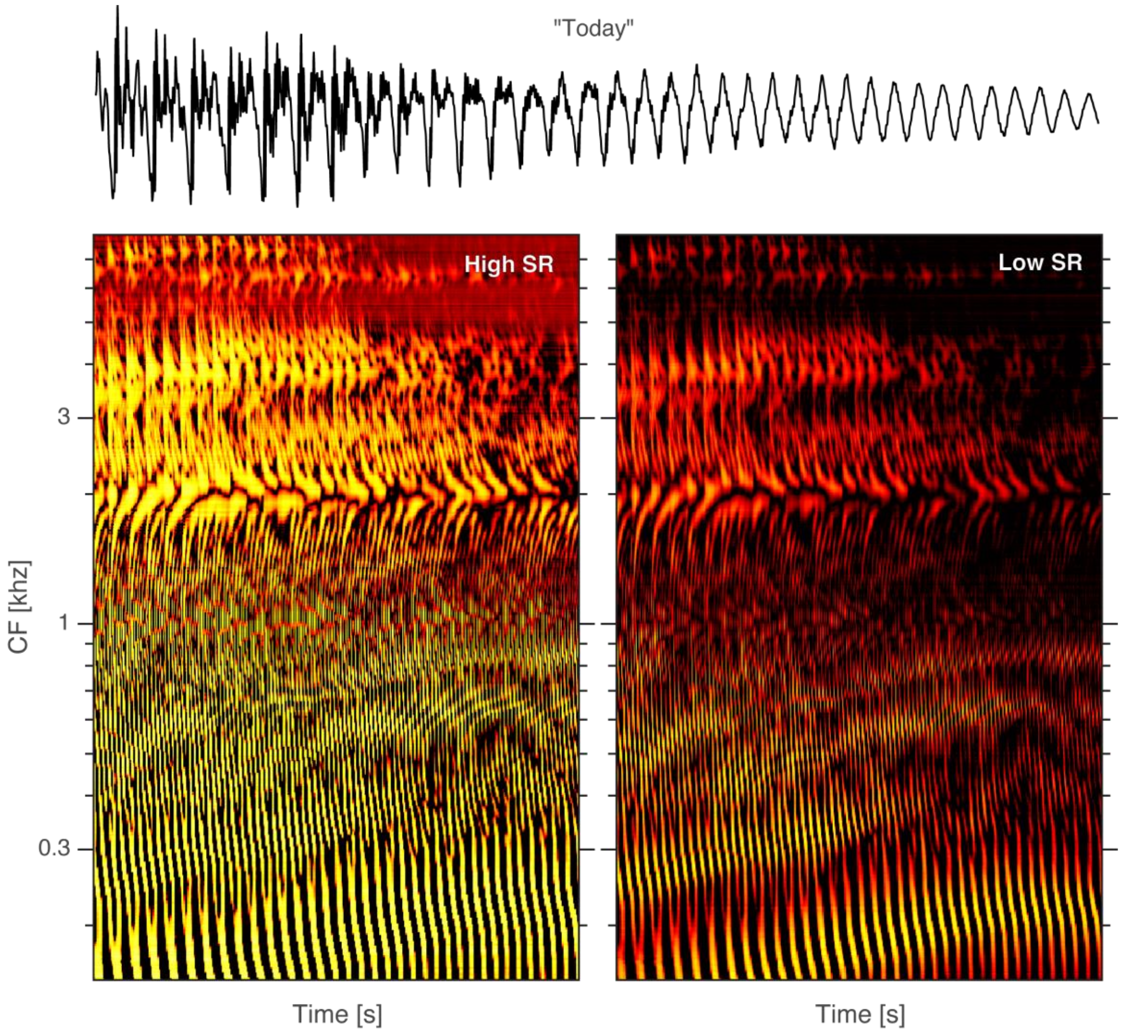
Auditory Nerve Properties: Timing (phase locking)



A neurogram is a population response of AN fibers (a cochlear STFT)



A neurogram is a population response of AN fibers (a cochlear STFT)

