AN ANALYSIS OF QUEFRENCY SELECTIVE TEMPORAL SMOOTHING OF THE CEPSTRUM IN SPEECH ENHANCEMENT

Dirk Mauler, Timo Gerkmann, and Rainer Martin

{Dirk.Mauler, Timo.Gerkmann, Rainer.Martin}@rub.de
Institute of Communication Acoustics (IKA), Ruhr-Universität Bochum, 44780 Bochum, Germany

ABSTRACT

Smoothing selected cepstral coefficients over time has been recently shown to be a powerful method in the enhancement of noisy speech signals. A difficulty that arises in this context is that averaging a random variable in the log-domain changes its mean in the linear domain. The knowledge of this bias is indispensable for most temporal cepstrum smoothing applications. To date, the bias is known only for Gaussian distributed signals and infinite averaging length. This contribution presents an analytic solution for signals that are recursively smoothed in the cepstral domain with arbitrary effective averaging lengths. Additionally, the correct bias is computed also for a quefrency dependent setting of the recursive averaging parameters.

Index Terms— cepstrum, temporal smoothing, quefrency, bias compensation, time-frequency plane

1. INTRODUCTION

In many speech processing algorithms careful smoothing of spectral parameters is applied, e.g. to reduce disturbances like musical noise that would otherwise be audible in the output signal. Temporal smoothing must be adaptive to prevent distortion of the speech signal in a noisy mixture.

Recently, temporal smoothing of selected cepstral coefficients was proposed as a robust and powerful method to selectively smooth noisy speech with very little distortion of the speech signal [1], [2]. The method relies on the observation that in the cepstral domain noisy speech is decomposed into coefficients that are dominated by the clean speech spectral envelope, the excitation or the noise. Hence, different degrees of smoothing can be applied to the noise and the speech process in the cepstral domain.

A difficulty that emerges when random variables are temporally smoothed in the cepstral domain is that not only their variance is reduced, which is the desired effect, but also their mean. The latter observation is plausible since the logarithm used in the definition of the cepstrum transforms the arithmetic mean into a geometric mean which is known to be always less or equal to the arithmetic mean. It is the purpose of this paper to derive and discuss this bias w.r.t. the mean value. After a discussion of the general effect of temporal cepstrum (TC) smoothing we give an explicit solution to the bias of TC-smoothed χ^2-distributed random variables in the spectral domain.

This paper starts with a review of temporal cepstrum smoothing. Then, in Sec. 3, we rewrite the TC-average in order to get a new and illustrative insight into TC-smoothing. The main focus of the paper is on the derivation of a bias compensation factor in Sec. 4. Experimental results are reported in Sec. 5.

2. TEMPORAL CEPSTRUM (TC) SMOOTHING

We assume a zero-mean complex-valued signal $Y(k,l)$ in the domain of the short-time discrete Fourier transform (DFT) with normally distributed and uncorrelated real and imaginary parts, each with variance $\sigma_Y^2/2 = 0.5E\{|Y(k,l)|^2\}$. The parameters k,l denote the discrete frequency index and the frame index of a block processing spectral analysis system, respectively.

The cepstrum is defined as the inverse DFT (IDFT) of the natural logarithm of $|Y(k,l)|^2$

$$|Y(q,l)|^2_{ceps} = \mathrm{IDFT}\{\ln(|Y(k,l)|^2)\}. \tag{1}$$

In temporal cepstrum smoothing [1], [2], cepstral coefficients are smoothed over time, e.g. in an efficient manner by using a first order recursive average

$$\overline{|Y(q,l)|^2_{ceps}} = \beta(q)\overline{|Y(q,l-1)|^2_{ceps}} + (1-\beta(q))|Y(q,l)|^2_{ceps}, \tag{2}$$

where $\beta(q) \in [0,1[$ is a quefrency dependent recursive smoothing parameter. Eventually, the corresponding TC-smoothed squared magnitude is obtained after DFT and after applying the exponential function,

$$\overline{|Y(k,l)|^2} = \exp\left(\mathrm{DFT}\{\overline{|Y(q,l)|^2_{ceps}}\}\right). \tag{3}$$

3. REFORMULATION OF THE TC-AVERAGE

The TC-smoothing defined by (1), (2), and (3) is summarized in (4) (see bottom of next page) for an arbitrary frequency bin,

$k = k_0$. Herein, the recursive average (2) has been replaced by its closed form representation

$$\overline{|Y(q,l)|^2_{ceps}} = (1-\beta(q)) \sum_{\lambda=0}^{\infty} \beta^{\lambda}(q) \left(|Y(\kappa, l-\lambda)|^2\right). \quad (5)$$

Furthermore the notation $w_K^{xy} = \exp\left(j2\pi \frac{xy}{K}\right)$ is used where K denotes the transformation length.

