Estimation of signal-to-noise ratios in realistic sound scenarios Supplement: The uncertainty of SNR estimates

Karolina Smeds, Florian Wolters, Martin Rung

Introduction

The SNR is defined as

$$SNR = 10 \log_{10} \frac{P_s}{P_n} \tag{1}$$

(Equation 1 in the printed paper), where P_s is the speech power and P_n is the noise power in a recorded segment of noisy speech. However, neither P_s nor P_n can be observed directly. As the speech and noise signals are statistically independent, the speech power is $P_s = P_{ns} - P_n$, where P_{ns} is the measured power of the noisy speech. In the current study, the power $P_{n'}$ of corresponding noise-only segments was used as an estimate of P_n. Thus, a point estimate of the SNR was calculated as

$$\widehat{SNR} = 10 \log_{10} \frac{P_{ns} - P_{n'}}{P_{n'}} \tag{2}$$

(Equation 3 in the printed paper). As with any other physical measurement, it is important to also estimate the uncertainty of the presented results. Previous studies did not consider a method to estimate this uncertainty, and therefore a new method had to be developed for this purpose. Fig. 6 in the printed paper indicates the uncertainty by error bars showing the point estimate plus/minus its standard error, i.e., one standard deviation, σ_{SNR}, of the estimate. The following sections describe the theory and methods used to estimate the variance of the point estimate,

$$\sigma^2_{SNR} = \mathrm{var}\left[\widehat{SNR}\right] \tag{3}$$

Contents

1 Analysis model **2**

2 Logarithmic transformation **2**

3 Estimating the noise variance **3**

4 Simulations **4**

5 Summary **7**

References **7**

1 Analysis model

The estimation relies on the assumption that the noise is stationary to the extent that the mean power, the power density spectrum, and the modulation spectrum of the noise are the same in the recorded noisy speech segments and in the corresponding noise-only segments.

Although only the two power measures, P_{ns} and $P_{n'}$, were available for each noisy situation in the present study, it is necessary to consider the infinite "space" of possible measurement results that might have occurred in experiments where the actual power of the speech in the noisy speech segment, P_s, and the true mean power of the noise, μ_n, are both fixed but unknown quantities.

In this theoretical framework, because of the natural modulations of the noise signal, the measured power of the noisy speech, P_{ns}, can be seen as a single sample drawn from a random variable with mean $\mu_{ns} = P_s + \mu_n$ and some variance σ_n^2. Similarly, the measured power in the noise-only segments, $P_{n'}$, is a single sample drawn from another random variable with mean μ_n and variance $\sigma_{n'}^2$. The two underlying random variables are statistically independent, because they represent different time segments of the noise. The mean μ_n is the same in both cases, because of the stationarity of the noise, but the variances σ_n^2 and $\sigma_{n'}^2$ are not necessarily equal, because the two measurement results are based on recordings with different durations.

The variance of a noise power measure, averaged over the whole duration of a signal segment, is inversely proportional to the duration. Thus, with a duration T for the noisy speech segment and duration T' of the corresponding noise-only segment,

$$\sigma_n^2 = \frac{c}{T} \tag{4}$$

$$\sigma_{n'}^2 = \frac{c}{T'} \tag{5}$$

where the constant c is the same in both equations, since it is determined by the same time-invariant statistical characteristics of the noise. Thus,

$$\sigma_n^2 = \frac{T'}{T}\sigma_{n'}^2 \tag{6}$$

Consequently, if the variance $\sigma_{n'}^2$ of the noise-only power can be somehow estimated, the variance of the noise in the noisy speech, σ_n^2, would also be known. A method to estimate $\sigma_{n'}^2$ is discussed later, in Sec. 3.

