CEPSTRAL WEIGHTING FOR SPEECH DEREVERBERATION WITHOUT MUSICAL NOISE

Timo Gerkmann

Sound and Image Processing Lab, Royal Institute of Technology (KTH), 10044 Stockholm, Sweden
gerkmann@kth.se

ABSTRACT

We present an effective way to reduce musical noise in binaural speech dereverberation algorithms based on an instantaneous weighting of the cepstrum. We propose this instantaneous technique, as temporal smoothing techniques result in a smearing of the signal over time and are thus expected to reduce the dereverberation performance. For the instantaneous weighting function we compute the *a posteriori* probability that a cepstral coefficient represents the speech spectral structure. The proposed algorithm incorporates *a priori* knowledge about the speech spectral structure by training the parameters of the respective likelihood function offline using a speech database. The proposed algorithm employs neither a voiced/unvoiced detection nor a fundamental period estimator and is shown to outperform an algorithm without cepstral processing in terms of a higher signal-to-interference ratio, a lower bark spectral distortion, and a lower log kurtosis ratio, indicating a reduction of musical noise.

1. INTRODUCTION

The intelligibility of speech signals that are captured by digital communication devices such as portable phones and hearing aids is often degraded by additive background noise and, in enclosed spaces, reverberation. The negative effect of reverberation increases with an increasing room volume and an increasing distance between speech source and microphones [1], [2]. As an example, for portable phones the negative effect of the reverberation significantly increases when the user is talking hands-free.

While in the literature many algorithms are proposed to reduce additive background noise [3], [4], speech dereverberation received much less attention until recently [2].

If only one microphone signal is present, both additive noise and reverberation can be reduced by weighting the short-time discrete Fourier coefficients with a spectral gain function that may result e.g. from spectral subtraction or the Wiener filter. For speech dereverberation, spectral weighting has been applied e.g. in [5], [1], [6].

In case two or more microphones are present, the spectral gain function can be obtained based on the empirical coherence between microphone pairs [7]. The same real valued spectral gain function can then be applied to all microphone signals, such that the inter-microphone level and time differences remain unchanged (e.g. [8]). This is desirable for instance for binaural hearing aids, where the inter-aural level and time difference are used by the auditory system to localize sources.

The research leading to these results has received funding from the European Communitys Seventh Framework Programme under Grant Agreement PIAP-GA-2008-214699

Spectral enhancement based on the weighting of short-time Fourier spectra often leads to undesired artifacts, often referred to as musical noise. These artifacts occur, if the spectral gain function is locally underestimated in the time-frequency domain. Recently it has been shown that musical noise can be effectively removed by smoothing the cepstral representation of the gain function [9]. This technique employs the observation that the spectral structure of speech is compactly represented by a small set of cepstral coefficients. Further, this set of speech-related cepstral coefficients is mostly disjoint from the set of cepstral coefficients that represents non-speech-like spectral artifacts. Thus, the speech related cepstral coefficients can be preserved while the remaining coefficients are smoothed. It has been shown that this technique does not only remove processing artifacts, but also leads to a better trade-off between speech distortions and noise reduction as compared to competing algorithms.

However, a temporal smoothing of the gain function may result in a slight noise shaping at the end of words which makes the speech sound slightly more reverberant [9]. Clearly, such an additional reverberant effect is not desirable for a dereverberation algorithm. Thus, in this work we propose to use instantaneous cepstral techniques, rather than temporal smoothing, in order to reduce spectral outliers without decreasing the dereverberation performance. For the cepstral weighting function, we employ the *a posteriori* probability that a cepstral coefficient represents the speech spectral structure. We incorporate *a priori* knowledge about speech by training parameters of the *a posteriori* probability offline using a speech database. Further, as opposed to [9], the proposed cepstral weighting technique employs neither a voiced/unvoiced detector, nor a fundamental period estimator.

We apply the proposed cepstral weighting to a coherence based speech dereverberation algorithm, similar to [7]. We describe the reference algorithm for coherence based dereverberation in Section 2 and propose an algorithm for cepstral weighting in Section 3. In Section 4, we show that the proposed algorithm outperforms the reference algorithm, while Section 5 concludes this paper.

