Speech Enhancement Using Kalman Filtering in the Logarithmic Bark Power Spectral Domain

Nikolaos Dionelis and Mike Brookes
Department of Electrical and Electronic Engineering
Imperial College London
London, UK
{nikolaos.dionelis11, mike.brookes}@imperial.ac.uk

***Abstract*—We present a phase-sensitive speech enhancement algorithm based on a Kalman filter estimator that tracks speech and noise in the logarithmic Bark power spectral domain. With modulation-domain Kalman filtering, the algorithm tracks the speech spectral log-power using perceptually-motivated Bark bands. By combining STFT bins into Bark bands, the number of frequency components is reduced. The Kalman filter prediction step separately models the inter-frame relations of the speech and noise spectral log-powers and the Kalman filter update step models the nonlinear relations between the speech and noise spectral log-powers using the phase factor in Bark bands, which follows a sub-Gaussian distribution. The posterior mean of the speech spectral log-power is used to create an enhanced speech spectrum for signal reconstruction. The algorithm is evaluated in terms of speech quality and computational complexity with different algorithm configurations compared on various noise types. The algorithm implemented in Bark bands is compared to algorithms implemented in STFT bins and experimental results show that tracking speech in the log Bark power spectral domain, taking into account the temporal dynamics of each subband envelope, is beneficial. Regarding the computational complexity, the percentage decrease in the real-time factor is 44% when using Bark bands compared to when using STFT bins.**

***Index Terms*—Speech enhancement, phase-sensitive observation model, phase factor, Bark bands, Kalman filter**

I. Introduction

Single-channel speech enhancement in non-stationary noise environments remains a challenging task. In the Short-Time Fourier Transform (STFT) time-frequency domain, inter-frame speech correlation exists and modulation-domain Kalman filtering refers to imposing temporal constraints on spectra such as the amplitude spectrum, [1], the power spectrum or the log-power spectrum. Due to noise, the temporal characteristics of the trajectories of the speech amplitude spectrum are distorted. Enhancement algorithms can benefit from modeling inter-frame speech correlation with a Kalman filter (KF) with a state of low dimension and a number of authors have found that the performance of a speech enhancer can be improved by using modulation-domain Kalman filtering, [1], [2], [3].

Modulation-domain Kalman filtering operates in a spectral time-frequency domain and changes the modulation spectrum. Inter-frame speech correlation modeling with a KF has been addressed in [4] [5]. Modulation-domain Kalman filtering, [2], is different from Kalman filtering in the time domain, [6].

In this paper, we present a phase-sensitive enhancement algorithm based on a KF that estimates speech and noise in the low-dimensional logarithmic Bark power spectral domain. By exploiting the temporal dynamics of speech, we track the evolution of the speech spectral log-power in each Bark-spaced frequency band. We use inter-frame linear and nonlinear relationships for the KF prediction and update steps, respectively. The nonlinear KF update step models the phase factor, [7], i.e. the cosine of the phase difference between speech and noise, in order to use that speech and noise are additive in the complex STFT domain. We approximate the posterior of the speech and noise spectral log-powers as a two-dimensional Gaussian distribution with a full covariance matrix using the probability distribution of the phase factor in Bark bands. The phase-sensitive KF update step computes the first two moments of the posterior distribution, [8], [9], thus suppressing noise.

As a main contribution, we create a phase-sensitive Kalman filtering algorithm to track speech and noise in the logarithmic Bark power spectral domain. We extend the speech enhancer in [7] to work in Bark bands and we take into account the phase factor in Bark bands using time-varying frequency-dependent weighted sigma points for its sub-Gaussian distribution, [10]. The modulation-domain Kalman filtering algorithm uses perceptually-motivated Bark bands that take into account the ear resolution. The use of low-dimensional Bark bands, instead of STFT bins, reduces the computational complexity of the KF algorithm in [7] and improves the speech quality.

II. The speech enhancement algorithm

The flowchart of the KF-based enhancement algorithm is shown in Fig. 1. The input of the algorithm is the noisy speech in the time domain. The algorithm's first step is to perform the STFT and to obtain both the noisy spectral amplitude, $|Y|$, and the noisy phase, θ. Next, the algorithm performs three different actions. First, it performs speech amplitude spectrum pre-cleaning and estimates an autoregressive (AR) model of order p for the spectral log Bark power of speech. Second, the algorithm performs noise amplitude spectrum pre-cleaning, using voice activity detection, and estimates an AR model of order q for the spectral log Bark power of noise. Third, the algorithm performs modulation-domain Kalman filtering for every Bark-spaced frequency band. For the modulation-domain KF, the observation is the spectral log Bark power of noisy speech. The KF state is the spectral log Bark power of speech together with the spectral log Bark power of noise.

