Instrumental Musical Tones Measurement of Arbitrary Noise Reduction Systems

Huajun Yu, Tim Fingscheidt

Technische Universität Braunschweig, Institut für Nachrichtentechnik, Schleinitzstr. 22, 38106 Braunschweig, Germany

Email: {yu, fingscheidt}@ifn.ing.tu-bs.de

Introduction

Instrumental noise distortion measurements of a noise reduction algorithm in terms of musical tones can be conducted in two manners. The first is a *white box* test, requiring knowledge of the internal parameters. Different to the white box test, no such knowledge is needed for a *black box* test. Recently, an instrumental musical tones measurement based on the (log) kurtosis ratio of the noise input signal and the processed noise signal has been reported in [1, 2], showing a high correlation of the perceived amount of musical tones. The musical tones measurements presented in [1, 2] utilize the assumption of gamma-distributed squared *noise* spectral amplitudes and the knowledge of internal variables of the noise reduction scheme (i.e., white box test) to obtain an analytical function for calculating the (log) kurtosis ratio. However, the derivation of such analytical function is difficult in general and still cannot be solved for noise reduction algorithms applying the commonly used decision-directed approach to *a priori* SNR estimation [3]. Moreover, the requirement to know the internal variables of the noise reduction algorithm prevents its use in a black box test.

In this paper, we improve the modified log kurtosis ratio [4, 5] to a *weighted* log kurtosis ratio based on the input noise signal and the processed noise signal. We show that the formerly required voice activity detection in [4] can be completely omitted, since *noise-only* signals can be processed yielding an instrumental measure related to musical tones only. The proposed measure does neither require any assumption about squared spectral amplitude statistics, nor is it related to a specific noise reduction algorithm. It does not require knowledge of internal variables, which makes it applicable for arbitrary noise reduction systems.

New Black Box Musical Tones Measurement

Investigations led us to the finding, that the results of the log kurtosis ratio measure in [4, 5] are inconsistent when applying wideband noise signals with a sampling frequency of 16 kHz and narrowband noise signals with a sampling frequency of 8 kHz. To deal with this problem, we define a new weighted log kurtosis ratio as $\Delta\Psi^{\mathrm{w}}_{\log} = \ln\left(\frac{\Psi^{\mathrm{w}}_{\tilde{n}}}{\Psi^{\mathrm{w}}_{n}}\right)$ with Ψ^{w}_{n} and $\Psi^{\mathrm{w}}_{\tilde{n}}$ being the *weighted* kurtosis of the noise signal and the processed noise signal, respectively. In order to calculate Ψ^{w}_{n}, an *instantaneous weighted* kurtosis of squared amplitude noise DFT coefficients for each frame ℓ can be computed as

$$\Psi^{\mathrm{w}}_{n}(\ell)=\frac{\frac{1}{K}\sum\limits_{k=0}^{K-1}\Big[\alpha_n(k)\cdot|N(\ell,k)|^2-\overline{\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2}\Big]^4}{\left(\frac{1}{K}\sum\limits_{k=0}^{K-1}\Big[\alpha_n(k)\cdot|N(\ell,k)|^2-\overline{\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2}\Big]^2\right)^2}, \quad (1)$$

with $\overline{\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2}=\frac{1}{K}\sum\limits_{\kappa=0}^{K-1}\alpha_n(\kappa)\cdot|N(\ell,\kappa)|^2$. The weighting factor $\alpha_n(k) = \left(\frac{1}{L}\sum\limits_{\ell=1}^{L}|N(\ell,k)|^2\right)^{-1}$ is being calculated for each frequency bin as the inverse mean value of $|N(\ell,k)|^2$ across all L frames. The instantaneous weighted kurtosis $\Psi^{\mathrm{w}}_{\tilde{n}}(\ell)$ for $\tilde{n}(n)$ can straightforwardly be computed in the same manner. The respective terms Ψ^{w}_{n} and $\Psi^{\mathrm{w}}_{\tilde{n}}$ can then be calculated using $\Psi^{\mathrm{w}}_{n} = \frac{1}{L}\sum_\ell \Psi^{\mathrm{w}}_{n}(\ell)$, $\Psi^{\mathrm{w}}_{\tilde{n}} = \frac{1}{L}\sum_\ell \Psi^{\mathrm{w}}_{\tilde{n}}(\ell)$. With $\Delta\Psi^{\mathrm{w}}_{\log}$, we can achieve consistent results both for wideband and narrowband instrumental musical tones measurements.

