BLACK BOX MEASUREMENT OF MUSICAL TONES PRODUCED BY NOISE REDUCTION SYSTEMS

Huajun Yu, Tim Fingscheidt

Technische Universität Braunschweig, Institute for Communications Technology,
Schleinitzstr. 22, 38106 Braunschweig, Germany
{yu, fingscheidt}@ifn.ing.tu-bs.de

ABSTRACT

In the context of noise reduction algorithms, three instrumental measures are of major interest: the speech component quality, the level of noise attenuation, and noise distortion in terms of musical tones. As several proposals are made for the first two, the amount of musical tones is commonly still subjectively evaluated. Recent exploration of the log-kurtosis ratio for instrumentally measuring musical tones has led to white box test methodologies requiring specific information about the particular noise reduction algorithm. In this paper we propose a simple yet robust instrumental musical tones measurement, which is applicable to arbitrary unknown noise reduction systems, i.e., a black box measurement. A subjective listening test has been conducted to verify the proposed instrumental measure. Our measurement methodology has been proposed as part of an ITU-T Recommendation in Study Group 12, FG CarCOM.

Index Terms— musical tones, instrumental measurement

1. INTRODUCTION

An important issue for successful development of noise reduction algorithms is an effective quality assessment. Due to the high costs and efforts of subjective tests, instrumental measurements are often conducted in practice. In this paper we differentiate the instrumental measurements into two categories: The white box test, which (in the context of musical tones measurements) often mandates a specific spectral weighting rule and knowledge of some internal parameters. Secondly, the black box test, which requires no knowledge of the noise reduction system at all. Three quality measures are typically of interest for a noise reduction system: the speech component quality, the level of noise attenuation, and the noise distortion, e.g., in terms of the amount of musical tones. The first two can be well evaluated in an instrumental manner, see, e.g., ITU-T Recommendations P.1100 [1] and P.1110 [2]. The instrumental N-MOS (noise mean opinion score) utilizing a psychoacoustical hearing-model based relative approach [3] is employed in [4] measuring the total noise transmission quality. N-MOS is calculated from the non-speech segments of a noisy speech signal and the enhanced signal, respectively. The clean speech signal is required as a reference signal to identify the non-speech segments by a VAD (voice activity detection). In our context, N-MOS is not useful because it combines noise attenuation and noise distortion in a single measure. However, we are concerned with noise distortion only, particularly with musical tones, so that noise distortion can be measured separately from noise attenuation. Recently, a high correlation of the perceived amount of musical tones with an instrumental log kurtosis ratio measure has been reported in [5] requiring specifically the spectral subtraction approach to noise reduction. Moreover, the noisy speech signal and the enhanced signal are used for computing the log kurtosis ratio in [5], making this ratio also dependent on the level of noise attenuation. The kurtosis ratio instead of the *log* kurtosis ratio of an input noise signal and the output (processed) noise signal has been further investigated in [6, 7]. Using the assumption of gamma-distributed squared speech *and noise* spectral amplitudes and assuming knowledge of internal variables of the noise reduction scheme (i.e., a white box test methodology), an analytical function can be obtained to calculate the (log) kurtosis ratio in [5–7]. However, the derivation of this analytical function is difficult and is still unavailable for noise reduction algorithms using the widely employed decision-directed approach to *a priori* SNR estimation [8]. Moreover, the requirement to know the internal variables of the noise reduction scheme prevents its use as black box measurement as required in practice.

In this paper we improve a modified log kurtosis ratio [9] based on input noise signal segments and respective output (i.e., processed) noise signal segments. We show that for the noisy speech input signals the formerly required VAD can now be completely omitted, and noise-only signals can be processed yielding an instrumental measure related to noise distortion only (more specifically: related to musical tones). We focus on noise-only signals, since we consider this to be sufficient for musical tones' analysis purposes. The proposed measure does neither require any assumption about squared spectral amplitude statistics, nor

does it mandate a specific noise reduction algorithm or even knowledge of internal variables; it is a black box measurement approach.

