Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2pSCb: Speech Intelligibility (Poster Session)

2pSCb11. Can physical metrics identify noise reduction settings that optimize intelligibility?

Gaston Hilkhuysen* and Mark Huckvale

***Corresponding author's address: Speech Hearing and Phonetic Sciences, University College London, Chandler House, London, WC1N 1PF, (No states in the UK you silly system), United Kingdom, g.hilkhuysen@ucl.ac.uk**

Five intelligibility metrics (SII, CSIImid, STOI, sEPSM and fAI) were used to identify the optimal settings of two parameters of a noise reduction system. The change in each metric's values across parameter settings were compared to listeners' performance on an intelligibility test with the enhanced signals as measured by Hilkhuysen et al. (in press, doi:10.1044/1092-4388). Metrics SII, CSII_{mid} and STOI predicted benefits for settings that in fact deteriorated intelligibility. Metric sEPSM identified some optimal settings but ignored the effect of one of the parameters. Metric fAI successfully identified all optimal settings.

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Noise reduction algorithms attempt to attenuate unwanted noise in an audio recording, while preserving the wanted signal. They are commonly found in many audio systems, such as mobile phones, hearing aids and digital audio workstations (DAWs). Most algorithms include various parameters that allow fine tuning of the noise reduction to a particular audio environment. But although these parameter settings may influence the intelligibility of the noise-reduced noisy speech (Lim, 1978, Jorgensen & Dau, 2011), little guidance is available for determining settings that optimize or at least conserve intelligibility. Consequently, the choice of parameter settings is left to the developer of the algorithm, the dispenser of the hearing aid or the audio engineer working on the DAW. But even experts find it hard to choose settings which optimize intelligibility. Hilkhuysen *et al.* (in press) found that the settings experts chose while adjusting two parameters of a commercial noise reduction system were different to one another and at times detrimental to intelligibility.

Various authors have attempted to predict the intelligibility of noise-reduced noisy speech on the basis of the physical properties of the audio signals involved. Physical metrics could potentially provide an objective guide to optimize parameters. Goldsworthy and Greenberg (2004) assessed eight intelligibility metrics in the context of noise reduction and found all metrics to be poor predictors. Ma *et al.* (2009) evaluated 25 metrics on the basis of speech distorted with combinations of four noise types, two signal-to-noise ratios and eight noise reduction algorithms. They found that the Coherence based Speech Intelligibility Index based on mid levels of speech (CSII_{mid}) (Kates & Arehart, 2005) gave the highest correlation between predicted and observed intelligibilities. Taal *et al.* (2011) introduced a method to determine the intelligibility of noisy speech processed with binary masks, the Short Time Objective Intelligibility (STOI) metric. Besides predicting the effects of binary masks, the authors reported that STOI also worked well for a single microphone noise reduction algorithm. Jorgensen and Dau (2011) adapted a model developed by Dau *et al.* (1997) to account for the effects of noise reduction, and labeled their new metric the speech-based envelope power spectrum model (sEPSM). While studying the effects of changing one parameter of a noise reduction algorithm, they found a high correlation between predictions and observed intelligibilities. In another attempt to account for the intelligibility effects of noise reduction, Loizou and Ma (2011) formulated the fractional articulation index (fAI) inspired by the Articulation Index (AI), proposed by French and Steinberg (1947). Using the same data set as Ma *et al.* (2009), they found that fAI gave rise to correlations similar to the ones previously obtained with CSII_{mid}.

Given the promising results reported by Ma *et al.* (2009), Taal *et al.* (2011), Jorgensen and Dau (2011), as well as Loizou and Ma (2011) the current paper explores whether these metrics can be used to determine optimal parameter settings. Using the data reported by Hilkhuysen *et al.* (in press), we examined whether SII, CSII_{mid}, STOI, fAI, and sEPSM are able to locate settings that optimize intelligibility. Additionally results from the Speech Intelligibility Index (SII) (ANSI, 1997) are included, to provide more information about the behavior of the noise reduction system being tested.

