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

Nursadul Mamun[1], *Sabbir Ahmed*[2], *John H.L. Hansen*[1]

[1]Cochlear Implant Processing Laboratory, Center for Robust Speech Systems (CRSS-CILab),
Department of Electrical & Computer Engineering, The University of Texas at Dallas
[2] Department of ETE, Chittagong University of Engineering and Technology, Bangladesh
nursadul.mamun@utdallas.edu

Abstract

This paper introduces a speech intelligibility prediction metric, the neurogram orthogonal polynomial measure (NOPM), to address the growing concern of Sensorineural Hearing Loss (SNHL). The proposed NOPM metric utilizes orthogonal moments applied to the auditory neurogram to predict speech intelligibility for both listeners with normal hearing and those experiencing hearing loss. To achieve this, the model simulates the responses of auditory-nerve fibers to speech signals, considering both quiet and noisy conditions. Neurograms are generated using a physiologically based computational model of the auditory periphery. Applying the well-known orthogonal polynomial measure, Krawtchouk moments facilitates the extraction of relevant features from the auditory neurogram. To validate the metric's accuracy, the predicted intelligibility scores were compared with subjective results. Encouragingly, NOPM demonstrated a strong correlation with the subjective scores for both normal listeners and individuals with hearing loss.

Index Terms: Speech intelligibility, Orthogonal moment, Auditory-nerve model, Neurogram, Sensorineural hearing loss.

1. Introduction

Assessing speech quality and intelligibility through traditional listening tests (behavioral studies) is a resource-intensive undertaking, demanding substantial expenses, time, and complex logistics due to its reliance on subject feedback and controlled laboratory conditions. Nevertheless, recent advancements have introduced promising alternatives for objective evaluation.

One such approach involves predicting listeners' performance objectively by mathematically comparing features extracted from the original (clean) and processed (degraded) speech, like the articulation index. This method provides valuable insights into speech quality and intelligibility without necessitating direct human feedback.

Another viable option entails estimating subjective scores by employing a reliable model of the auditory system, such as the neurogram similarity index measure (NSIM) [1]. Utilizing the NSIM circumvents the need for human listeners, enhancing efficiency while maintaining a high level of assessment accuracy.

Building upon these advancements, this study introduces a novel objective metric for evaluating speech quality and intelligibility: the neurogram orthogonal polynomial measure (NOPM)[2]. Leveraging orthogonal polynomials applied to the auditory neurogram, the NOPM enables the extraction of specific features known as orthogonal moments—proven successful in image quality assessment and now adapted for speech analysis.

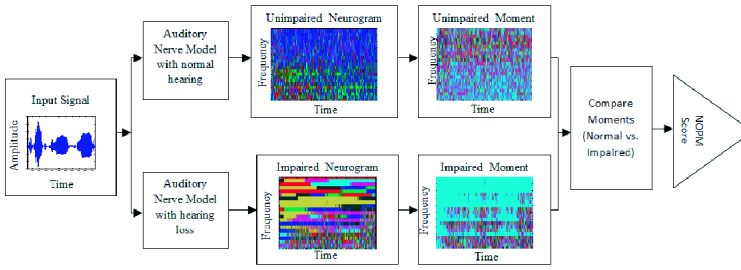

Figure 1: *Block diagram summarizing the methodology of NOPM*

By incorporating the NOPM, this study aims to streamline the evaluation process, making it more accessible, cost-effective, and precise. The proposed metric holds significant promise for further enhancing our understanding of speech perception and optimizing communication systems.

In conclusion, the integration of advanced mathematical approaches, such as the NOPM, and reliable auditory system models like NSIM, offers a compelling avenue for objective speech quality and intelligibility assessment. These developments pave the way for a more efficient and accurate evaluation of speech processing technologies and facilitate the improvement of speech communication systems.

2. Methods

The following sections briefly describe the computation of the proposed metric, the NOPM. In addition, existing computational model-based metrics, as well as the components of the proposed metric, will be briefly explained.

Figure 1 illustrates the procedure of the neurogram orthogonal polynomial measure (NOPM) for listeners with hearing loss. Initially, the speech stimulus is fed into a computational model of the auditory periphery. Neurograms are generated for both normal and impaired auditory systems. Next, the neurogram features are extracted using the orthogonal moment transform. Subsequently, the computed moments for impaired and unimpaired conditions are compared to derive an NOPM score, encompassing a range of SNHLs.

Conversely, for predicting speech intelligibility scores under noisy conditions, both the clean speech and its corresponding noisy signals are applied to the auditory system model. This process results in the construction of neurograms for both the clean (normal) and distorted (equivalent to impaired) conditions.