2 Logarithmic transformation

As the final SNR results are presented in dB units, the consequence of the logarithmic transformation of speech and noise power values in (2) must be considered. If a random variable Z is formed by a non-linear mapping from two other random variables Y and X, as $Z = g(X, Y)$, the variances are transformed as (e.g., NIST, 2014, Sec. 2.5.5)

$$\mathrm{var}\,[Z] \approx \left[\frac{\partial g}{\partial X}(\bar{X}, \bar{Y})\right]^2 \mathrm{var}\,[X] + \left[\frac{\partial g}{\partial Y}(\bar{X}, \bar{Y})\right]^2 \mathrm{var}\,[Y] \tag{7}$$

using the partial derivatives of the mapping function, evaluated at the means $\bar{X}$ and $\bar{Y}$. This general relation is valid regardless of the probability distributions of X and Y, and the approximation gets more and more accurate with decreasing variances $\mathrm{var}\,[X]$ and $\mathrm{var}\,[Y]$. (A slightly more complex expression is needed if X and Y are correlated, but this is not the case in the present application.)

The power-to-dB transformation (2) can be expressed as

$$\widehat{SNR} = 10\log_{10}\frac{P_{ns} - P_{n'}}{P_{n'}} = 10\log_{10}\frac{P_s + P_n - P_{n'}}{P_{n'}} \tag{8}$$

Here, the mean of the nominator is $P_s + \mu_n - \mu_n = P_s$, and the mean of the denominator is μ_n. Therefore, the variance σ^2_{SNR} can be approximated as

$$\sigma^2_{SNR} \approx \left[\frac{\widehat{\partial SNR}}{\partial P_n}(P_s, \mu_n)\right]^2 \sigma^2_n + \left[\frac{\widehat{\partial SNR}}{\partial P_{n'}}(P_s, \mu_n)\right]^2 \sigma^2_{n'} \tag{9}$$

where

$$\frac{\widehat{\partial SNR}}{\partial P_n}(P_s, \mu_n) = \frac{10\log_{10} e}{P_s} \tag{10}$$

$$\frac{\widehat{\partial SNR}}{\partial P_{n'}}(P_s, \mu_n) = -\frac{10\log_{10} e}{P_s} - \frac{10\log_{10} e}{\mu_n} \tag{11}$$

and $e \approx 2.718$ is the base of the natural logarithm. Finally, applying the variance relation (6) in (9), the desired variance is

$$\sigma^2_{SNR} \approx \left(\frac{10\log_{10} e}{P_s}\right)^2 \frac{T'}{T}\sigma^2_{n'} + \left(\frac{(P_s + \mu_n)10\log_{10} e}{P_s \mu_n}\right)^2 \sigma^2_{n'} \tag{12}$$

Simulation experiments have shown that the accuracy of this approximation is good when $\sigma_{n'} \leq 0.25\mu_n$ and $\sigma_{SNR} \leq 2$ dB. This is generally fulfilled in the present study.

To obtain a numerical estimate of the variance, the unknown parameters P_s and μ_n are replaced by their corresponding point estimates $\widehat{P_s} = P_{ns} - P_{n'}$, and $\widehat{\mu_n} = P_{n'}$. The estimation of the remaining quantity $\sigma^2_{n'}$ is more complicated, as discussed in detail in the following section.

3 Estimating the noise variance

In the current calculations, $\sigma^2_{n'}$ is estimated from analysis of the noise-only recording by dividing it into a number of sub-segments. The noise-only recording with duration T' is divided into J sub-segments, each having the duration T'/J. If the j-th sub-segment has the power $P_{n',j}$, the conventional sample variance across the sub-segments is calculated as

$$s^2 = \frac{1}{J-1}\sum_{j=1}^{J}\left(P_{n',j} - P_{n'}\right)^2; \quad P_{n'} = \frac{1}{J}\sum_{i=1}^{J} P_{n',i} \tag{13}$$

If and only if the power measures $P_{n',j}$ are independent samples drawn from the same distribution, the measured s^2 is an unbiased estimate of the true variance of the power of sub-segments. Then, the variance of the power $P_{n'}$ averaged across all J sub-segments can be estimated as

$$\widehat{\sigma^2_{n'}} = \frac{s^2}{J} \tag{14}$$

However, for real noise signals, the power values measured in different sub-segments might be correlated. The correlation can be caused by spectral shaping and/or envelope modulations of the signal. The correlation is exactly zero only if the noise is truly white.

