On the Predictability of the Intelligibility of Speech to Hearing Impaired Listeners

Mark Huckvale, Gaston Hilkhuysen

Speech, Hearing & Phonetic Sciences, University College London, U.K.

m.huckvale@ucl.ac.uk, g.hilkhuysen@ucl.ac.uk

Abstract

What information do we need to know about listeners to predict their performance on a speech intelligibility task and how well can we predict intelligibility anyway? This paper performs a meta-analysis on two speech intelligibility studies of hearing-impaired listeners in which we evaluate different approaches to building a predictive model of intelligibility. The model has two components: a cochlear loss term based on a number of psychoacoustic measures of hearing, and a supra-cochlear loss term to explain residual performance variation. These models are trained using a method of cross-validation to determine how well they might perform on new listeners and new tasks. We found that cochlear loss could only explain 40% of the variability in performance across hearing-impaired listeners, while the supra-cochlear loss can account for a further 20-40% depending on the task. The combined cochlear and supra-cochlear loss terms allow good estimates of intelligibility scores in the data, with speech reception thresholds on a novel listening task being predictable to within 1dB on average.

Index Terms: hearing impairment, speech intelligibility, psychoacoustics, metrics.

1. Introduction

The availability of increasingly powerful computational resources in miniaturized form allows for more advanced signal processing algorithms to be applied within hearing aids. These techniques hold promise for improved speech intelligibility for hearing impaired (HI) users in everyday noisy and reverberant environments.

However, the development of some advanced signal processing algorithm is not enough. It is also necessary to ensure that processing adapts to the requirements of the listener and to the requirements of the listening situation [1]. For a given impaired listener, the advanced aid needs to choose between the relative benefits of equalization, compression, noise reduction, dereverberation, beam-forming or speech enhancement for every listening situation. The challenge is not just to find good signal processing approaches, but to understand how they benefit the listener to ensure they are optimized for the listener and the listening environment.

In previous work we have investigated the utility of speech intelligibility metrics for predicting the impact of speech signal processing on intelligibility for normally-hearing (NH) listeners [2]. We evaluated predictions from intrusive signal metrics of intelligibility against the actual performance of listeners. We showed that metrics like STOI [3] and NCM+ [4] gave fair predictions of the likely speech intelligibility to a listener from analysis of the differences between the clean signal and the processed noisy/reverberant signal. Typically intelligibility could be predicted within 2dB SNR [2].

For these metrics to be useful for finding the best signal processing approaches in hearing aids, they need to be developed in two directions: firstly they need to be made non-intrusive, that is capable of working from the noisy signal alone, and secondly they need to take into account the impact of hearing impairment. Well-established means for converting intrusive metrics to non-intrusive use statistical learning methods applied to large databases of speech materials rated by the intrusive metric [5]. How best to modify these metrics to make predictions for HI listeners, however, is still an active area of research.

A typical approach to take hearing loss into account within intelligibility metrics is to incorporate information about the listener into the front-end signal processing: for example a front-end filterbank might be modified to accommodate degradations in frequency sensitivity (auditory thresholds), frequency selectivity (auditory filter bandwidths) and dynamic range (recruitment) [6]. While this approach seems sensible, it relies on the assumption that the difference between NH and HI listeners is well predicted by characteristics of their hearing loss. In turns out that this is not the whole story. When a group of HI listeners are assessed (as we show later in this paper) there is considerable residual variation in performance compared to NH listeners even after taking their hearing loss into account. There is an echo of the Anna Karenina principle: normally hearing listeners are all alike; every hearing impaired listener is hearing impaired in their own way.

A number of explanations are proposed for why HI listeners are more variable than NH listeners. It could be to do with correlations between hearing loss and cognitive decline [7], or that imperfect auditory representations require more cognitive effort to process which tests the ability of the listener to recruit additional processing capacity or working memory [8]. Or it could be that some of the phonological fine tuning used by NH listeners to discriminate phonemes is degraded to different degrees in different listeners.

HI listener variation that is not predictable from characteristics of their hearing loss is a problem for speech intelligibility metrics, since manipulation of the front-end of the metric may not be enough. For the design of metrics to predict the benefit of speech enhancement to HI listeners this is a real problem since the accuracy of a metric based only on hearing loss could be worse than the likely differences between processing approaches (e.g. approaches may only vary by 1dB in effective SNR but estimated intelligibility for a listener might vary by 2dB).

