



Using Neurogram Similarity Index Measure (NSIM) to Model Hearing Loss and Cochlear Neural Degeneration

Ahsan Cheema^{1,2,3}, Sunil Puria^{1,2,3}

¹Speech and Hearing Bioscience and Technology Program, Harvard University, Cambridge, USA

²Eaton Peabody Laboratories, Massachusetts Eye and Ear Infirmary (MEEI), Boston, USA

³Department of Otolaryngology, Harvard Medical School, Boston, USA

ahsancheema@g.harvard.edu, sunil.puria@meei.harvard.edu

Abstract

Trouble hearing in noisy situations remains a common complaint for both individuals with hearing loss and individuals with normal hearing. This is hypothesized to arise due to condition called: cochlear neural degeneration (CND) which can also result in significant variabilities in hearing aids outcomes. This paper uses computational models of auditory periphery to simulate various hearing tasks. We present an objective method to quantify hearing loss and CND by comparing auditory nerve fiber responses using a Neurogram Similarity Index Measure (NSIM). Specifically study 1, shows that NSIM can be used to map performance of individuals with hearing loss on phoneme recognition task with reasonable accuracy. In the study 2, we show that NSIM is a sensitive measure that can also be used to capture the deficits resulting from CND and can be a candidate for noninvasive biomarker of auditory synaptopathy.

Index Terms: neurogram similarity index measure, cochlear neural degeneration, synaptopathy, hearing aids, image-similarity

1. Introduction

Difficulty understanding speech in the presence of background noise is one of the most common complaints of patients with sensorineural hearing loss (SNHL) [1]. Extensive research has shown that the death of cochlear hair cells leads to hearing loss, but it is often preceded by a loss of the synapses linking the hair cells to the auditory-nerve fibers (cochlear neural degeneration). Cochlear neural degeneration (CND) can contribute to difficulty hearing in noise and goes undetected during standard audiometric evaluations [2]. Hearing aids (HAs) remain the standard treatment for hearing loss, but they primarily address hearing loss due to death of cochlear cells, leaving neural deficits untreated. This limits the effectiveness of hearing aids and can result in variability in patient outcomes. Therefore, development of an objective measure to study the effects of hearing loss and CND remains an essential challenge which can help improve the design and gain compensation strategies for hearing aids.

One of the promising approaches to arrive at an objective measures for hearing loss and CND is to simulate the response of auditory nerve fibers (ANF) from a normal cochlea and a cochlea with hearing loss using physiologically inspired auditory nerve model (ANM) of cochlea [3] and compare differences between the discharge patterns for the two cases. In order to study the differences in discharge patterns of ANF Hines and Harte [4, 5] developed a "Neurogram Similarity Index Measure (NSIM)" which is similar to Structural Similarity Index Measure (SSIM) used widely in the image processing literature [6]. This measure views the time-frequency discharge patterns from auditory

nerve fibers as an image pattern called a 'neurogram' and can be used to quantify the differences between ANFs from a normal hearing cochlea with no hearing loss and a cochlea with hearing loss. Previous studies have established NSIM as a quantitative measure to simulate performance of normal hearing individuals on /CVC/ phoneme recognition tasks for a range input stimulus levels [5]. Mamun et. al. [7] improved the dynamic range for neurogram based measures using Neurogram orthogonal polynomial to map performance of normal hearing individuals, but this study only included 1 hearing loss individual and thus was not generalizable to study the effect of hearing loss. No previous study has comprehensively (on a large enough datasets of individuals with hearing loss) explored the utility of neurogram based objective measures to study the effects of hearing loss on ANF responses and link that to the performance on speech recognition tasks. Similarly, the effects of CND on neurogram based metrics has also not been explored in previous literature. In this work we present two studies that evaluate NSIM as a metric. In Study 1, NSIM is used to evaluate performance of individuals with varying degrees of hearing loss. In Study 2, we use NSIM to model varying degrees of CND.

