EFFECTS OF REPLAY ON THE INTELLIGIBILITY OF NOISY SPEECH

GASTON HILKHUYSEN[1], JENNIFER LLOYD[1], AND MARK HUCKVALE[1]

[1] *Department of Speech Hearing and Phonetic Sciences, University College London, London, United Kingdom*
g.hilkhuysen|jennifer.lloyd|m.huckvale@ucl.ac.uk

This study investigates whether repeated listening to a spoken utterance can improve its intelligibility. Forty sentences were distorted with noise and, in one condition, enhanced by noise reduction. Each sentence was presented 8 times. Thirty participants repeated these sentences verbatim and judged how correct they thought their responses were. Outcomes showed that performance improved from 36 to 51% (0.9 Berkson, approximately 1.3 dB) during the first 5 presentations and then saturated. Listeners initially underestimated their performance, but believed it continued to improve after 5 presentations, leading to overestimation. We conclude that replay can improve intelligibility performance but may lead to overconfidence in one's perceptions.

INTRODUCTION

Recordings of speech collected in law enforcement often contain regions that have poor intelligibility. A common strategy to help uncover what was said is to play the audio more than once. Given its widespread use, surprisingly few studies have addressed whether intelligibility is actually improved on replay [1, 2, 3, 4, 5].

In one study [1] words were presented in noise at various signal-to-noise ratios (SNRs). A speaker uttered each word either once or three times consecutively and each utterance was mixed with a fresh fragment of white noise. The authors determined intelligibility performance by measuring the ability of listeners to reproduce the words correctly. Results showed that presenting the word three times had the same effects on performance as lowering the SNR by about 2 dB for a word that was only presented once. The work of [2] replicated this work, but had speakers say the same word up to four times. It was found that there was no benefit of hearing the word more than twice. Additional presentations of the same words had little effect on percentage correct scores.

Both of studies [1,2] used new utterances of the same word mixed to a fresh noise fragment, and thus may not make good predictions about the effects of replay as used in law enforcement settings. In law enforcement recordings speech and noise fragments do not vary across replays. Pollack [3] contrasted the two situations, playing the same word up to six times. In contrast to the previous studies where the benefit of replay saturated after the second presentation, he found that with fresh utterances and fresh noise fragments word-correct scores continued to improve until the sixth presentation. Replaying a single recording of a noisy word resulted in less improvement, but word scores also continued to increase up to the sixth presentation.

In the light of the intelligibility improvements of replayed speech, one could hypothesize that effects of replay may be different for noise suppressed speech. Although at best noise reduction appears to have little effects on intelligibility performance [6, 7], it has been reported that noise reduction can reduce listening effort [8]. If noise reduction does not affect the information in the audio signal, but facilitates human processing of noisy speech, one could imagine that the effects of replay might be stronger for enhanced noisy speech and that the effect might saturate earlier.

All of the previous studies on the effects of replay investigated whether listeners were able to reproduce the presented speech. Because the presented words were known, intelligibility performance could be measured. However, in law enforcement audio the words are not known, since one attempts to decipher the message from the recording. The decision about the correct perception of the message and the decision to replay the recording again or not will be based on a belief held by the listener, i.e., on opinions.

Various studies have addressed the question whether intelligibility opinions correspond to intelligibility performance, e.g. [9, 10]. Outcomes show that opinions can be measured reliably and that listeners' opinions correspond well with their performance.

Intelligibility opinions have been addressed in the context of replay [3]. Besides repeating the perceived words, listeners rated their confidence in their responses. The author assumed that in real life listeners would stop replaying once they had reached a certain level of confidence. Simulations showed that placing the stop criterion at lower confidence levels reduced intelligibility performance. Unfortunately, the article

provides no information about how confidence ratings changed with increasing presentations.
Given the outcomes of the studies mentioned above, this paper focuses on the following questions: 1) Does intelligibility performance increase when presenting the same noise distorted speech recording multiple times, and if so, when does the effect saturate? 2) What are the effects of noise reduction in the context of the intelligibility of replayed speech? 3) Can intelligibility opinions be used to decide when intelligibility performance has saturated?