The paper is organized as follows: In Section 2.1 we review the previous published (log) kurtosis ratio calculation for white box musical tones measurements. Our modified noise kurtosis ratio measure, which can be applied in black box tests, is then described in Section 2.2. Finally, Section 3 presents our experimental setup and results. Furthermore, subjective listening test results will be provided.

2. MUSICAL TONES MEASUREMENT USING (LOG) KURTOSIS RATIO

2.1. Specific White Box Test Approaches

In this Section, we briefly review the state-of-the-art (log) kurtosis ratio for instrumentally measuring musical tones [5–7]. In the discrete Fourier transform domain, the microphone signal at frame ℓ and frequency bin k can be formulated as $Y(\ell,k) = S(\ell,k)+N(\ell,k)$, with $N(\ell,k)$ being the additive noise and $S(\ell,k)$ being the clean speech signal. In [5], the spectral subtraction approach for noise reduction is investigated, which is given as

$$\hat{S}(\ell,k)=\sqrt{|Y(\ell,k)|^2-\nu\cdot\hat{\phi}_{NN}(\ell,k)}\cdot e^{j\arg(Y(\ell,k))} \quad (1)$$

with $\hat{S}(\ell,k)$ being the enhanced signal, $\hat{\phi}_{NN}(\ell,k)$ being the estimated noise power spectral density, ν being a subtraction coefficient and $\arg(Y(\ell,k))$ being the phase of $Y(\ell,k)$, respectively. Parameter ν controls how much of the estimated noise power spectrum will be subtracted from the noisy microphone signal power spectrum.

A high correlation of the perceived amount of musical tones with the log kurtosis ratio has been shown in [5]. The log kurtosis ratio here is calculated as the log ratio between the kurtosis of $|Y(\ell,k)|^2$ and the kurtosis of $|\hat{S}(\ell,k)|^2$. In the theory of higher-order statistics [10], the kurtosis Ψ_x of a random variable x is defined as

$$\Psi_x = \frac{E\{[x-E\{x\}]^4\}}{(E\{[x-E\{x\}]^2\})^2}, \quad (2)$$

where $E\{\cdot\}$ is the expectation operator. In order to calculate the kurtosis, the authors of [5] assume the squared speech and noise spectral amplitudes as gamma-distributed. The pdf of $|Y(\ell,k)|^2$ can then be formulated as a function $f(\alpha,\theta)$ with α and θ being estimated from $|Y(\ell,k)|^2$. Subsequently, the kurtosis of $|Y(\ell,k)|^2$ can be calculated as a function $f(\alpha)$ with α being the only parameter. In the same way and by knowing the subtraction coefficient ν, the kurtosis of $|\hat{S}(\ell,k)|^2$ can be calculated as a function $f(\alpha,\nu)$. Finally, the log kurtosis ratio of $|Y(\ell,k)|^2$ and $|\hat{S}(\ell,k)|^2$ can be formulated as an analytical function controlled by the parameters α and ν only. The same method of formulating an analytical function for calculating the kurtosis ratio has been applied also in [6,7]. Please note, in [6,7] the kurtosis ratio instead of the log kurtosis ratio is calculated based on the noise components only, which makes it independent of speech distortion and noise attenuation.

However, all (log) kurtosis ratio calculations in [5–7] need internal access to the noise reduction algorithm, e.g., the subtraction coefficient ν for spectral subtraction [5, 6] and also for the Wiener filter family in [7], which is implemented in a spectral subtraction-like manner[1]. This makes it an improper musical tones measure for black box measurements of arbitrary and internally unknown noise reduction systems.