2. COHERENCE BASED DEREVERBERATION

In coherence based speech dereverberation [7], the magnitude squared coherence between two microphone signals is measured. For this, the time-domain microphone signals are segmented, windowed and transformed into the frequency domain to obtain the short-time discrete Fourier coefficients $Y_{k,m}(l)$. Here k is the frequency index, m is the microphone index, and l is the segment index. We choose the segments to be of length 32 ms, a normalized Hann spectral analysis window, a sampling frequency of 16 kHz and a discrete

Fourier transform (DFT) length of $32\,\text{ms} \cdot 16\,\text{kHz} = 512$. For the dual-channel case, the spectral gain function is given by

$$G_k(l) = \frac{|\Phi_{k,12}(l)|^2}{\Phi_{k,11}(l)\,\Phi_{k,22}(l)}, \qquad (1)$$

where $\Phi_{k,nm}(l)$ is the cross-power spectral density between microphone signals n and m for $n \neq m$, and $\Phi_{k,mm}(l)$ is the auto-power spectral density of microphone signal m. Often, the gain function $G_k(l)$ is bound to be larger than a lower limit $G_{\min}$, where typically $-25\,\text{dB} < 20\log_{10}(G_{\min}) < -5\,\text{dB}$. $G_{\min}$ reduces the amount of speech distortions and processing artifacts, but also the amount of noise and reverberation reduction. The same real-valued gain function is then applied to disturbed speech of both channels to obtain the enhanced speech $\widehat{S}_{k,m}(l)$ (e.g. [8]):

$$\widehat{S}_{k,m}(l) = G_k(l)\,Y_{k,m}(l). \qquad (2)$$

The power spectral densities are estimated using temporal recursive averaging, as

$$\Phi_{k,nm}(l) = \alpha\,\Phi_{k,nm}(l-1) + (1-\alpha)\,Y_{k,n}(l)\,Y_{k,m}(l)^*. \qquad (3)$$

where we set the smoothing constant to $\alpha = 0.6$ and $n,m \in \{1,2\}$. The low value of $\alpha = 0.6$ results in only little smearing of the speech signal, however also in quite some spectral outliers in the estimate $\Phi_{k,nm}(l)$. Thus, when employing $\Phi_{k,nm}(l)$ in speech enhancement via (1) and (2), the enhanced speech $\widehat{S}_{k,m}(l)$ contains spectral outliers which may be perceived as annoying musical noise.

3. CEPSTRAL WEIGHTING

In this section we propose to reduce outliers in $\Phi_{k,nm}$ by an instantaneous cepstral weighting.

3.1 Definition of the Cepstrum

The cepstral transform is defined as the inverse Fourier transform of the logarithm of squared spectral quantities like Φ_k and denoted by $\text{CEPS}_q\{\cdot\}$, while its inverse is denoted as $\text{ICEPS}_k\{\cdot\}$. We thus have

$$\phi_q = \quad \text{CEPS}_q\{\Phi_k\} = \quad \text{IDFT}_q\{\log(\Phi_k)\}, \qquad (4)$$

$$\Phi_k = \quad \text{ICEPS}_k\{\phi_q\} = \quad \exp\big(\text{DFT}_k\{\phi_q\}\big), \qquad (5)$$

where q is the cepstral index, $\text{DFT}_k\{\cdot\}$ and $\text{IDFT}_q\{\cdot\}$ denote the discrete Fourier transform and its inverse, while the natural logarithm $\log(\cdot)$ denotes the inverse of $\exp(\cdot)$.

We compute the cepstrum of $\Phi_{k,11}(l)$, as

$$\phi_q(l) = \frac{1}{N}\sum_{k=0}^{N-1} \log\big(\Phi_{k,11}(l)\big)\, e^{j\frac{2\pi kq}{N}} = \text{CEPS}_q\big\{\Phi_{k,11}(l)\big\}, \qquad (6)$$

where and $N = 32\,\text{ms} \cdot 16\,\text{kHz}$ is the length of the inverse discrete Fourier transform (IDFT). In (6), we limit $\Phi_{k,11}(l)$ to be larger than 10^{-4} to avoid numerical difficulties when computing the natural logarithm $\log(\cdot)$.