2.1. Neurogram Orthogonal Polynomial Measure (NOPM)

The neurogram orthogonal polynomial measure is an objective intelligibility measurement metric designed to accurately assess changes in signal information, including small variations in magnitude (pixel intensity) and phase (location) of the neurogram. Structural information plays a crucial role in determining the quality of the neurogram, with the local phase (discharge timing) containing more vital structural details than magnitude. Changes in the neurogram structure can result in shifts in phase, making phase information essential for precise distortion prediction. Moreover, the location of distortion within a signal is equally significant as its magnitude. Discrete orthogonal moments serve as effective signal descriptors, representing image information with minimal redundancy and efficiently capturing even subtle differences in pixel intensity. As a result, the computed moment values exhibit variations in response to small changes in pixel intensities, enabling a robust assessment of signal distortion.

Noise-induced impairment in the cochlea causes damage to both the IHC and OHC stereocilia [3]. Damage to the OHC stereocilia causes both elevated threshold and broadened tuning of AN-fibers, whereas IHC stereocilia damage results only in the elevation of the tuning curve without any substantial effect on the bandwidth. The effects of the OHC and IHC status are incorporated in the model by introducing a scaling factor COHC ($0 \sim 1$) to the control path output and CIHC ($0 \sim 1$) to the IHC transduction function, respectively. COHC = 1 simulates normal functioning and COHC = 0 indicates complete impairment in the OHC. Similarly, the normal functioning of the IHC is represented by CIHC = 1, whereas CIHC = 0 corresponds to complete impairment in the IHC. These two scaling factors successfully capture the phenomena reported for damage to the OHC and IHC stereocilia.

2.1.1. Auditory-Nerve Model

The AN model utilized in this study accurately simulates the responses of the cochlea, inner hair cells (IHCs), outer hair cells (OHCs), and the IHC-AN synapse, up to the responses of the auditory nerve (AN) fiber. The model effectively captures various nonlinearities, including compression, two-tone rate suppression, frequency selectivity, level-dependent rate, phase responses, and the shift in the best frequency observed at higher levels within the AN. These model responses have been extensively validated against a broad range of physiological recordings from AN fibers using both simple (tone-like) and complex (speech-like) stimuli. For this study, the AN model introduced by Zilany and colleagues is employed to predict human speech-recognition performance, and its schematic diagram can be found in Figure 1 of Zilany et al.'s work [3].

2.1.2. Neurograaml

In this study, the neurogram is like to a spectrogram, but it differs in that neural responses are simulated for a range of characteristic frequencies (CFs) instead of analyzing the acoustic waveform. Neurograms were constructed by simulating the responses of 32 auditory nerve (AN) fibers, logarithmically spaced from 250 to 8000 Hz, to phonemes and words from the databases. The AN model introduced by Zilany and colleagues [3] was used to predict human speech recognition performance. To align with physiological characteristics, responses of three types of AN fibers (high, medium, and low spontaneous rates) were simulated and weighted based on the distribution of spontaneous rates (high = 0.6, medium = 0.2, and low = 0.2 of the total population).

To simulate responses for hearing-impaired AN fibers, the model parameters for the inner hair cell (CIHC) and outer hair cell (COHC) were adjusted from 1 to 0, corresponding to the degree of hearing loss. For normal hearing, both parameters were set to 1, while complete impairment in the IHC and OHC was represented by setting both parameters to 0.

2.1.3. Orthogonal moments

Orthogonal moments, employing orthogonal polynomials as basis functions, offer superior feature representation capability compared to other real transforms like the discrete cosine transform and exhibit enhanced robustness to noise [4]. These moments effectively transform 1-D or 2-D signals from time or spatial domains into the moment domain, providing more compact representations for signals. Within the moment domain, lower-order moments capture low-frequency components of a signal, while higher-order moments represent their high-frequency components. Typically, a significant portion of the signal energy in the moment domain is concentrated in the lower-order moments, while most of the noise energy resides in the higher-order moments. However, in the presence of colored noise, both lower- and higher-order moments contribute to describing the noise characteristics.

3. Conclusions

The proposed metric showed a reasonably good quality of linear fitting between the predicted and subjective scores for all cases. The neural response-based proposed metric also showed a realistic and wider dynamic range compared to NSIM based on the responses from a model of the auditory periphery.

4. References

[1] A. Hines and N. Harte, "Speech intelligibility prediction using a neurogram similarity index measure," *Speech Communication*, vol. 54, no. 2, pp. 306–320, 2012.

[2] N. Mamun, W. A. Jassim, and M. S. Zilany, "Prediction of speech intelligibility using a neurogram orthogonal polynomial measure (nopm)," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 760–773, 2015.

[3] M. S. Zilany and I. C. Bruce, "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1446–1466, 2006.

[4] C.-Y. Wee, R. Paramesran, R. Mukundan, and X. Jiang, "Image quality assessment by discrete orthogonal moments," *Pattern Recognition*, vol. 43, no. 12, pp. 4055–4068, 2010.