2. Methods

2.1. Phenomenological Model of Cochlea and Auditory Nerve Fiber Response

This study used ANM [3] to simulate the responses of ANF and generate neurogram for a given sound (.wav file) and for a specific hearing loss profile. The model allows the simulation of hearing loss by decreasing OHC gain and making the frequency response of the cochlear filters broader [3]. The output of the ANM is in the form of post stimulus histogram (PSTH) which represents the spiking activity of the auditory nerve fibers in time bins. Each time bin in the PSTH can be converted to frequency response and generate a neurogram, which maps the spiking activity on a time-frequency map. The ANM allowed simulation of the three different types of auditory nerve fibers: 1) Low Spontaneous (LS) rate fibers with high thresholds, 2) Medium Spontaneous (MS) rate Fibers with medium thresholds, and 3) High Spontaneous (HS) rate fibers with low thresholds. The summed spiking activity of all fibers can be used to get one neurogram that represents activity for all three ANF types in response to an acoustic stimulus, or three separate neurograms representing each ANF fiber types. Additionally, to capture long term and short-term properties of auditory nerve, the model allows control of the time bins used for the PSTH which in turn allows generation of fine timing (FT) neurograms and mean rate (MR) neurograms.

2.2. Neurogram Similarity Index Measure:

To quantify the similarity between the normal hearing neurograms and degraded hearing neurograms (neurograms from a hearing loss cochlea), methods previously developed [6, 4] were used to calculate NSIM. The image processing based method Structural Similarity Index Measure (SSIM) compares luminance(l), contrast(c) and structure(s) between a reference image and degraded image using a suitable window size [6] using Eq. 1. The constants α , β , and γ are the weighting coefficients that are applied to weigh the different components of the similarity equation. The coefficients C_1 , C_2 , C_3 have negligible effect on the results and are added to prevent unstable results at boundaries. Hines et. al [5] applied the same method and derived an equivalent similarity measure for comparing neurograms. They showed that for the neurograms, in the Eq. 1 contrast and component weightings have negligible effect. This reduces Eq. 1 to Eq. 2 which is then calculated over a 3X3 gaussian window of radius 0.5 [5, 8] and averaged over all the points in the neurogram to calculate the NSIM (eq. 3). The general design of the experiments is given in Fig. 1.

$$SSI(r, d) = [l(r, d)]^\alpha \cdot [c(r, d)]^\beta \cdot [s(r, d)]^\gamma \quad (1)$$

where:

$$l(r, d) = \frac{2\mu_r\mu_d + C_1}{\mu_r^2 + \mu_d^2 + C_1}, c(r, d) = \frac{2\sigma_{rd} + C_2}{\sigma_r^2 + \sigma_d^2 + C_2},$$

$$s(r, d) = \frac{\sigma_{rd} + C_3}{\sigma_r\sigma_d + C_3},$$

μ_r = mean of reference image intensity over window,

μ_d = mean of degraded image intensity over window,

σ_r = standard deviation of reference image over window,

σ_r^2 = variance of reference image over window,

σ_d = standard deviation of degraded image over window,

σ_d^2 = variance of degraded image over window,

σ_{rd} = covariance matrix over the window,

$$C_1 = 0.01L, \quad C_2 = (0.03L)^2, \quad C_3 = \frac{C_2}{2},$$

$L = \text{intensity range}$

$$NSI(r, d) = \frac{2\mu_r\mu_d + C_1}{\mu_r^2 + \mu_d^2 + C_1} \cdot \frac{\sigma_{rd} + C_3}{\sigma_r\sigma_d + C_3} \quad (2)$$

$$NSIM(r, d) = \frac{1}{N \cdot M} \sum_{f=1}^N \sum_{t=1}^M NSI(r, d) \quad (3)$$