3.2 A posteriori probability weights

While a selective temporal smoothing of the cepstrum has proven to be very effective to reduce processing artifacts in noise reduction algorithms, for speech dereverberation we propose to apply an instantaneous weighting function in the cepstral domain. As the operation is done in each segment l independently, in the sequel we omit the segment index l for brevity.

We assume that the cepstral representation ϕ_q consists of coefficients that represent the speech spectral structure, denoted by the hypothesis $\mathscr{H}_1$, and coefficients that are not speech related, denoted by the hypothesis $\mathscr{H}_0$. In particular, non-speech related coefficients include those cepstral coefficients that represent artifacts, such as undesired spectral outliers. The speech related cepstral coefficients $\overline{\phi}_q$ are obtained using the *a posteriori* probabilities that a cepstral coefficient is speech related, given by $P(\mathscr{H}_1|\phi_q)$, or not, given by $P(\mathscr{H}_0|\phi_q)$, as

$$\begin{aligned}\overline{\phi}_q &= P(\mathscr{H}_1|\phi_q)\phi_q + P(\mathscr{H}_0|\phi_q)\cdot 0 \\ &= P(\mathscr{H}_1|\phi_q)\phi_q. \end{aligned} \qquad (7)$$

Thus, the speech related cepstral coefficients are maintained, while the non-speech related coefficients are attenuated.

With Bayes' theorem, the *a posteriori* probability that a given cepstral coefficient is speech related is given by

$$P(\mathscr{H}_1|\phi_q) = \frac{P(\mathscr{H}_1)p(\phi_q|\mathscr{H}_1)}{P(\mathscr{H}_1)p(\phi_q|\mathscr{H}_1) + P(\mathscr{H}_0)p(\phi_q|\mathscr{H}_0)}. \qquad (8)$$

While the *a priori* probabilities allow to bias the posterior distribution in favor of one of the hypotheses, we set them equal $P(\mathscr{H}_1) = P(\mathscr{H}_0) = 0.5$. Further, a model for the likelihood functions $p(\phi_q|\mathscr{H}_1)$, $p(\phi_q|\mathscr{H}_0)$ is needed. Cepstral coefficients are commonly assumed to be Gaussian distributed [10] due to the central limit theorem and the summation in (6). Thus, we obtain

$$p(\phi_q|\mathscr{H}_1) = \frac{1}{\sqrt{2\pi\sigma_{q,1}^2}}\exp\left(-\frac{(\phi_q-\mu_{q,1})^2}{2\sigma_{q,1}^2}\right) \qquad (9)$$

$$p(\phi_q|\mathscr{H}_0) = \frac{1}{\sqrt{2\pi\sigma_{q,0}^2}}\exp\left(-\frac{(\phi_q-\mu_{q,0})^2}{2\sigma_{q,0}^2}\right). \qquad (10)$$

The mean $\mu_{q,1}$ and variance $\sigma_{q,1}^2$ of speech like cepstral coefficients are trained offline by taking empirical long-term averages of the cepstral coefficients of 15 minutes of clean speech of different male and female speakers from the TIMIT database [11]. By doing so, we incorporate *a priori* knowledge about the speech spectral structure.

On the other hand, we assume the cepstral coefficients that represent non-speech like spectral structures to have zero-mean, $\mu_{q,0} = 0$ and the variance in [12, Eq. (14)], which is derived for cepstral coefficients obtained using Hann-spectral analysis windows, as

$$\sigma_{q,0}^2 = \begin{cases} \frac{2}{N}\left(\kappa_0 + 2\sum_{\nu=0}^{2}\kappa_\nu\cos\left(\nu\frac{2\pi}{N}q\right)\right) & ,q\in\left\{0,\frac{N}{2}\right\} \\ \frac{1}{N}\left(\kappa_0 + 2\sum_{\nu=0}^{2}\kappa_\nu\cos\left(\nu\frac{2\pi}{N}q\right)\right) & ,\text{else}\end{cases}, \qquad (11)$$

with $\kappa_0 = \zeta(2,\eta)$ and Riemann's zeta function $\zeta(2,\eta) = \sum_{n=0}^{\infty} 1/(\eta+n)^2$. For the derivation of (11) it is assumed that the magnitude squared spectral coefficients are χ^2-