1 METHODS

1.1 Participants

The study included 30 listeners (16 female, 14 male) with ages ranging from 18 to 56 years (median 22). All had attended primary school in the UK and always spoken English as their first language. Their pure-tone hearing thresholds at octave frequencies ranging from 0.125 to 8 kHz were at 20 dB HL or below [11]. None of the listeners had ever been exposed to the speech materials used in this study.

1.2 Materials

Stimuli were a selection of IEEE sentences [12] spoken by a male speaker [13]. Each of these sentences contains five keywords. Stimuli were generated in Matlab with a sampling rate of 16 kHz using 64-bit floating point representations, and presented diotically over HDA-200 headphones connected to a RME Fireface 400 D-A converter. The speech level before additive noise and eventual subsequent noise reduction was fixed at 65 dB SPL. This level was calibrated using an artificial ear (B&K 4153) equipped with a ½" microphone (B&K 4192), a microphone power supply (B&K 2804) and a spectrum analyzer (OnoSokki cf-350 z).

1.3 Stimuli

The intelligibility was deteriorated by adding car-cabin noise at -15dB SNR or babble at -6 dB SNR. A detailed description of both noises and the process that generated the stimuli has been given elsewhere [7]. In short, a target sentence was embedded into two other sentences. This sentence triplet was mixed to a randomly selected noise fragment with a duration equivalent to the triplet. For enhanced speech, the triplet was processed with a variant of the minimum means squared estimator of the log-spectrum [14]. The noise estimator, required by this noise reduction algorithm, was based on optimal smoothing and minimal statistics [15, 16]. All noise suppression parameters were set as in [7]. But in contrast to that study, the signal after noise reduction was mixed to the original noisy speech with a 1:1 ratio. This resulted in a less aggressive variant of noise reduction. Finally, the target sentence was extracted from the triplet. Using a triplet instead of just a target sentence gave the noise estimator time to stabilize before processing of the target.

1.4 Design

The two noise types combined with the noise reduction switched off or on gave rise to four experimental conditions. Forty IEEE sentences were divided into 10 subsets of 4 sentences each. Per subset sentences were assigned to the four experimental conditions following a Latin square. The presentation order of these subsets was also varied according to a Latin square.

1.5 Procedures

Measurements took place in a sound attenuated listening booth. The experimenter first informed the participant about the study's objectives. Following consent, the participant's audiogram was measured. Subsequently, the experimenter instructed the participant about the listening task: i.e. repeating verbatim the sentence played over the headphones. The intelligibility of the sentence would be poor, but the same stimulus would be presented eight times consecutively. The participant was asked to respond after each presentation, encouraged to provide correct responses as early as possible and to guess even if this would result in an incomplete or nonsense sentence. Additionally, the participant was asked to estimate the accuracy of each response. The participant rated the percentage of the words that the participant thought to have heard correctly, using a rating scale from 0 to 100% with 20% steps. The experimenter scored whether the verbatim response contained the keywords of the sentence, outcomes that will be addressed as performance scores (φ). The experimenter also noted the accuracy as estimated by the participant. These will be addressed as opinion scores (ψ). As proposed by [7] a logit transform was applied to both the performance and opinion scores, as in:

$$\Phi = \log_2(\varphi/(100-\varphi)) \quad (1)$$

and

$$\Psi = \log_2(\psi/(100-\psi)) \quad (2)$$

where Φ and Ψ denote intelligibility performance and opinion, respectively. Both quantities can take values from $-\infty$ to ∞ and have Berkson (Bk) units. One Bk increase signifies that for a fixed number of incorrect words, the number of correct words has doubled. Statistical significance of shifts in Φ and Ψ was addressed with mixed-effects logistic regression (e.g. [17]) using the lme4 package [18] available for R [19].