Therefore, for the estimation method in Eqs. (13) and (14) to give good results, the sub-segments must be long enough for their powers to vary independently, and the length should be short enough to result in a sufficiently large number of sub-segments.

If J is chosen too large, the sub-segments will be so short that there is considerable correlation between the sub-segment power values. Then the final variance will be under-estimated, because the effective number of independent samples is less than J. Conversely, if J is chosen too small, to allow longer sub-segments, the final variance estimate will suffer from large random variability, because it is based on too few samples.

Simulations and analyses of a wide range of noise signals with known characteristics indicated a sub-segment duration of 250 ms to give a good compromise between these requirements. The simulations

are presented in the next section, and the results indicate that the exact choice of sub-segment duration is not very critical. With noise-only segments of 2 s (or more), the chosen duration results in 8 sub-segments or more. In the present study, the duration of noise-only segments ranged from 1.3 s to 16 s, with an average duration of 5.4 s.

4 Simulations

The effect of correlation on the estimation of the variance of the noise power was studied by simulations using special noise signals for which the true variance could be determined, because this was not possible with the real-life noise examples of the study. As the amount of correlation depends on the spectral shaping and/or envelope modulations of the noise, the simulations were done using several different types of noise, including types with spectra and modulation patterns similar to those of the real noise examples.

The following types of noise were studied:

- unmodulated white noise,
- white noise, band-limited within $100 - 3000$ Hz,
  - unmodulated,
  - with random envelope modulations low-pass filtered with 5 Hz and 30 Hz cut-off frequencies,
  - with random envelope modulations band-pass filtered in the range $5 - 30$ Hz,
- recorded car engine noise,
- recorded multi-talker babble

For each of these noise types, 32 examples with 15-s duration were analyzed. For each example, the average power $P_{n'}$ was measured, and the variance of the noise power was estimated by (13) and (14) with the number of sub-segments J ranging from 2 to a maximum about 650000, which was the total number of digital samples in each example. These results were calculated for each noise example, as if the other 31 similar examples had not existed. Furthermore, a single variance estimate was also calculated across all 32 examples. As these examples were known to be independent, this result is guaranteed to be unbiased, and it is therefore used as a reference for comparison.

The results for the (unrealistic) unmodulated white noise are shown in Fig. 1, confirming that the variance estimates converge towards the true variance for any choice of sub-segment duration. In this type of noise there is no correlation even if J is chosen arbitrarily large, even so large that each sub-segment includes only one single digital signal sample.

The results for car-engine noise and for multi-talker babble are shown in Figs. 2 and 3. The results for the other types of modulated synthetic noise signals were rather similar. As expected, these results indicate that the variance estimated from sub-segments of a single noise recording tends to be biased towards under-estimating the true variance, if J is so large that sub-segments are very short. Using about $30 - 100$ sub-segments of the 15-s example, i.e., with sub-segment durations about $150 - 500$ ms, gave fairly unbiased results with acceptably small variations.

Fig. 4 shows an over-estimation of the variance with J in the range of about 500 sub-segments, i.e., with durations of about 30 ms. Here, the modulation pattern probably interacted with the sub-segment duration, which exaggerated the measured power variations. Although an over-estimation of the variance would be on the safe side for the purpose of the present study, these examples indicate that caution is necessary if the present method is applied to other types of noise.

