Prediction of Behavioral Speech Intelligibility using a Computational Model of the Auditory System

Nursadul Mamun, Sabbir Hossain, and John H.L. Hansen

Cochlear Implant Laboratory
Center for Robust Speech Systems (CRSS)
Erik Jonsson School of Engineering & Computer Science
Department of Electrical Engineering
University of Texas at Dallas
Richardson, Texas 75083-0688, U.S.A.

ISCA INTERSPEECH 2023
Dublin, Ireland, August 20–24, 2023.

This work was supported by the grant R01DC016839 from the National Institutes of Health, National Institute on Deafness, and Other Communication Disorders

Motivations

- Listening tests are **expensive, time-consuming and impractical.**
- Objective intelligibility measures:
  - Acoustic signal based
  - Computational model based
- Can be successfully used in designing algorithm for hearing aids and cochlear implants.

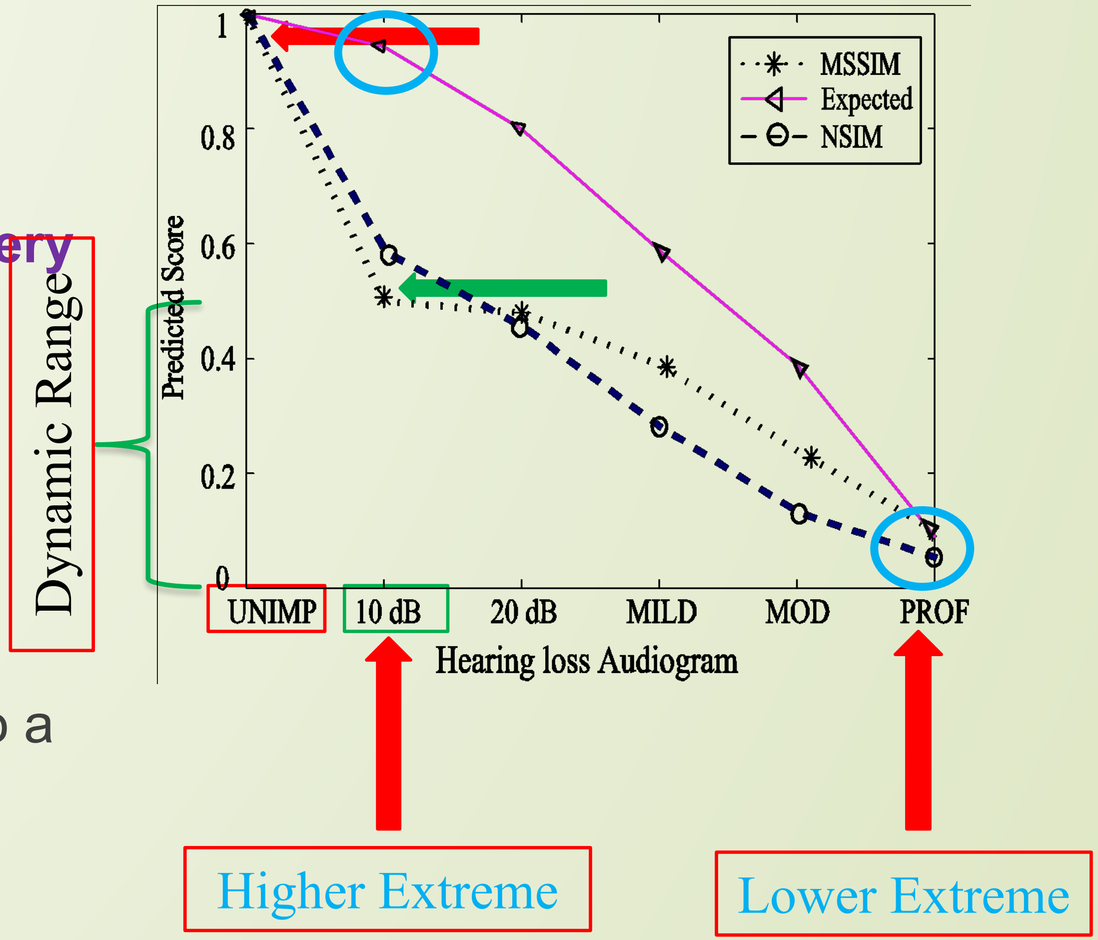
Outlines

- Background study
- Objectives
- Methodology
- Results
- Conclusions

Background Study

MSSIM & NSIM

- Function of luminance, contrast, and structure.
- Processing acoustic signal by **auditory-periphery** gives time-frequency spectrogram-like output (**neurogram**)



Limitations:

- Small **dynamic ranges**.
- A **sharp decrease** in the scores from unimpaired to a flat 10 dB hearing loss.

MSSIM: Mean Structural Similarity Index.
NSIM: Neurogram Similarity Index Measure

Objectives

- Main goal is **"To develop a speech intelligibility prediction metric using physiologically-based model of the auditory system".**

It includes

- To predict speech intelligibility for listeners **with and without hearing loss.**
- To predict speech intelligibility under both **quiet and noisy condition.**
- Compare the predicted intelligibility scores with the subjective scores.

Methodology

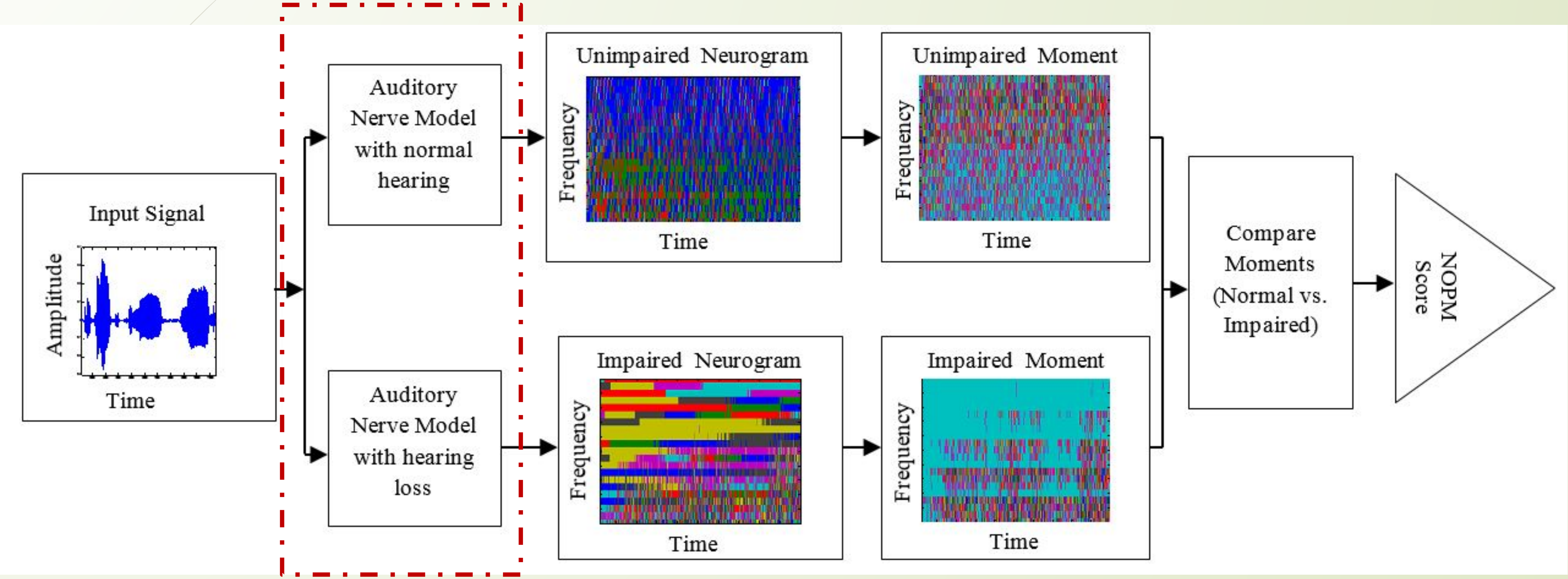
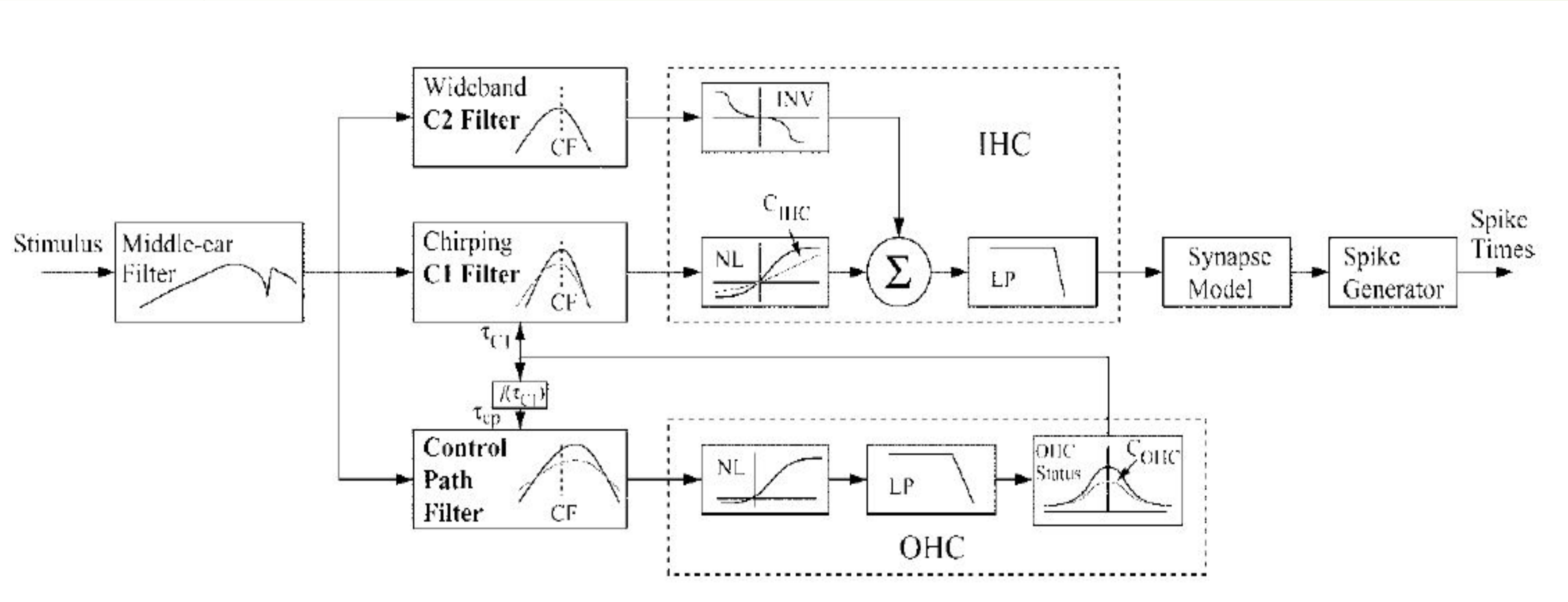


