Simulated Performance Intensity Functions

Andrew Hines and Naomi Harte

***Abstract*— Measuring speech intelligibility for different hearing aid fitting methods in a simulated environment would allow rapid prototyping and early design assessment. A simulated performance intensity function (SPIF) test methodology has been developed to allow experimentation using an auditory nerve model to predict listeners' phoneme recognition. The test discriminates between normal hearing and progressively degrading levels of sensorineural hearing loss. Auditory nerve discharge patterns, presented as neurograms, can be subjectively ranked by visual inspection. Here, subjective inspection is substituted with an automated ranking using a new image quality metric that can quantify neurogram degradation in a consistent manner. This work reproduces the test results of a real human listener with moderate hearing loss, in unaided and aided scenarios, using a simulation. The simulated results correlate within comparable error margins to the real listener test performance intensity functions.**

I. INTRODUCTION

Developing improved hearing aid algorithms is an intensive process in terms of labour, test subjects and time. A simulated environment would allow rapid prototyping and basic assessment of new fitting algorithms. The ability to test and quantitatively compare the speech intelligibility improvements offered by different hearing aid fitting methods would not replace listener tests but could significantly reduce development costs and times. To realise this, a quantitative simulation and test methodology is needed to discriminate between normal hearing auditory systems and those with a variety of progressively degraded levels of sensorineural hearing loss (SNHL).

A simulated performance intensity function (SPIF) test methodology has been developed to allow experimentation using an auditory nerve (AN) model to predict the phoneme recognition of listeners - both unimpaired and with SNHL. This work sought to reproduce the results for a human listener with moderate hearing loss that were presented by Boothroyd [1]. Using the same dataset the simulations investigate whether the AN model and human listeners produce comparable results. Experiments were carried out with three hearing profiles - an unaided normal auditory system, and one with moderate SNHL tested in unaided and aided scenarios.

Auditory nerve discharge patterns can be represented visually as neurograms, illustrating the neural discharge intensity for a given time and frequency band. Neurograms for speech sounds from normal and impaired listeners can be subjectively ranked by visual inspection [2]. Manual subjective visual inspections are substituted with a new image quality metric which is used as a quantitative rank. It has been shown to quantify neurogram degradation in a consistent manner that correlates closely with real test data for normal hearing subjects [3].

Section 2 introduces the AN model, the NSIM image quality metric, listener test simulation, hearing profiles and hearing aid fitting algorithm used. Section 3 describes the simulation methodology and how the tests were designed to reproduce real listener tests. Section 4 presents the simulated results and a comparison to the human listener test results [1]. Section 5 discusses the results, continuing work and potential applications.

A. Hines and N. Harte with Sigmedia Group, Departement of Electrical and Electronic Engineering, Trinity College Dublin, Ireland hinesa@tcd.ie

II. BACKGROUND

A. Auditory Nerve Model

The Zilany et al. AN model used in this study is the latest version developed in ongoing research [4] and has been extended and enhanced over the last decade [5]. Physiological data was matched to a wide variety of inputs including speech, noise and pure tones in extensive testing. The latest model adds power-law dynamics as well as exponential adaptation in the synapse model. The AN model covers the middle and inner ear, so a pre-filter based on measurements from Wiener and Ross [6] is used to model the outer ear when simulating free field listener tests.

B. Neurogram Assessment

A neurogram is analogous to a spectrogram. It presents a pictorial representation of a signal in the time-frequency domain using colour to indicate activity intensity. An example signal, the word 'ship' presented at 65 dB SPL, is presented in Fig. (1). The top row shows the time domain signal. Below it, the spectrogram presents the sound pressure level of a signal for frequency bands in the y-axis against time on the x-axis. Three neurograms, created from AN model outputs for signals presented at progressively lower presentation levels (65, 30 and 15 dB SPL), are then shown. The colour represents the neural firing activity at a given CF band in the y-axis over time in the x-axis. The neural activity is binned into time bins (100μs) to create post stimulus time histogram (PSTH) information. The neurogram smoothes the information and presents the average discharge rate (equivalent to the signal envelope) by convolving them with 50% overlap, 128 sample Hamming window. As in prior work [7], [8], neurograms with 30 characteristic frequencies (CFs) were used, spaced logarithmically between 250 and 8000 Hz. The neural response at each CF was created

from the PSTH of 50 simulated AN fibres with varying spontaneous rates.