$N = \text{total frequency bins}$, $M = \text{total time bins}$

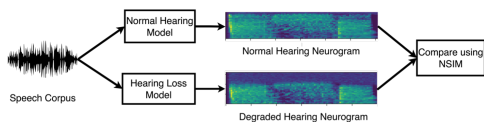


Figure 1: Design of experiment to simulate Neurogram Similarity Index Measure

2.3. Study 1: Correlating phoneme-recognition-task performance with NSIM

For this study, speech material selected and tested by Harris et. al. [9, 10] was used. The speech level was administered monaurally at 65 dB SPL. The test consisted of presenting /VCV/ speech stimuli with the vowel /a/ and 10 different consonants combined with speech shaped noise at different signal to ratios (SNR) found to be the most sensitive to predict hearing loss [9]. In their study there were a total of 94 individuals with hearing loss ranging between Normal hearing (PTA = \leq 15 dB HL) to Moderate Hearing loss (PTA = 41 – 55 dB HL). Corpus developed by Harris et. al. [10] was used here and processed through ANM to simulate 94 hearing loss profiles that were clinically measured by Harris et. al. Neurograms were generated for each hearing loss profile and compared with corresponding normal hearing neurogram using NSIM. For each hearing loss profile/individual, MR-NSIM and FT-NSIM were calculated for all 10 /VCV/ from Harris et. al. and averaged to generate average MR-NSIM and FT-NSIM for each hearing loss profile/individual.

Average MR and FT NSIM values across all 10 phonemes and Pure Tone Average Thresholds (PTA in dB) were used to train support vector regression(SVR) based models to predict performance of individuals as recorded by Harris et. al. [10] using a threefold cross validation paradigm (Fig. 2).

2.4. Study 2: Extending NSIM to study Cochlear Neural Degeneration (CND)

CND is characterized by loss of auditory nerve synapses which ultimately results in less information reaching the higher order processing centers in the brain. To simulate the performance that can be used to predict and study CND, it is important to select a type of stimulus that can encode the effects of CND at the neurogram level. Additionally, all three types of fibers have different response characteristics and are present in unequal numbers in an AN bundle. Therefore, to accurately study the effect of CND, it may be important to look at neurograms for each fiber type independently and then compare them with the corresponding fiber type neurogram from a normal hearing cochlea. The resulting similarity maps for each fiber type can then be summed and averaged to get one value of similarity for the ease of analysis. Another hypothesis related to effects of CND is the decrease in information content in the neural signal which becomes more evident only during complex listening tasks [11]. Therefore, in order to simulate these conditions, 20 lists consisting of 10 iso-phonemic, monosyllabic common /CVC/ words were used. There were no word repetitions across the lists. These words were generated using Google text-to-speech api in python [12]. After generating the wav files for the words, they were normalized using the RMS value of the waveforms (16 bit) and converted to pressure values using:

$$p = \frac{\text{wav}}{\text{rms}(\text{wav})} \times 20 \times 10^{-6} \times 10^{\frac{L}{20}} \quad (4)$$

where $L \in \{50, 65, 80, 95\}$ dB SPL.

These words were presented in a range of listening levels from 50–95 dB SPL, and difficulty conditions: 1) without any compression or reverberation, 2) with 65% time compression, and 3) with 65% time compression and reverberation. Audiometric profile shown in Table 1 and different degrees of CND as shown in the Table 2 were used in this study. The NSIM for each fiber type were calculated separately for all 200 words and then averaged to get a single NSIM values for FT-neurogram

and MR-neurograms for each fiber type (eq. 5). As a next step, the NSIM values for the sloping loss profile with no CNL were used as baseline values and then the effect of introducing CNL was quantified by normalizing with CNL to hearing loss profile without any CNL (eq. 6).

$$NSIM_{\text{overall}} = \frac{1}{3} (NSIM_{\text{LS}} + NSIM_{\text{MS}} + NSIM_{\text{HS}}) \quad (5)$$

$$CND_{\text{effect}} = \frac{NSIM_{\text{HL no CNL}} - NSIM_{\text{HL and CNL}}}{NSIM_{\text{HL no CNL}}} \quad (6)$$

Table 1: Hearing Loss profile used in the study and simulated using OHC loss in the cochlear model

Audiometric Profile	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Sloping Loss	0	0	10	20	23	45	75

Table 2: CNL Profiles used for the the age related hearing loss profile listed in Table 1. The distribution [Low-SR, Med-SR, High-SR] = [5, 5, 12] refers to the number of low-spontaneous-rate (LS), medium-spontaneous-rate (MS), and high-spontaneous-rate (HS) fibers, respectively, for each sensory inner hair cell for the normal cochlea with no CNL.