Experimental Setup and Results

18 in-car background noise signals as the only input signals for noise reduction systems are chosen from the ETSI noise database [6], each having a length of 8s. Since our experimental evaluation is done for narrowband only, all signals are downsampled to 8 kHz and scaled to -26 dBov (see to ITU-T Rec. P.56 [7]). A DFT with length $K = 256$ and a frame shift of 50% are applied, using the square root Hann window as analysis and synthesis windows, respectively. Four noise reduction algorithms are then applied in the DFT domain: the MMSE-SA (SA) estimator [3], the MMSE-LSA (LSA) estimator [8], the *a priori* SNR-driven Wiener filter (WF) [9], and the super-Gaussian joint MAP (SG) estimator [10]. For all weighting rules, an estimation of the *a priori* SNR defined as $\xi(\ell,k) = \frac{E\{|S(\ell,k)|^2\}}{E\{|N(\ell,k)|^2\}}$ is needed, being addressed by Ephraim and Malah in their decision-directed (DD) approach [3] as

$$\begin{aligned}\xi'(\ell,k) &= \beta\cdot\frac{|\hat{S}(\ell-1,k)|^2}{\hat{\phi}_{NN}(\ell-1,k)}+(1-\beta)\cdot P[\gamma(\ell,k)-1], \quad (2)\\ \xi(\ell,k) &= \max\{\xi'(\ell,k),\ \xi_{\min}\},\end{aligned}$$

with a smoothing factor β, the *a posteriori* SNR $\gamma(\ell,k) = \frac{|Y(\ell,k)|^2}{\hat{\phi}_{NN}(\ell,k)}$, $\xi_{\min}=-15\,\mathrm{dB}$, and $\hat{\phi}_{NN}(\ell,k)$ being the estimated noise power spectrum. A simulation over the full range of $0 \leq \beta < 1$ for generating less or more musical tones is performed to evaluate the weighted log kurtosis ratio measure. For the weighted log kurtosis ratio, we use a DFT length $K = 256$, a frame shift of 50%, and a square root Hann analysis window, respectively.

Once β increases beyond 0.9 in Fig. 1, $\Delta\Psi^{\mathrm{w}}_{\log}$ will decrease very fast towards zero for all weighting rules, meaning that the weighted kurtosis of $\tilde{n}(n)$ becomes more similar to the weighted kurtosis of $n(n)$, showing higher *statistical* similarity of $n(n)$ and $\tilde{n}(n)$, or, less musical tones. This matches the observation in [11] well, where β has been increased from 0.98 to 0.998. However, we found that smoothing comes only into effect for $\beta \geq 0.9$. This corresponds to our above mentioned observation with the weighted log kurtosis ratio. On the contrary, changing β from 0 to 0.9 for LSA and SA, the perceived amount of musical tones hardly changes, likewise $\Delta\Psi^{\mathrm{w}}_{\log}$ of LSA and SA remaining nearly unchanged within $0 < \beta \leq 0.9$. However, $\Delta\Psi^{\mathrm{w}}_{\log}$ increases with increasing β in $0 \leq \beta < 0.9$ for SG and WF. This is illustrated in Fig. 3 employing SG with three different βs: Using $\beta = 0$, much more peaks (red points) are seen in the spectrum of the weighted processed noise signal as compared to using $\beta = 0.9$. However, since for $\beta = 0$ too many such peaks are generated within each frame ℓ, they

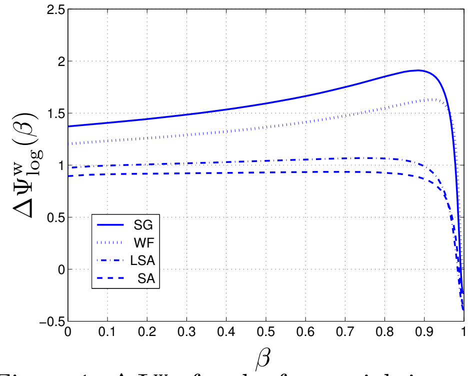

Figure 1: $\Delta\Psi^{\mathrm{w}}_{\log}$ for the four weighting rules

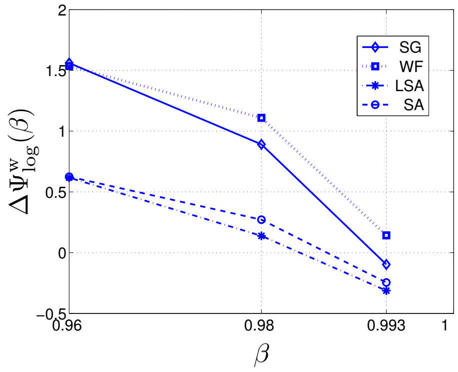

Figure 2: Weighted log kurtosis ratio for the four weighting rules with $\beta = 0.96, 0.98, 0.993$

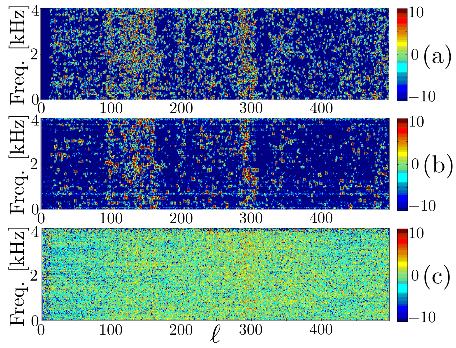

Figure 3: Spectrum of a weighted processed noise signal using SG with (a) $\beta = 0$, (b) $\beta = 0.9$, and (c) $\beta = 0.999$

