Predictive measures of the intelligibility of speech processed by noise reduction algorithms

Karolina Smeds[1], Florian Wolters[1,2], Arne Leijon[3], Anders Nilsson[1,3], Sara Båsjö[1], and Sofia Hertzman[1]

[1]Widex A/S, ORCA Europe, Stockholm, Sweden; karolina.smeds@orca-eu.info

[2]Univeristy of Applied Sciences, Oldenburg, Germany, [3]KTH, Stockholm, Sweden

ABSTRACT

A number of theoretical measures were evaluated in terms of their ability to predict the effect on speech intelligibility of different types of noise reduction (NR) for listeners with impaired and normal hearing.

Twenty listeners with hearing impairment and ten listeners with normal hearing participated in a blinded laboratory study. An adaptive speech test, with sentences in babble noise, was used. The speech test produces results in terms of physical signal-to-noise ratios that correspond to equal speech recognition performance with and without the NR algorithms, which facilitates a direct statistical test of how well the predictive measures agree with the experimental results.

Three NR algorithms and a reference condition were compared. The experimental results were used to evaluate a number of predictive measures, including a standard Speech Intelligibility Index (SII) method, two time-variable SII methods, and one coherence-based SII method. Further, one measure based on the correlation between band envelope magnitudes of clean and processed noisy speech was evaluated.

Short-time analysis of the SNR and methods based on correlation of the clean speech and the processed noisy speech did best in the comparison.

BACKGROUND

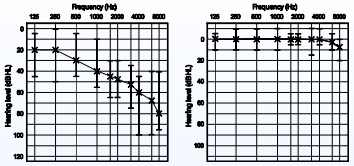
Noise reduction (NR) is commonly used in modern hearing aids. It would be of great value if predictive measures could be used to indicate the effect of various NR algorithms prior to laboratory or field testing with listeners.

The now reported work was part of a larger study, where both speech intelligibility and sound quality of NR processed speech were evaluated. The sound-quality work has been reported by Smeds et al. (2010), and further details can be found in two theses (Nilsson, 2010; Wolters, 2010).

METHOD

Participants

- 20 listeners with impaired hearing (HI), 62-82 years (mean: 72 years)
- 10 listeners with normal hearing (NH), 19-28 years (mean: 23 years)



Hearing aids

- Inteo 9, Widex A/S
- Bilateral fittings
- Linearly programmed according to NAL-R – 6 dB
- Tight earmoulds
- Real-ear and coupler-gain measurements

Noise Reduction Algorithms

- WEDM – Bayesian noise estimator based on the Weighted Euclidean Distortion Measure, implemented by Loizou (2007)
- Wiener – Wiener filtering based on a priori SNR estimation, implemented by Loizou (2007)
- PSSLP – Perceptually tuned Spectral Subtraction algorithm with Low-Pass filtered spectral filter coefficients, (Luts et al., 2010)

Hagerman speech test

- Swedish adaptive sentence test
- 5-word sentences with a fixed syntax
- Female talker
- 8-talker artificial babble noise derived from the ISTS signal
- Result: SNR at 80% correctly repeated key-words.

Evaluation procedure

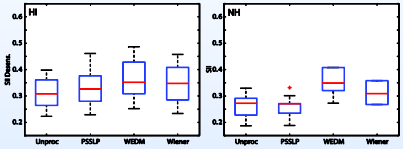
- Individual results from the speech test (SNR at 80% correct) entered in all calculations.
- A good predictive measure will give the same calculated value for all conditions.
- Tested with Friedman's analysis of variance by ranks

Pros:

- No assumptions about transfer functions from calculated scores to predicted speech recognition scores necessary
- Easy to statistically test if a predictive measure shows good performance, i.e., gives the same results for all conditions.

SPEECH INTELLIGIBILITY INDEX, SII

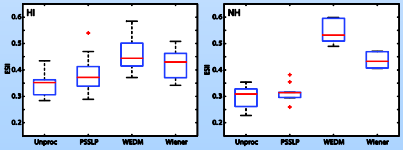
- Calculated according to ANSI S3.5 (1997)
- Long-term average speech and noise spectra
- Using a desensitization factor suggested by Pavlovic et al. (1986)



For both HI and NH listeners, the results for the four listening conditions differ (Friedman, p<0.05), i.e., the SII is not a good predictor of the effect the NR algorithms have on speech recognition.