After regrouping the complex exponentials we obtain the alternative representations of the TC-average

$$\overline{|Y(k_0,l)|^2} = \exp\left(\frac{1}{K}\sum_{\kappa=0}^{K-1}\sum_{\lambda=0}^{\infty} b_{k_0}(\kappa,\lambda)\ln(|Y(\kappa,l-\lambda)|^2)\right) \quad (6)$$

$$= \prod_{\kappa=0}^{K-1}\prod_{\lambda=0}^{\infty}\left(|Y(\kappa,l-\lambda)|^2\right)^{\frac{1}{K}b_{k_0}(\kappa,\lambda)}. \quad (7)$$

Therein,

$$\frac{1}{K}b_{k_0}(\kappa,\lambda) = \frac{1}{K}\sum_{q=0}^{K-1}(1-\beta(q))\beta^{\lambda}(q)w_K^{q(\kappa-k_0)} \quad (8)$$

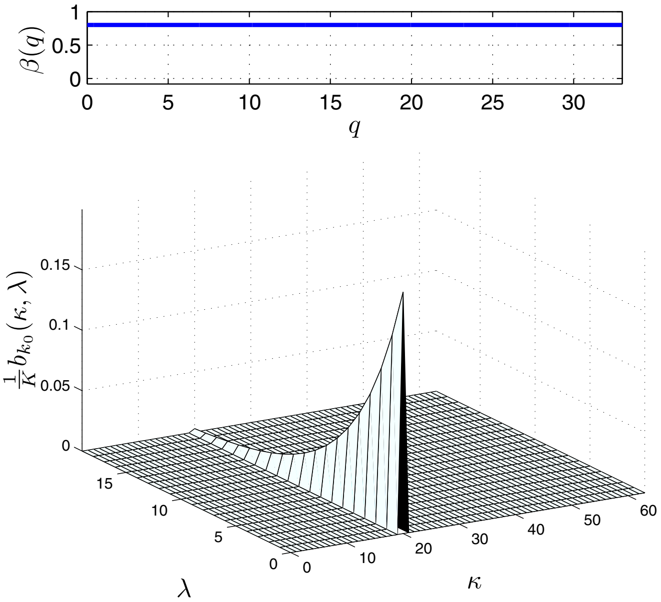
is the IDFT of the expression $(1-\beta(q))\beta^{\lambda}(q)$ modified by the cyclic shift k_0 in the spectral domain. Therefore we have $b_{k_0}(\kappa,\lambda) = b_0(\kappa-k_0,\lambda)$, which shows that only K spectral parameters need to be computed. Furthermore, since $(1-\beta(q))\beta^{\lambda}(q)$ is symmetric, the factors b_{k_0} are real-valued. Note also that $\sum_{\lambda=0}^{\infty}\sum_{\kappa=0}^{K-1}\frac{1}{K}b_{k_0}(\kappa,\lambda) = 1$.

The representations (6) and (7) allow for a new interpretation of the TC-average: According to (6) the TC-average of the spectral parameter at frequency bin k_0 and time index l is obtained via the weighted sum of the logarithm of all preceding data in the time-frequency plane. Similarly, in (7) the data is exponentially weighted before being multiplied over the entire time-frequency plane. Apparently, the selection of data that essentially contribute to the TC-average is determined by the time-frequency weighting function $b_{k_0}(\kappa,\lambda)$, which is discussed next.

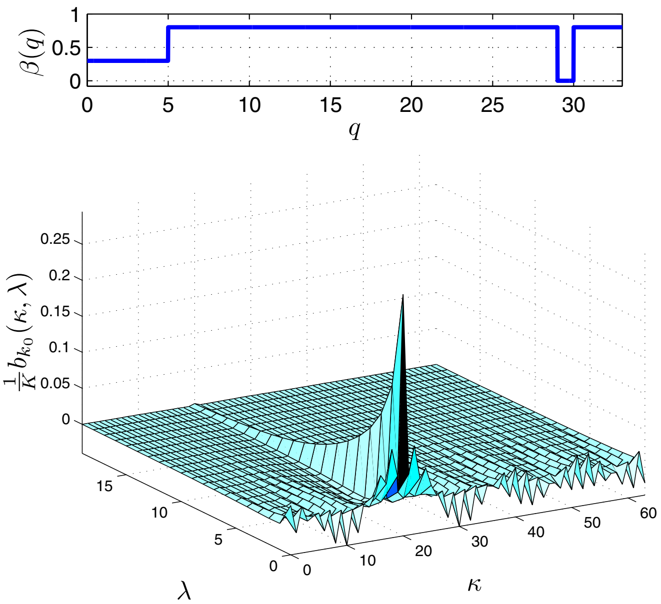
3.1. Behavior of $\frac{1}{K}b_{k_0}(\kappa,\lambda)$