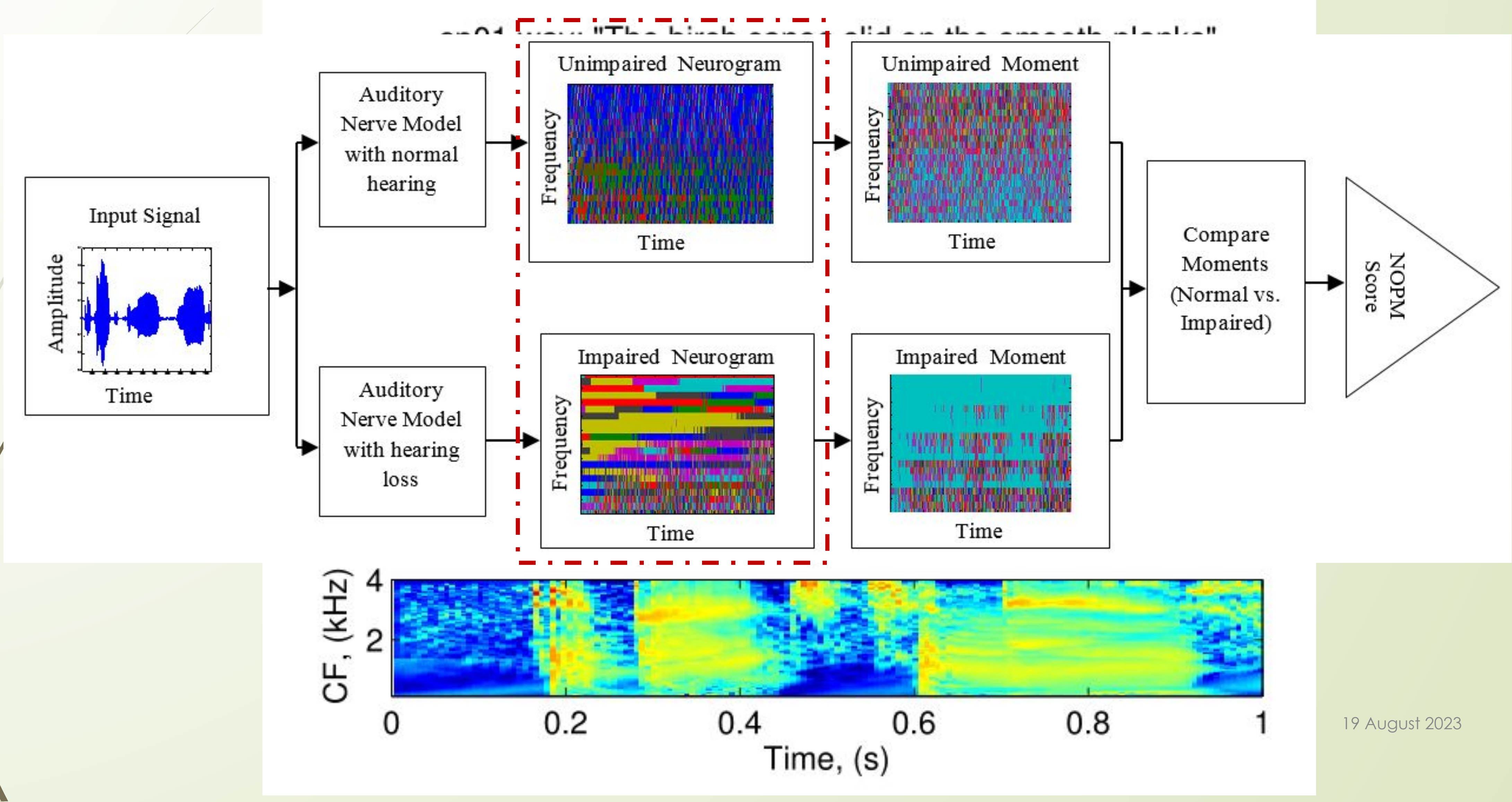
Figure: Simple block diagram of the proposed metric.

Model of the Auditory Periphery, 2014

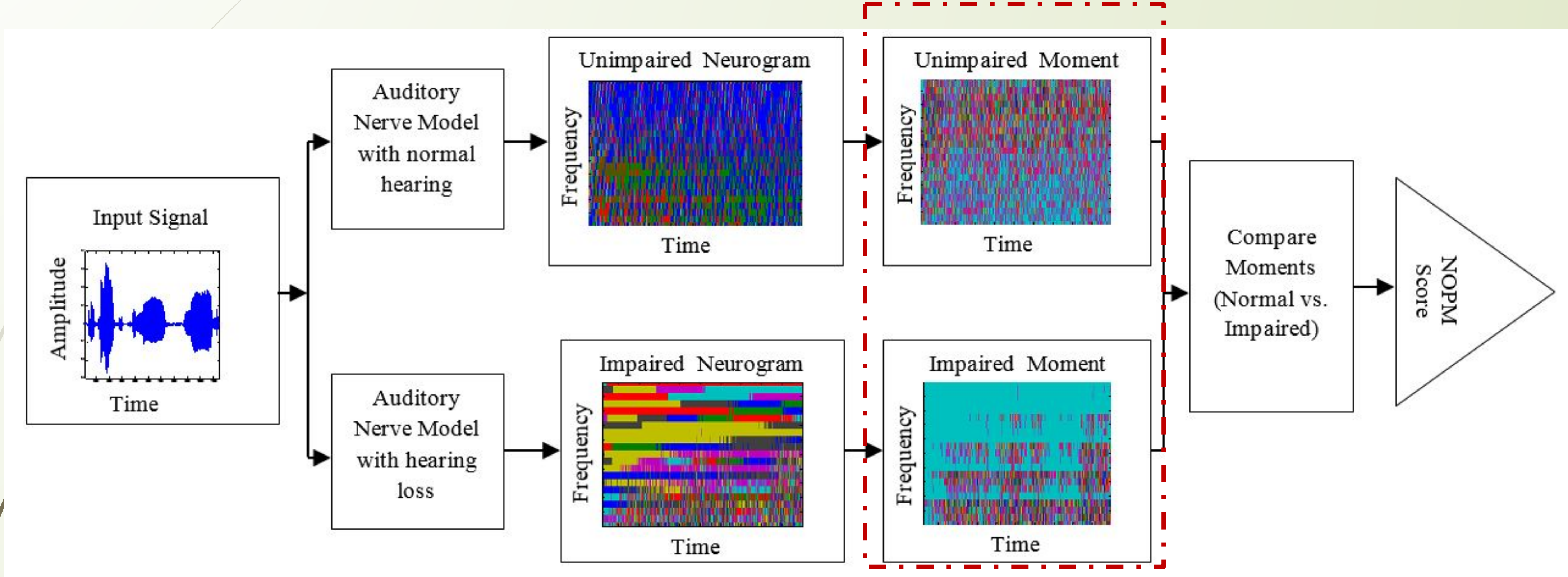


- Realistic temporal response properties and average discharge rates
- Model incorporates most of the nonlinearities of the cochlea
- Effects of acoustic trauma (impairment in OHC and IHC)