ISBN 978-90-827970-1-5

According to Fig. 1, the noisy amplitude spectrum is used in two ways: (a) it is converted to the spectral log Bark power domain and used as the KF observation, and (b) it is pre-cleaned, [8], and used for speech and noise AR(p) and AR(q) modeling respectively in the spectral log Bark power domain. Speech amplitude spectrum pre-cleaning is used because the algorithm's input is the noisy speech signal and we want to perform AR modeling on the speech signal. For the same reasons as for speech, noise pre-cleaning is also performed.

The main part of the algorithm is the KF. The modulation-domain Kalman filtering algorithm performs a nonlinear phase-sensitive KF update step to track the spectral log Bark power of speech. In the end, using the enhanced speech STFT amplitude spectrum, the algorithm performs the inverse STFT (ISTFT) to obtain the speech signal in the time domain.

A. Signal model and Bark bands

In the complex STFT domain, the noisy speech is given by

$$|Y_t(k)|\ e^{j\theta_t(k)} = |X_t(k)|\ e^{j\phi_t(k)} + |N_t(k)|\ e^{j\psi_t(k)} \qquad (1)$$

where the time-frame index is denoted by t and the STFT frequency bin index by k for $1 \leq k \leq K$. The spectral amplitudes of the noisy speech, clean speech and noise are respectively denoted by $|Y_t(k)|$, $|X_t(k)|$ and $|N_t(k)|$ while the corresponding phases are $\theta_t(k)$, $\phi_t(k)$ and $\psi_t(k)$.

The phase factor in STFT bins, α_k, is given by $\alpha_k = \cos(\phi(k) - \psi(k))$, as in [9], [7] and [8]. For clarity, we omit the time-frame index, t, below and we only include it in equations involving multiple frames. We use the Bark band index, l, for $1 \leq l \leq L$, $1 \leq k \leq K$ and $K > L$. A filterbank comprising triangular filters, which are similar to those used for Mel bands in [11] and [12], is used to transform the power spectrum of each frame from a number of STFT bins, K, to a reduced number of Bark subbands, L. The noisy speech in the log Bark power spectral domain, $y(l)$, is given by

$$y(l) = \log\left(\sum_{k=1}^{K} W_{k,l}\,|Y(k)|^2\right) \qquad (2)$$

where $W_{k,l}$ are the overlapping triangular filter weights used to go to the Bark power spectral domain, [13]. The speech and the noise in the log Bark power spectral domain, $x(l)$ and $n(l)$, are defined similarly to (2), using $|X(k)|$ and $|N(k)|$.

Considering the relation between noisy speech, speech and noise in the logarithmic Bark power spectral domain, [10], and using l and y, x and n, the nonlinear distortion equation in the logarithmic Bark power spectral domain is given by

$$y(l) = \log\left(e^{x(l)} + e^{n(l)} + 2\beta_l \times e^{0.5(x(l)+n(l))}\right) \qquad (3)$$

where β_l is the Bark phase factor. This β_l is given by

$$\beta_l = \frac{\sum_k W_{k,l}\,|X(k)|\,|N(k)|\,\alpha_k}{\sqrt{\sum_k W_{k,l}|X(k)|^2} \times \sqrt{\sum_k W_{k,l}|N(k)|^2}}, \qquad (4)$$

$$c_{k,l} = \frac{W_{k,l}\,|X(k)|\,|N(k)|}{\exp\left(0.5(x(l)+n(l))\right)}, \qquad \beta_l = \sum_{k=1}^{K} c_{k,l}\,\alpha_k \qquad (5)$$

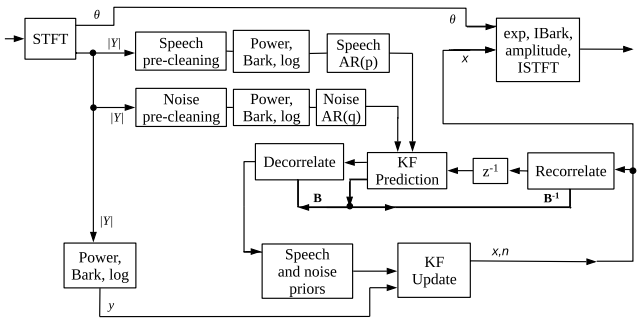


Figure 1. The flowchart diagram of the KF-based enhancement algorithm that tracks and estimates speech in the logarithmic Bark power spectral domain.