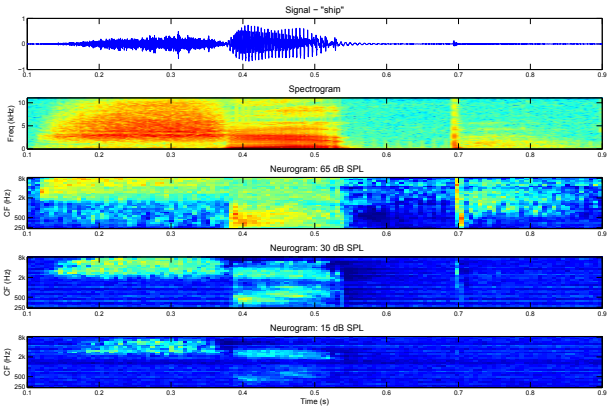


Fig. 1. *A sample signal, the word "ship". The top row shows the time domain signal, with the time-frequency spectrogram below it. Three sample neurograms for the same signal presented to the AN model at 65, 30 and 15 dB SPL signal intensities are presented.*

Neurograms for each phoneme are assessed as an image comparison between the neurogram being assessed and a reference neurogram from a normal hearing AN model for the same input signal. The structural similarity measure, SSIM [9], is an image quality metric originally designed to measure the impact of compression techniques on the quality of jpeg images. It measures the similarity over a window or 'patches' rather than a simple point to point comparison and takes into account perceived changes in luminance, contrast and structure. It can provide a quantitative measure of neurogram degradation to predict phonemic discrimination. The use of SSIM as a ranking measure with phoneme neurograms from a wide variety of speakers and accents was previously demonstrated [7]. SSIM has been shown to be superior to other simple point to point measures such as a relative mean squared error assessed per neurogram element. It was established that for the purposes of neurogram comparisons the optimal window size was a 3x3 pixel square covering 3 CF bands and a 12.8ms time window. SSIM was further tuned and it was established that the contrast component provided negligible value when comparing neurograms and that closer fitting to listener test data occurred using only a luminance and structural comparison [3].

The Neurogram Similarity Index Measure (NSIM) is a simplified version of SSIM and is defined as

$$N(r,d) = l(r,d)\cdot s(r,d) = \frac{2\mu_r\mu_d + C_1}{\mu_r^2 + \mu_d^2 + C_1}\cdot\frac{\sigma_{rd} + C_2}{\sigma_r\sigma_d + C_2} \quad (1)$$

The NSIM between two neurograms, the reference, r, and the degraded, d, is constructed as a weighted function of intensity (l), and structure (s) as in eqn. (1). Intensity looks at a comparison of the mean (μ) values across the two neurograms. The structure uses the variance (σ) and is equivalent to the correlation coefficient between the two neurograms. As with SSIM, each component also contains constant values ($C_1 = 0.01L$ and $C_2 = (0.03L)^2$, where L is the intensity range, as per [9]) which have negligible influence on the results but are used to avoid instabilities at boundary conditions. See [7] for further information on neurogram ranking with SSIM. A simulated performance intensity function (SPIF) can be produced by using NSIM to rank a large number of neurogram comparisons, over a range of intensity levels.

C. Performance Intensity Function

The performance intensity (PI) function is the basis for standard listener tests. Evaluation of a test subject's speech reception threshold (SRT) and word recognition in lists of phonetically balanced words allow validation of pure tone thresholds and estimating auditory resolution respectively. The PI function has been shown to be useful for comparative tests of aided and unaided speech recognition results and it has been proposed as a useful method of evaluation of the performance improvement of subjects speech recognition under different hearing aid prescriptions or settings [1].

The test corpus used came from the Computer Aided Speech Perception Assessment (CASPA; [10]) software package which was developed to simplify the data recording and analysis for performance intensity listener tests. It contains 20 word lists of 10 phonemically balanced consonant-vowel-consonant (CVC) words. Words are not repeated within lists and lists are designed to be isophonemic, i.e. to contain one instance of each of the same 30 phonemes. Word lists comprising 10 words are presented over a range of intensity levels. The tester records the subject's responses with the CASPA software. It automatically scores results in terms of words, phonemes, consonants, and vowels correct and generates separate PI functions for each analysis.