We discuss now how the weights $\frac{1}{K}b_{k_0}(\kappa,\lambda)$ are distributed over the time-frequency plane for a specific pattern of smoothing parameters $\beta(q)$.

In Figure 1 two examples of patterns $\beta(q)$ are given (2-D plots) for which the respective weights $\frac{1}{K}b_{k_0}(\kappa,\lambda)$ over the (truncated) time-frequency plane are shown (3-D plots). As argued before, changing k_0 results in a circular shift of



(a) Using constant-$\beta(q)$-pattern.



(b) Using staircase $\beta(q)$-pattern with notch.

Fig. 1. $\frac{1}{K}b_{k_0}(\kappa,\lambda)$ over the time (λ) -frequency (κ) plane (bottom) for two different patterns of $\beta(q)$ (top) and $k_0 = 20$. A positive value of λ indicates frames in the past.

$$\overline{|Y(k_0,l)|^2} = \exp\left(\sum_{q=0}^{K-1}\left((1-\beta(q))\sum_{\lambda=0}^{\infty}\left(\beta^{\lambda}(q)\frac{1}{K}\sum_{\kappa=0}^{K-1}\left(\ln\left(|Y(\kappa,l-\lambda)|^2\right)w_K^{q\kappa}\right)w_K^{-qk_0}\right)\right)\right) \quad (4)$$

the factors $\frac{1}{K}b_{k_0}(\kappa, \lambda)$ along κ. In these examples we have $k_0 = 20$ and $K = 64$.

If, as in Fig. 1(a), the same recursive averaging parameter is applied for all K quefrencies, $\beta(q) = \beta$, (8) simplifies

$$\frac{1}{K}b_{k_0}(\kappa, \lambda) = \begin{cases} (1-\beta)\beta^{\lambda} & \text{for } \kappa = k_0 \\ 0 & \text{else,} \end{cases} \quad (9)$$

constituting now a delta-function in frequency $\kappa = k_0$ with decaying amplitude over time λ. The smaller the smoothing parameters $\beta(q)$, the faster decays the exponential weighting in (8), therefore limiting the influence of past data on the TC-average.

In case of a quefrency dependent pattern $\beta(q)$, like in Fig. 1(b), we first discuss $\lambda = 0$, which denotes the index of the actual frame. In this case, $\frac{1}{K}b_{k_0}(\kappa, 0) = \mathrm{IDFT}\{1 - \beta(q)\}$ which shows that the pattern of the weights along frequency κ is a superposition of the negative IDFT of $\beta(q)$ and a delta impulse at $\kappa = k_0$. Past frames ($\lambda > 0$) are weighted with the IDFT of the expression $(1 - \beta(q))\beta(q)^{\lambda}$. This expression exponentially decays towards past frames with a decaying constant depending on $\beta(q)$.

4. BIAS OF TC-AVERAGES

The expected value of a logarithmically distorted random variable is smaller than the expected value of the undistorted random variable

$$\ln\left(E\left\{|Y(k,l)|^2\right\}\right) = E\left\{\ln\left(|Y(k,l)|^2\right)\right\} + B, \quad (10)$$

where B is a constant which equals the Euler constant $\gamma = 0.5772\ldots$ for complex Gaussian distributed $Y(k,l)$ [3]. For stationary signals, the expectation operator can be approximated by recursive smoothing with a large smoothing constant $\beta(q) \to 1$. Then, due to (10), TC-smoothing results in $E\{\overline{|Y(k,l)|^2}\} = E\{|Y(k,l)|^2\} \cdot C_{bias}$ [2]. We now derive an analytic solution for the bias C_{bias} for arbitrary $\beta(q)$.