Fixed factors were coded with treatment contrasts, unless stated differently. Model selection was based on the comparisons of nested models. Starting with the model that included all the interactions of the factors of interest, interactions and potentially main effects were removed in subsequent steps only if their exclusion had no significant consequences for the model's log likelihood. Statistical significance of levels within the factors was addressed with the Wald statistic. 95% confidence intervals will be reported by values placed within brackets.

2 RESULTS

2.1 Performance

Averaged across all participants, noise types, and suppressor conditions, Φ attained -0.8, -0.2, 0.1, 0.1, 0.1, 0.2, 0.2 and 0.2 Bk at succeeding presentations. Intelligibility performance seems to improve up to the sixth presentation of the stimulus, but to plateau with additional presentations. Fig. 1 specifies the effects of replay for the various combinations of noise types and suppressor. Circles and squares denote car-cabin noise and babble, respectively. Open and filled markers differentiate between conditions without and with noise suppression. Intelligibility performance in Berksons is expressed on the left hand axis. In previous research [7] it has been found that in conditions similar to the ones used here Φ is almost linearly related to SNR, i.e., an increase of 0.7 Bk corresponds approximately to an increase of 1 dB SNR. Vertical dotted lines mark such 1 dB SNR shifts. Labels on the right express performance in traditional word correct scores.

Intelligibility performance was lowest for car-cabin noise without noise reduction and highest for car-cabin noise with noise reduction. Noise reduction had little effect on the intelligibility of speech in babble. Effects of replay appear similar across noise types and suppressor conditions, i.e., curves in Fig. 1 have similar shapes and seem to be shifted along the ordinate.

Statistical significance of the effects visible in Fig. 1 was addressed by mixed effects logistic regression including random factors for Participants {1..30} and Sentences {1..40} as well as fixed factors for Noise type {car-cabin, babble}, Suppressor {off, on} and Presentations {1..7}. Stepwise elimination of insignificant interactions resulted in a model that contained the main effects Participants, Sentences, Noise type, Suppressor, Presentations and the interaction Noise type × Suppressor. The fact that the interactions Noise type × Presentations and Suppressor × Presentations did not attain significance confirms the observation that curves in Fig. 1 are shifted among the ordinate. Noise suppression shifted the curve for car-cabin noise 0.3 Bk [0.3, 0.4] upwards. For speech without noise suppression but using babble instead of car-cabin noise intelligibility performance improves by 0.2 Bk [0.1, 0.2]. With noise suppression, replacing car-cabin noise by babble deteriorates intelligibility performance by -0.4 Bk [-0.5, -0.3].

The plateau effect was addressed with treatment coding, such that the performances in the first to seventh presentation were compared to the intelligibility in the eighth presentation. The regression model addressed these shifts for speech in car-cabin noise without noise reduction. Intelligibility was significantly lower for the first to fourth presentation (b_1 = -1.1 [-1.2, -1.0]; b_2 = -0.5 [-0.6, -0.4]; b_3 = -0.3 [-0.4, -0.2]; b_4 = 0.1 [-0.3, -0.0]). Intelligibility in the fifth to seventh presentation did not differ significantly from the eighth presentation (b_5 = 0.1 [-0.2, 0.0], b_6 = 0.0 [-0.2, 0.1], b_7 = 0.0 [-0.1, 0.1]). Hence the effects of replay appear to stabilise after presenting the stimulus five times. Since all interactions with Presentation were not significant, the effects are similar for the babble and conditions with noise suppression.

2.2 Opinion

Averaged across all participants, noise types, and suppressor conditions, Ψ attained -1.3, -0.6, 0.2, 0.1, 0.3, 0.4, 0.5 and 0.5 Bk at succeeding presentations. Intelligibility opinion appeared to increase up to the seventh presentation of the stimulus. Fig. 2 displays the effects for the various listening conditions. Markers correspond to Fig. 1. Intelligibility opinion is highest for speech in car-cabin noise with noise suppression. For all car-cabin noise and babble without as well as for babble with noise reduction intelligibility opinions are about 0.3 Bk less. Effects of replay seem stable across all listening condition: curves in Fig. 2 coincide or are shifted along the ordinate.