2.2. Generic Approach for Black Box Tests

Now we investigate our modified noise log kurtosis ratio [9]

$$\Delta\Psi_{\log} = \ln\left(\frac{\Psi_{\tilde{n}}}{\Psi_n}\right), \quad (3)$$

where Ψ_n and $\Psi_{\tilde{n}}$ are the kurtosis related to the noise signal and to the filtered noise signal, respectively. We use $\Delta\Psi_{\log}$ defined in (3) to quantify the amount of musical tones. Different from [5–7], where $|N(\ell,k)|^2$ are assumed to be gamma distributed in the power spectral domain, no such assumption is needed here. Similar to (2), an *instantaneous kurtosis* of *squared amplitude* noise DFT coefficients for each frame ℓ can be computed as

$$\Psi_n(\ell)=\frac{\frac{1}{K}\sum_{k=1}^{K}\left[|N(\ell,k)|^2-\overline{|N(\ell,k)|^2}\right]^4}{\left(\frac{1}{K}\sum_{k=1}^{K}\left[|N(\ell,k)|^2-\overline{|N(\ell,k)|^2}\right]^2\right)^2}, \quad (4)$$

with $\overline{|N(\ell,k)|^2} = \frac{1}{K}\sum_{k=1}^{K}|N(\ell,k)|^2$. The kurtosis $\Psi_{\tilde{n}}(\ell)$ can straightforwardly be computed by applying $|\tilde{N}(\ell,k)|^2$ in (4). The respective terms Ψ_n and $\Psi_{\tilde{n}}$ can then be calculated as

$$\Psi_n = \frac{1}{L}\sum_{\ell=1}^{L}\Psi_n(\ell),\ \Psi_{\tilde{n}} = \frac{1}{L}\sum_{\ell=1}^{L}\Psi_{\tilde{n}}(\ell). \quad (5)$$

It is important to note that we *only* process noise, i.e., $y(n)=n(n)$. Inserting Ψ_n and $\Psi_{\tilde{n}}$ into (3), the log kurtosis ratio $\Delta\Psi_{\log}$ can finally be computed without any assumption about (speech and) noise probability distribution functions.

Please note, our proposed noise log kurtosis ratio measure defined in (3) is calculated only from the input noise signal $n(n)$ and the output (processed) noise signal $\tilde{n}(n)$, which allows it to be applicable for all noise reduction algorithms, also those using the decision-directed approach for estimating the *a priori* SNR. Furthermore, the calculation of $\Delta\Psi_{\log}$ needs no extra knowledge of the noise reduction scheme and its internal parameters, which means that it can be considered as a black box measurement.

[1]Note that it is further stated in [6,7], that since the derivation of an analytical function for calculating the (log) kurtosis ratio is difficult, a solution cannot be given for noise reduction algorithms applying the decision-directed approach for *a priori* SNR estimation. Therefore, the referenced method for calculating the kurtosis ratio is not applicable to a wide range of state-of-the-art noise reduction algorithms.

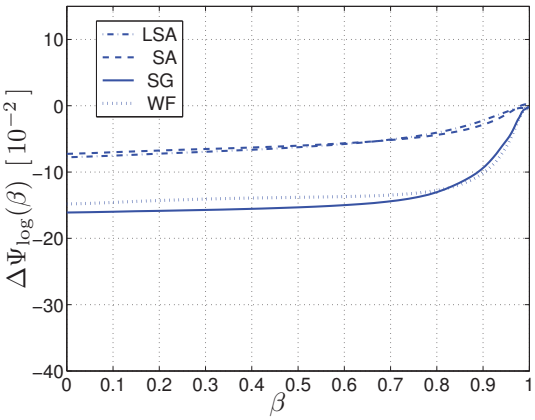

Fig. 1. Noise log kurtosis ratio for the four weighting rules

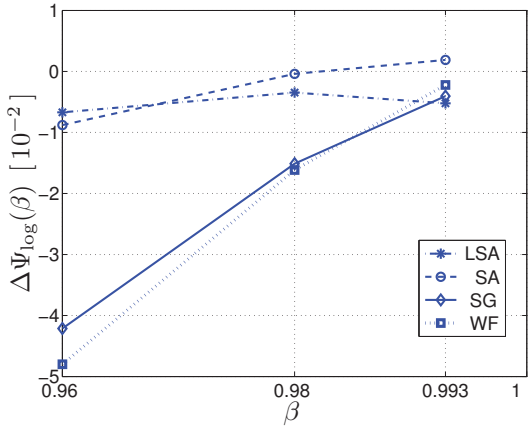

Fig. 2. Noise log kurtosis ratio for the four weighting rules with $\beta = 0.96, 0.98, 0.993$