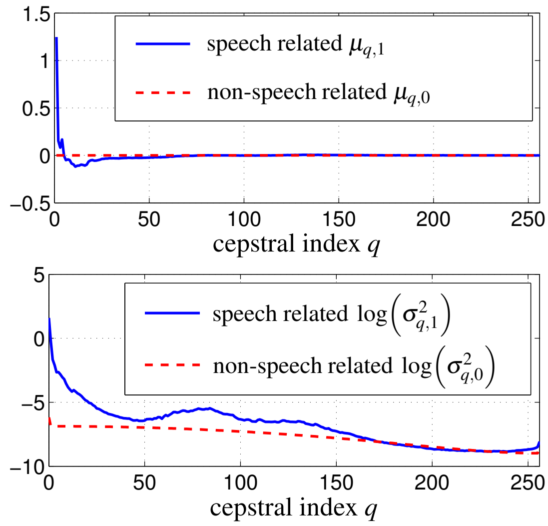

Figure 1: Comparison of the means (top) and variances (bottom) of the cepstral coefficients representing speech-like spectral structures (solid), and non-speech-like spectral structures (dashed).

distributed with 2η degrees-of-freedom. The smoothing of the spectra in (3) results in an increase of the degrees-of-freedom [13]. Assuming that $|Y_{k,m}(l)|^2$ is χ^2-distributed with $\eta = 1$, the increase of the degrees-of-freedom is approximated by $\eta = \frac{1+\alpha}{1-\alpha}$, where α is the smoothing constant in (3). The spectral correlation introduced by using Hann spectral analysis windows is reflected by κ_1 and κ_2, which are derived in [12, Eq. (16)]. For $\alpha = 0.6$ they result in $\kappa_1 = 0.1166$ and $\kappa_2 = 0.0070$.

The resulting values for the means and variances are compared in Figure 1.

3.3 Applying the *a posteriori* probability

Assume we have the cepstral representations

$$\phi_{q,nm}(l) = \mathrm{CEPS}_q\left\{|\Phi_{k,nm}(l)|\right\}, \tag{12}$$

with $n,m \in \{1,2\}$. Then, the corresponding speech related cepstral coefficients can be obtained similar to (7), as $\overline{\phi}_{q,nm} = P(\mathscr{H}_1|\phi_{q,nm})\phi_{q,nm}$. However, instead we apply the same *a posteriori* probability $P(\mathscr{H}_1|\phi_q)$ to $\phi_{q,11}$, $\phi_{q,22}$, and $\phi_{q,12}$, as

$$\overline{\phi}_{q,nm} = P(\mathscr{H}_1|\phi_q)\phi_{q,nm}, \tag{13}$$

where $\phi_q = \phi_{q,11}$ is defined in (6). This simplification is done for two reasons: first, the computational complexity is reduced as fewer cepstral transforms are needed. Secondly, if we apply the same relative cepstral variance reduction to all three power spectral density estimates, $\Phi_{k,12}$, $\Phi_{k,11}$, $\Phi_{k,22}$, the bias introduced by cepstral variance reduction cancels out in (1), as will be explained next.

It has been shown in [12] that each cepstral coefficient has a certain mean that depends on the spectral shape, and a cepstral variance that decreases with an increase of the degrees-of-freedom of χ^2-distributed spectral coefficients. Further, it has been shown that a change of the cepstral variance by temporal cepstrum smoothing or cepstral nulling results in a biased spectral estimate $\widetilde{\Phi}_k = \mathrm{ICEPS}_k\left\{\overline{\phi}_q\right\}$, and that this bias is additive in the log-spectral domain and multiplicative in the linear-spectral domain. The unbiased coefficients $\overline{\Phi}_k$ are then obtained using the bias correction $\mathscr{B}$ as

$$\overline{\Phi}_k = \mathscr{B}\cdot\widetilde{\Phi}_k = \mathscr{B}\cdot\mathrm{ICEPS}_k\left\{\overline{\phi}_q\right\}. \tag{14}$$