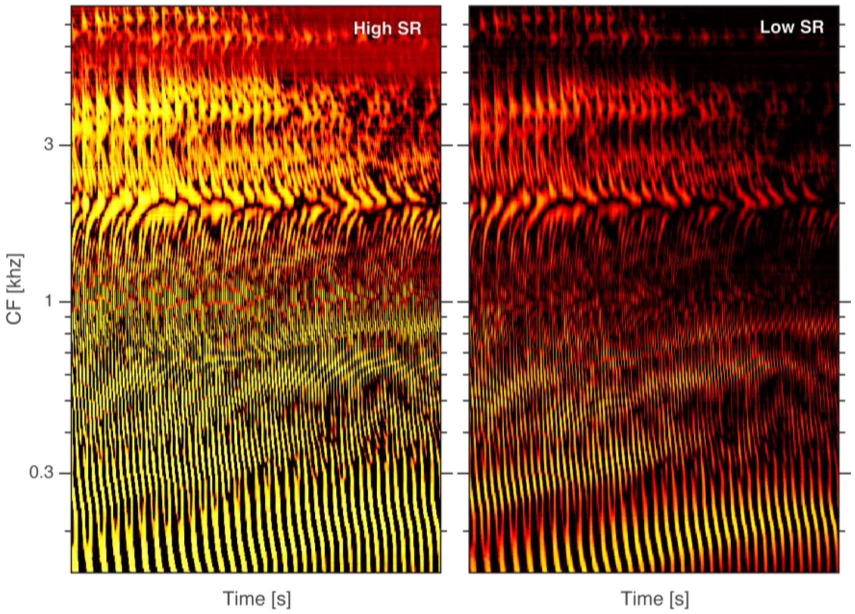


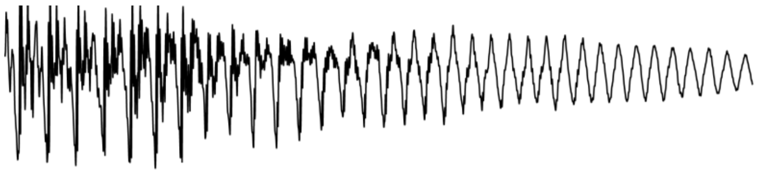
Not encoding phase of CF

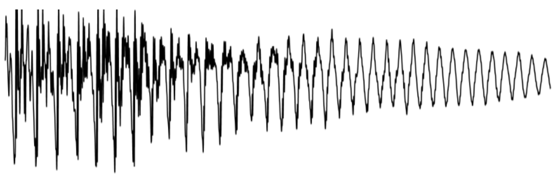
Encoding envelope

Encoding phase of CF

Invert the Auditory Nerve Model (ANM)





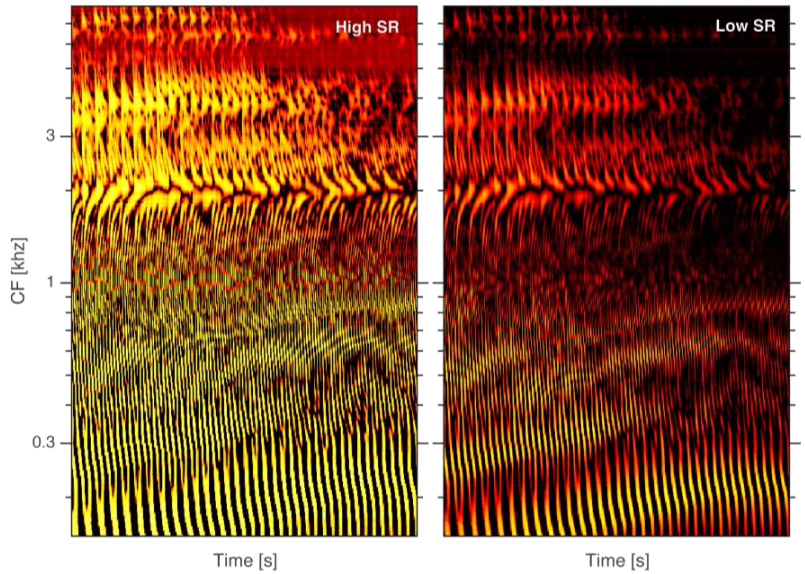


ANM with HL

→

iANM with no HL

←



?

Hearing Hearing Loss

Data

LibriSpeech Audio Corpus

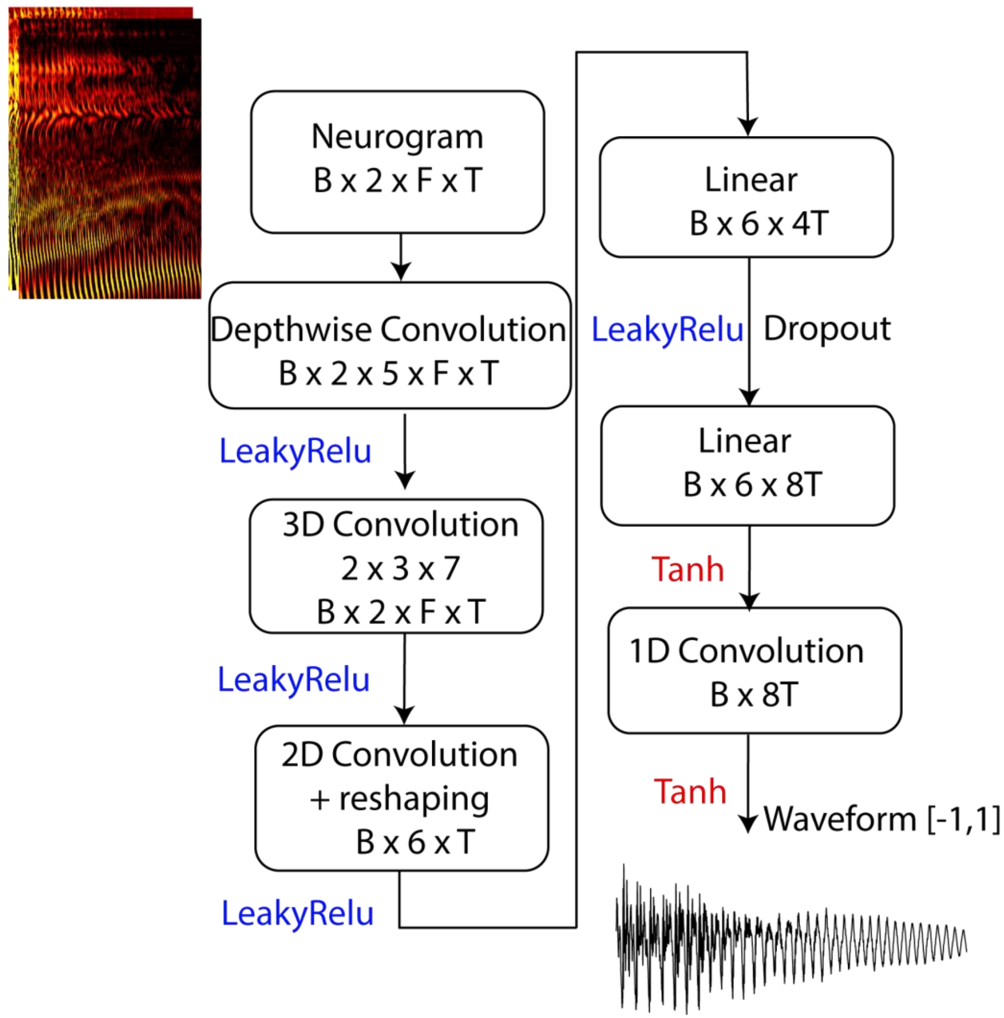
2 hours of English speech (40 speakers, all different sentences) at 16kHz

