Estimation of signal-to-noise ratios in realistic sound scenarios
Supplement: Automatic noise estimation

Karolina Smeds, Florian Wolters, Martin Rung

Introduction

In an early stage of the current project, a number of automatic noise estimation algorithms were evaluated. These methods estimate the noise spectrum in a sound file that contains speech and noise. For the current pilot test, sound files where speech and noise were mixed with known signal-to-noise ratios (SNRs) were used. Noise power was automatically estimated (using the methods described below), an estimate of the speech power was calculated by subtracting the estimated noise power from the power of the noisy speech signal and the SNR was estimated using Equation 3 in the main article. The calculated SNRs were compared to the known, nominal SNRs. The manual noise estimation method, which was actually used in the study, was evaluated in a similar way.

Contents

1 Automatic noise estimation procedures

Seven of the evaluated noise estimation algorithms are described in a textbook by Loizou (2007), and the Matlab code on the CD accompanying the textbook was used (Doblinger 1995; Hirsch and Ehrlicher 1995; Martin 2001; Cohen 2002, 2003; Sorensen and Andersen 2005; Rangachari and Loizou 2006). Further, one algorithm of Hendriks et al. (2010) was evaluated. An implementation of this method, by Hendriks et al., can be found on Matlabs file exchange website (`http://www.mathworks.se/matlabcentral/fileexchange/27312-mmse-based-noise-psd-tracking-algorithm`).

Test material

Female and male speech (Swedish HINT sentences) was mixed with five types of noise:

- Artificial babble noise created by superimposing (with variable starting points) the International Speech Test Signal (ISTS) (Holube et al. 2010) with 1, 2, 4 or 8 talkers.
- Car noise
- Stationary speech-shaped HINT noise
- Party noise
- Street noise

Speech and noise were mixed to eight SNRs (-12, -6, -3, 0, 3, 6, 12, 24 dB). SNRs were mixed using both C-weighted and un-weighted speech and noise levels.

Results

The algorithm of Hendriks et al. (2010) performed best. Using a fairly generous success criterion, that the calculated SNR (based on the estimated noise power) should be within 3 dB of the nominal SNR, this algorithm did well for the stationary noises, in particular the Car and the stationary speech-shaped noise, for all but the highest SNR (24 dB). For the party noise the algorithm did well for all but the highest and lowest SNRs. However, for the babble noise the algorithm did not do well. Only for a couple of SNRs did the algorithm produce acceptable results, irrespective of the number of talkers included in the babble. The other algorithms produced acceptable results only for a couple of SNRs and a couple of background noises without any particular pattern (noise type and SNR).

2 Manual noise estimation procedure

Based on this pilot study, it was decided that a manual SNR estimation method should be used. This method was evaluated in a similar way as the other algorithms, but with a reduced number of noise types and SNRs.

Test material

Female and male speech (Swedish HINT sentences) was mixed with five types of noise:

- Artificial babble noise created by superimposing (with variable starting points) the International Speech Test Signal (ISTS) (Holube et al. 2010) with 4 or 8 talkers.
- Car noise
- Stationary speech-shaped HINT noise
- Party noise

Speech and noise were mixed to four SNRs (-3, 0, 3, 6 dB) except for the Car noise which also included SNRs at -12 and -6 dB. SNRs were mixed using un-weighted speech and noise levels.

Results

For all noise types and SNRs, the estimation error was within 1.5 dB. For the Car noise, the stationary speech-shaped noise and the two Babble noises, the estimation error was within 1 dB. Generally, the SNR tended to be overestimated by the procedure, but with less than 0.5 dB for SNRs above 0 dB except for the party noise, where the SNR was overestimated by 1 dB when the actual SNR was 0 dB, gradually decreasing to 0.5 dB when the actual SNR was 6 dB.

References

Cohen I. (2002) Noise estimation by minima controlled recursive averaging for robust speech enhancement. *IEEE Signal Processing Letters* 9: 12-15.

Cohen I. (2003) Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Trans Speech Audio Processing* 11: 466-475.

Doblinger G. (1995) Computationally efficient speech enhancement by spectral minima tracking in subbands. *Proc Eurospeech* 2: 1513-1516.

Hendriks RC, Heusdens R, Jensen J. (2010) MMSE based noise psd tracking with low complexity. *Acoustics Speech and Signal Processing (ICASSP), IEEE.* 4266 - 4269

Hirsch H, Ehrlicher C. (1995) Noise estimation techniques for robust speech recognition. *IEEE Int Conf Acoust Speech Signal Processing.* 153-156.

Holube I, Fredelake S, Vlaming M, Kollmeier B. (2010) Development and analysis of an International Speech Test Signal (ISTS). *Int J Audiol* 49: 891-903.

Loizou PC. (2007) *Speech enhancement theory and practice.* Taylor ampersand Francis Group, Boca Raton, FL, USA.

Martin R. (2001) Noise power spectral density estimation based on optimal smoothing and minimum statistic. *IEEE Trans Speech Audio Processing* 9: 504-512.

Rangachari S, Loizou P. (2006) A noise estimation algorithm for highly nonstationary environments. *Speech communication* 28: 220-231.

Sorensen K, Andersen S. (2005) Speech enhancement with natural sounding residual noise based on connected time-frequency speech presence regions. *EURASIP J Appl Signal Process* 18: 2954-2964.