EXTENDED SII, ESII

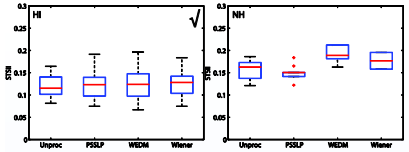
- Rhebergen and Versfeld (2005)
- Long-time average speech spectrum and short-time (9-35 ms) noise spectrum analysis
- Takes forward masking into account



For both HI and NH listeners, the results for the four listening conditions differ (Friedman, p<0.05), i.e., the ESII is not a good predictor of the effect the NR algorithms have on speech recognition.

SHORT-TIME SII, STSII

- Short-time (25 ms) spectrum analysis
- ANSI S3.5 standard (1997)
- Densensitisation factor (Pavlovic et al. 1986)

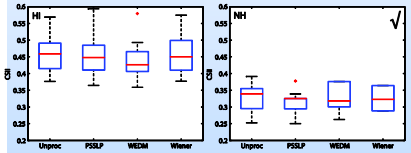


For the HI listeners, the results for the four conditions do *not* differ (Friedman, p>0.05). For the NH listeners, the results for the listening conditions differ (Friedman, p<0.05)

.

THREE-LEVEL COHERENCE SII, CSII

- Kates and Arehart (2005)
- Short-time (8 ms) spectrum analysis
- Three level regions based on RMS values
- Signal-to-distortion ratio: coherence between clean speech and processed noisy speech
- CSII calculated for each level region separately
- Final measure a weighted sum of the contributions from the three level regions.



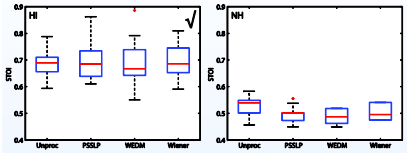
For the NH listeners, the results for the four conditions do *not* differ (Friedman, p>0.05). For the HI listeners, the results differ (Friedman, p=0.044).

SHORT-TIME OBJECTIVE INTELLIGIBILITY MEASURE, STOI

- Taal, Hendriks, Heusdens, and Jensen (2010)
- Short-time (13 ms) spectrum analysis
- Correlation between band envelope magnitudes of clean speech and processed noisy speech
- Linear correlation coefficient calculated within running overlapping time segments of about 400 ms (after scaling and clipping)
- Average across time and frequency bands

Modifications to incorporate impaired hearing

- Short-time spectra adjusted to represent the sound-field pressure values actually presented to the listeners
- Absolute hearing thresholds simulated by adding an internal masking noise floor with a spectrum corresponding to the individual hearing thresholds



For the HI listeners, the results for the four conditions do *not* differ (Friedman, p>0.05). For the NH listeners, the results for the listening conditions differ (Friedman, p<0.05).

CONCLUSIONS

None of the theoretical measures was able to predict the speech test results for both groups of listeners. Short-time analysis of the SNR and methods based on correlation of the clean speech and the processed noisy speech seems most promising.

REFERENCES

ANSI-S3.5 1997. American national standard methods for the calculation of the speech intelligibility index.

Kates & Arehart. 2005. Coherence and the speech intelligibility index. *JASA*, 117, 2224-2237.

Loizou. 2007. *Speech enhancement theory and practice*.

Luts et al. 2010. Multicenter evaluation of signal enhancement algorithms for hearing aids. *JASA*, 127, 1491-1505.

Nilsson. 2010. Evaluation of theoretical measures to assess their ability to predict the performance of noise reduction algorithms. MSc thesis, KTH Stockholm.

Pavlovic et al.986. An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. *JASA* 80, 50-7

Rhebergen & Versfeld. 2005. A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *JASA*, 117, 2181-2192.

Smeds et al. 2010. Objective measures to quantify the perceptual effects of noise reduction in hearing aids. *AES 38th International Conference - Sound quality evaluation*. 101-107.

Taal et al. 2010. A short-time objective intelligibility measure for time-frequency weighted noisy speech *ICASSP 2010*. 4214-4217.

Wolters F. 2010. Predictive ability of objective methods in terms of speech intelligibility and sound quality for noise reduction systems in hearing aids. BSc thesis, University of Applied Sciences, Oldenburg.