At first we determine an expression for the expected value of TC-smoothed squared magnitudes. Applying the expectation operation to (7) and assuming spectral and temporal uncorrelateness of the squared magnitudes the expectation operation may be applied to each factor in the product separately. For the following computations the analysis window is rectangular with no frame overlap.

$$E\left\{\overline{|Y(k_0,l)|^2}\right\} = \prod_{\kappa=0}^{K-1}\prod_{\lambda=0}^{\infty} E\left\{\left(|Y(\kappa, l-\lambda)|^2\right)^{\frac{1}{K}b_{k_0}(\kappa,\lambda)}\right\}.$$

With the Gaussian assumption it follows that the squared magnitudes $|Y(k,l)|^2$ are χ^2-distributed with with $N = 1$ degree of freedom for frequency bin $k = 0$ and for the Nyquist-bin and with $N = 2$ degrees of freedom otherwise. For simplicity and since the transformation length K is usually large we proceed assuming a χ^2-distribution of degree $N = 2$ for all frequency bins. Using theorem [4, 3.381,4.], the expectation integral can be solved

$$E\left\{\overline{|Y(k_0,l)|^2}\right\} = \prod_{\kappa=0}^{K-1}\prod_{\lambda=0}^{\infty}\left\{\left(\sigma_Y^2(\kappa, l-\lambda)\right)^{\frac{1}{K}b_{k_0}(\kappa,\lambda)} \cdot\Gamma\left(\frac{1}{K}b_{k_0}(\kappa,\lambda)+1\right)\right\}. \quad (11)$$

We now assume that the recursive averaging parameters $\beta(q)$ are chosen so as to perform the TC-smoothing on short-time stationary sections of the process. Then

$$E\left\{\overline{|Y(k_0,l)|^2}\right\} = \prod_{\kappa=0}^{K-1}\left(\sigma_Y^2(\kappa,l)\right)^{\sum_{\lambda=0}^{\infty}\frac{1}{K}b_{k_0}(\kappa,\lambda)} \cdot \prod_{\kappa=0}^{K-1}\prod_{\lambda=0}^{\infty}\Gamma\left(\frac{1}{K}b_{k_0}(\kappa,\lambda)+1\right). \quad (12)$$

Inserting (8) into the exponent of the variances, then changing the order of the summations, and finally using $\sum_{\lambda=0}^{\infty}(1-\beta)\beta^{\lambda} = 1$ shows that

$$\sum_{\lambda=0}^{\infty}\frac{1}{K}b_{k_0}(\kappa,\lambda) = \begin{cases} 1 & \text{for } \kappa = k_0 \\ 0 & \text{else,} \end{cases} \quad (13)$$

which simplifies (12) and gives the final result

$$E\left\{\overline{|Y(k_0,l)|^2}\right\} = \sigma_Y^2(k_0,l)\cdot C_{bias},$$
$$\text{with } C_{bias} = \prod_{\kappa=0}^{K-1}\prod_{\lambda=0}^{\infty}\Gamma\left(\frac{1}{K}b_{k_0}(\kappa,\lambda)+1\right) \quad (14)$$
$$\text{and } b_{k_0}(\kappa,\lambda) \text{ as in (8).}$$

$C_{bias} \in [0,1]$ is the factor by which the expected value of the TC-average is biased as compared to the power $\sigma_Y^2(k_0,l)$ of the unsmoothed spectral amplitudes. The value of k_0 in (14) can be arbitrarily set, e.g. $k_0 = 0$, as according to (8) changing k_0 results in a cyclic shift of the factors $b_{k_0}(\kappa,\lambda)$ along frequency κ which has no effect on the total product (14). Hence the factor C_{bias} is a function of the smoothing parameters only. For a fix pattern $\beta(q)$ C_{bias} can thus be computed offline.

In the following we discuss special cases of sets of smoothing parameters $\beta(q)$ and the resulting factors (14).

4.1. Equal smoothing along all quefrency bins q

In case of equal and constant $\beta(q) = \beta$ the result in (9) is used again, yielding the factor

$$C_{bias} = \prod_{\lambda=0}^{\infty}\Gamma\left((1-\beta)\beta^{\lambda}+1\right).$$

In Fig. 2 we show the natural logarithm of the inverse of this factor, normalized by the Euler constant, γ. For $\beta \to 1$,

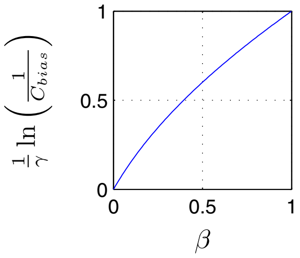


Fig. 2. Necessary additive correction in the log-domain to achieve unbiased power estimates. The log-compensation factor is normalized to the Euler constant, $\gamma = 0.5772\ldots$.