Preprocessing

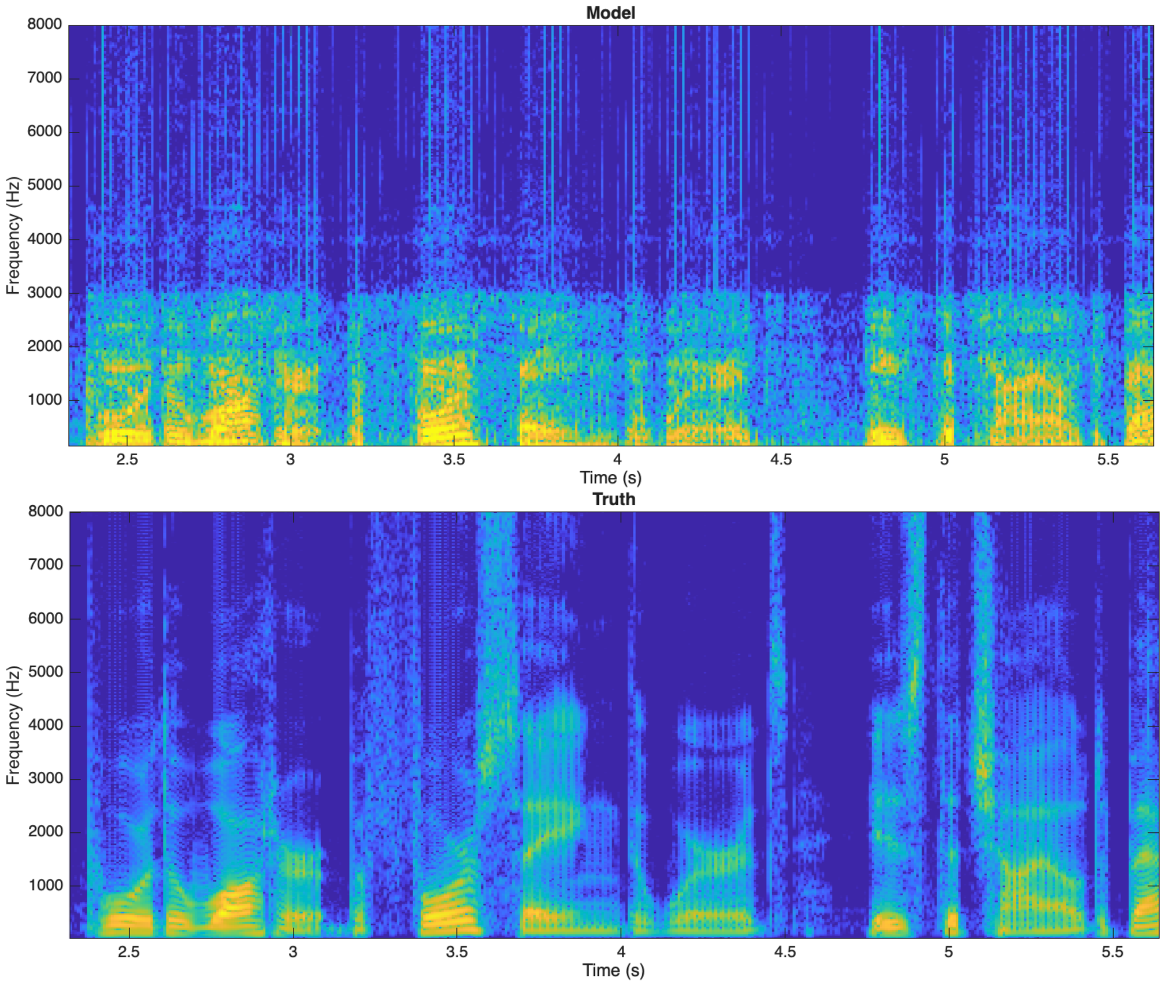
Bruce, Erfani & Zilany (2018) Auditory Nerve Model