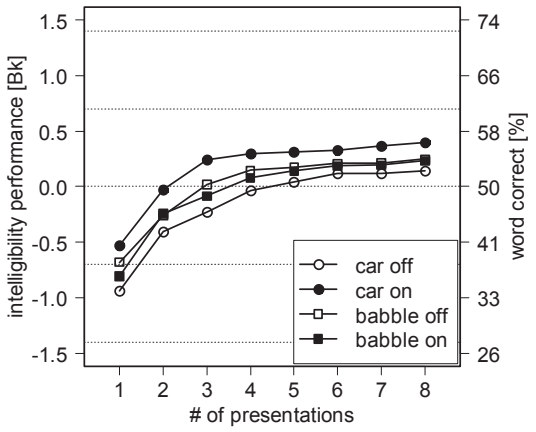

Figure 1: intelligibility performance across presentations. Open and closed markers for conditions without and with noise reduction, respectively. Dotted lines correspond to shifts of 1 dB SNR.

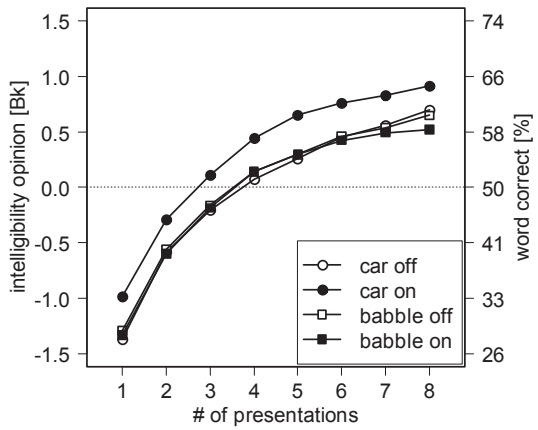

Figure 2: intelligibility opinions across presentations. Open and closed markers for conditions without and with noise reduction, respectively.

The effects visible in Fig. 2 were addressed with mixed effects logistic regression, including the random effects Participants and Sentences, and the fixed effects Noise type, Suppressor and Presentations. Stepwise elimination removed the following interactions: Noise type × Presentations; Suppressor × Presentations; and Noise type × Suppressor × Presentations. Noise reduction shifted the intelligibility opinion for car-cabin noise upward by 0.4 Bk [0.3, 0.4]. For the conditions without noise reduction, changing the car-cabin noise into babble had no effect (0.0 Bk [-0.1, 0.1]. With noise reduction, this change reduced the intelligibility for babble by -0.4 Bk [-0.5, 0.3]. In other words, the 0.4 Bk effect of noise reduction for car-cabin noise was absent for babble. The plateau effect was addressed by comparing the intelligibility at first to seventh presentation with the eighth presentation. Except for seventh presentation, all these differences were significant (b_1 = -2.1 Bk [-2.2, -2.0]; b_2 = -1.3 Bk [-1.4, -1.2]; b_3 = -0.9 Bk [-1.0, -0.8]; b_4 = -0.5 Bk [-0.6, -0.4]; b_5 = -0.3 Bk [-0.5, -0.3]; b_6 = -0.2 Bk [-0.3, -0.1]; b_7 = -0.1 Bk [-0.2, 0.0]). Hence, intelligibility opinion reached its plateau at the seventh presentation.