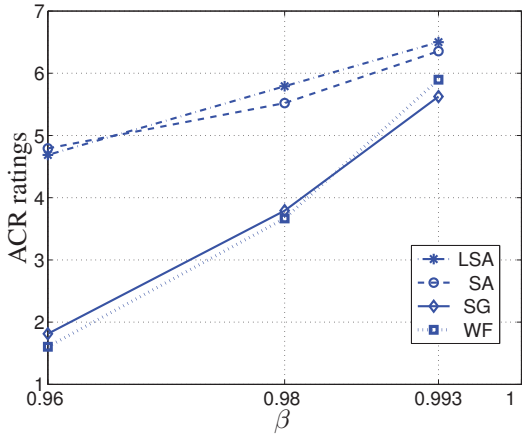

Fig. 3. ACR listening test results for the four weighting rules with $\beta = 0.96, 0.98, 0.993$

3. EXPERIMENTS

In this section, we will evaluate the proposed instrumental musical tones measure from Section 2.2 by processing only noise signals with different noise reduction systems. Furthermore, a subjective listening test verifying the proposed measure will be presented.

3.1. Simulation Setup

We evaluate four state-of-the-art noise reduction algorithms with *noise signals only* as input signals. Our experiments are performed with 18 in-car background noise signals from the ETSI noise database [11], each sampled with 16 kHz and having a length of 8s. All noise signals are at -26 dBov according to ITU-T Recommendation P.56 [12]. The following setup is used: A DFT with length $K = 512$ and a frame shift of 50% are applied, using the square root Hann window as analysis and synthesis windows, respectively. Four state-of-the-art noise reduction algorithms are tested: the MMSE-SA (SA) estimator [8] and the MMSE-LSA (LSA) estimator [13], the *a priori* SNR-driven Wiener filter (WF) [14], and the super-Gaussian joint MAP (SG) estimator [15]. 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 successfully addressed by Ephraim and Malah in their decision-directed (DD) approach [8] as

$$\xi'(\ell,k) = \beta \cdot \frac{|\hat{S}(\ell-1,k)|^2}{\hat{\phi}_{NN}(\ell-1,k)} + (1-\beta)\cdot P[\gamma(\ell,k)-1], \quad (6)$$
$$\xi(\ell,k) = \max\{\xi'(\ell,k),\ \xi_{\min}\},$$

with a smoothing factor β, the enhanced signal of the previous frame $\hat{S}(\ell-1,k)$, the *a posteriori* SNR $\gamma(\ell,k) = \frac{|Y(\ell,k)|^2}{\hat{\phi}_{NN}(\ell,k)}$, and $\xi_{\min} = -15$ dB. The estimated noise power spectrum $\hat{\phi}_{NN}(\ell,k)$ is obtained by minimum statistics [16].

Setting β close to unity yields a strong smoothing of the *a priori* SNR estimate, which helps to significantly reduce musical tones [17]. To demonstrate the proposed instrumental musical tones measurement, an evaluation with the full range of $0 \leq \beta < 1$ should be performed. Please note, we change β from 0 to 1 only to show the noise log kurtosis measurement results for different values of β, however, no information of β is needed for calculating the noise log kurtosis measure according to (3).

3.2. Simulation Results

The results of the noise log kurtosis ratio $\Delta\Psi_{\log}$ for SA, LSA, WF and SG are shown in Fig. 1. Using the proposed noise log kurtosis ratio (3), we observe: With increasing β, $\Delta\Psi_{\log}$ will accordingly increase towards zero, meaning that the kurtosis of $\tilde{n}(n)$ becomes more similar to the kurtosis of $n(n)$, which means higher *statistical* similarity of $n(n)$ and $\tilde{n}(n)$, or, less musical tones. We found that by changing β in (6), the higher the noise log kurtosis ratio is, the less musical tones are observed. If β is chosen to be greater than 0.9, WF and SG show a more rapid $\Delta\Psi_{\log}$ increase than SA and LSA .