This bias also applies for cepstral weighting and is, in the context of this work, a function of the *a posteriori* probability $P(\mathscr{H}_1|\phi_q)$. However, if the same weight $P(\mathscr{H}_1|\phi_q)$ is applied to $\phi_{q,11}$, $\phi_{q,22}$, and $\phi_{q,12}$, under the assumption that $|\Phi_{k,nm}(l)|$ is χ^2-distributed with the same shape parameter η for all n,m (but possibly different spectral means and variances), the multiplicative bias cancels out in (1)

$$\overline{G}_k = \frac{|\overline{\Phi}_{k,12}|^2}{\overline{\Phi}_{k,11}\,\overline{\Phi}_{k,22}} = \frac{|\mathscr{B}\widetilde{\Phi}_{k,12}|^2}{\mathscr{B}\widetilde{\Phi}_{k,11}\,\mathscr{B}\widetilde{\Phi}_{k,22}} = \frac{|\widetilde{\Phi}_{k,12}|^2}{\widetilde{\Phi}_{k,11}\,\widetilde{\Phi}_{k,22}}. \tag{15}$$

Further, if the same *a posteriori* probability is applied to all $\phi_{q,nm}$, the *a posteriori* probability can also be applied directly to the cepstral transform of the gain function. Starting from (15) with (13) and (14), we have

$$\begin{aligned}
\overline{G}_k &= \frac{\left(\mathrm{ICEPS}_k\left\{\phi_{q,12}P(\mathscr{H}_1|\phi_q)\right\}\right)^2}{\mathrm{ICEPS}_k\left\{\phi_{q,11}P(\mathscr{H}_1|\phi_q)\right\}\mathrm{ICEPS}_k\left\{\phi_{q,22}P(\mathscr{H}_1|\phi_q)\right\}}\\
&= \exp\left\{2\mathrm{DFT}_k\left\{\mathrm{IDFT}_q\left\{\log\left(|\Phi_{k,12}|\right)\right\}P(\mathscr{H}_1|\phi_q)\right\}\right.\\
&\quad -\mathrm{DFT}_k\left\{\mathrm{IDFT}_q\left\{\log\left(|\Phi_{k,11}|\right)\right\}P(\mathscr{H}_1|\phi_q)\right\}\\
&\quad \left.-\mathrm{DFT}_k\left\{\mathrm{IDFT}_q\left\{\log\left(|\Phi_{k,22}|\right)\right\}P(\mathscr{H}_1|\phi_q)\right\}\right\}\\
&= \exp\left\{\mathrm{DFT}_k\left\{P(\mathscr{H}_1|\phi_q)\mathrm{IDFT}_q\left\{\log\left(\frac{|\Phi_{k,12}|^2}{\Phi_{k,11}\Phi_{k,22}}\right)\right\}\right\}\right\}\\
&= \mathrm{ICEPS}_k\left\{P(\mathscr{H}_1|\phi_q)\,\mathrm{CEPS}_q\{G_k\}\right\}.
\end{aligned} \tag{16}$$

Thus, for this multichannel algorithm, in total three cepstral transforms are needed. One to compute ϕ_q in (6), and two for (16). The overall algorithm is summarized in Algorithm 1.

Algorithm 1 Proposed algorithm

Learn and store the mean $\mu_{q,1}$ and the variance $\sigma^2_{q,1}$ of speech cepstral coefficients offline using e.g. [11].
Set $\mu_{q,0} = 0$, and $\sigma^2_{q,0}$ according to (11).
for all signal segments l **do**
  Estimate the power spectral densities $\Phi_{k,nm}(l)$ using (3) and $\alpha = 0.6$.
  Compute the spectral gain function (1)
    $G_k(l) = |\Phi_{k,12}(l)|^2 / \left(\Phi_{k,11}(l)\,\Phi_{k,22}(l)\right)$.
  Limit $G_k(l)$ to be larger than -22 dB.
  Compute $\phi_q(l) = \mathrm{CEPS}_q\left\{\Phi_{k,11}(l)\right\}$ (6).
  Compute $P(\mathscr{H}_1|\phi_q(l))$ using (8), (9), and (10).
  Obtain the smoothed gain (16)
    $\overline{G}_k(l) = \mathrm{ICEPS}_k\left\{P(\mathscr{H}_1|\phi_q(l))\cdot\mathrm{CEPS}_q\{G_k(l)\}\right\}$
  Limit $\overline{G}_k(l)$ to be larger than -17 dB.
  Apply the gain function to the microphone spectra
    $\widehat{S}_{k,1}(l) = \overline{G}_k(l)Y_{k,1}(l)$
    $\widehat{S}_{k,2}(l) = \overline{G}_k(l)Y_{k,2}(l)$.
end for