2.3 Performance and opinion

For both intelligibility opinion and performance no significant Noise type × Presentations or Suppressor × Presentations interactions were observed. Hence, to investigate the difference between opinions and performance, both were calculated while averaging φ and ψ across noise types and noise reduction conditions, and transforming percentage word correct scores into Berksons following equations 1 and 2. The resulting difference between opinions and performance (Ψ - Φ) is visualized in Fig. 3. It seems that performance is underestimated for first, second and third presentation, while overestimated for the fourth to eighth presentation. Statistical significance of this effect was addressed with mixed effects logistic regression predicting word correct scores. The analysis included the fixed factor Intelligibility assessment with two levels {opinion, performance}, and all other factors used previously. Noise type and Suppressor were coded using summary contrasts; hence their regression coefficients represented shifts from overall means. Treatment coding of Presentations was such that coefficients represented effects relative to the fourth presentation. A similar coding strategy was used for Intelligibility assessment: its regression coefficient represented Ψ minus Φ. Elimination of non-significant interactions resulted in a regression equation including all main effects as well as the interactions: Noise type × Suppressor; Noise type × Intelligibility assessment; and Presentations × Intelligibility assessment.

Inspection of the Noise type × Suppressor and Noise type × Intelligibility assessment interactions confirmed the observations made from the separate analysis of Ψ and Φ. The main effect of Intelligibility assessment was non-significant (b_{op} = 0.1 Bk [0.0, 0.2]), indicating that performance and opinion were equal at the fourth presentation. Coefficients for the Presentations × Intelligibility assessment interaction were: $b_{1,op}$ = -0.6 Bk [-0.8, -0.5]; $b_{2,op}$ = 0.4 Bk [-0.5, -0.2]; $b_{3,op}$ = -0.2 Bk [-0.3, -0.0]; $b_{5,op}$ = 0.1 Bk [0.0, 0.3]; $b_{6,op}$ = 0.3 Bk [0.1, 0.4]; $b_{7,op}$ = 0.3 Bk [0.2, 0.5]; and $b_{8.op}$ = 0.4 [0.2, 0.5]. Hence for the first, second and third presentation, intelligibility opinion is smaller than performance. The reverse holds for the sixth, seventh and eighth presentation.

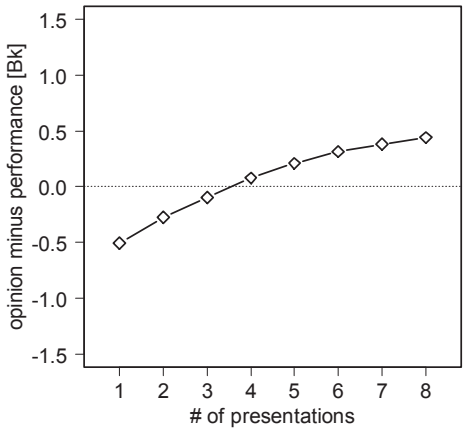

Figure 3: opinion-performance mismatch across presentations.

3 DISCUSSION

Across the first five presentations, intelligibility performance increased by 0.9 Bk. Expressed in percentages, words scores went up from 36 % to 51 %. Assuming the linear relation between Berkons and SNR [7], the benefit of replay approximates 1.3 dB. In other words, after the five presentations, performance was equal as for a speech signal with a 1.3 dB higher SNR during its first presentation. The effect is smaller than the benefit of about 2 dB reported by [1]. This is probably due to the fact in the earlier studies utterances and noise fragments changed across replays. As argued [3], different speech sound may become available on successive presentations while presenting independent utterances and noise fragments. This is less likely to occur with recorded noisy speech. For fresh utterances and noise, Pollack [3] reported a shift from 32% at the first presentation up to about 55% at the sixth presentation. With his recorded speech, words scores only improved to 41%. This is also substantially less than found here, which could be due to the size of the response set. In [3], the stimuli and responses were restricted to a set of 32 words. In the current study the response set approximated infinity, i.e., participants could respond with any word. It has been found that benefit of replay increased with the number of possible responses [1].

Most remarkable is the intelligibility performance after the first presentation at -0.9 Bk and 0.7 Bk for speech in car-cabin noise and babble, respectively. Using identical noise and speech materials at equivalent SNRs while presenting the stimuli only once, secondary analysis of the data reported by [7] revealed intelligibilities of -0.3 Bk and -0.1 Bk, respectively. This gives the impression that part of the replay benefit is caused by the fact that listeners perform less well at the first presentation when knowing they are going to listen to the recording multiple times. But besides replay, other differences between the current study and [7] may have contributed to the difference in intelligibility performance at the first presentation. The study by [7] also included stimuli at higher SNRs that are easier to respond to correctly. From our unpublished work we know that participants perform less well when only subjected to poorly intelligible speech.