D. Simulated Performance Intensity Function

In a standard performance intensity listener test, CVC words are presented to the test subject who listens and repeats the words, which are manually scored, per phoneme correctly identified, by the tester. This is repeated at a progressive range of intensity levels and a PI function is measured.

The Simulated Performance Intensity Function (SPIF) replaces the listener with the AN model and scoring is based on automated comparisons of the neurograms produced by the nerve firing simulations from the model. Neurograms from the AN model with normal hearing thresholds are used to create a baseline set of neurograms at a comfortable speech level for normal listeners. A 65 dB SPL reference is used as it represents a mean sound field pressure for conversational speech [11].

NSIM scores are calculated by comparing neurograms from a given listener's phoneme recognition threshold (PRT) level. This establishes a neurogram phoneme recognition threshold (NPRT) which is used to establish the percentage recognition at each sound intensity level and allow a SPIF to be plotted.

III. SIMULATED TESTS

Three Simulated Performance Intensity Function listener tests were carried out using the AN model to simulate

an unimpaired, normal hearing listener, and listener with a moderate SNHL in unaided and NAL-RP aided scenarios. For this experiment, a software model of the NAL-RP linear fitting method was developed to pre-filter the input signals with the output gains prescribed using the formula for the fitting method is outlined in [12]. The hearing loss thresholds and prescribed insertion gains are presented in Table. I. The thresholds are a mean of the left and right ear values for the human listener test subject where there were slight differences in the left/right ear thresholds [1].

f(Hz)	250	500	1k	2k	4k	6k	8k
dB HL	37.5	40	45	35	42.5	55	60
IG (dB SPL)	2	10	20	16	17	21	-

TABLE I

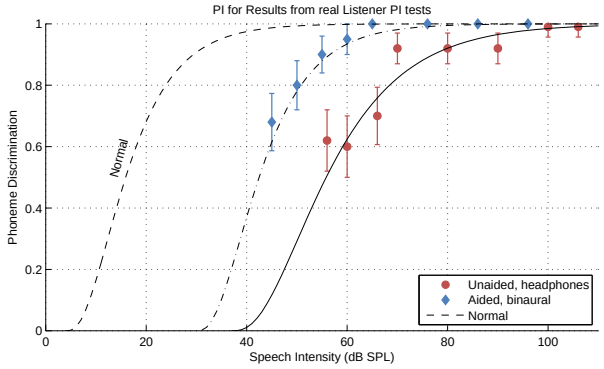
HEARING LOSS THRESHOLDS AND PRESCRIBED NAL-RP INSERTION GAINS TO THE NEAREST DB SPL.

The SPIF procedure mimics that of a real listener test. The human listener with the AN model and the NSIM scores are used to assess neurogram degradation and to predict phoneme discrimination. Word lists from the CASPA dataset [10] were used. Timing label files marking the phoneme boundaries were created for the 200 words. For each word, the time was split into 5 portions, a leading and trailing silence, and 3 distinct phonemes.

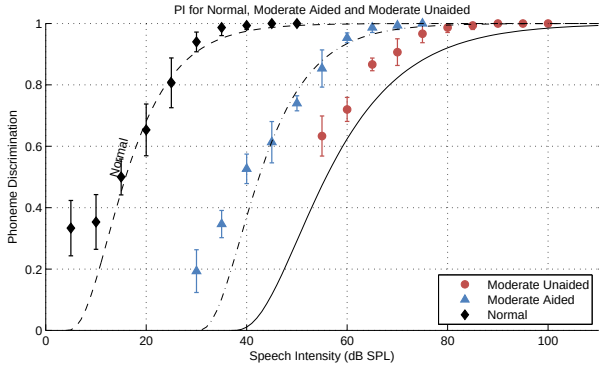
For normal hearing listeners, the Phoneme Recognition Threshold (PRT) is the level in dB SPL at which the listener scores 50% of their maximum and is analogous to their Speech Reception Threshold (SRT). The modal value of this was set at 15 dB SPL for normal hearing listeners as per Boothroyd [1]. A level of 65 dB SPL was taken as the standard level to generate reference neurograms to test against.