4. EVALUATION

To evaluate the proposed Algorithm 1, we use recordings inside a cafeteria with a reverberation time of 1.76 s. The recordings have been obtained using two hearing aid dummies on an artificial head, where we use one microphone of each hearing aid, respectively. Babble noise of the crowded cafeteria has been recorded, and the acoustic impulse responses to the two hearing aids has been measured in the empty cafeteria. With the measured impulse responses, we synthesize 4 minutes of speech from 3 female and 5 male speakers that are in frontal direction to the artificial head, and in a distance of 1.5 m. We shorten the measured impulse response to 2 seconds, resulting in a direct-to-reverberation ratio [2, (2.33)] of 10 dB. Besides babble noise, we also create diffuse pink Gaussian noise with an inter-microphone correlation typical for human heads using [14].

We compare the *proposed algorithm*, Algorithm 1, to the coherence based dereverberation (1) without cepstral weighting which we refer to as the *reference algorithm* in the sequel. For both approaches we set $\alpha = 0.6$ in (3). For this low value of α the reference algorithm yields little speech distortions but severe spectral outliers that are perceived as musical noise. For both the proposed algorithm and the reference algorithm, the gain function (1) is bounded to be larger than $G_{\min} = -17$ dB. The training for the proposed method is speaker independent.

In Figure 2 the segmental signal to interference ratio (SIR) [2, (2.45)] and the segmental bark spectral distortion (BSD) [15], [2, (2.38)] are given for several segmental input signal-to-noise ratios (SNRs). All quality measures are evaluated on microphone signal one. The segmental input SNR and the segmental SIR are only evaluated where reverberant speech is present. For the input SNR, the "signal" is the reverberant speech of one input channel, and the "noise" is the difference between the noisy reverberant speech and the reverberant speech of the same channel. For the SIR the "signal" is the non-disturbed direct speech component, while the "interference" is the difference between the enhanced signal and the non-disturbed direct speech component,. The SIR thus increases with an increasing noise reduction, increasing dereverberation, and decreasing speech distortions. The BSD predicts the perceived quality of speech coders [15] and speech enhancement algorithms [2]. For this, it measures the difference between the loudness-equalized bark-scaled non-disturbed direct speech component and the loudness-equalized bark-scaled processed signal. This difference is then normalized on the loudness-equalized bark-scaled non-disturbed direct speech component.

From Figure 2, it can be seen that the proposed method that includes cepstral weighting yields a larger SIR improvement as opposed to the reference algorithm. The spectral distortion measured using BSD is lower for low input SNRs, and similar to the reference for high SNRs.

Informal listening revealed, that both for high and low input SNRs cepstral weighting results in more natural sounding results and far less processing artifacts. For a more formal assessment of musical noise, we use the log kurtosis ratio as proposed in [16]. For this we apply the proposed algorithm to 8 minutes of diffuse white Gaussian noise, compute the ratio of the empirical kurtosis before and after processing in each frequency band, and then the log of this ratio. The resulting log kurtosis ratio is given in Figure 3. For the reference algorithm, we obtain a mean log kurtosis ratio of 1.89 while for the proposed method we obtain a mean of 0.86. Thus, the proposed algorithm reduces the log kurtosis ratio by a factor of 2.20, indicating an effective reduction of musical noise.

