Increasing Speech Intelligibility by DeNoising: What can be achieved?

Jan Rademacher[1], *Joerg Bitzer*[2], *and Joerg Houpert*[3]

[1]*Medical Physics – Signal Processing Group, University of Oldenburg, Oldenburg, Germany*
jan.rademacher@uni-oldenburg.de

[2]*Institute for Hearing Technology and Audiology, University of Applied Science Oldenburg / Ostfriesland / Wilhelmshaven, Germany (*joerg.bitzer@fh-oldenburg.de*)*

[3]*Cube-Tec International GmbH, Anne-Conway Str. 1, 28359 Bremen Germany*
*(*j.houpert@cube-tec.com*)*

A lot of forensic material is of poor recording quality. A variety of disturbances is added to the desired speech signal. A typical way to increase the quality of the signal is to apply algorithms that can be called DeNoisers, which filter out the background noises. However, increasing the quality does not imply that the speech intelligibility is increased, which is the essential goal of the process.
In (Bitzer et al. 2005) the authors show that no significant enhancement is possible if the background noise is a perfect masker for the desired speech signal. In their study the noise signal was computed by a mixture of 500 sentences randomly mixed by the same speaker. Therefore, the long-term noise spectrum is exactly the same as the spectrum of the desired speech signal. Finally, the speech reception threshold is computed, which is the Signal-to-Noise-Ratio (SNR) at which 50% of all words can be understood. The methodology is described in (Kollmeier and Wesselkamp 1997 and Wagener 2003). The test is called Oldenburger sentence test and is based on (Hagerman 1982). In real world recordings, however, the masker has a different spectrum.

Prediction of enhancement

The differences of the long term spectrum of the desired signal and the disturbance can be used as a measure of how well a denoiser could work for this combination of signals. Of course this is a coarse prediction since stationarity has to be assumed. Our new measure is based on the so called “optimal-” or “Wiener- filter” (Boll 1979, Wiener 1949) which is well-known for denoising tasks. Based on this filter, we suggest to calculate the masking thresholds (MT) (Zwicker 1982) in a way closely related to the MPEG standard (ISO/IEC 11172 1993) both for the original noise spectrum and for the noise spectrum denoised by the Wiener-filter. These MTs are then used to find the spectral components of the speech signal which are audible before and after the denoising process. Taking into account the “band importance function” (American National Standard 1997) reflecting the importance of each frequency band for intelligibility, the increase of audible frequency bands after denoising will be a good approximation of the forthcoming improvement of the intelligibility achieved by denoising. In order to show this behaviour subjective listening tests will be used for comparison.

Conclusions:

Increasing speech intelligibility is possible, if the disturbing signal is not masking all important frequencies of the desired signal. Since denoising algorithms are able to filter signal-bands individually a significant improvement in quality and intelligibility can be achieved. Furthermore, we introduced a new measure that can be used as a coarse predictor of the possible improvement.

References

American National Standard, “Methods for calculation of the speech intelligibility index”, ANSI S3.5-1997

Bitzer, Simmer, Holube, Schaer, “Some Experiments on Short-Time Spectral Attenuation (STSA) Algorithms and Speech Intelligibility” , International Workshop on Acoustic Echo and Noise Control (IWAENC), 2005, pp. 193-196

Boll, “Suppression of acoustic noise in speech using spectral subtraction”, IEEE Trans. Acoust., Speech, Signal Process. , ASSP-27, pages 113–120, 1979

Hagermann, “Sentences for Testing Speech Intelligibility in Noise”, Scandinavian Audiology 11, 1982, pp 79-87

ISO/IEC 11172, “Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: audio”, BSI, London, First edition, October 1993. Implementation of ISO/IEC 11172-3:1993. ISBN 0-580-22594-1

Kollmeier, Wesselkamp, “Development and evaluation of a German sentence test for objective speech intelligibility assessment”. J. Acoust. Soc. Am., 102, 2412–2421.

Wagener, “Factors Influencing Sentence Intelligibility in Noise”. PhD-thesis, University of Oldenburg, 2003

Wiener, “Extrapolation, interpolation and smoothing of stationary time series: With engineering applications”, In Principles of Electrical Engineering Series, Cambridge, Massachusetts and New York, 1949. MIT Technology Press and John Wiley & Sons Inc.

Zwicker, “Psychoakustik”, Springer Verlag, Berlin, 1982