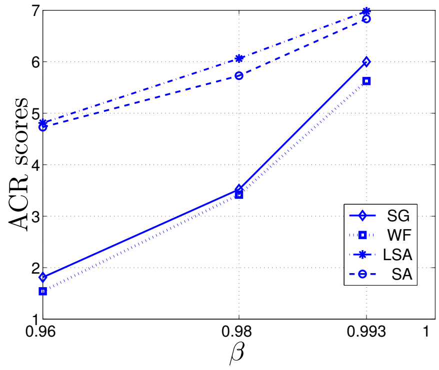

Figure 4: ACR listening test results for the four weighting rules with $\beta = 0.96, 0.98, 0.993$

are not isolated any more, leading to spectral distortions of broader bandwidth. For $\beta = 0.9$, we observe less peaks, but in a more isolated fashion, clearly indicating the effect of musical tones. That's why $\Psi^{\mathrm{w}}_{\tilde{n}}$ and consequently $\Delta\Psi^{\mathrm{w}}_{\log}$ increase for $\beta \to 0.9$. When we compare the case of using $\beta = 0.9$ and $\beta = 0.999$, strong smoothing comes into effect for $\beta \to 0.999$, indicating removal of isolated peaks (musical tones) also by a decrease of $\Delta\Psi^{\mathrm{w}}_{\log}$. To validate $\Delta\Psi^{\mathrm{w}}_{\log}$, a subjective listening test with 16 test persons (experts and non-experts) is conducted in an absolute category rating (ACR) fashion to judge the audible level of musical tones with 1 meaning intolerably audible musical tones and 7 meaning inaudible musical tones. Three noise signals have been randomly chosen with $\beta = 0.96,\ 0.98,\ 0.993$ for all weighting rules, respectively. Altogether 36 output (processed) noise signals had to be rated by each subject. The related instrumental $\Delta\Psi^{\mathrm{w}}_{\log}$ measurements are shown in Fig. 2 for comparison with the subjective listening test results shown in Fig. 4. Accordingly, the Pearson's correlation is computed for each weighting rules and all weighting rules combined as $\rho_{\mathrm{SG}} = -0.99$, $\rho_{\mathrm{WF}} = -0.98$, $\rho_{\mathrm{LSA}} = -0.99$, $\rho_{\mathrm{SA}} = -0.99$, and $\rho_{\mathrm{all}} = -0.98$, respectively. The overall correlation of $|\rho_{\mathrm{all}}| = 0.98$ between instrumental and subjective results verifies that the weighted log kurtosis ratio is an adequate instrumental measure for musical tones within a black box test environment.

Conclusions

In this paper, we propose the weighted log kurtosis ratio as a new black box musical tones measure, yielding a high overall correlation $|\rho_{\mathrm{all}}| = 0.98$ with the subjective listening test.

References

[1] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Musical Noise Generation Analysis for Noise Reduction Methods Based on Spectral Subtraction and MMSE STSA Estimation," in *Proc. of ICASSP'09*, Taipei, Taiwan, Apr. 2009.

[2] T. Inoue, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Theoretical Analysis of Musical Noise in Wiener Filtering Family via Higher-Order Statistics," in *Proc. of ICASSP'11*, Prague, Czech Republic, May 2011.

[3] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[4] H. Yu and T. Fingscheidt, "A Figure of Merit for Instrumental Optimization of Noise Reduction Algorithms," in *Proc. of 5th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems*, Kiel, Germany, Sep. 2011.

[5] H. Yu and T. Fingscheidt, "Black Box Measurement of Musical Tones Produced by Noise Reduction Systems," in *Proc. of ICASSP'12*, Kyoto, Japan, Mar. 2012.

[6] "ETSI EG 202 396-1, Speech Processing, Transmission and Quality Aspects (STQ), Speech Quality Performance in the Presence of Background Noise; Part 1: Background Noise Simulation technique and Background Noise Database," 2008.

[7] "ITU-T Recommendation P.56, Objective Measurement of Active Speech Level," ITU-T, Mar. 1993.

[8] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[9] P. Scalart and J.V. Filho, "Speech Enhancement Based on A Priori Signal to Noise Estimation," in *Proc. of ICASSP'96*, Atlanta, GA, USA, May 1996, pp. 629–632.

[10] T. Lotter and P. Vary, "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1110–1126, May 2005.

[11] O. Cappé, "Elimination of the Musical Noise Phenomenon With the Ephraim and Malah Noise Suppressor," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, Apr. 1994.