In this paper we study the variability of speech intelligibility performance by HI listeners in terms of two loss functions. The first relates to those aspects related to psychoacoustic measurements of their hearing loss. We call this their cochlear loss. The second relates to everything else, we call this their supra-cochlear loss. The paper then has three goals:

a) What proportion of the variability of HI listener performance on speech intelligibility tasks is predictable from their cochlear loss?
b) What proportion of the variability of an HI listener on a new intelligibility task is predictable from an estimate of their supra-cochlear loss obtained from another intelligibility task?
c) To what degree is supra-cochlear loss independent of the nature of the listening task?

Our approach is a meta-analysis of two existing data sets in which both psychoacoustic measures and speech intelligibility scores are available for a group of HI listeners. We assume that the psychoacoustic measures reflect cochlear processing. We build the best predictive models of performance from these psychoacoustics and interpret the remaining performance variation on some task as supra-cochlear loss. We then explore how estimates of listeners' supra-cochlear losses vary with intelligibility task.

The structure of the paper is as follows: in section 2 we shortly describe the contents of the data sets in terms of the speech intelligibility scores and the psychoacoustic descriptors available in each. We refer to the original papers for details. In section 3 we describe the modelling approach and the performance measures used. In section 4 we present the results of the meta-analysis and in section 5 discuss their implications.

2. Data sets

2.1. Bethesda Data Set

The Bethesda data set was collected by Summers et al [9] at the Walter Reed National Military Medical Center, Bethesda, MD. The listeners on the test comprised 10 normally-hearing and 18 hearing-impaired subjects. It includes the following psychophysical measurements (code in brackets).

- Pure-tone hearing thresholds at 250, 500, 1000, 1500, 2000, 3000, 4000, 6000 & 8000Hz. (H)
- Degree of peripheral amplitude compression at 500, 1000, 2000 & 4000Hz. (C)
- Auditory filter bandwidths at 500, 1000, 2000 & 4000Hz measured at both 70 and 80 dB SPL. (B)
- Frequency modulation detection thresholds measured at 500, 1000, 2000 & 4000Hz. (F)

The intelligibility of speech-in-noise to each listener was measured using IEEE sentences with both speech-shaped noise and amplitude modulated speech-shaped noise at signal-to-noise ratios (SNRs) of -6, -3, 0 and +3 dB. The speech, presented at 92 dB SPL for all listeners, was not equalized to match auditory thresholds.

For subsequent analysis the speech test scores for each listener were converted to Speech Reception Thresholds (SRT). The % scores were first converted to log-odds ratios and then linear regression was used to find the SNR value for the listener which gave a log-odds of 0 (i.e. 50%). In addition all frequency measurements were converted to log Hertz before modelling.

Since our analysis is focused on the HI listeners, the data points of the NH listeners were combined into one average listener.

2.2. Salamanca Data Set

The Salamanca data set was collected by Johannesen et al [10] at the Universidad de Salamanca, Spain. It consists of test scores on 68 hearing-impaired listeners. The following measurements were made of each listener's hearing ability (code in brackets):

- Pure-tone hearing thresholds at 500, 1000, 2000, 4000 & 6000Hz (H)
- Estimate of cochlear mechanical gain loss (also referred to as outer-hair cell loss, OHC) expressed in decibels (dB). (O)
- Basilar-membrane compression exponent (BMCE). It was defined as the slope (in dB/dB) of an inferred cochlear input/output curve over its compressive segment. (C)
- Frequency modulation detection thresholds (FMDTs), defined as the minimum detectable excursion in frequency for a pure tone carrier at 1500Hz. (F)

Speech intelligibility was assessed for speech-shaped noise (SSN) and a time-reversed two-talker masker (R2TM) using HINT sentences. Performance was recorded in terms of SRT score. Speech materials were presented with linear, frequency-specific amplification to compensate for listeners' audiometric losses.

3. Method

Our goal is to model the effects of cochlear and supra-cochlear deficits on speech intelligibility performance as measured in terms of speech reception threshold. To build a model of cochlear loss we perform a regression on the psychoacoustic measurements of each listener to predict their SRT score for each listening task. Since we do not know the form of that function we use support-vector regression (SVR) [11] that makes no assumptions about the form of the function other than listeners with similar psychoacoustics are likely to have similar scores. SVR determines a subset of the data set that can be used as examples (support vectors) against which a new listener can be compared to best predict their score. The final score is then just the linear combination of support vector scores weighted by their distance to the new vector. Since any given listener may be chosen to be one of the support vectors, we must model the data set using cross-validation, where each listener is left out in turn and a predicted score is made from a model trained from the remaining listeners. To build the cochlear loss model, the psychoacoustic measures were divided into four sets as coded in section 2, and each feature set was tested in isolation and in combination with all other feature sets. Features are normalised before modelling. A grid-search is used to find the best SVR hyper-parameters, and final predictions are computed from 10 modelling runs.