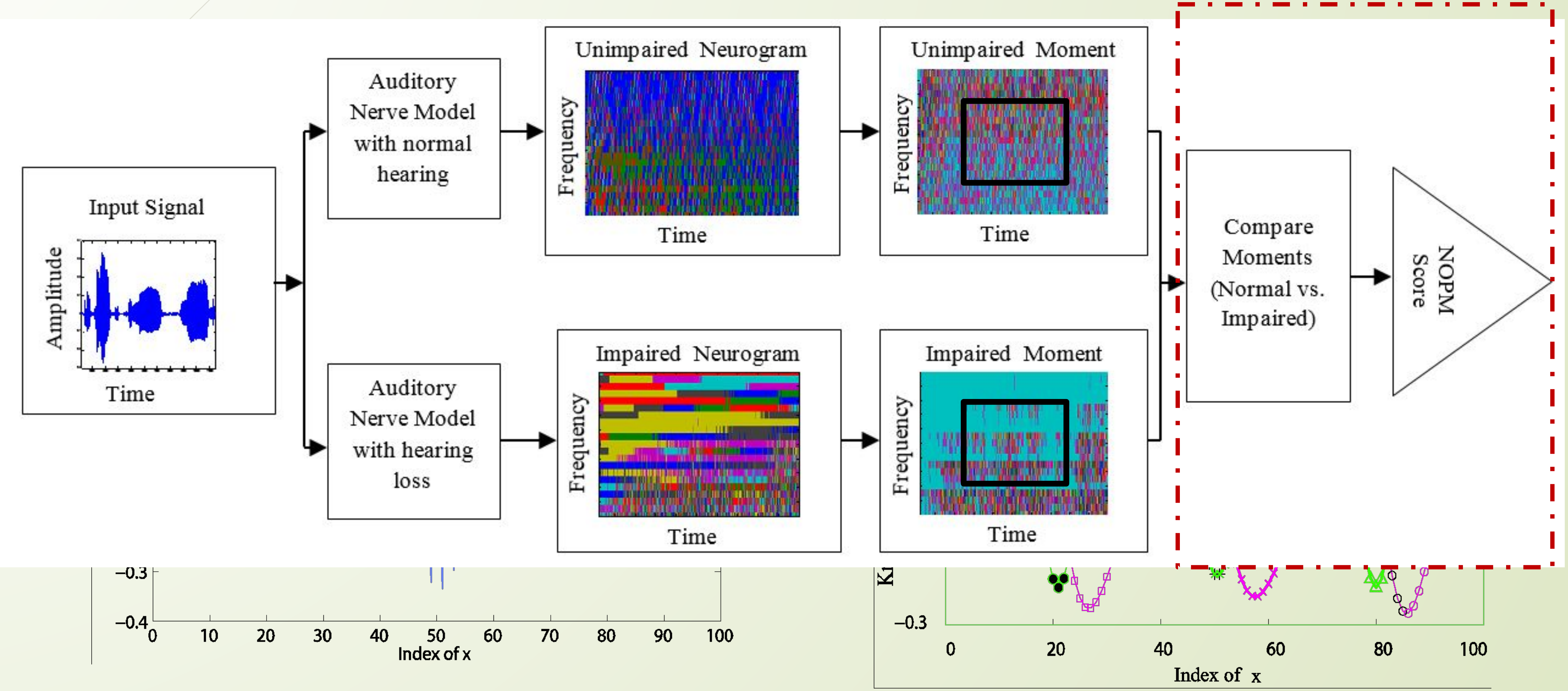
Neurogram


Input Signal
Amplitude
Time
Auditory Nerve Model with normal hearing
Auditory Nerve Model with hearing loss
Unimpaired Neurogram
Frequency
Time
Impaired Neurogram
Frequency
Time
Unimpaired Moment
Frequency
Time
Impaired Moment
Frequency
Time
Compare Moments (Normal vs. Impaired)
NOPM Score
CF, (kHz)
4
2
0
0.2
0.4
0.6
0.8
1
Time, (s)

Why Orthogonal Polynomials?



Orthogonal polynomial

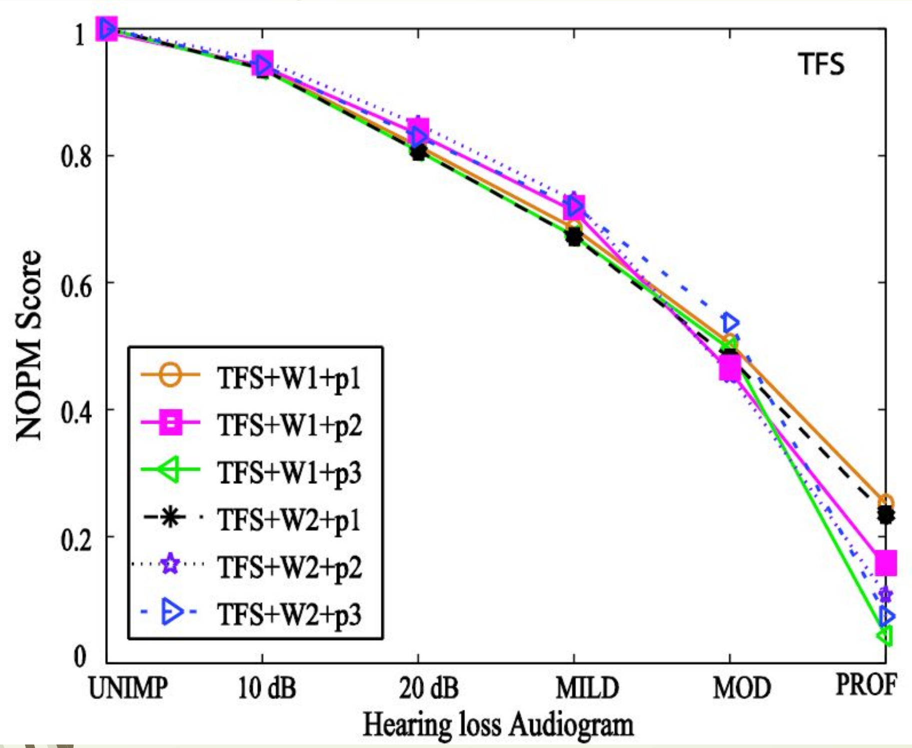


Results:

- The intelligibility scores
  - For the **normal listener**
  - For the **listener with hearing loss**.
- The effect on intelligibility scores due to
  - Signal to noise ratio (**SNR**)
  - Sound presentation level (**SPL**)
- Comparison between NOPM scores with
  - Subjective scores
  - Scores from existing metrics