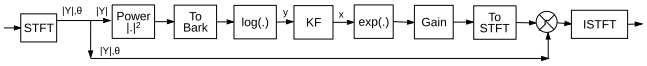


Figure 2. The flowchart diagram shows how spectral amplitude estimation is performed using Bark-spaced frequency bands, as explained in Sec. II.B.

where β_l follows a sub-Gaussian distribution, [10]. The phase factor in STFT bins, α_k, follows the arcsine distribution

$$f_{\alpha_k}(\alpha_k) = \begin{cases} \dfrac{1}{\pi \times \sqrt{1-\alpha_k^2}} & \text{if } |\alpha_k| < 1 \\ 0 & \text{otherwise.} \end{cases} \qquad (6)$$

In [10], $|X(k)|$ and $|N(k)|$ are assumed constant within a Bark band. Here, the two cases, listing them with increasing complexity, are: (a) assume $|X(k)|$ and $|N(k)|$ are constant within a Bark band, and (b) make no assumptions and use $|X(k)|$ and $|N(k)|$. Using (b), from (4) and (5), we obtain

$$\mathbb{E}\left\{\beta_l^2\right\} = 0.5\sum_k c_{k,l}^2,\ \ \mathbb{E}\left\{\beta_l^4\right\} = 3\mathbb{E}\left\{\beta_l^2\right\}^2 - 0.375\sum_k c_{k,l}^4$$

$$\mathbb{E}\left\{\beta_l^6\right\} = 1.25\sum_k c_{k,l}^6 - 5.625\mathbb{E}\left\{\beta_l^2\right\}\sum_k c_{k,l}^4 + 15\mathbb{E}\left\{\beta_l^2\right\}^3$$

and the odd moments of β_l are zero. The latter equation for the sixth moment, $\mathbb{E}\{\beta_l^6\}$, is not included in [10] and is computed using the relation between cumulants and moments, [14].

B. Spectral amplitude estimation using Bark bands

Figure 2 depicts how speech spectral log-power tracking is performed in Bark bands. The "Bark" block in the flowchart diagram in Fig. 1 converts the STFT power spectral domain to the Bark power spectral domain using (2). The "IBark" block in Fig. 1 performs an inverse Bark (IBark) operation to go from Bark bands to STFT bins for the amplitude spectrum.

As shown in Fig. 2, the algorithm first performs the STFT, then goes to the power spectral domain, to the Bark power spectral domain and to the log Bark power spectral domain. The KF operates in the log Bark power spectral domain. Using the noisy spectral power and the exponential of the KF output, a gain is computed. To map a number of Bark bands onto a larger number of STFT bins, linear interpolation in frequency is performed on the gain to impose a smoothness constraint. The gain in a STFT bin is a weighted average of the gains in

ISBN 978-90-827970-1-5 © EURASIP 2018

the Bark bands whose centre frequencies are either side of it and the weights are inversely proportional to the differences between the Bark band and the STFT bin centre frequencies.

C. The KF state and the KF prediction step

The KF state consists of the speech KF state, $\mathbf{x}_t$, and the noise KF state, $\mathbf{x}_t^{(\mathrm{n})}$. The speech KF prediction step uses the equations in (7), [7]. The speech KF transition matrix from speech AR modeling of order p is $\mathbf{A}_t$, the speech KF transition noise covariance matrix is $\mathbf{Q}_t$ and the speech KF transition noise, which is zero-mean with covariance matrix $\mathbf{Q}_t$, is $\mathbf{w}_t$.

$$\begin{aligned} &\mathbf{x}_t = (x_t \; x_{t-1} \; \ldots \; x_{t-p+1})^T, \qquad \mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{w}_t \\ &\mathbf{x}_t \in \Re^p, \; \mathbf{A}_t \in \Re^{p \times p}, \qquad \mathbf{w}_t \in \Re^p, \; \mathbf{Q}_t \in \Re^{p \times p} \end{aligned} \tag{7}$$

In (7), $\mathbf{A}_t$ and $\mathbf{Q}_t$ are found from AR modeling on the pre-cleaned speech, [7]. As in (7), noise tracking based on AR(q) modeling is performed for $\mathbf{x}_t^{(\mathrm{n})} = (n_t \, n_{t-1} \, \ldots \, n_{t-q+1})^T$, [7].