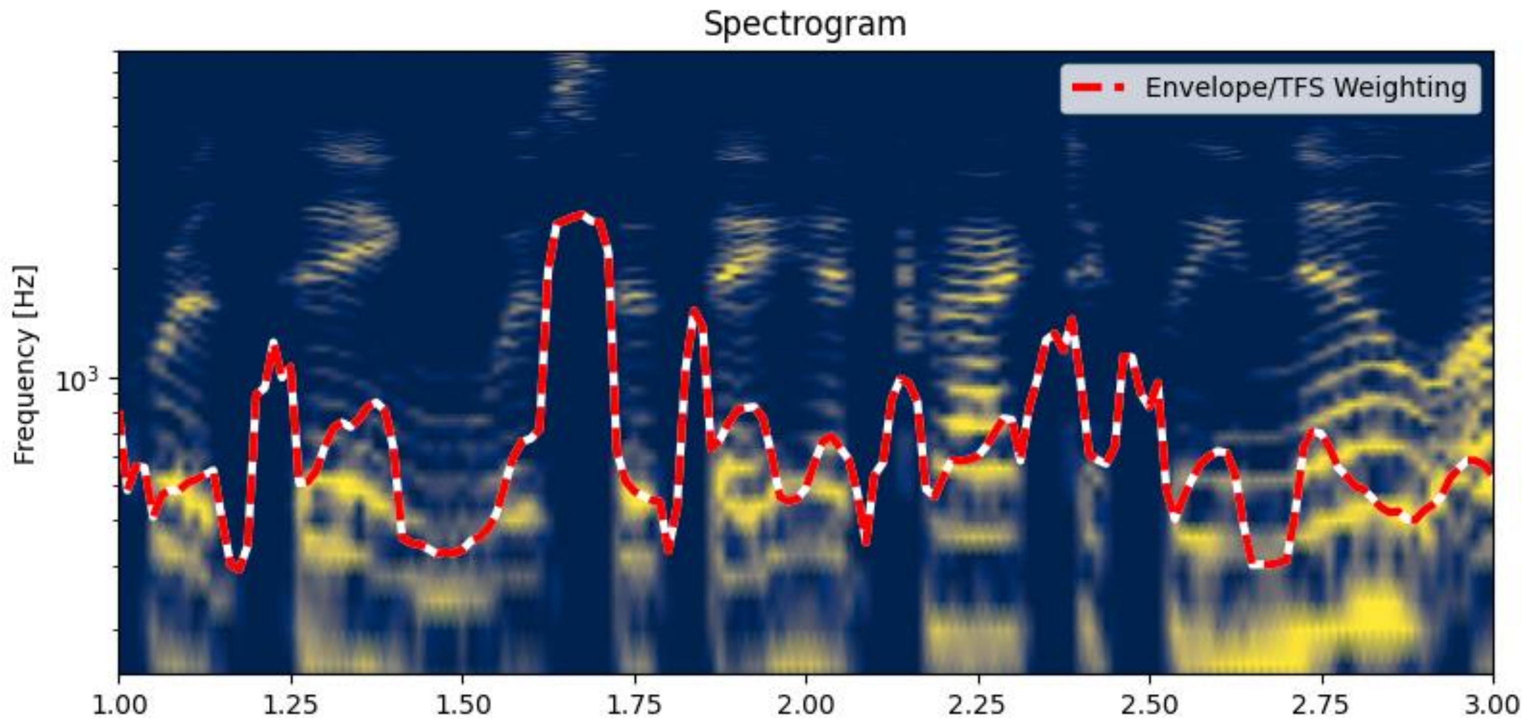
Neurograms have 2 channels (fiber types) and 512 frequency channels (150Hz – 8kHz) at fixed 65 dB SPL

Few minutes of speech take 1 hour of processing when run in parallel on 20 CPUs, 90GB RAM

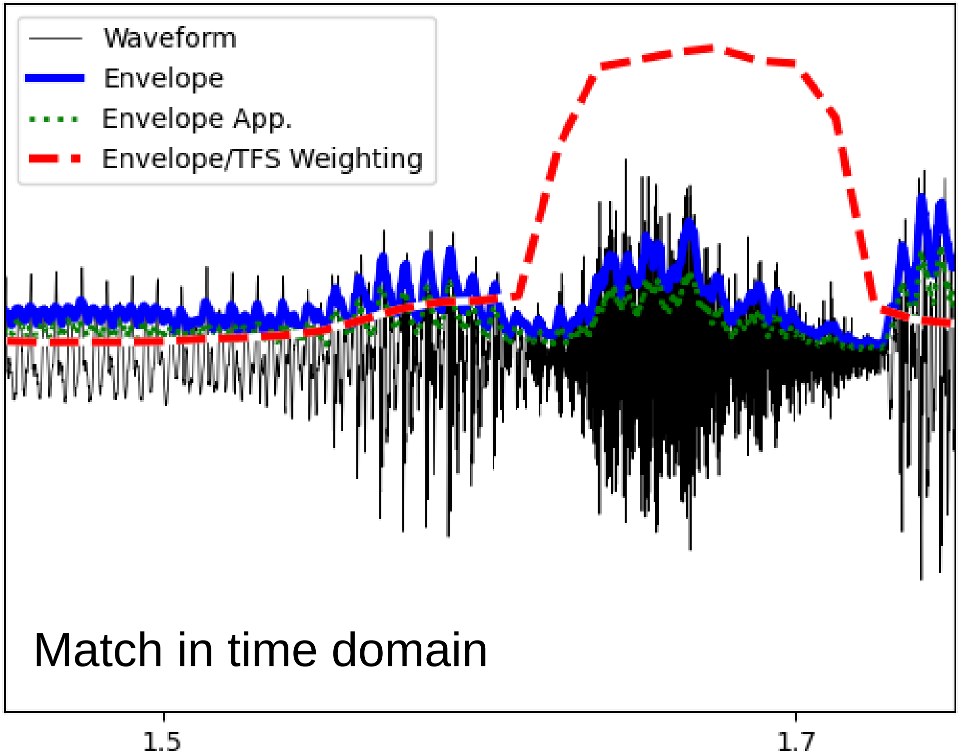

Neurogram
B x 2 x F x T
Depthwise Convolution
B x 2 x 5 x F x T
LeakyRelu
3D Convolution
2 x 3 x 7
B x 2 x F x T
LeakyRelu
2D Convolution
+ reshaping
B x 6 x T
LeakyRelu
Linear
B x 6 x 4T
LeakyRelu
Dropout
Linear
B x 6 x 8T
Tanh
1D Convolution
B x 8T
Tanh
Waveform [-1,1]