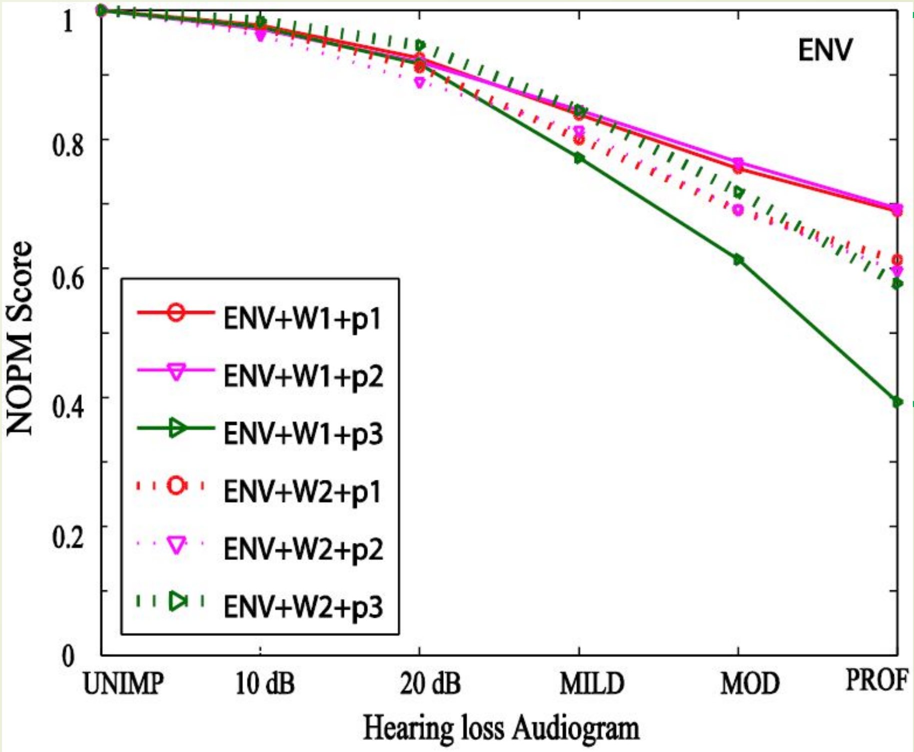
Results: Using TFS and ENV responses

TFS



Dynamic Range

ENV



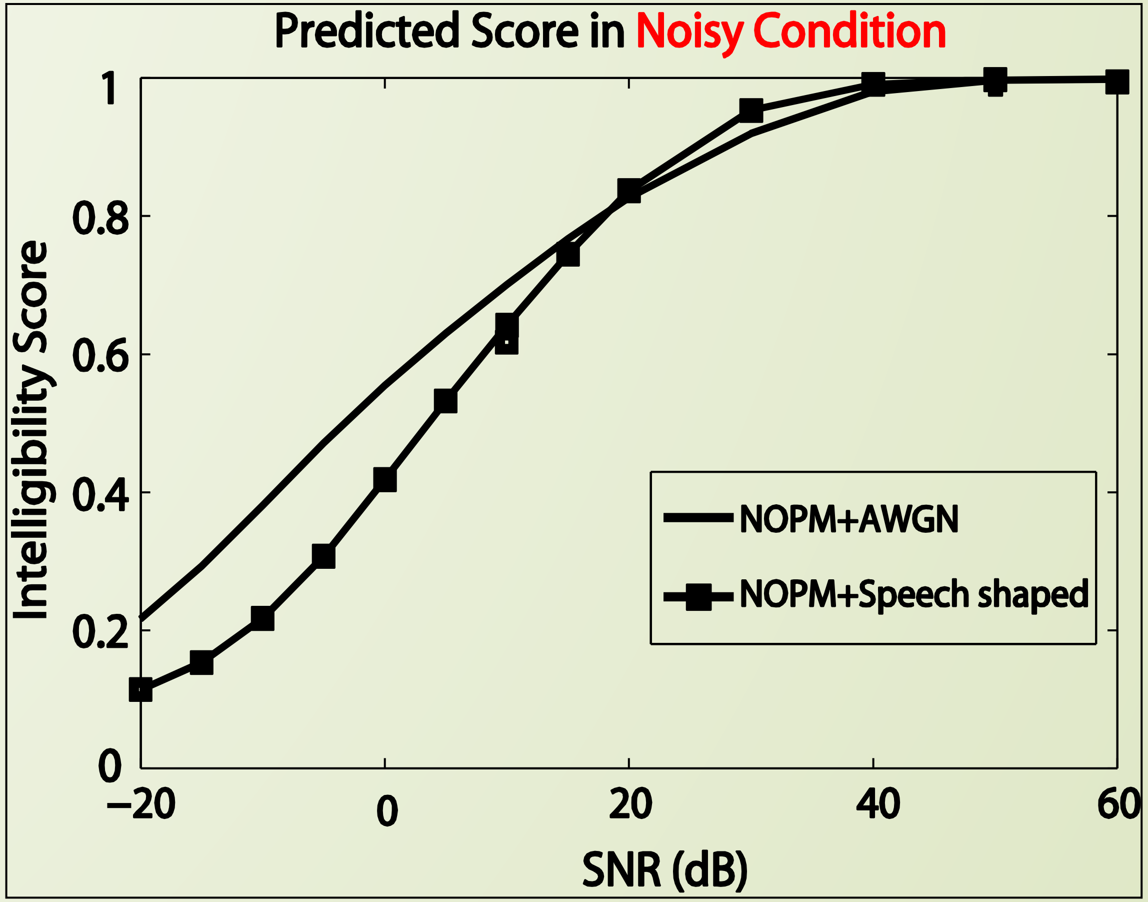
Dynamic Range

Results: Normal Hearing(NH)

- NOPM scores as a function of signal-to-noise ratio(SNR)

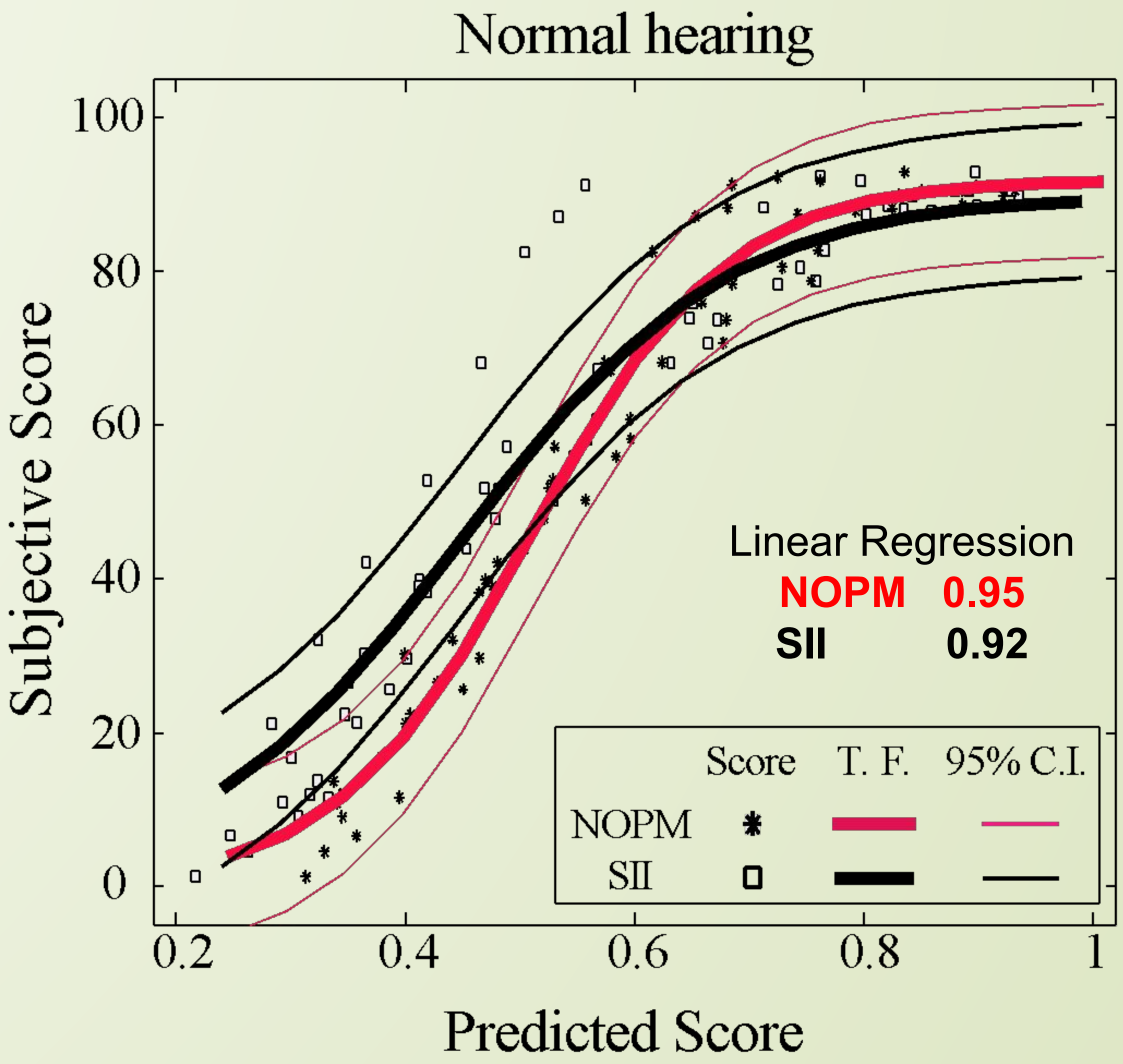
Two types of noise:

- Additive white Gaussian noise
- Speech-shaped Noise



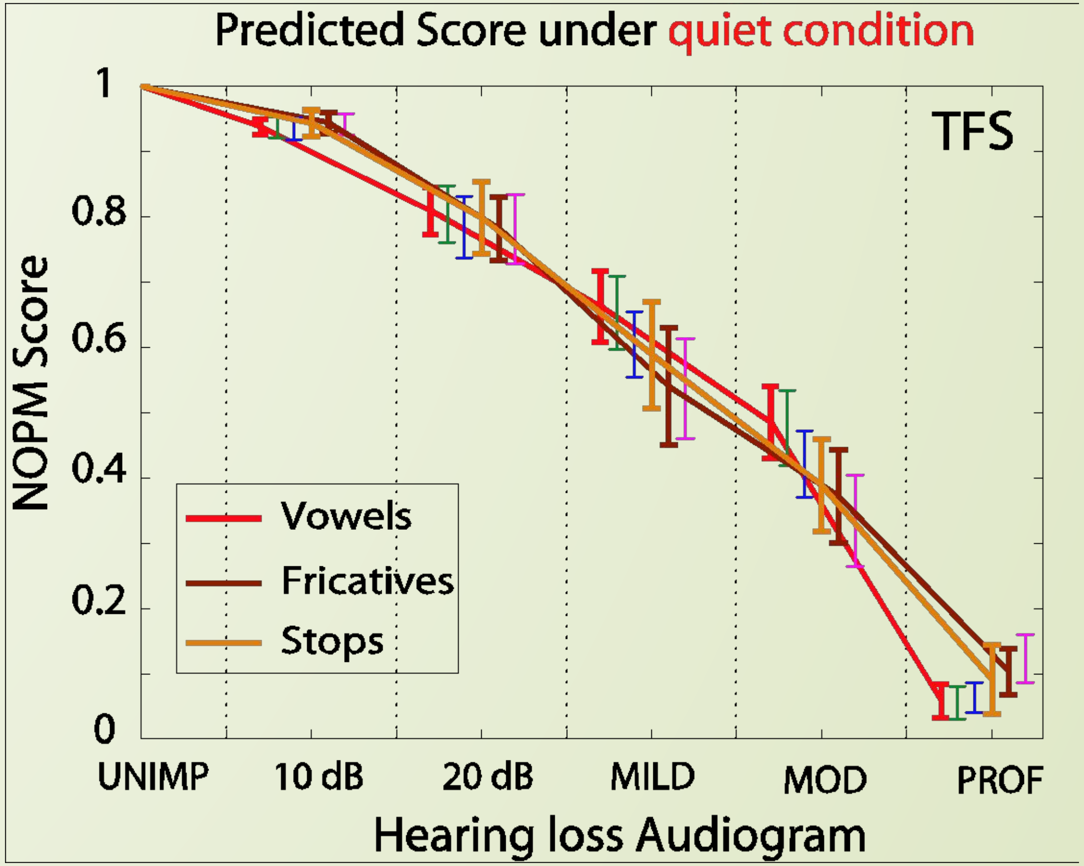
Results: NOPM vs. Subjective scores (NH)

- Subjective scores are from Studebaker et al. (1999)
- NU#6 words with speech-shaped noise
- Sound presentation level from **64 to 99 dB** SPL
- SNRs varied from **-4 to +28** dB in steps of 4 dB



Results: NOPM for Listeners with Hearing Loss (HL)

- NOPM scores for phonemes as a function of Hearing Loss

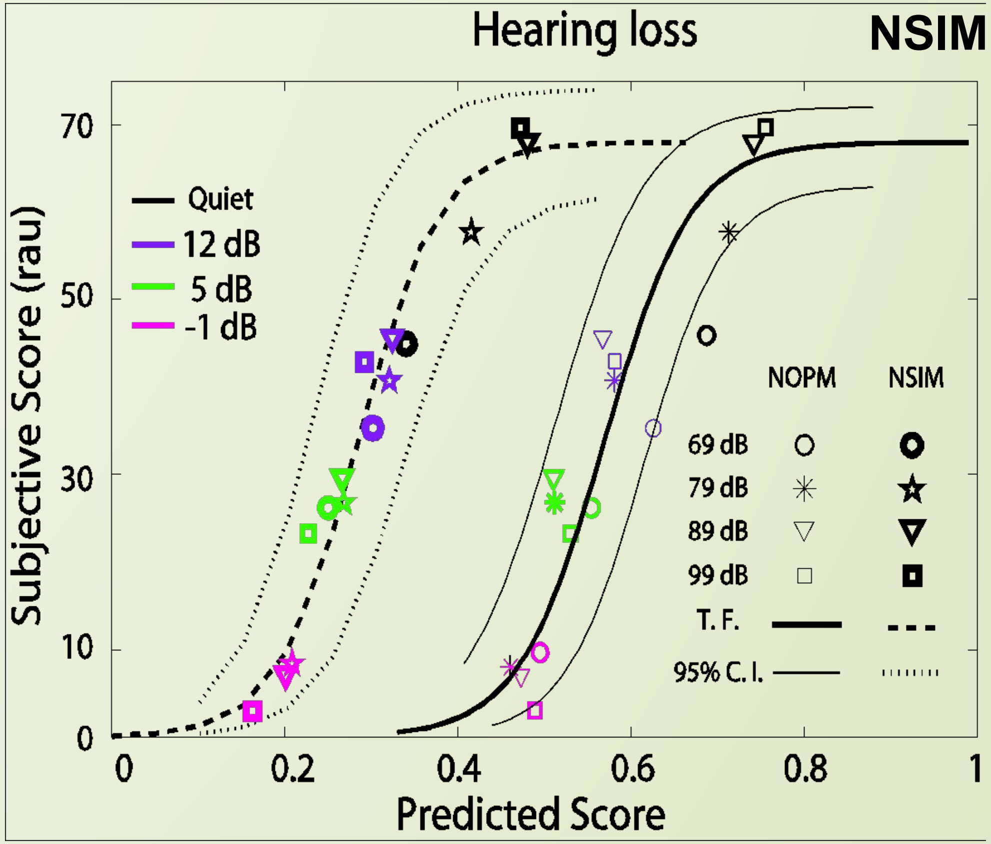


Results: NOPM vs. Subjective score (HL)