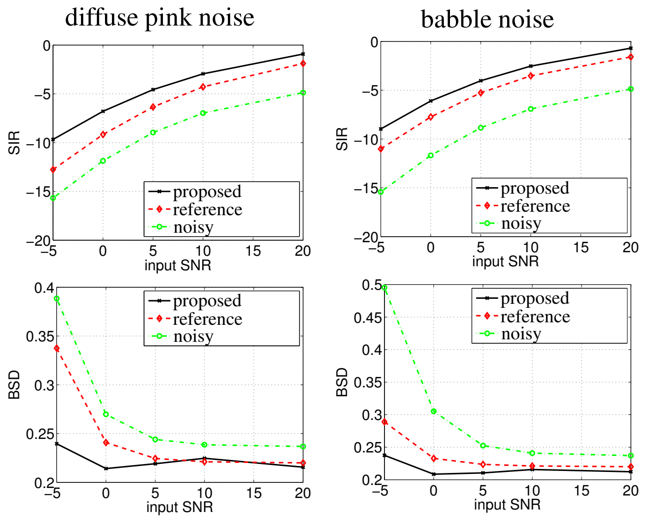

Figure 2: Signal to interference ratio (SIR) (top) and bark spectral distortion (BSD) (bottom) for the proposed and the reference algorithm, evaluated on 4 minutes of speech disturbed by diffuse pink noise (left) and babble noise (right). It can be seen that the proposed algorithm results in a higher SIR improvement, and lower or similar spectral distortions.

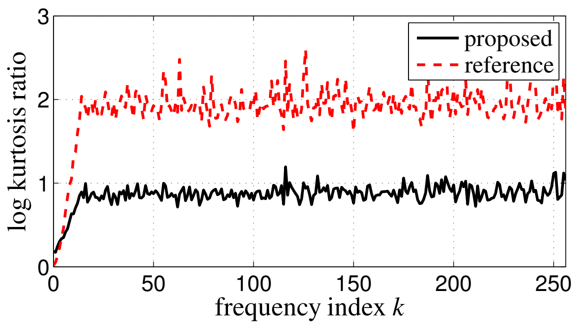

Figure 3: The log kurtosis ratio is an indicator for the amount of musical noise [16]. In this example we processed a spectrally white, spatially diffuse, Gaussian distributed random noise-only signal. It can be seen that the proposed method effectively reduces the log kurtosis by a factor of two, indicating a reduction of spectral outliers and thus of musical noise.

In Figure 4 spectrograms of clean, noisy and processed speech are given, where it can be seen that the proposed algorithm reduces spectral outliers.

5. CONCLUSIONS

In this work we propose to reduce musical noise in dereverberation algorithms by instantaneously weighting the cepstral coefficients of parameters of the dereverberation algorithm. As opposed to temporal smoothing techniques, the proposed instantaneous weighting does not smear the signal over time. This is particularly beneficial for dereverberation