Presenting the noisy speech more than five times did not improve intelligibility performance further. This finding corresponds to impressions from the work of [3], but contrasts with the finding by [2]. The latter reports saturation just after the second presentation. However, his study used new utterances and new noise fragments in each presentation, which could account for a faster saturation than reported here. Our previous observation that listeners initially perform less well when presented to the same recording multiple times, provides a second explanation. It is possible that performance saturated earlier in the study by [2], because the stimuli were only presented four times. While replaying a recording less often, the benefit of replay may saturate faster.

Only few studies [20, 21] showed improved intelligibility performance due to noise suppression. The current results show a small but significant increase of 0.3 Bk for speech in car-cabin noise only, corresponding to an improvement of 0.5 dB SNR. Other work [7] employed the same algorithm, car-cabin noise and SNR but did not mix the enhanced speech to the noisy speech. They reported a non-significant deterioration in performance of 0.2 Bk. The current results reconfirm that effects of single microphone noise reduction are noise dependent [6, 7]. Deteriorations due to speech enhancement or absence of any effects on intelligibility performance could be due to too much gain reduction. Mixing the enhanced speech with the original noisy speech limits the gain reduction, and improved performance in car-cabin noise. Nevertheless, the hypothesized interaction of noise reduction with replay did not occur. The idea that listeners process enhanced noisy speech differently while replaying recordings could not be confirmed.

Pollack [3] asked his participants to rate the confidence in their responses. In the current study participants were asked to estimate the percentage of words perceived correctly. This latter approach corresponds to other studies on intelligibility opinions and performance, [9, 10].

Noise reduction had similar effect on performance and opinions: with car-cabin noise both increased by about 0.3 Bk; in babble no effects could be observed. But performance in babble was 0.2 Bk better than in car-cabin noise without noise reduction, while intelligibility opinion showed no difference.

Participants underestimated performance during the first three presentations when performance was below 0 Bk. Additionally, they overestimated intelligibility performance once the stimulus had been presented more than five times and performance was above 0 Bk. It could be that in general listeners underestimate low performance and overestimate high performance. Such an expansion of opinions relatives to performance would account for the effects visualized in Fig. 3. However, this explanation is not supported by studies that addressed the relation between intelligibility opinions and performance. In one study no systematic under- or overshoot depending on intelligibility performance was observed [9]. Other work [10] found some evidence for overshoot when performance was high, but observed no undershoot while performance was low. The current data provide some evidence against an explanation based on expanded intelligibility opinions. After the fifth presentation, performance saturates while opinion still improves. Consequently we think that the improving opinions are not related to performance as such, but result from replay. Presenting

a recording multiple times gives listeners the false impression that they are hearing more words correctly.

4 SUMMARY

In the current study, listeners improved their performance in a listening task, while repeatedly being presented the same distorted noisy speech recording. Compared to their performance at the first presentation, performance intelligibility improved by 0.9 Bk, comparable to a 1.3 dB reduction in SNR. This improvement was observed during the first five presentations. Presenting the same signal more and up to eight times gave no significant improvement. These effects were independent from noise type or additional noise reduction, Noise reduction improved the performance for speech distorted by car-cabin noise, an improvement that was equivalent across presentations.
Although performance saturated after five presentations, listeners believed that additional presentations still improved their ability to reproduce the speech correctly. While initially underestimating their performance, it was overestimated after the fifth presentation. Hence, although replay of noisy speech recordings can improve their factual reproduction, it augments the risk that listeners think they heard right what they factually heard wrong.

REFERENCES

[1] G. A. Miller, G. A. Heise, and W. Lichten, "The Intelligibility of Speech as a Function of the Context of the Test Materials" *Journal of Experimental Psychology*, vol. 41, pp. 329-335 (1951).