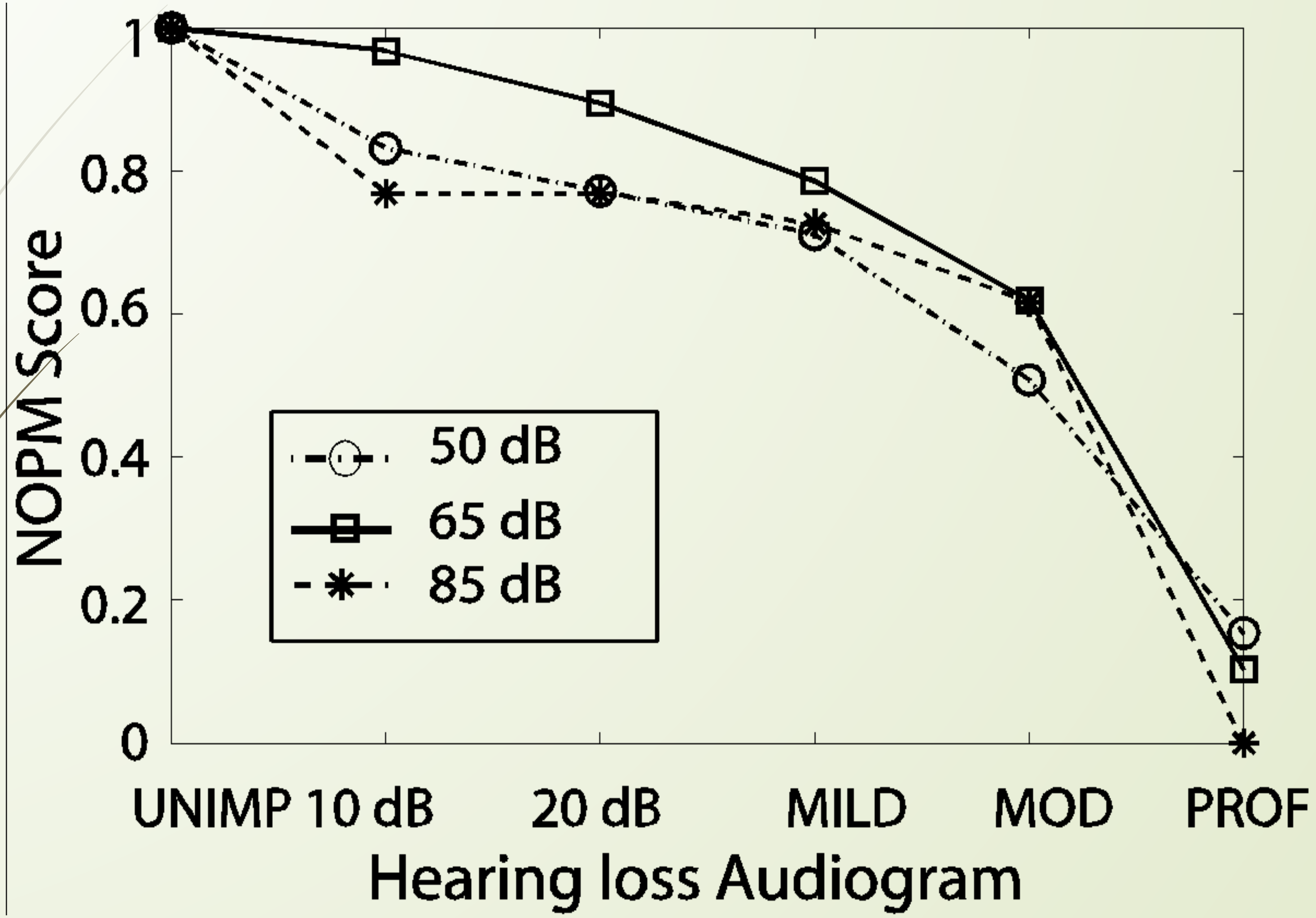
- Subjective scores are from Studebaker et al. (1999)
- Four sound levels and four SNRs
- Only the **moderately severe** hearing loss profile was considered

Linear Regression

NOPM	0.97
NSIM	0.93



Results: Effects of SPL



Results: Clarity Challenge

RMSE	**39.80**
Correlation	**0.329**

Conclusions

- Proposed metric **can predicts reliably** the subjective scores for NH and HL people.
- Solved the problems faced with the NSIM.
- The proposed metric
  - has a realistic and **wider dynamic range**.
  - scores are also **well-separated** as a function of hearing loss.
  - worked for different types of noise
- Scores using **TFS** neurogram had a wider dynamic range.

Thank you

Question and Answer

Related Publications

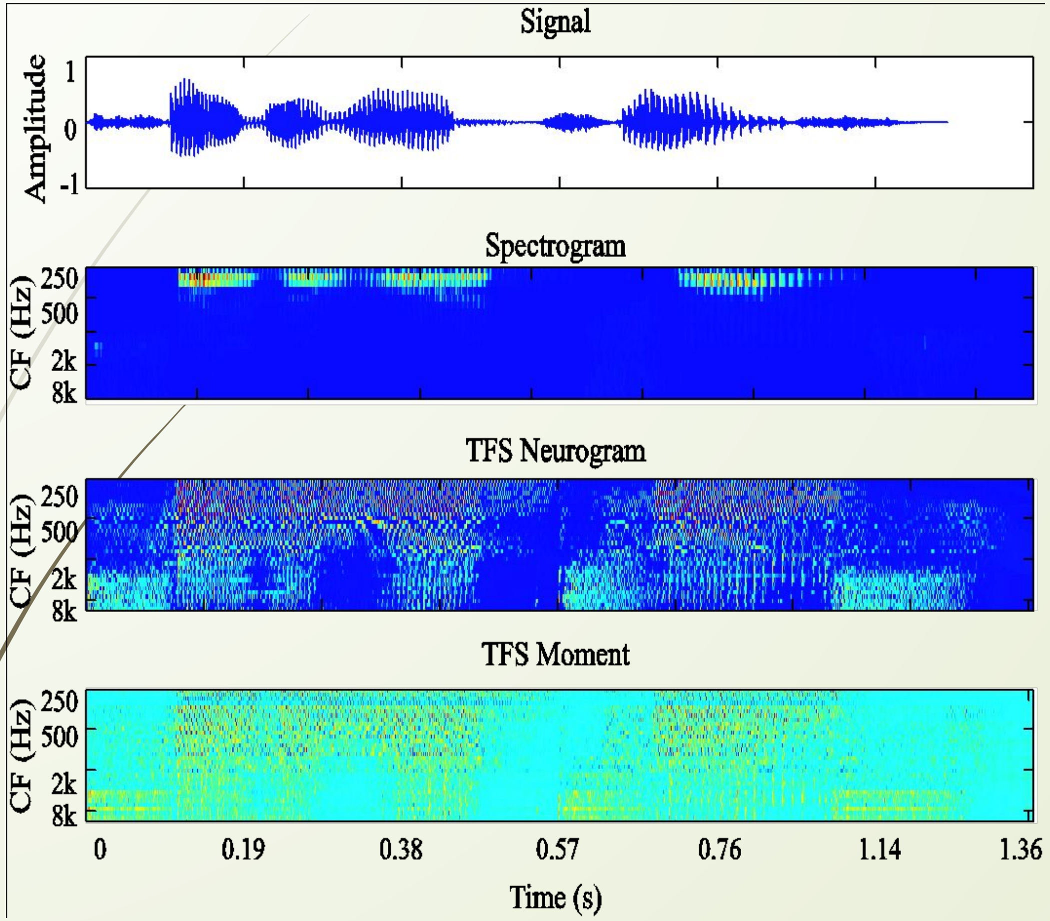
- **Journal paper**

  N. Mamun, W. Jassim, and M. S. Zilany, "Prediction of Speech Intelligibility Using a Neurogram Orthogonal Polynomial Measure (NOPM)," Audio, Speech, and Language Processing, IEEE/ACM Transactions on, vol. 23, pp. 760-773, 2015.