D. The phase-sensitive KF update step

The KF update step computes the first two moments of the posterior distribution of speech and noise in the log Bark power spectral domain. As shown in Fig. 1, decorrelation is performed before the KF update using $\mathbf{B} \in \Re^{(p+q)\times(p+q)}$, as in [8] and [3]. Decorrelation and recorrelation are performed before and after the nonlinear KF update step, respectively.

The KF update step uses the equations presented in [7] and [9] but utilises the Bark phase factor, β_l, instead of the phase factor, α_k, and the log Bark power spectral domain instead of the log-power spectral domain. The algorithm tracks speech and noise using correlated priors in the log Bark power spectral domain. The KF update considers the two-dimensional Gaussian prior for speech and noise from the KF prediction step, the distribution of β_l from Sec. II.A and the observation constraint surface in the three-dimensional space (x, n, β_l).

The nonlinear distortion equation is given by (3). To calculate the posterior, we need to obtain the conditional distribution of (x, n, β_l) subject to the observation constraint, y. Applying the observation constraint reduces the dimension of the distribution from three to two. To impose this constraint, we make a transformation of variables from (x, n, β_l) to (u, y, β_l). Since the transformed parameterization includes y, it becomes straightforward to impose the constraint. For $0 \leq a + b \leq 2$, using $u = n - x$, $x = x(u, y, \beta_l)$ and $n = n(u, y, \beta_l)$, the first two moments of the posterior of (x, n) are computed using

$$\mathbb{E}\left\{x^a n^b \mid y\right\} \propto \int_{\beta_l} p(\beta_l) \int_u x^a n^b \; p(x, n) \; du \; d\beta_l \tag{8}$$

where the outer integration over the Bark phase factor, β_l, is performed using G sigma points. $\mathbb{E}\{\beta_l^z\}$ for $z \in \mathbb{Z}_{\geq 0}$, as computed in Sec. II.A, is needed for the sigma points, [7].

E. Discussion of the KF algorithm

We denote the presented KF algorithm by BSNT to indicate Bark speech and noise tracking. We denote the speech tracking algorithm implemented in Bark bands by BST, which is a simpler version of the BSNT algorithm. We note that BST does not track the log Bark power spectrum of noise. BSNT and BST use the correct sub-Gaussian distribution for the phase factor in Bark bands, β_l, [10], as described in Sec. II.A. The sigma points for β_l in BSNT and BST change in every time-frequency cell and depend on estimates of $|X(k)|$ and $|N(k)|$, using $c_{k,l}$ in (5). BSNT and BST do not assume that $|X(k)|$ and $|N(k)|$ are constant within each Bark band and use time-varying and frequency-dependent sigma points and weights.

The BSNT and BST algorithms compute the first moments of the sub-Gaussian distribution for the phase factor in Bark bands, β_l, in every time-frequency cell. This approach differs from the offline-training approach that is followed in [15] [16], where Gaussian distributions are used to model the phase factor in Mel bands invoking the central limit theorem.

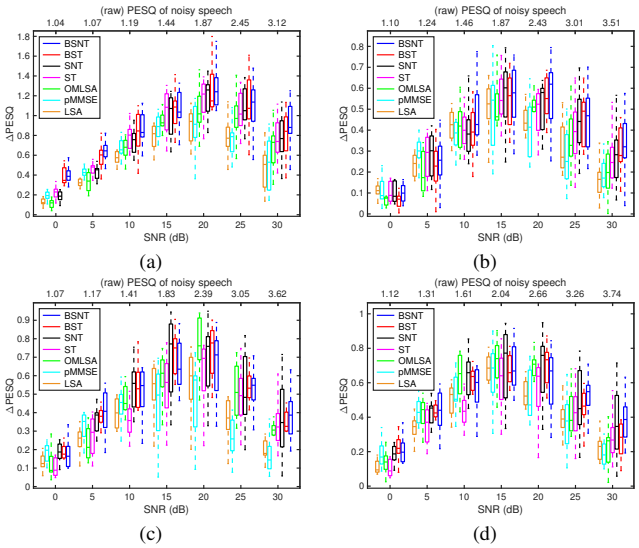


Figure 3. Boxplots of ΔPESQ scores at SNRs from 0 dB to 30 dB for: (a) white noise, (b) babble noise, (c) F16 noise, and (d) factory noise.