Initial Loss: Time Domain MSE



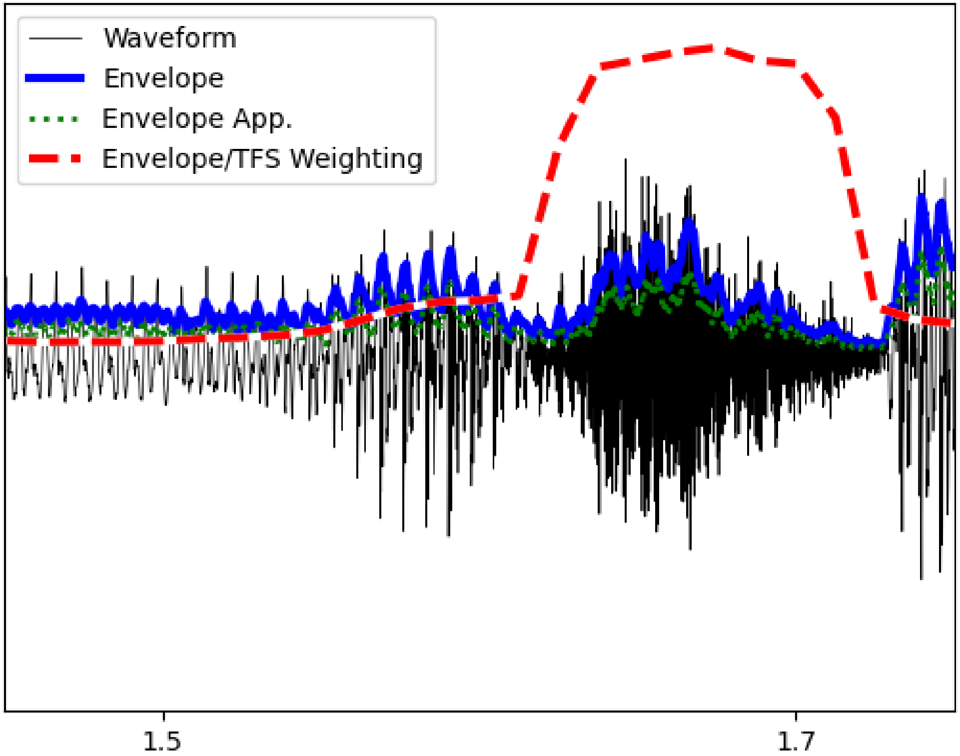


$$\xi(t) = \frac{\bar{S}_{xx} * f}{\int \bar{S}_{xx} df}$$



$$L_1 = \sqrt{\frac{1}{N}\sum(y - \hat{y})^2}$$