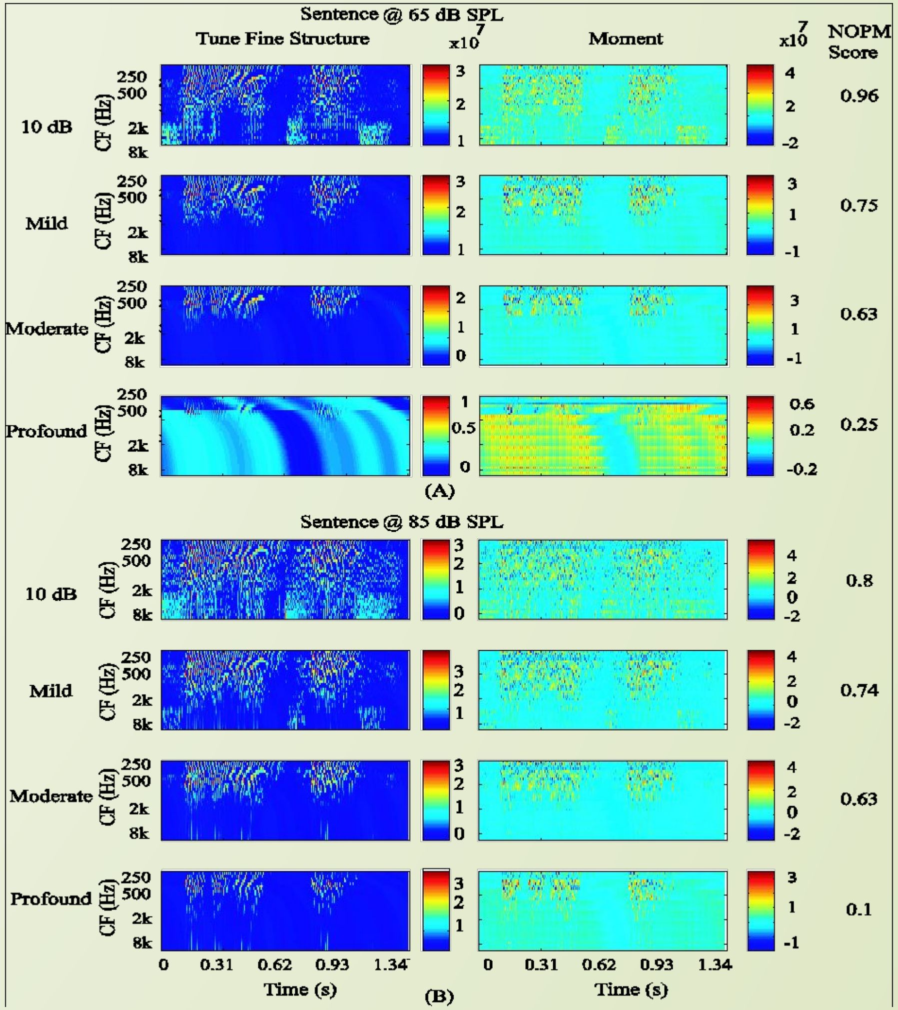
- **Conference papers**

  - N. Mamun, Wissam Jassim, and Muhammad S. Zilany. "Robust gender classification using neural responses from the model of the auditory system."Functional Electrical Stimulation Society Annual Conference (IFESS), 2014 IEEE 19th International. IEEE, 2014.
  - N. Mamun, W. A. Jassim, S. A. Zilany. "Speech-based Gender Classification Using Neurogram Orthogonal Polynomial measure" Net Regional Conference on Electrical Engineering (RCEE), 04 Mar 2014 to 05 Mar 2014, University Malaya and JICA Project for AUN/Speed-Net, (National)