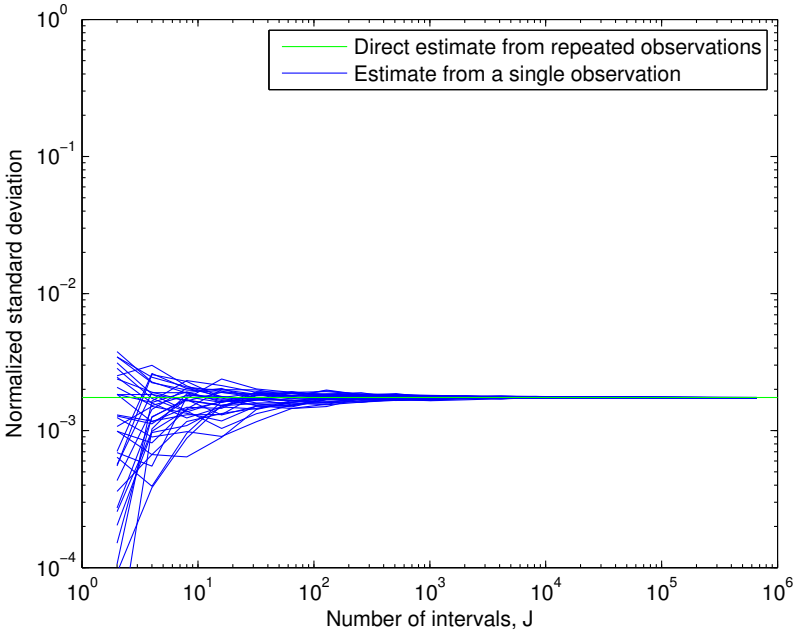

Figure 1: Simulation results for estimating the variance of the noise power for unmodulated white noise as defined in Eqs. (13) and (14) in the text, by segmenting each recording into sub-segments. Results are normalized to show $\sqrt{\widehat{\sigma^2_{n'}}}/P_{n'}$ as a function of the number of sub-segments, J. The horizontal (green) line shows an unbiased reference estimate, and the other (blue) curves show the results obtained from each of 32 independent examples.

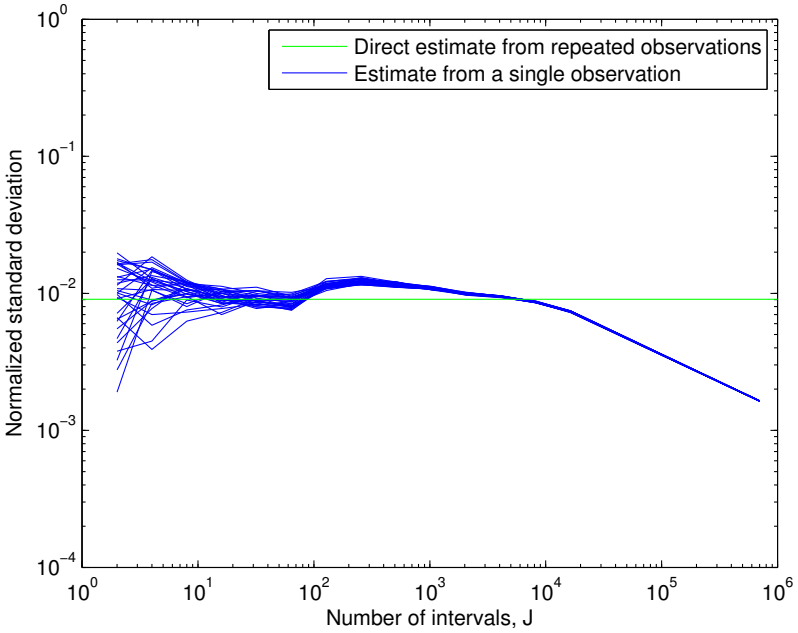

Figure 2: Simulation results for estimating the variance of the noise power for Car-Engine noise, shown in the same format as in Fig. 1.