Once we have obtained a predicted score for each listener we can compare the prediction against the actual score and determine two performance measures: R^2, the proportion of variance in scores explained by the prediction, and mean absolute error (MAE) of prediction, which answers the question how far away on average is the prediction from the correct answer.

The difference between actual and predicted scores for a listener was used as an estimate of their supra-cochlear loss.

To explore the size and variability of the supra-cochlea loss term, we can calculate this for each one of the listening tasks in the data sets, and evaluate it on the other. We compare actual scores and the prediction from the estimated cochlear loss on each task together with the estimated supra-cochlear loss from the other task in terms of R^2 and MAE.

Finally we can calculate how much the estimate of the supra-cochlear loss varies across the two listening tasks to explore the extent to which the supra-cochlear loss is dependent upon the nature of the task.

4. Results

4.1. Prediction of Cochlear Loss

Table 1 shows the MAE of prediction of the SRTs for speech-shaped noise and modulated speech-shaped noise for hearing impaired listeners in the Bethesda data set for each combination of psychoacoustic features. Table 2 shows the MAE of prediction of the SRTs for speech-shaped noise masker and for a reversed two-talker masker for hearing impaired listeners in the Salamanca data set.

Table 1. SRT Prediction from Psychoacoustics for Bethesda data in MAE (dB)

Group	Features	SSN	Modulated SSN
Baseline	None	1.982	3.215
Single	H	1.462	2.121
Single	C	1.851	2.871
Single	B	1.848	2.957
Single	F	1.539	2.524
Double	H+C	1.493	2.129
Double	H+B	1.477	***2.027***
Double	H+F	***1.434***	2.192
Double	C+B	1.903	3.037
Double	C+F	1.664	2.718
Double	B+F	1.717	2.507
Triple	H+C+B	1.476	2.135
Triple	H+C+F	1.505	2.251
Triple	H+B+F	1.486	2.085
Triple	C+B+F	1.736	2.628
All	H+C+B+F	1.557	2.174

Tables 1 and 2 show that incorporation of psychoacoustic features into the model can improve the prediction of speech intelligibility scores over a baseline prediction based on the mean of the other listeners. For the Bethesda data set, the MAE reduces from 1.982 to 1.434dB for SSN, and from 3.215 to 2.027 for Modulated SSN. The reduction on the Salamanca data set is much smaller, from 1.137 to 1.018dB for SSN, and from 1.244 to 1.006dB for reversed two-talker masker.

The best feature set combinations were different for the different data sets and tasks; these are indicated in bold in the tables. The SRT predictions for the best performing models on the Bethesda data set are plotted in Figure 1. The SRT predictions for the best-performing models on the Salamanca data set are plotted in Figure 2. The proportion of variance explained by the best performing models is shown in the plots..

Table 2. SRT Prediction from Psychoacoustics for Salamanca data in MAE (dB)

Group	Features	SSN	R2TM
Baseline	None	1.137	1.244
Single	H	1.115	1.083
Single	O	1.129	1.155
Single	F	1.101	1.036
Single	C	1.022	1.219
Double	H+O	1.119	1.111
Double	H+F	1.061	1.010
Double	H+C	1.045	1.046
Double	O+F	1.085	1.060
Double	O+C	1.073	1.118
Double	F+C	***1.018***	1.090
Triple	H+O+F	1.092	1.034
Triple	H+O+C	1.049	1.035
Triple	H+F+C	1.040	1.009
Triple	O+F+C	1.060	1.031
All	H+O+F+C	1.041	***1.006***

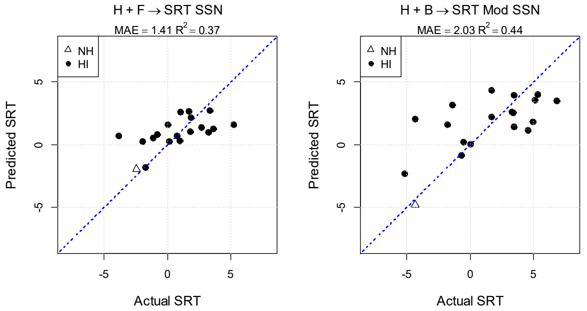

Figure 1. Prediction of SRT from best psychoacoustic features for Bethesda data set (left = SS noise, right = modulated SS noise)