METHOD

SII, CSII_{mid} and STOI were calculated as respectively described in the standard (ANSI, 1997), Kates and Arehart (2005) and Taal *et al.* (2011). sEPSM was implemented as presented by Jorgensen and Dau (2011). In the final steps of its calculation a combined SNR in the envelope domain is transformed into a percentage correct score. This resulted in near-perfect performance for all the conditions considered here. The combined SNR in the envelope domain nevertheless showed considerable variance; hence we report outcomes on the basis of this measure, labeling it sEPSM. In the presentation of the fAI, Loizou and Ma (2011) used signals with telephone bandwidth and restricted the metric's sensitivity to frequency bands up to 3.6 kHz. Since the speech used in Hilkhuysen *et al.* (in press) contained frequencies up to 8 kHz the fAI was extended to include frequency bands up to 6.25 kHz. The output of every metric should be monotonically related to intelligibility with higher values of a metric indicating better intelligibility.

The SII (ANSI, 1997) is based on the long-term average speech and noise levels. Hagerman and Olofsson (2004) proposed a method to determine these levels after noise reduction, which was used here. The noisy speech signal was processed twice: once with a mixture of speech and noise having their original phase relation and once with a 180 degrees phase shift on the noise only. Adding or subtracting the two noise-reduced signals gave estimates of the noise-reduced speech and noise-reduced noise signals, respectively. The noise-reduced noise was also used in calculation of sEPSM to determine the modulations in the noise after noise reduction.

Speech sentences were equalized in level, concatenated, mixed to car noise at -12 dB SNR, and processed by the commercial noise reduction system, using the same audio files and noise reduction system as Hilkhuysen *et al.* (in press). Two out of the eight parameters available on the system were varied across four different settings equally spaced across their ranges. This resulted in 16 different combinations of parameter settings. The two parameters addressed as X and Y had ranges of 0 to 40 dB and 0 to 100%, respectively. Set(x,y) will be used to denote a particular parameter combination, where x and y indicate the values of parameters X and Y. The noisy speech was processed by the noise reduction system for each of the 16 parameter combinations. Hilkhuysen *et al.* (in press) provide a more detailed description of the noise reduction system. Metrics were calculated using 45 s fragments of the signals involved.

RESULTS

Panel A in Figure 1 shows the intelligibility effect of noise reduction as reported by Hilkhuysen *et al.* (in press). Numbers represent the difference in percentage word correct scores before and after noise reduction as measured in a listening experiment, with negative numbers indicating a loss of intelligibility due to noise reduction. Intelligibility with set(0,33) is higher than for unprocessed speech, but this difference did not attain statistical significance. When both parameters were set at nonzero values, intelligibility deteriorated. When one of the parameters was set at zero, the system showed no effects on intelligibility, which were the optimal settings for noise reduction system under study.