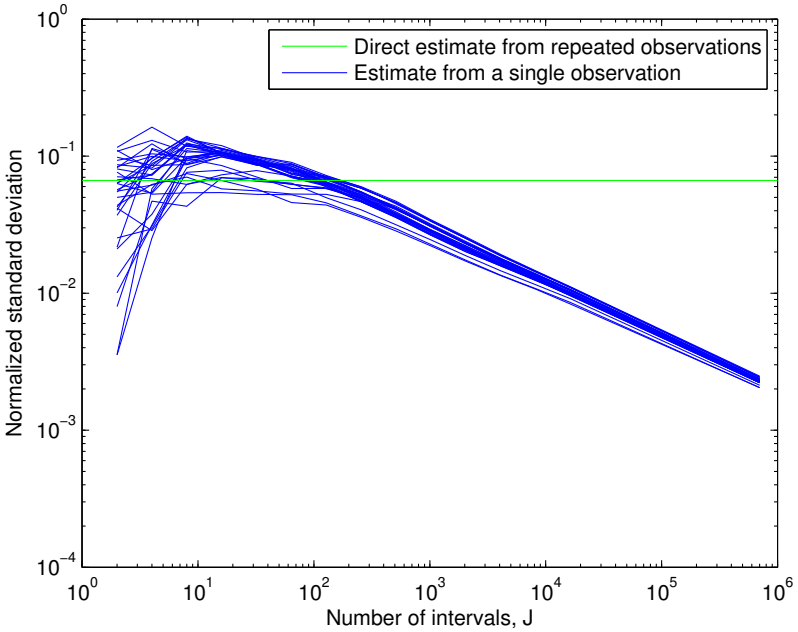

Figure 3: Simulation results for estimating the variance of the noise power for Multi-talker Babble, shown in the same format as in Fig. 1.

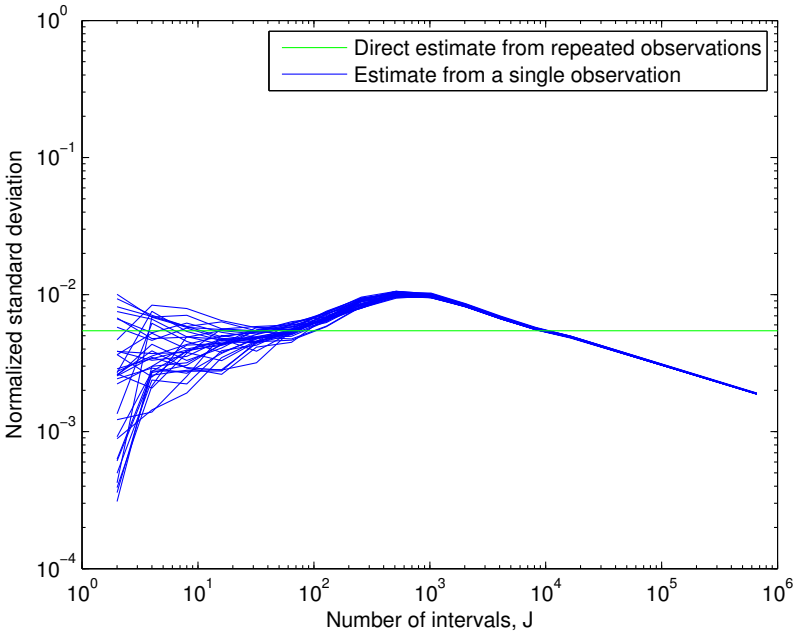

Figure 4: Simulation results for estimating the variance of the noise power for white noise, band-pass filtered ($100 - 3000$ Hz) and modulated by band-pass filtered ($5 - 30$ Hz) noise, shown in the same format as in Fig. 1.

5 Summary

A method for estimating the uncertainty of SNR measurements was developed and verified by theoretical arguments combined with simulation studies. This uncertainty measure has been used in the printed paper. The validity of the uncertainty measure is affected by the correlation patterns of the noise signals, and caution is therefore necessary if the present method is applied to other types of noise.

References

NIST (2014). NIST/SEMATECH e-Handbook of Statistical Methods, Sec. 2.5.5. http://www.itl.nist.gov/div898/handbook/mpc/section5/mpc55.htm.