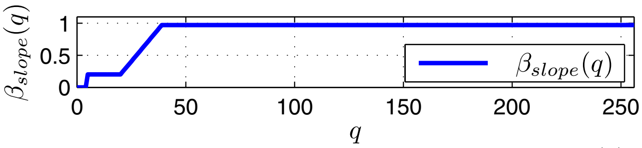


Fig. 3. Example of quefrency dependent smoothing, $\beta(q)$.

the TC-smoothing approximates the expectation operator, and $\ln(1/C_{bias}) \to \gamma$. When no TC-smoothing is performed, i.e. for $\beta = 0$, no bias occurs and $\ln(1/C_{bias}) \to 0$. For all other values of β the compensation term should be set according to the concave function in Fig. 2.

4.2. Quefrency dependent smoothing

In some applications a quefrency dependent setting of the recursive averaging parameters $\beta(q)$ is required. As an example, we consider the vector of smoothing parameters $\beta_{slope}(q)$ in Fig. 3 which applies strong smoothing to the fine-structure (large q) and less smoothing to the envelope and to the low-quefrency structure of a signal. The necessary additive correction in the log-domain to achieve unbiased estimates is $\ln(1/C_{bias}) = 0.87\gamma$.

5. EXPERIMENTAL RESULTS

We performed TC-smoothing on the squared magnitudes of real data ($K = 512$) with and without the proposed bias compensation (14) and computed the relative error between the mean of the unsmoothed and the mean of TC-smoothed squared magnitudes. The pattern of the averaging parameters $\beta(q)$ is as in Section 4.1. Stationary signals are used to allow temporal averaging in the evaluation.

In Fig. 4 we show the relative error over β and for white and for pink Gaussian noise signals. The prediction of the bias is accurate for white and also for pink Gaussian noise. The latter shows that even if the squared magnitudes are weakly correlated over the time-frequency plane, thus violating the assumptions, still good predictions can be obtained.

If the frequency dependent smoothing parameters $\beta(q)$ from Fig. 3 are used, the relative errors are without bias compensation -0.39 (white) and -0.38 (pink) and are $9 \cdot 10^{-4}$

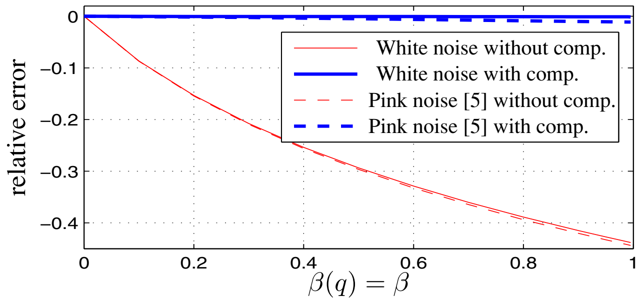


Fig. 4. Relative error between mean of unsmoothed, $\overline{|Y(k,l)|^2}$, and mean of TC-smoothed squared magnitudes, $\overline{|Y(k,l)|^2}$, with (thick line) and without (thin line) bias compensation (14) for two signal types.

(white) and 0.019 (pink) if the bias compensation is used.

6. CONCLUSIONS AND OUTLOOK

We developed a representation of temporal cepstrum (tc) smoothing of random variables in the spectral domain that shows how smoothing in the cepstral domain affects the squared spectral magnitudes of the input signal. Based on this the bias has been derived by which a TC-smoothed random variable is on average smaller than the unsmoothed random variable. The bias factor is a function of the smoothing parameters along quefrency and is accurate for uncorrelated and weakly correlated signals. Further work will concentrate on improvements in case of strongly correlated data.

7. REFERENCES

[1] Colin Breithaupt, Timo Gerkmann, and Rainer Martin, "Cepstral Smoothing of Spectral Gains for Speech Enhancement Without Musical Noise," *IEEE Signal Proc. Letters*, vol. 14, no. 12, pp. 1036–1039, Dec. 2007.

[2] Colin Breithaupt, Timo Gerkmann, and Rainer Martin, "A Novel A Priori SNR Estimation Approach Based on Selective Cepstro-Temporal Smoothing," in *ICASSP*, 2008, pp. 4897–4900.

[3] Yariv Ephraim and Mazin Rahim, "On Second-Order Statistics and Linear Estimation of Cepstral Coefficients," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 162–176, March 1999.

[4] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 6. edition, 2000.

[5] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, July 1993.