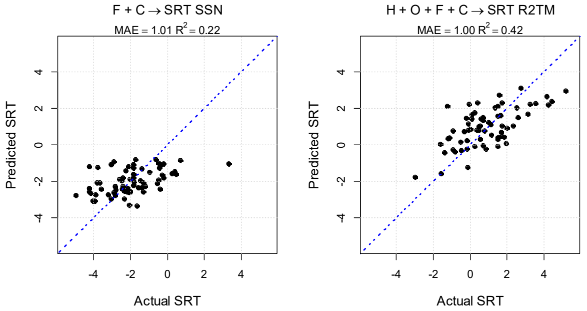

Figure 2. Prediction of SRT from best psychoacoustic features for Salamanca data set (left = SS Noise, right = reversed 2-talker masker)

4.2. Prediction of Supra-Cochlear Loss

The supra-cochlear loss term for each listener for each task is then calculated as the difference between the actual SRT and the SRT predicted from the best feature set for the task. Figures 3 and 4 show the predicted SRT after inclusion of the supra-cochlear loss term. In each case the loss term is computed for the other task. In both data sets and for both tasks, the prediction error is reduced by the inclusion of the supra-cochlear loss, with the MAE reducing to about 1dB for the Bethesda data set and 0.8dB for the Salamanca data set.

Figure 5 shows the correlation between the supra-cochlear loss terms across the two tasks for each of the two data sets. The graphs suggest that the supra-cochlear loss varies by around 1dB on average across the pair of tasks.

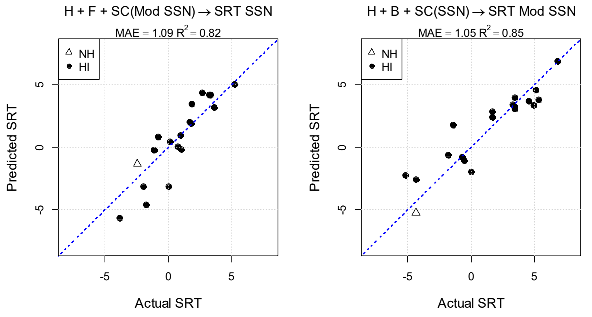

Figure 3. SRT prediction after supra-cochlear loss included in Bethesda data set. Left: SSN score after MSN calibration, right: Mod SSN score after SSN calibration.

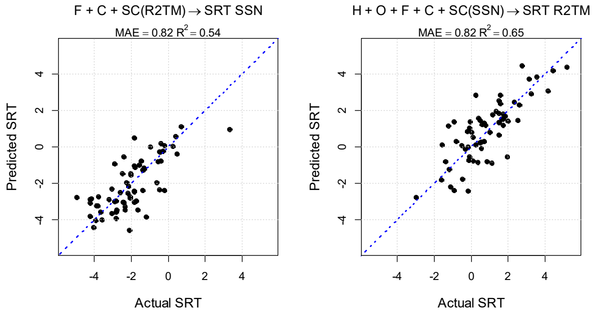

Figure 4 . SRT Prediction after supra-cochlear loss included in Salamance data set. Left: SSN after R2TM calibration, right R2TM after SSN calibration.

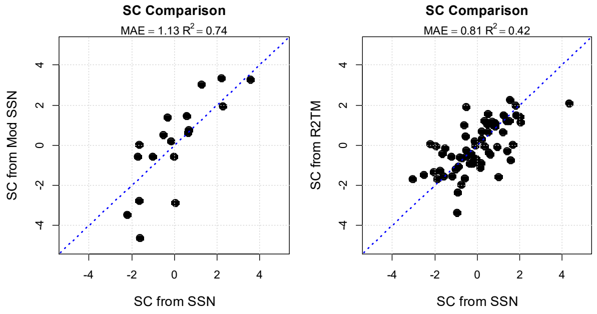

Figure 5. Comparison of supra-cochlear loss across listening tasks. Left Bethesda data set, right Salamanca data set.

5. Discussion

This paper has shown how SRT predictions for HI listeners may be significantly improved using an SVR model of cochlear loss based on the available psychoacoustic measures. Prediction accuracy was better in the Salamanca data set probably because the intelligibility scores were collected with equalization for listener thresholds and so were less variable to begin with. This equalization also explains the different importance given to features in the model, with thresholds being very important features for the Bethesda data set, while for the Salamanca data set, the choice of features made little difference in terms of MAE.