The NSIM was measured between a reference neurogram at 65 dB SPL and a degraded neurogram at 15 dB SPL (PRT level) over a large sample of phonemes gives a neurogram PRT (NPRT). The NPRT value was calculated as the median NSIM score of the 300 phonemes (evaluated using ten lists, #11-20, of CVC words). For the normal hearing test, the word lists were presented to the AN model at speech intensity levels of 5 through to 50 dB SPL in 5 dB increments and neurograms were created. The same procedure that was used for evaluation of the NPRT was repeated at each speech intensity level using 5 other word lists (150 phonemes). The results were recorded and a phoneme discrimination score was calculated by counting the number of phonemes scoring above the NPRT value. A simulated performance intensity function was calculated from the results.

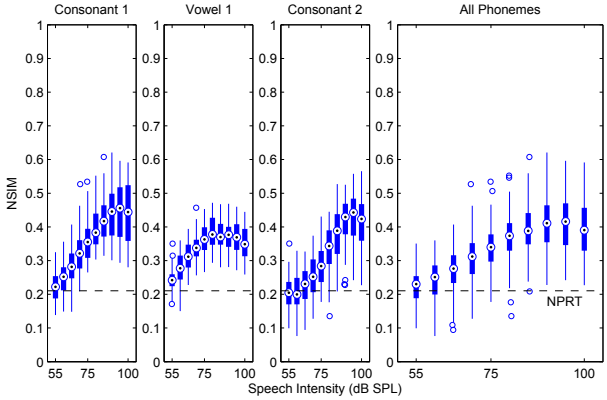
The procedure was repeated for the moderate SNHL unaided and aided scenarios. For the unaided case, as per Boothroyd's results, the PRT was set at 54 dB SPL and measurements were taken with input speech signals presented at 5 dB intervals between 55 and 100 dB SPL. For the aided tests, the PRT was 42 dB SPL and measurements were taken at 5 dB intervals between 35 and 75 dB SPL.



(a) Human listener results reproduced from Boothroyd[1].



(b) Simulated PI function results calculated from NSIM results.



(c) NSIM results for unaided moderate SNHL, with NPRT level marked.

Fig. 2.

IV. RESULTS

The results from Boothroyd's real listener tests for a listener with moderate hearing loss are reproduced for reference in Fig. 2(a). The corresponding results for the simulated PI function tests are presented in Fig. 2(b). In both cases the error bars indicate one standard error above and below.

Fig. 2(a) shows three plots, a normal listener result which has been normalised to a PRT of 15 dB SPL and the unaided and aided results for a listener with moderate SNHL. The hearing aid shifts the PI curve by around 15-20 dB for the moderate hearing loss tested, which from Table (I) has a threshold loss ranging from 35 to 60 dB HL.

Fig. 2(c) presents the raw NSIM scores for the simulation of unaided moderate SNHL. It is broken down by phoneme

position (i.e. initial consonant, vowel, final consonant) and grouping the phonemes together. The bars mark the central median and inter quartile range with whiskers extending to extremes and outliers plotted individually. The NPRT line was calculated across all phonemes together as the basic PI function does not differentiate between recognition by phoneme type. The breakdown is shown to illustrate the variance in results by phoneme position and type. The corresponding plots for normal and aided moderate are not shown are not presented due to space constraints.

Fig. 2(b) shows the three SPIF functions, a normal listener and the results for a listener with moderate SNHL. The results for normal and moderate aided hearing track very well to the actual listener PI functions. The unaided results are a close match to the trend but are offset and over predicting the phoneme recognition. The PI curves that are plotted are redrawn from Fig. 2(a) to allow a comparison in the data fit between the human listener and simulated tests.

V. DISCUSSION

A. Simulation and Clinical Test Comparison

Comparing the results in Fig. 2(a) for the real listener results to those in Fig. 2(b) for the simulated results from the AN model, the overall correlation is very promising. The key area of interest is between the 50% phoneme discrimination (%P.D.) and the level where it plateaus. The results for the normal hearing listener show a very close fit through this area. The %P.D. for 5 and 10 dB SPL presentation levels is indicating higher recognition than the listener PI curve would predict.