Audiometric Profile	CNL Profile	Low-SR Fibers	Med-SR Fibers	High-SR Fibers
Sloping Loss	No CNL	5	5	12
	20% LS MS loss	4	4	12
	40% LS MS loss	3	3	12
	60% LS MS loss	2	2	12
	80% LS MS loss	1	1	12
	100% LS MS loss	0	0	12
	100% LS MS loss, 20% HS loss	0	0	10

3. Results

3.1. Study 1: Correlating phoneme-recognition-task performance with NSIM

We trained support vector regression-based models (Fig. 3) to predict performance on a phoneme recognition task using the MR-NSIM, FT-NSIM and PTA (dB) values (Table 3). The results from the best performing model are presented in the Fig. 2C. The model presented achieved a mean squared error of 0.015 and R^2 value of 0.64. The models which incorporated both MR-NSIM and FT-NSIM features performed the better as shown in the Table 3. There is a discrepancy between the variation of MR-NSIM and the actual performance when plotted against pure-tone average hearing loss (dB) (Fig. 2B), particularly below 25 dB HL. MR-NSIM degrades monotonically as hearing loss increases, whereas actual performance only begins to decline beyond approximately 25 dB HL. We attribute this difference to the way NSIM is calculated, which uses the ideal (no-loss) neurogram as the reference (eq. 3) and compares it to degraded neurograms. To address this, we do not use raw NSIM values directly as a measure of performance; instead, we treat them as features in our downstream SVR models (Table 3).

3.2. Study 2: Extending NSIM to study Cochlear Neural Degeneration

Figure 3 shows the CND_{effect} using Eq. 6 (with degree of fiber loss defined in Table 2) plotted against the signal level (dB SPL), for the three different speech types in each panel. For every step increase in CNL by 20%, there is a change in the values

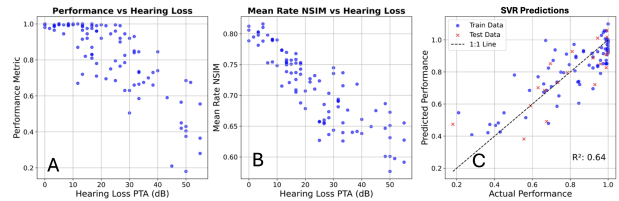


Figure 2: (A) Percent-correct performance measured on 94 subjects. (B) Mean Rate NSIM on the phoneme-recognition task plotted against the Pure Tone Average (PTA) threshold hearing loss (dB) for 94 subjects. (C) Performance of the support-vector regression (SVR) predictions trained using the Mean Rate NSIM, Fine Timing NSIM, and PTA as features resulting in $R^2=0.64$.

Table 3: Performance of best Performing Support Vector Regression Models trained on various feature combinations to predict performance on phoneme recognition task

Features	Hyperparameters	MSE	R^2
MR NSIM	{C: 1, epsilon: 0.075, gamma: auto, kernel: rbf}	0.023	0.485
FT NSIM	{C: 0.25, epsilon: 0.1, gamma: auto, kernel: rbf}	0.028	0.352
MR NSIM \times FT NSIM	{C: 10, epsilon: 0.05, gamma: auto, kernel: rbf}	0.018	0.588
MR NSIM \times FT NSIM \times PTA	{C: 2.5, epsilon: 0.05, gamma: scale, kernel: linear}	0.0157	0.639

the CND_{effect} (different colored lines) at all the stimulus levels tested and all three speech types of speech material. The greatest effect due to CNL was observed at 95 dB SPL for all speech materials. The overall effect of CNL is highest for the 65% compressed speech for profiles with total loss of LS and MS fibers (light purple). Additional loss of 20% HS fibers did not affect the results significantly (near overlap between light and darker purple lines).