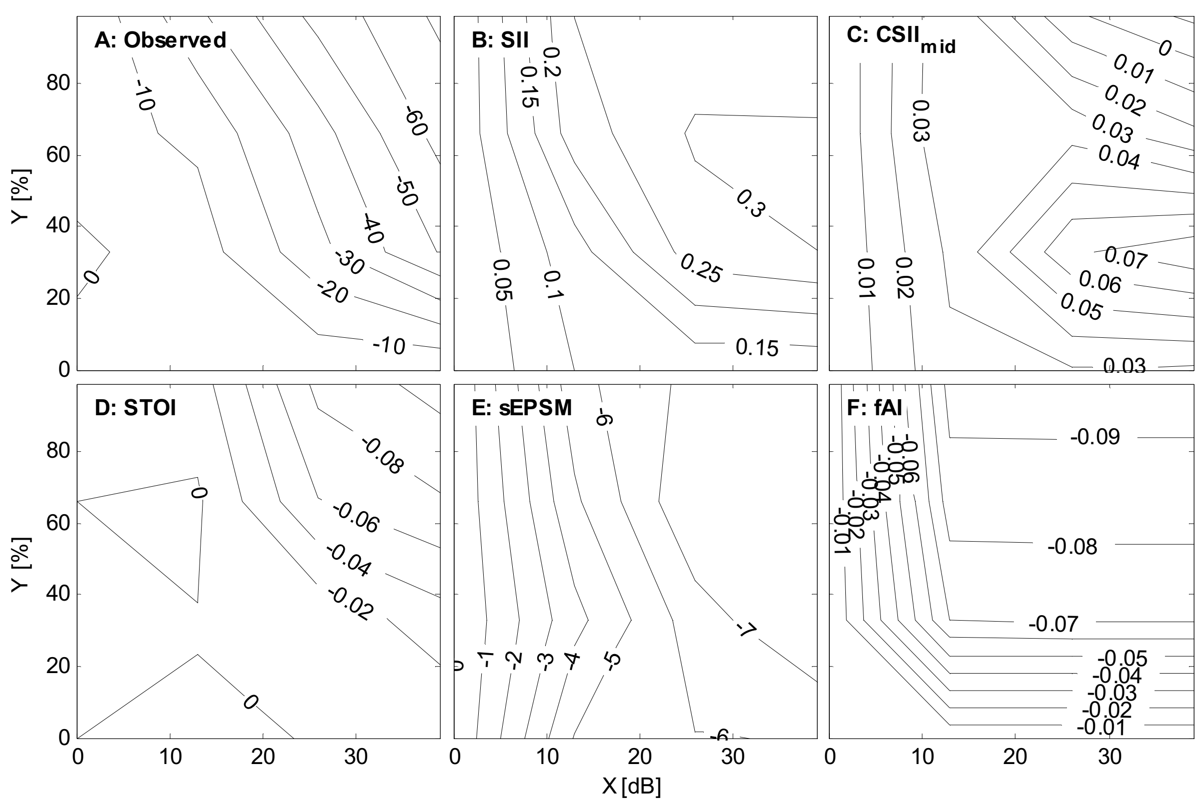

FIGURE 1. Contour plots of observed and predicted effects of noise reduction settings. Panel A: shifts in percentage word scores as observed in Hilkhuysen *et al.*(in press). Panels B to F: shifts in intelligibility metrics as a function of two noise reduction parameters for SII, CSII$_{mid}$, STOI, sEPSM and fAI, respectively.

Panels B and C show the changes caused by noise reduction in the values of SII and $CSII_{mid}$, respectively. Numbers indicate the difference between the metrics' values before and after noise reduction. All contour lines exhibit positive values, indicating that both metrics predict intelligibility improvements. Out of the 16 combinations considered, SII and $CSII_{mid}$ consider set(39,66) and set(39,33) to optimize intelligibility, respectively. Results for STOI are represented in Panel D. This metric successfully predicts deteriorating intelligibilities when both parameters are set at high values, but falsely foresees intelligibility improvements for set(13,0) and set(13,66). The latter is considered to give optimal intelligibility. Panel E displays the outcomes for sEPSM. This metric anticipates deteriorations in intelligibility, but wrongly foresees little effect of parameter Y, whereas changes in parameter X are assumed to provoke the strongest effects. fAI, whose results are displayed in Panel F correctly predicts deterioration in intelligibility when both parameters are set at nonzero values. The metric accurately predicts that settings with one or both parameters at zero will optimize intelligibility

DISCUSSION

SII and $CSII_{mid}$ incorrectly predict improvements in intelligibility where Hilkhuysen *et al.* (in press) observed deteriorations. Set(39,66) and set(39,33) as suggested by SII and $CSII_{mid}$, respectively, factually deteriorate intelligibility by 60 and 53 percent. This makes these metrics inappropriate for parameter optimization. We nevertheless consider the finding with SII informative. Based on the long-term average levels of speech and noise, its positive values indicate that the noise suppressor successfully removed more energy from the noise than from the speech.