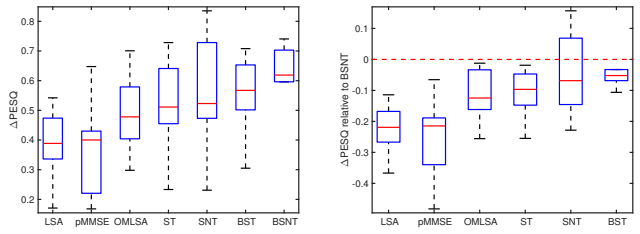
III. Implementation, Results and Evaluation

We use 32 ms acoustic frames, an 8 ms acoustic frame hop, 64 ms modulation frames and an 8 ms modulation frame hop. In Fig. 1, for speech amplitude spectrum pre-cleaning, we use the Log-MMSE estimator, [17], [13]. The KF state dimensions for speech and noise in Sec. II.C are respectively $p = 2$ and $q = 2$. We use the external noise estimator from [18], [13]. In Sec. II.D, $G = 3$ sigma points are utilised, as in [7] [8].

For evaluation, the TIMIT database [19], sampled at 16 kHz, and the RSG-10 noise database [20] are used. From the TIMIT core test set, 50 speech utterances are chosen. The proposed BSNT algorithm is examined at SNRs from 0 to 30 dB.

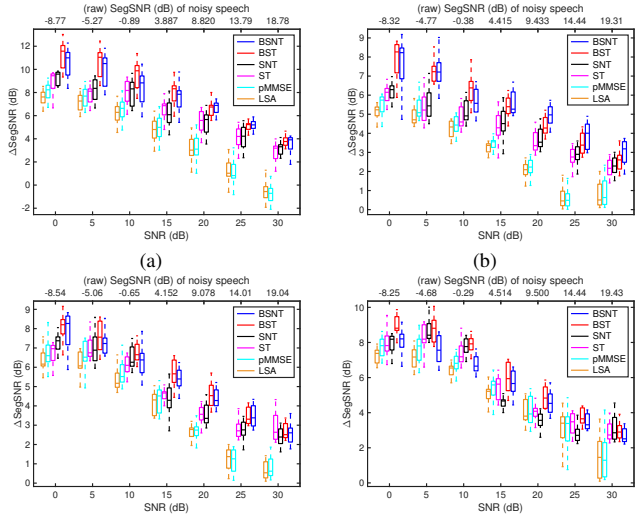
Contrary to expectations, there is no penalty in speech quality when using Bark bands. The tradeoff between computational complexity and speech quality is not apparent. The use of low-dimensional Bark bands, instead of STFT bins, reduces the computational complexity of the KF algorithm that jointly tracks the speech and noise spectral log-powers and also improves the speech quality. The frequency precision is higher when using STFT bins than when using Bark bands;

ISBN 978-90-827970-1-5 © EURASIP 2018



(a) (b)

Figure 4. (a) ΔPESQ and (b) ΔPESQ relative to BSNT averaged over the noise types of white, babble, F16 and factory when the noisy PESQ is 2.8.



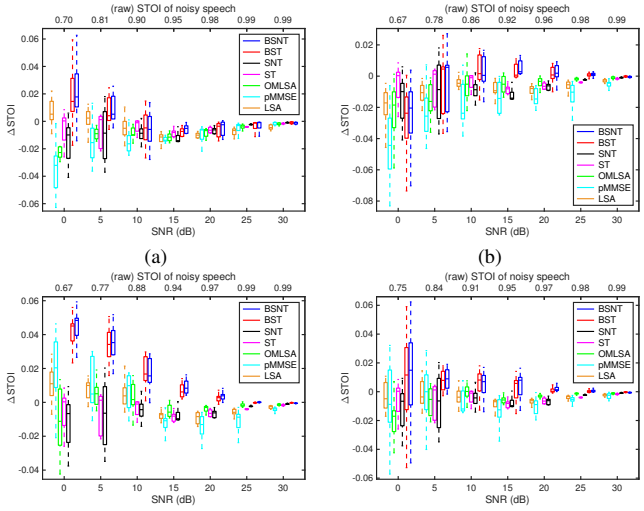
(a) (b) (c) (d)

Figure 5. Boxplots of ΔSegSNR scores at SNRs from 0 dB to 30 dB for: (a) white noise, (b) babble noise, (c) F16 noise, and (d) factory noise.

nevertheless, the algorithm implemented in Bark bands has comparable to and/or better speech quality results than the algorithm implemented in STFT bins. In terms of the real-time factor, R, the computational complexity of the proposed algorithm is lower with Bark bands than with STFT bins. When the STFT length is 512, we use 20 Bark bands instead of 257 STFT bins. According to our implementation using parallel computation for processing frequency components in parallel, using two cores, the real-time factor is $R = 16$ for the STFT-bin-based algorithm while $R = 9$ for the Bark-band-based algorithm. The percentage decrease in R is 44% when using Bark bands compared to when using STFT bins.