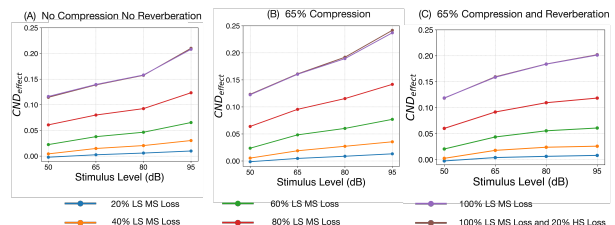


Figure 3: Average CND_{effect} calculated from MR-NSIM for the entire speech corpus (indicated in the title of subplots) and plotted across various signal levels (dB SPL) for Hearing Loss profiles with varying degrees of CNL as indicated by the legend.

4. Discussion

Zaar and Carney [13] predicted speech-reception thresholds (SRTs) for subjects with normal hearing or hearing loss using different types of speech material and speech-in-noise conditions using neurograms fed into an approximate model for the inferior colliculus (IC). Our strategy differs from [13] in that our backend decision is based on NSIM rather than correlation on approximate responses of IC. As such, our approach only uses responses from the auditory periphery, which is relatively well understood, and is free from the potential drawbacks of assumptions and hypotheses about processing strategies used by central auditory brain mechanisms. Another important difference from previous work is that the combined ANM and NSIM methods

have not been applied to quantify the effects and contributions of CND.

4.1. Study 1: Correlating phoneme-recognition-task performance with NSIM

In Study 1 we were able to show that SVR models trained using both the MR-NSIM, which represents the long-term similarity in the response, and FT-NSIM, which represents short term similarity between neurograms, performed the better when predicting performance on the phoneme recognition task. This shows that for a phoneme recognition task, individuals rely on both short-term spiking differences to detect phoneme transitions particularly between vowel and consonant and long term spiking activity to track the low-frequency components in the speech. The R^2 value for the best trained model was 0.64. The performance of the model was worst for region where there were fewest points in the training data (Fig 2). The results from this study show that NSIM is a useful measure to study and simulate performance of individuals with hearing loss and given enough data, it can generate reliable predictions for performance on phoneme recognition task. However, one caveat remains: we observed that for lower degrees of hearing loss, the MR NSIM values degrade but the actual performance does not degrade (Fig. 2A,B) which suggests that there could be certain thresholds for the raw NSIM values which must be exceeded before the performance begins to degrade.

4.2. Study 2: Extending NSIM to study Cochlear Neural Degeneration (CND)

In Study 2, we demonstrated that when using speech recognition tests to detect the effect of CND or synaptopathy, the highest differences between profiles with and without CND were observed when the presentation level of the stimulus was high (above 80 dB SPL), which follows from our understanding of that at higher stimulus levels, all three types of fibers are required to encode the information, since the HS fibers with low thresholds will be saturated and to increase the total number of spikes, LS/MS fibers will need to be recruited. However, in the absence of these fibers (typical CND case) no additional spiking activity is produced resulting in lower NSIM and hence greater difference between profiles with and without CND. It was also observed that the maximum effect of CND was observed for 65% compressed speech without reverberation, which means that 65% compressed speech test can be used to discriminate CND profile from the profile without CND. Increasing the difficulty of test further by adding reverberation slightly decreases the CND_{effect} . This means that even though the difficulty of the test will increase by adding reverberation but without any CND discriminability benefits. This would mean more false positives, if such a test were used as a metric to detect CND in patients.

Another interesting observation from the Fig. 3 is that for all speech material, the values of CND_{effect} are relatively small up to 60% LS MS Loss profile (green line), however 80% CND profile has approximately twice the values for CND_{effect} at all levels for all profiles. This observation is similar to what Grant et. al. [14] observed where they saw significant reduction in word recognition scores only after the individual had lost more than 60% neurons. This further suggests that NSIM is a reliable measure and can be employed as a sensitive metric for detecting CND.