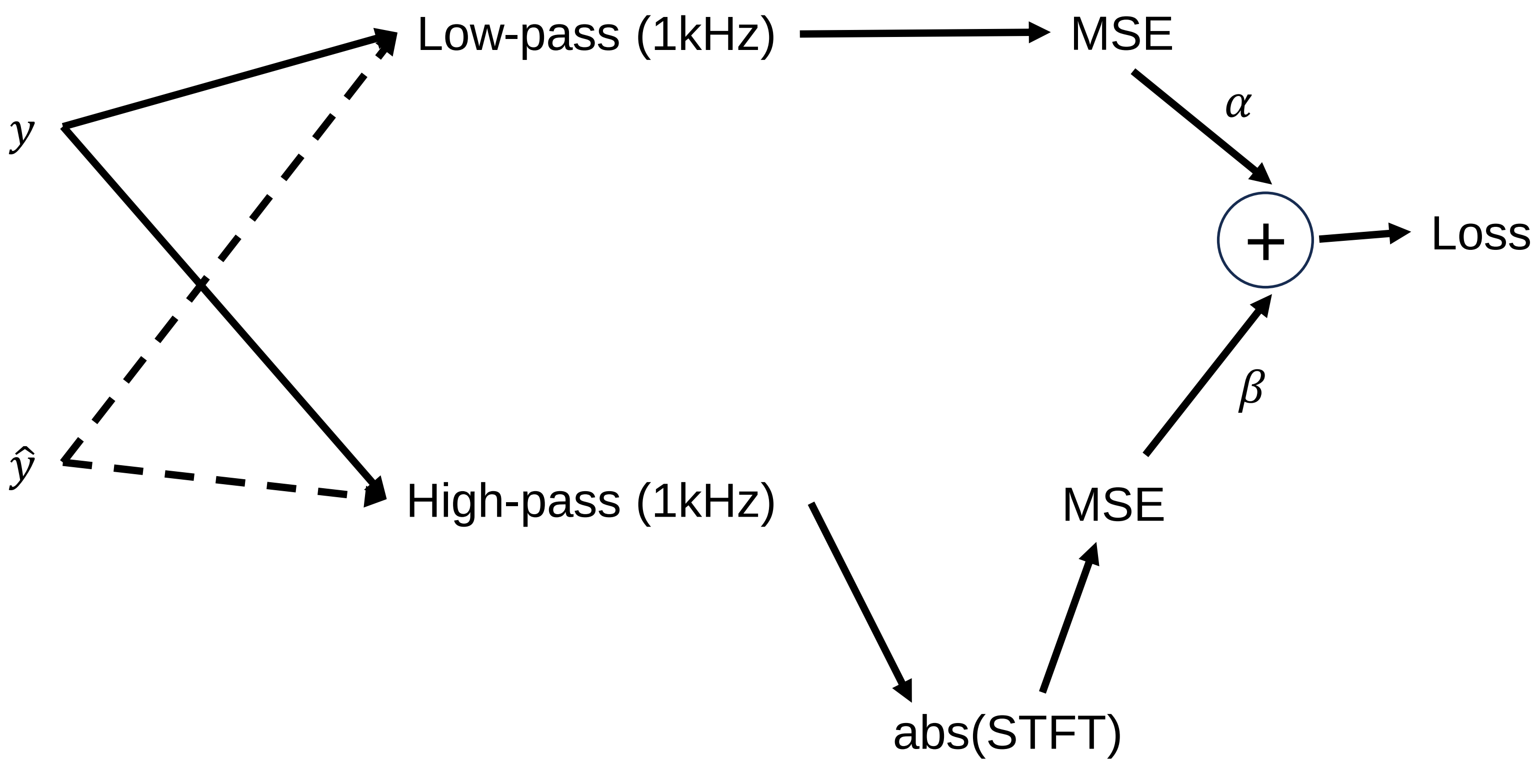
Match in envelope and variance

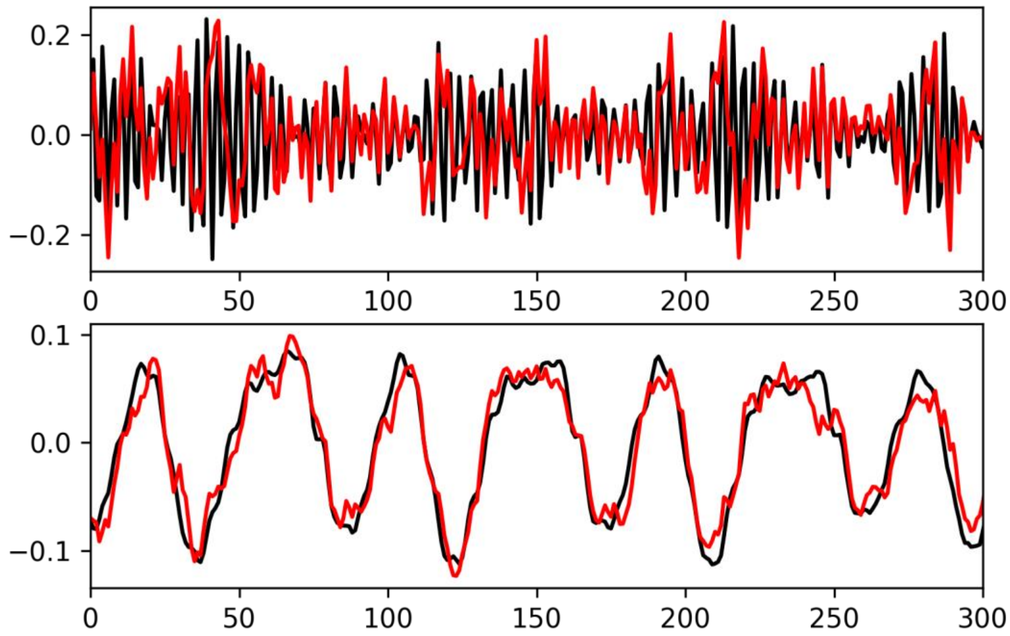


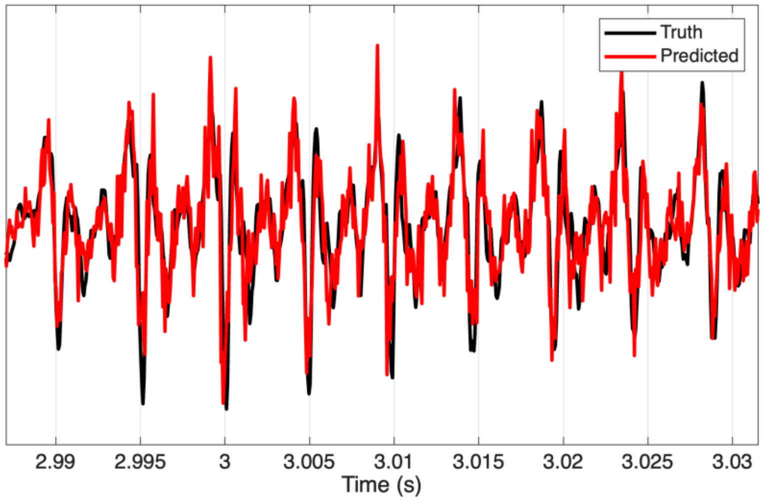
$$L_2 = \sqrt{\frac{1}{N}\sum(y_{env} - \widehat{y_{env}})^2}$$