Cochlear loss alone only explained at best 40% of the variability in test scores across HI listeners. Inclusion of a supra-cochlear loss term (calculated from the other task) into the model explains a further 40% of the variation for the Bethesda data set, and a further 20% for the Salamanca data set. The difference is explained by Figure 5, which shows that the two tasks in the Bethesda data set are more similar than those in the Salamanca data set.

Taking both loss terms together we have shown that we can predict second test score performance from first test score performance to within 1dB MAE. This seems within the likely prediction error of a speech signal intelligibility metric.

The residual variability in prediction might come from different sources: (i) experimental error in the collection of the psychoacoustic measures or the intelligibility scores; (ii) the effects of cochlear loss on the task other than that explained by the particular set of psychoacoustic measurements available, or (iii) the task dependency of supra-cochlear loss caused by interactions between the task and cognitive deficits. This interaction might also have arisen if variation in cognition had impact on the collection of psychoacoustic measurements themselves.

In the future, better modelling might arise from: (i) a wider range of psychoacoustic measures – although the evidence presented here suggests that such measures are highly correlated with one another; (ii) a wider range of intelligibility tests per listener to unpack the reasons why supra-cochlear loss is dependent on characteristics of the task; (iii) repeated testing of listeners to obtain estimates of measurement error.

Overall the analysis presented here seems promising for the development of speech signal intelligibility metrics for hearing impaired listeners provided these include a supra-cochlear calibration term for each listener. This might be estimated by incorporating a standardised speech intelligibility test alongside standard psychoacoustic tests in their clinical assessment. The study also makes clear that further work is required to understand the causes of variability in the intelligibility of speech to the hearing impaired.

6. Acknowledgements

Thanks to Van Summers and Enrique Lopez-Poveda for supplying us with the Bethesda and Salamanca data sets.

This work is supported by U.K. Engineering and Physical Sciences Research Council award: “Environment-aware Listener-Optimized Binaural Enhancement of Speech (E-LOBES)” to Imperial College London & University College London.

7. References

[1] B. Kollmeier & J. Kiessling "Functionality of hearing aids: state-of-the-art and future model-based solutions", International Journal of Audiology, DOI: 10.1080/14992027.2016.1256504

[2] G. Hilkhuysen, N. Gaubitch, M. Brookes, M. Huckvale, "Effects of noise suppression on intelligibility II: An attempt to validate physical metrics", J.Acoust.Soc.Am., 135 (2014) 439-50.

[3] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," IEEE Trans. Audio, Speech, Language Process., vol. 19, no. 7, pp. 2125–2136, 2011.

[4] G. Hilkhuysen, N. Gaubitch, M. Brookes, M. Huckvale, "Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios", J.Acoust.Soc.Am. 131 (2012) 531-539.

[5] D. Sharma, G. Hilkhuysen, N. Gaubitch, P. Naylor, M. Brookes, M. Huckvale, "Data driven method for non-intrusive speech intelligibility", 18th European Signal Processing Conference (EUSIPCO-2010) Aalborg, Denmark, August 23-27, 2010.

[6] J. Kates, K. Arehart, "The hearing aid speech perception index (HASPI)", Speech Communication 65, 75-93. doi: 10.1016/j.specom.2014.06.002

[7] C. Füllgrabe, B. Moore, M. Stone, "Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition", Front Aging Neurosci. 2015;6:347. doi: 10.3389/fnagi.2014.00347.

[8] P. Souza, K. Arehart, T. Neher, "Working memory and hearing aid processing: Literature findings, future directions, and clinical applications", Front. Psychol. 2015;6:1894. doi: 10.3389/fpsyg.2015.01894.

[9] V. Summers, M. Makashay, S. Theodoroff, M. Leek, "Suprathreshold auditory processing and speech perception in noise: hearing-impaired and normal-hearing listeners", J.Am. Acad. Audiol 24 (2013) 274-292.

[10] P. Johannesen, P. Perez-Gonzalez, S. Kalluri, J. Blanco, E. Lopez-Poveda, "The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners", Trends in Hearing 2016, Vol. 20: 1–14.

[11] A. Smola, B. Scholkopf, "A tutorial on support vector regression", Statistics and Computing 14: 199–222, 2004.