5. Conclusions and Future Directions:

Both the studies indicate that NSIM can be used as an objective metric to map performance on phoneme recognition tasks and to detect sensory hearing loss associated outer hair cell loss (Study 1) and it can be used to estimate CND (study 2). In study 1, we observed that for lower degrees of hearing loss, there is difference in the trend of actual performance and raw NSIM values (Fig. 2 A,B). Future work can explore deriving thresholds that can be applied on the raw NSIM values, so that it can be related to the performance on speech tests. In the study 2, only 1 hearing loss profile was simulated, and future studies can use the methods presented here to look at all commonly observed hearing loss profiles. Additionally, average values for NSIM were calculated by averaging NSIM for all fiber types, however, there is no physiological evidence that differences/correlations for all three types of fibers can be averaged. In this study average NSIM for all three fiber types simplified the analysis. Future studies can explore deriving weighting coefficients for the NSIM for different fiber types. Another possible future direction would be to conduct the speech tests on human subjects with material discussed in study 2 and relate the NSIM to performance measures on these tests. Having established NSIM as an objective measure can ultimately allow deriving hearing aids gain compensation strategy that can account for CND, which currently remains a major challenge in hearing healthcare.

6. Acknowledgements

The authors would like to thank Dr. Stephen Neely from Boys-town National Research Hospital for sharing the dataset used for this study. We would also like to acknowledge our funding sources: National Institutes of Health (NIH), National Institute on Deafness and Other Communication Disorders grant number R01DC007910-16, Massachusetts Eye and Ear Infirmary (MEEI) Shark Tank Research Award, and funding from Harvard Graduate School of Arts and Sciences (GSAS).

7. References

- [1] A. J. Vermiglio, S. D. Soli, D. J. Freed, and L. M. Fisher, "The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index," *J. Am. Acad. Audiol.*, vol. 23, no. 10, pp. 779–788, Nov. 2012.
- [2] S. G. Kujawa and M. C. Liberman, "Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss," *J. Neurosci.*, vol. 29, no. 45, pp. 14 077–14 085, Nov. 2009.
- [3] I. C. Bruce, Y. Erfani, and M. S. A. Zilany, "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites," *Hear. Res.*, vol. 360, pp. 40–54, Mar. 2018.
- [4] A. Hines and N. Harte, "Speech intelligibility from image processing," *Speech Commun.*, vol. 52, no. 9, pp. 736–752, Sep. 2010.
- [5] —, "Speech intelligibility prediction using a neurogram similarity index measure," *Speech Commun.*, vol. 54, no. 2, pp. 306–320, Feb. 2012.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [7] N. Mamun, W. A. Jassim, and M. S. A. 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.

- [8] C. Sloan, N. Harte, D. Kelly, A. C. Kokaram, and A. Hines, "Objective assessment of perceptual audio quality using visqolaudio," *IEEE Transactions on Broadcasting*, vol. 63, no. 4, pp. 693–705, 2017.
- [9] J. J. Hajicek, S. E. Harris, and S. Neely, "Using consonant confusion to predict unexplained hearing loss," *J. Acoust. Soc. Am.*, vol. 154, no. 4-supplement, pp. A34–A34, Oct. 2023.
- [10] S. Harris, P. Hajicek, and S. Neely, "Evaluation of hearing-aid benefit with a consonant-confusion test," Poster presented at the International Hearing Aids Conference 2024, 2024.
- [11] M. DiNino, L. L. Holt, and B. G. Shinn-Cunningham, "Cutting through the noise: Noise-induced cochlear synaptopathy and individual differences in speech understanding among listeners with normal audiograms," *Ear Hear.*, vol. 43, no. 1, pp. 9–22, Jan. 2022.
- [12] gTTS Contributors, "gtts: Google text-to-speech," <https://github.com/pndurette/gTTS>, 2024, python library for interfacing with Google Translate's text-to-speech API.
- [13] J. Zaar and L. H. Carney, "Predicting speech intelligibility in hearing-impaired listeners using a physiologically inspired auditory model," *Hear. Res.*, vol. 426, no. 108553, p. 108553, Dec. 2022.
- [14] K. J. Grant, A. Parthasarathy, V. Vasilkov, B. Caswell-Midwinter, M. E. Freitas, V. de Gruttola, D. B. Polley, M. C. Liberman, and S. F. Maison, "Predicting neural deficits in sensorineural hearing loss from word recognition scores," *Sci. Rep.*, vol. 12, no. 1, p. 8929, Jun. 2022.