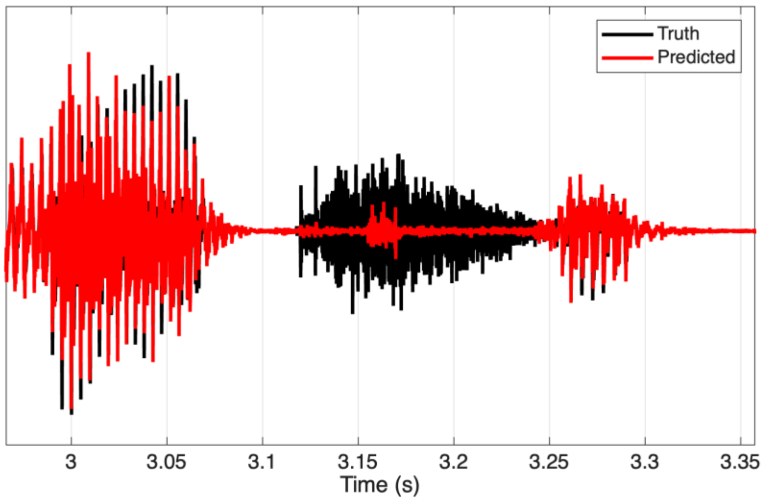
$$L_3 = \sqrt{\frac{1}{N}\sum(\sigma^2 - \widehat{\sigma^2})^2}$$