[2] E. J. Thwing, "Effect of Repetition on Articulation Scores for Pb Words" *Journal of the Acoustical Society of America*, vol. 28, pp. 302-303 (1956).

[3] I. Pollack, "Message Repetition and Message Reception" *Journal of the Acoustical Society of America*, vol. 31, pp. 1509-1515 (1959).

[4] M. Haggard, "Selectivity versus summation in multiple observation tasks: evidence with spectrum parameter noise in speech" *Acta Psychologica*, vol. 37, pp. 285-99 (1973).

[5] J. E. Clark, P. Dermody, and S. Palethorpe, "Cue Enhancement by Stimulus Repetition - Natural and Synthetic Speech Comparisons" *Journal of the Acoustical Society of America*, vol. 78, pp. 458-462 (1985).

[6] Y. Hu, and P.C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms" *Journal of the Acoustical Society of America*, vol. 122, pp. 1777-1786 (2007).

[7] G. Hilkhuysen, N. Gaubitch, M. Brookes, and M. Huckvale, "Effects of Noise Suppression on Intelligibility: Dependency on Signal-to-Noise Ratios" *Journal of the Acoustical Society of America*, vol. 131, pp. 531-539 (2012).

[8] A. Sarampalis, S. Kalluri, B. Edwards, and E. Hafter, "Objective measures of listening effort: effects of background noise and noise reduction" *Journal of Speech Language and Hearing Research*, vol. 52, pp. 1230-1240 (2009).

[9] K. M. Cienkowski, and C. Speaks, "Subjective vs. objective intelligibility of sentences in listeners with hearing loss" *Journal of Speech Language and Hearing Research*, vol. 43, pp. 1205-1210 (2000).

[10] C. M. Rankovic, and R. M. Levy, "Estimating articulation scores" *Journal of the Acoustical Society of America*, vol. 102, pp. 3754-3761 (1997).

[11] International Organization for Standardization. *ISO 389-8:2004, Reference Zero for the Calibration of Audiometric Equipment—Article 8: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Circumaural Earphones*, Geneva CH, (2004).

[12] E. H. Rothauser, W. D. Chapman, N. Guttman, H. R. Silbiger, M. H. L. Hecker, G. E. Urbanek, K. S. Nordby, and M. Weinstock, IEEE Recommended Practice for Speech Quality Measurements *IEEE Transactions on Audio and Electroacoustics*, vol. AU17, pp. 225-246 (1969).

[13] M. W. Smith, and A. Faulkner, "Perceptual adaptation by normally hearing listeners to a simulated "hole" in hearing" *Journal of the Acoustical Society of America*, vol. 120, pp. 4019-4030 (2006).

[14] Y. Ephraim, and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator" *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 33, pp. 443-445 (1985).

[15] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and

Minimum Statistics" *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 504-512 (2001).

[16] R. Martin, "Bias Compensation Methods for Minimum Statistics Noise Power Spectral Density Estimation" *Signal Processing*, vol. 86, pp. 1215-1229 (2006).

[17] T. F. Jaeger, "Categorical Data Analysis: Away From Anovas (Transformation or Not) and Towards Logit Mixed Models" *Journal of Memory and Language*, vol. 59, pp. 434-446, (2008).

[18] D. Bates, and M. Maechler, *lme4: Linear mixed-effects models using S4 classes,* http://CRAN.R-project.org/package=lme4 (2010).

[19] R Development Core Team. *R: A language and environment for statistical computing*, http://www.R-project.org (2009).

[20] K. H. Arehart, J. H. L. Hansen, S. Gallant, and L. Kalstein, "Evaluation of an Auditory Masked Threshold Noise Suppression Algorithm in Normal-Hearing and Hearing-Impaired Listeners" *Speech Communication*, vol. 40, pp. 575-592 (2003).

[21] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech Enhancement Based on Audible Noise Suppression" *IEEE Transactions on Speech and Audio Processing*, vol. 5, pp. 497-514 (1997).