The poor result for $CSII_{mid}$ may be unexpected given its high correlation of 0.92 with observed intelligibility scores reported in Ma *et al.* (2009). However these authors considered differences among noise suppression algorithms at various SNRs. Here effects of noise reduction relative to unprocessed noisy speech at one SNR are involved.

STOI correctly predicts deteriorations in intelligibility when both parameters have high values as observed by Hilkhuysen *et al.* (in press). The metric however incorrectly predicted some improvement in intelligibility, coincidentally suggesting similar settings to those proposed by the human experts (set(13,48)) in Hilkhuysen *et al.* (in press).

sEPSM identified some but not all optimal settings. Based on the spurious modulations that noise suppression may impose on the noise, the metric appears to capture some but not all signal properties that deteriorate speech intelligibility after noise reduction.

The contour plot of fAI differs strongly from the one based on the observed intelligibilities. The contours from fAI are primarily parallel to the two axes, while showing a more diagonal orientation in the observed intelligibilities. This implies that fAI might not be an accurate predictor of the absolute intelligibility, because different values of fAI correspond to similar intelligibilities. fAI is nevertheless able to identify settings that optimize intelligibility, i.e. settings in which one or both parameters are set to zero.

CONCLUSION

While trying to determine optimal parameter settings of a commercial noise reduction on the basis of speech intelligibility metrics, SII, $CSII_{mid}$ and STOI predicted intelligibility improvements for settings that in fact deteriorated intelligibility. sEPSM successfully suggested some settings, but failed to capture the effects of one of the two parameters considered. fAI was the only intelligibility metric to correctly identify all optimal settings.

ACKNOWLEDGMENTS

This study was performed within CLEAR, a research project run jointly by Imperial College London and University College London, funded by the UK Home Office.

REFERENCES

ANSI **(1997)**. “Methods for calculation of the speech intelligibility index,” S3.5–1997 (American National Standards Institute, New York).

Dau, T., Kollmeier, B. and Kohlrausch, A. **(1997)**. “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,” J. Acoust. Soc. Am. **102**, 2892-2905.

French, N.R. and Steinberg, J.C. **(1947)**. “Factors governing the intelligibility of speech sounds,” J. Acoust. Soc. Am. **19**, 90-119.

Goldsworthy, R.L. and Greenberg, J.E. **(2004)**. “Analysis of speech-based Speech Transmission Index methods with implications for nonlinear operations,” J. Acoust. Soc. Am. **116**, 3679-3689.

Hagerman, B. and Olofsson, A. **(2004)**. A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. Acta Acust. united Acust , **90**, 356-361.

Hilkhuysen, G., Gaubitch, N. and Huckvale, M. **(in press)**. “Effects of noise suppression on intelligibility: experts’ opinions and naïve normal hearing listeners’ performance,” J. Speech Lang. Hear. Res.

Jorgensen, S. and Dau, T. **(2011)**. “Predicting Speech Intelligibility Based on the Signal-to-Noise Envelope Power Ratio After Modulation-Frequency Selective Processing,” J. Acoust. Soc. Am. **130**, 1475-1487.

Kates, J.M. and Arehart, K.H. **(2005)**. “Coherence and the speech intelligibility index,” J. Acoust. Soc. Am. **117**, 2224-2237.

Lim, J. S. **(1978)**. “Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise,” IEEE Trans. Acoust., Speech, Signal Process. **26**, 471–472.

Loizou, P. C. and Ma, J. **(2011)**. “Extending the Articulation Index to Account for Non-Linear Distortions Introduced by Noise-Suppression Algorithms,” J. Acoust. Soc. Am. **130**, 986-995.

Ma, J., Hu, Y. and Loizou, P.C. **(2009)**. “Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions,” J. Acoust. Soc. Am. **125**, 3387-3405.

Taal, C.H., Hendriks, R.C., Heusdens, R. and Jensen, J. **(2011)**. “An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech,” IEEE Trans. Audio. Speech. Lang. Process. **19**, 2125-2136.