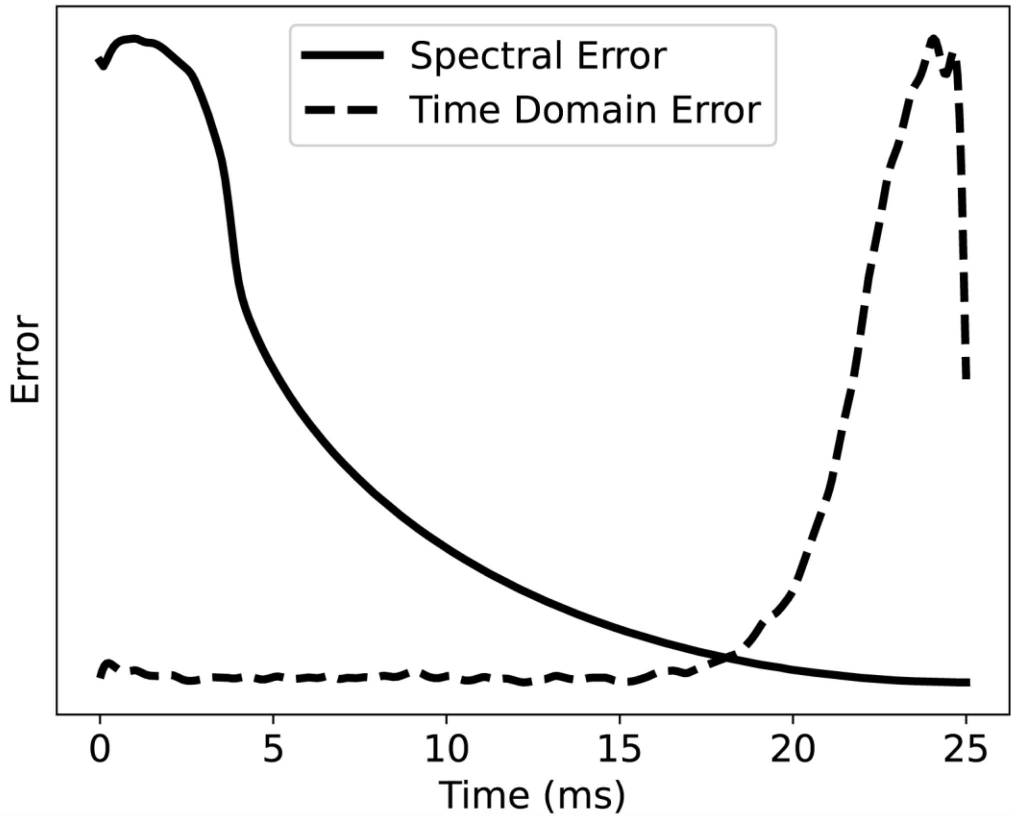
Final Loss Function



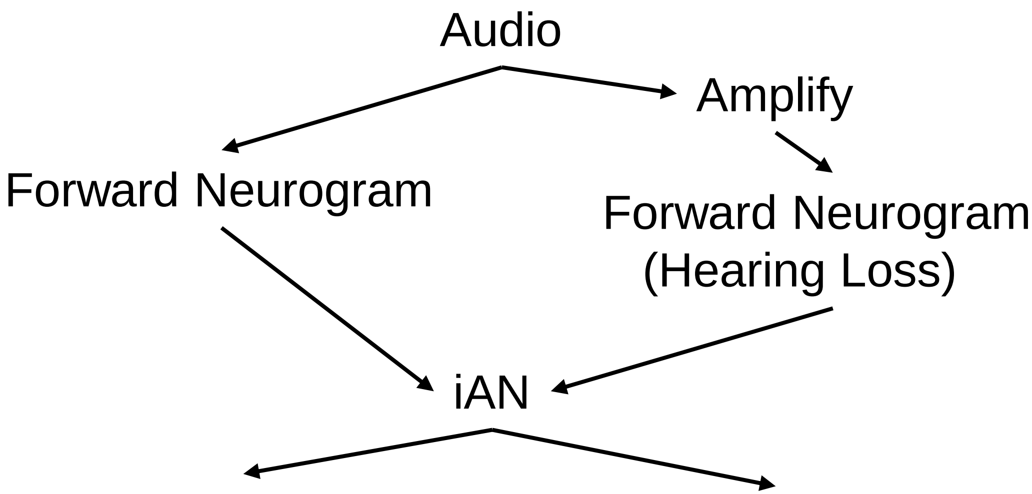

0.2
0.0
−0.2
0
50
100
150
200
250
300
0.1
0.0
−0.1
0
50
100
150
200
250
300

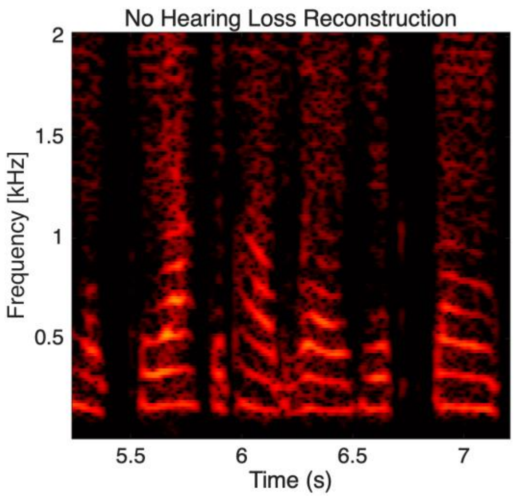

Truth
Predicted
2.99
2.995
3
3.005
3.01
3.015
3.02
3.025
3.03
Time (s)

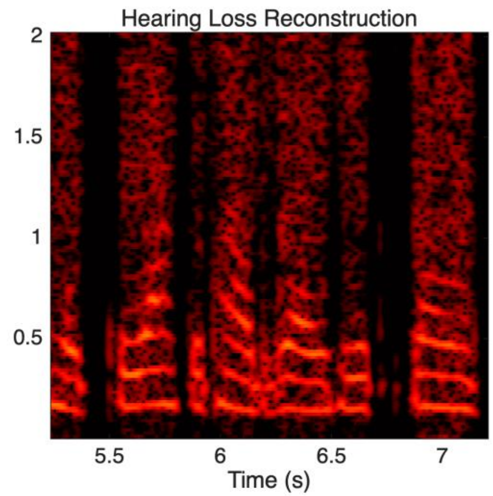

Truth
Predicted
3
3.05
3.1
3.15
3.2
3.25
3.3
3.35
Time (s)

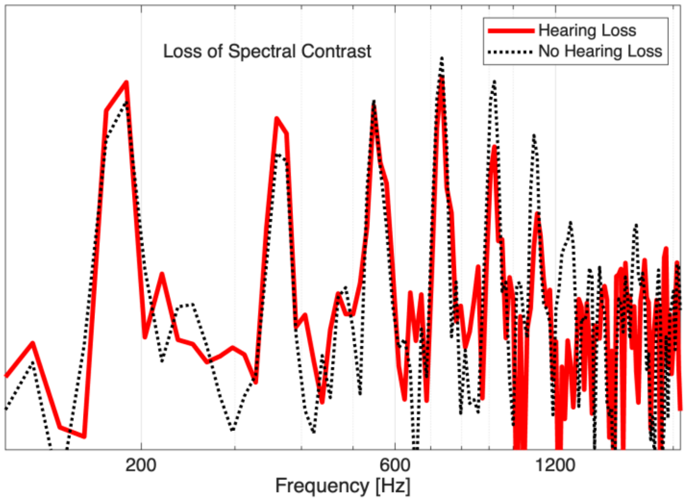


To use the middle most-sensitive region,
output should be windowed and hopped at 50%.


Audio
Amplify
Forward Neurogram
Forward Neurogram (Hearing Loss)
iAN


No Hearing Loss Reconstruction
Frequency [kHz]
2
1.5
1
0.5
5.5
6
6.5
7
Time (s)


Hearing Loss Reconstruction
2
1.5
1
0.5
5.5
6
6.5
7
Time (s)


Loss of Spectral Contrast
Hearing Loss
No Hearing Loss
200
600
1200
Frequency [Hz]

Two possibilities

1. The model degrades information more than what hearing loss would.

2. The model underestimates hearing loss.
   How do we think about this?

   Farhadi (2023) showed that the efferent feedback (from the brain to the ear) can affect the periphery responses (**it can increase spectral contrast**.).

Summary

- An inverse auditory model was created and trained with a custom loss function
- When evaluated on hearing loss neurograms, the model reduces spectral contrast.
- (Unavoidable?) obstacle to hearing hearing loss is the efferent feedback.

References

Souza P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. Trends in amplification, 6(4), 131–165.

Bruce, I. C., Erfani, Y., & Zilany, M. S. A. (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. Hearing research, 360, 40–54. https://doi.org/10.1016/j.heares.2017.12.016

Zilany, M. S., Bruce, I. C., Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America, 135(1), 283–286.

Farhadi, A., Jennings, S. G., Strickland, E. A., Carney, L. H. (2023). Subcortical auditory model including efferent dynamic gain control with inputs from cochlear nucleus and inferior colliculus. The Journal of the Acoustical Society of America, 154(6), 3644–3659.