The results for moderate SNHL (unaided) follow quite closely to the shape of the listener curve but are over predicting the %P.D. and have shifted by 5-10 dB. This will be looked at in more detail below. The simulated aided results fit closely to the predicted listener PI function.

The error bars (representing +/- 1 standard error) for the simulated results are smaller than those for the real listener tests. The real listener tests were for a single individual and were not tested with as many lists as used in the simulation so from a purely statistical perspective this would be expected as there is not as much data to establish the range and outliers. The size of the error bars do highlight the variance in results from a clinical environment.

Fig. 2(c) shows the raw NSIM data broken down by phoneme position and then a grouped scoring encompassing all phonemes. The breakdown by phoneme shows that with a moderate loss the vowels are performing better at low presentation level but that the consonants perform better at higher presentation levels. At high presentation levels the NSIM scores begin to drop, which may be a representation of rollover effects decreasing phoneme discrimination.

The all phoneme NSIM data shows the spread of results at each presentation level. It can be seen that the NPRT line crosses just below the inter quartile range at 55 dB SPL and that a very small increase in the NPRT level would cause a significant change to the %P.D. at 55 and 60 dB and would cross the whiskers on the higher presentation levels NSIM scores. Shifting the NPRT by 1dB improved the fit significantly for the unaided results, suggesting that for good correlation, the methodology is heavily dependent on an accurate PRT measurement.

This does not imply inconsistencies in the results. Significant testing to ensure reliability and repeatability were carried out previously [3]. To test whether there was a variability in the SPIF results based on calculated NPRT values, the results presented here were checked with NPRT values created using 10 lists (#11-20) and also using the 5 lists that were used at each presentation level (#6-10) and there was no significant difference.

VI. CONCLUSIONS

A review and comparison with other intelligibility indices was presented in prior work[7], where it was acknowledged that the methodology required validation with real listener tests. The results demonstrate that a Simulated Performance Intensity Function can predict speech intelligibility for normal and impaired listeners. These early results are promising, indicating that the AN model and hearing aid model can produce results that closely follow human test results, even for listeners with SNHL. This study was limited to a quiet environment, but the same methodology could be applied with speech in noise. Work is ongoing to validate the methodology with further SNHL profiles (e.g. severe hearing loss). Alternative hearing aid fitting algorithms (DSL) are also being investigated to assess whether the test differentiates between the phoneme discrimination performance of alternative fitting strategies.

REFERENCES

[1] A. Boothroyd, "The performance/intensity function: An underused resource," *Ear and Hearing*, vol. 29, no. 4, pp. 479–491, 2008.

[2] M. B. Sachs, I. C. Bruce, R. L. Miller, and E. D. Young, "Biological basis of hearing-aid design," *Ann. Biomed. Eng.*, vol. 30, p. 157168, 2002.

[3] A. Hines and N. Harte, "Speech intelligibility prediction using a neurogram similarity index measure," *Speech Communication*, 2011, under Revision.

[4] M. S. A. Zilany, I. C. Bruce, P. C. Nelson, and L. H. Carney, "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," *J Acoust Soc Am*, vol. 126, no. 5, pp. 2390–2412, 2009.

[5] X. Zhang, M. Heinz, I. Bruce, and L. Carney, "A phenomenological model for the responses of auditory-nerve fibers. i. non-linear tuning with compression and suppression," *J Acoust Soc Am*, vol. 109, pp. 648–670, 2001.

[6] F. Wiener and D. Ross, "The pressure distribution in the auditory canal in a progressive sound field," *J Acoust Soc Am*, vol. 18, no. 2, pp. 401–408, 1946.

[7] A. Hines and N. Harte, "Speech intelligibility from image processing," *Speech Communication*, vol. 52, no. 9, pp. 736–752, 2010.

[8] ——, "Error metrics for impaired auditory nerve responses of different phoneme groups," in *Interspeech 09*, Brighton, England, 2009, pp. 1119–1122.

[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE T Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[10] A. Boothroyd, "Computer-aided speech perception assessment (caspa) 5.0 software manual. san diego, ca." 2006.

[11] B. C. J. Moore, *Cochlear Hearing Loss - Physiological, Psychological and Technical Issues*, 2nd ed. J Wiley, 2007.

[12] H. Dillon, "Hearing Aids," *Thieme Medical Pub (NYC)*, 2001.