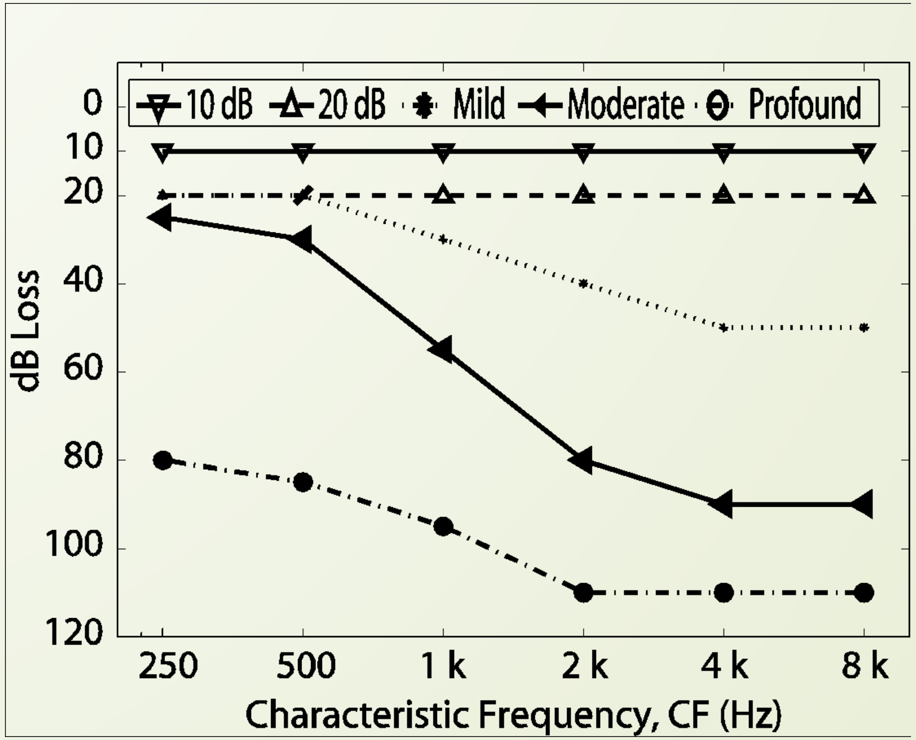
- Figure 4.1

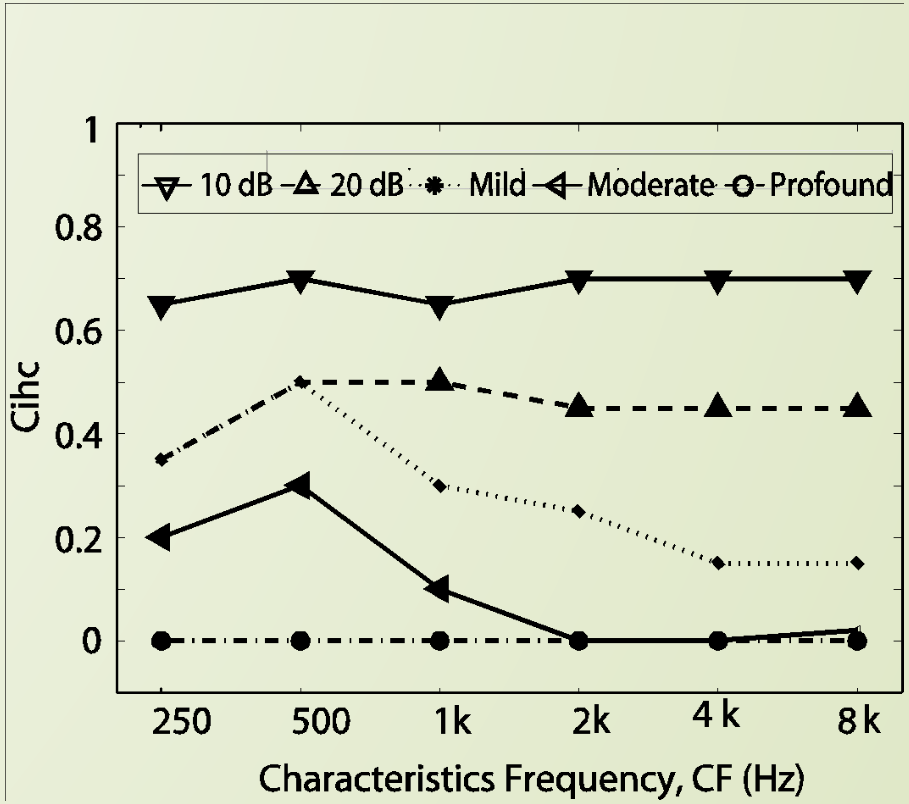


- Figure 4.2

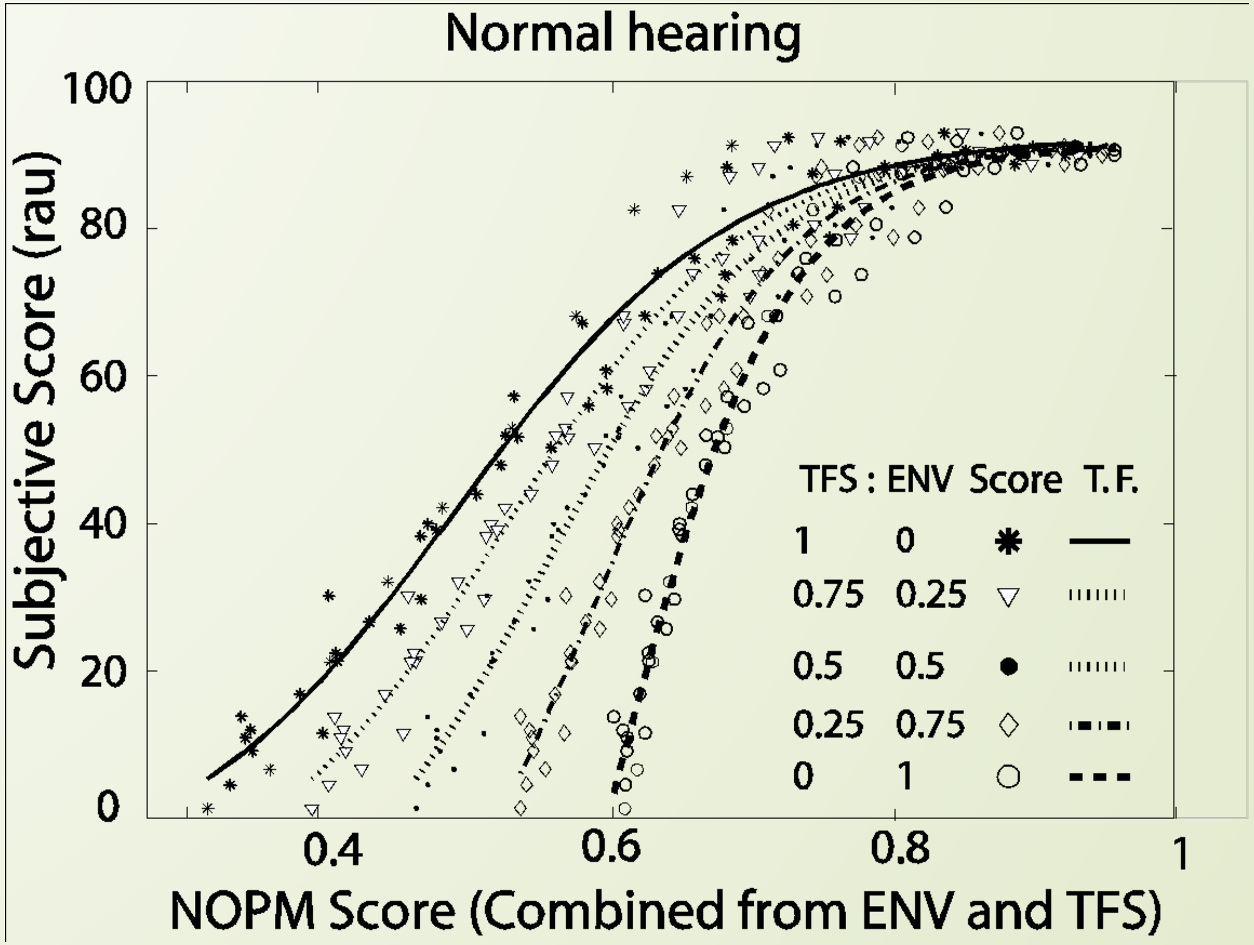


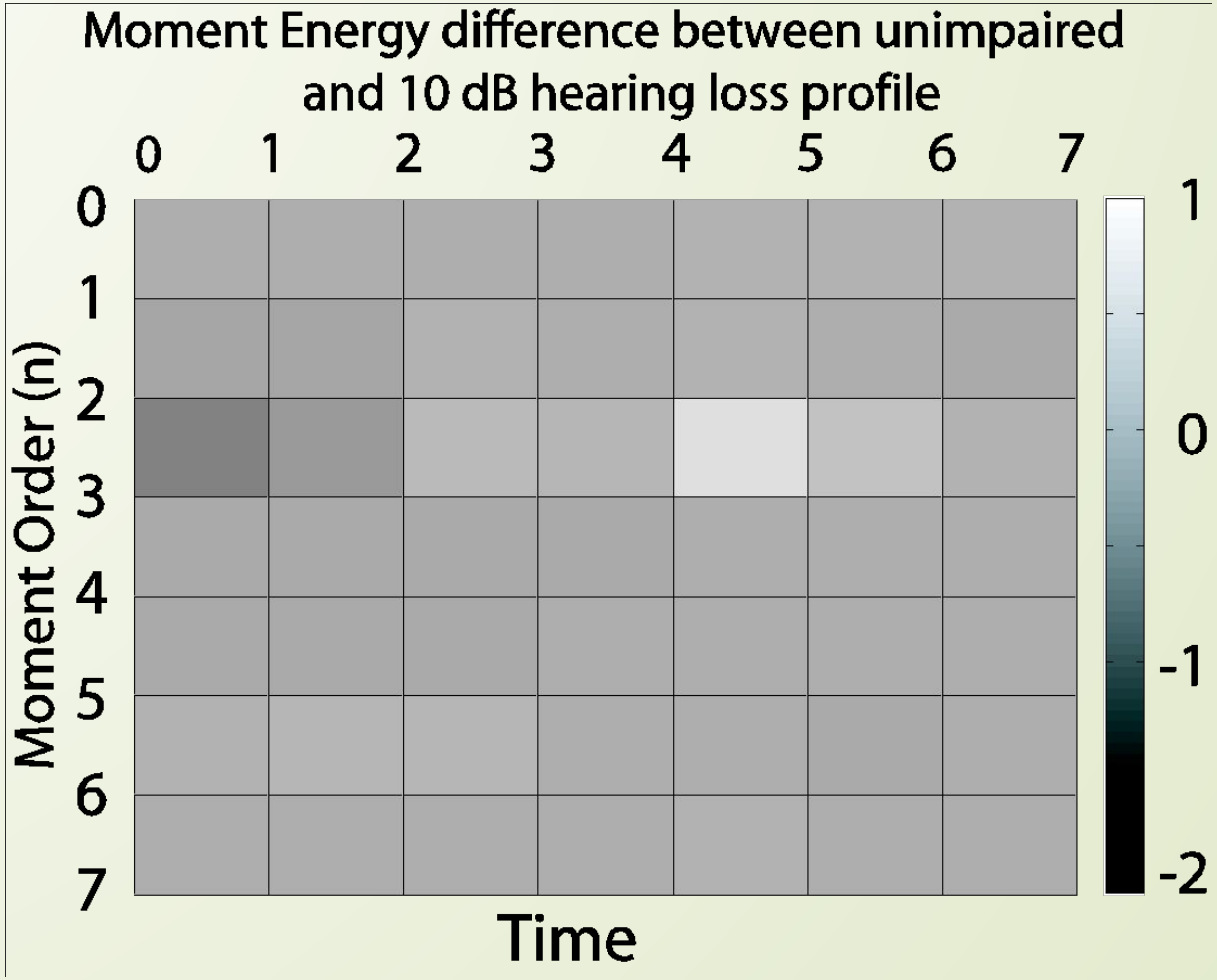
Hearing loss profile





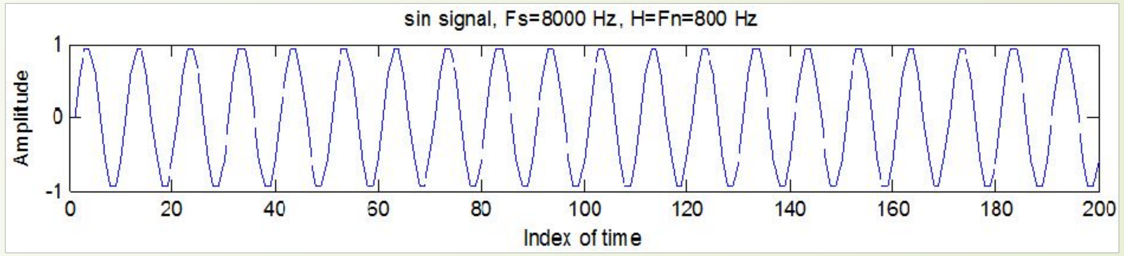
Different combinations of TFS and ENV neurogram



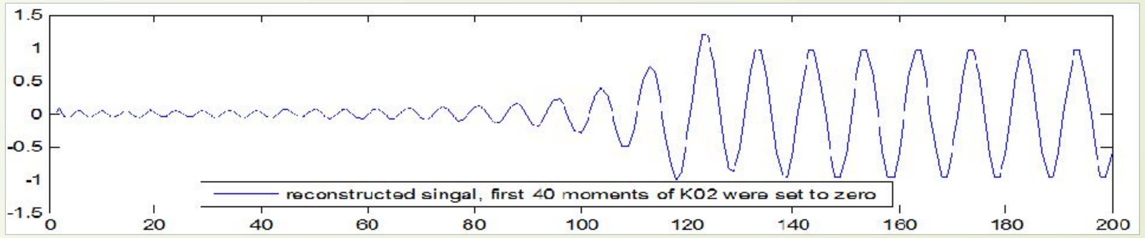


Moment Energy difference between unimpaired and 10 dB hearing loss profile
0
1
2
3
4
5
6
7
Moment Order (n)
Time
1
0
-1
-2

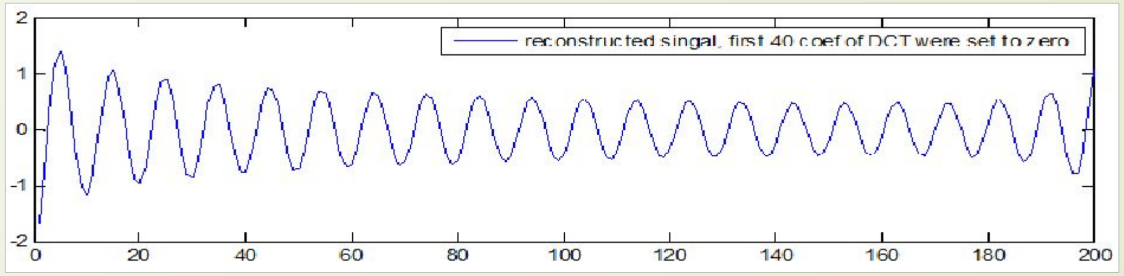
(A)


sin signal, Fs=8000 Hz, H=Fn=800 Hz
Amplitude
Index of time

(B)


reconstructed singal, first 40 moments of K02 were set to zero

(C)


reconstructed singal, first 40 coef of DCT were set to zero