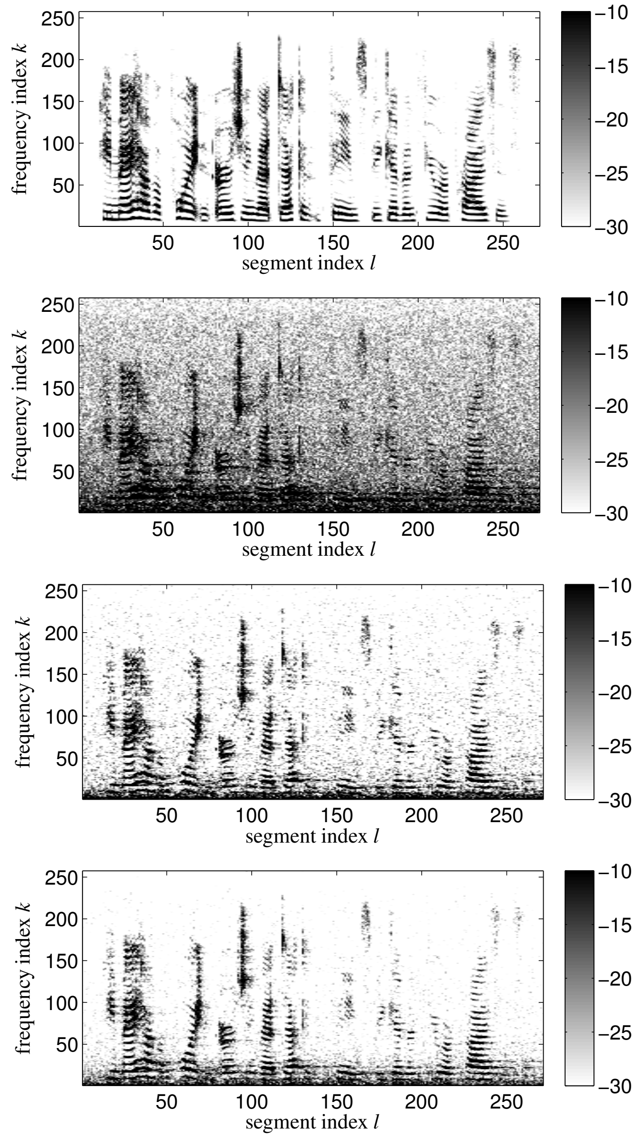

Figure 4: Spectrograms of clean non-reverberant speech (top), noisy reverberant speech (2nd from top), the reference's output (3rd from top), and the proposed algorithm's output (bottom) for 0 dB input segmental SNR, and diffuse spectrally pink Gaussian noise. It can be seen that the proposed algorithm effectively reduces spectral outliers.

algorithms, as a smearing over time is expected to decrease the dereverberation performance. Further, in contrast to temporal cepstrum smoothing, the proposed cepstral weighting does neither require a voiced/unvoiced detector nor fundamental period estimator.

For the weighting function we apply the *a posteriori* probability that a certain cepstral coefficient is speech related. The means and variances of the likelihood function of speech-related cepstral coefficients are obtained empirically from 15 minutes of speech from the TIMIT database.

We apply cepstral weighting to a coherence based dual channel dereverberation algorithm where the two microphone signals are obtained from two hearing aids worn by an artificial head inside a cafeteria. The proposed method in shown to result in larger signal to interference ratio (SIR) improvements and less spectral distortions as measured by the bark spectral distortion (BSD). Further, informal listening and the log kurtosis measure indicate an effective reduction of musical noise. The price for the improved performance is an increase in the computational complexity, dominated by three real-valued and symmetric DFTs.

REFERENCES

[1] E.A.P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, June 2007.

[2] P.A. Naylor and N.D. Gaubitch, Eds., *Speech Derverberation*, Springer Verlag, London, UK, 2010.

[3] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley & Sons, Chichester, West Sussex, UK, 2006.

[4] P.C. Loizou, *Speech Enhancement - Theory and Practice*, CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2007.

[5] K. Lebart, J. M. Boucher, and P.N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica*, vol. 87, pp. 359–366, 2001.

[6] H.W. Löllmann and P. Vary, "A blind speech enhancement algorithm for the suppression of late reverberation and noise," *IEEE ICASSP*, pp. 3989–3992, 2009.

[7] J.B. Allen, D.A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, vol. 62, no. 4, pp. 912–915, Oct. 1977.

[8] M. Jeub, M. Schäfer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio, Speech, Lang. Process*, vol. 18, no. 7, pp. 1732–1745, Sept. 2010.

[9] C. Breithaupt, T. Gerkmann, and R. Martin, "Cepstral smoothing of spectral filter gains for speech enhancement without musical noise," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1036–1039, Dec. 2007.

[10] P. Stoica and N. Sandgren, "Total-variance reduction via thresholding: Application to cepstral analysis," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 66–72, Jan. 2007.

[11] J.S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, 1988.

[12] T. Gerkmann and R. Martin, "On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4165–4174, Nov. 2009.

[13] R. Martin and T. Lotter, "Optimal recursive smoothing of non-stationary periodograms," *IWAENC*, pp. 167–170, Sept. 2001.

[14] C. Borß and R. Martin, "An improved parametric model for perception-based design of virtual acoustics," *AES 35th int. conference*, Feb. 2009.

[15] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. Selected Areas in Communications*, vol. 10, no. 5, pp. 819–829, June 1992.

[16] 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," *IWAENC*, 2008.