In order to further validate $\Delta\Psi_{\log}$ as an applicable instrumental musical tones measurement, a subjective listening test in an ACR (absolute category rating) fashion is conducted for judging the audible level of musical tones. Six-

teen test persons (experts and non-experts) had to rate the audibility of the musical tones according to an ACR listening test with seven categories: (1) intolerably audible, (2) loudly audible, (3) rather loudly audible, (4) moderately audible, (5) slightly audible, (6) just audible, (7) inaudible. Three in-car background noises from the 18 in-car background noises have been randomly chosen. Each noise has been processed by the spectral weighting rules SA, LSA, WF and SG. Three values of β with 0.96, 0.98 (being the optimal value for SA [8]), and 0.993 (being the optimal value for SG [9]) are chosen for each spectral weighting rule. Altogether 36 output (processed) noise signals had to be rated by each subject. The related instrumental $\Delta\Psi_{\log}$ measurements are shown in Fig. 2 for comparison with the subjective listening test results shown in Fig. 3. It can been seen that the ACR results match the $\Delta\Psi_{\log}$ results very nicely for all weighting rules, the correlations for SA, LSA, WF and SG are $\rho_{\mathrm{SA}} = 0.94$, $\rho_{\mathrm{LSA}} = 0.56$, $\rho_{\mathrm{WF}} = 0.97$ and $\rho_{\mathrm{SG}} = 0.98$, respectively. Please note that the outlier for LSA at the large value of $\beta = 0.993$ is responsible for the relatively low correlation value of $\rho_{\mathrm{LSA}} = 0.56$. However, the instrumental measure shows its optimal point (highest $\Delta\Psi_{\log}$) at the very typical value of $\beta = 0.98$, which is definitely a good parameter choice for LSA. In addition, when we observe LSA for the whole range of $0 \leq \beta < 1$ in Fig. 1, it can be seen that $\Delta\Psi_{\log}$ is still an almost monotonically increasing function of β. We have achieved an average correlation of $\rho = 0.86$ for the pool of all spectral weighting rules between the instrumental $\Delta\Psi_{\log}$ measure and the ACR listening test. Both instrumental results and subjective results reveal that the larger β is, the less musical tones are perceivable. These results verify that the noise log kurtosis ratio is an adequate instrumental measure for musical tones in a generic black box test environment.

4. CONCLUSIONS

We address a new black box instrumental musical tones measurement for arbitrary noise reduction systems. Compared to state-of-the-art musical tones measures requiring specific noise reduction algorithms, knowledge of internal variables, and assuming specific noise (and speech) distributions, our proposed noise log kurtosis ratio calculation does not require any such assumptions. Subjective tests have proven a good correlation to subjective musical tones rating for a variety of noise reduction approaches.

5. REFERENCES

[1] "ITU-T Recommendation P.1100, Narrow-Band Hands-Free Communication in Motor Vehicles," ITU-T, Oct. 2008.

[2] "ITU-T Recommendation P.1110, Wideband Hands-Free Communication in Motor Vehicles," ITU-T, Dec. 2009.

[3] K. Genuit, "Objective Evaluation of Acoustic Quality Based on a Relative Approach," in *Proc. of InterNoise'96*, Liverpool, UK 1996, pp. 3233–3238.

[4] ETSI EG 202 396-3, "Speech Processing, Transmission and Quality Aspects (STQ), Speech Quality Performance in the Presence of Background Noise; Part 3: Background Noise Transmission - Objective Test Methods," 2008.

[5] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Automatic Optimization Scheme of Spectral Subtraction based on Musical Noise Assessment via Higher-Order Statistics," in *Proc. of IWAENC'08*, Seattle, WA, Sep. 2008.

[6] 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.

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

[8] 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.

[9] 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.

[10] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing*, Dover Publications, Inc., New York, NY, 1972.

[11] 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.

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

[13] 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.

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

[15] 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, 2005.

[16] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[17] 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.