We denote the speech tracking (ST) algorithm from [7] by ST and the speech and noise tracking (SNT) algorithm from [7] by SNT. We compare the presented BSNT algorithm with the BST, SNT and ST KF baselines in terms of speech quality with the perceptual evaluation of speech quality (PESQ), [21], the segmental SNR (SegSNR) and the short-time objective intelligibility (STOI), [22], metrics. We compare the Bark and the full STFT implementations of the different algorithms.

We also compare the proposed KF-based BSNT algorithm with the non-KF baselines: (a) the Log-MMSE log-spectral amplitude (LSA) estimator, [17], implemented with the MMSE



(a) (b) (c) (d)

Figure 6. Boxplots of ΔSTOI scores at SNRs from 0 dB to 30 dB for: (a) white noise, (b) babble noise, (c) F16 noise, and (d) factory noise.

noise estimator, [18], with (b) the optimally modified LSA (OMLSA) estimator, [23], implemented with the improved minima controlled recursive averaging (IMCRA) noise estimator, [24], and with (c) the perceptually-motivated MMSE (pMMSE) estimator from [25], as implemented in [13], using weighted Euclidean distortion with a power exponent of -1.

Figure 3 shows the PESQ improvement, ΔPESQ, for the presented BSNT algorithm, the non-KF baselines, OMLSA, pMMSE and LSA, and the KF baselines, BST, SNT and ST, for the noise types of white, babble, F16 and factory. The boxplots show the median and the inter-quartile range, as well as the 5% and 95% points of the distribution, of ΔPESQ. The ordering of the legends matches that of the plots at high SNRs. According to Figs. 3(a)-3(d), the BSNT algorithm consistently improves the PESQ metric, approximately 0.14 on average compared to the non-KF baselines for all the examined SNRs and noise types. The BSNT algorithm consistently improves the PESQ metric at most SNRs compared to the ST and SNT algorithms, depending on the noise type. For middle SNRs, compared to the unprocessed speech signal, the PESQ improvement of BSNT is approximately 1.2 for white noise, 0.6 for babble and factory noises and 0.7 for F16 noise.

To examine the performance over a range of noise types, we evaluate the algorithms in Fig. 4(a) using the four examined noise types of white, babble, F16 and factory with the average SNR for each noise type chosen to give a mean PESQ score of 3.0 for the noisy speech. Compared to the unprocessed speech signal, Fig. 4(a) shows that the BSNT algorithm achieves a PESQ improvement of 0.62, while LSA and pMMSE achieve a PESQ improvement of approximately 0.4. OMLSA has a PESQ improvement of about 0.5, ST and SNT achieve a PESQ improvement of 0.5 and BST has a PESQ improvement of 0.55. According to Fig. 4(b), the BSNT algorithm has a higher PESQ improvement of approximately 0.22 compared to the LSA and pMMSE baselines, of 0.12 compared to OMLSA,

ISBN 978-90-827970-1-5 © EURASIP 2018

of 0.1 compared to ST and SNT and of 0.05 compared to BST. According to Figs. 4(a) and 4(b), considering speech and noise combinations that have a raw PESQ of 3.0, there is a consistent benefit in ΔPESQ when using Bark bands.

Figure 5 shows the SegSNR improvement, ΔSegSNR, of the presented BSNT algorithm compared to the non-KF baselines, LSA and pMMSE, and to the KF baselines, BST, SNT and ST. The BSNT algorithm has a consistently improved SegSNR compared to the non-KF baselines for all the examined noise types. BSNT achieves higher ΔSegSNR scores for white and babble noises, and smaller for F16 and factory noises. For most SNRs and noise types, the BSNT algorithm has a consistently improved SegSNR score compared to the KF baselines.

Figure 6 depicts the STOI improvement, ΔSTOI, of the presented BSNT algorithm and the non-KF and KF baselines. The BSNT algorithm shows marginal STOI improvements.

To sum up, with log-spectrum modulation-domain Kalman filtering in Bark bands, we reduce the computational complexity of the KF algorithm and also achieve better speech quality performance, while preserving intelligibility. The use of low-dimensional Bark bands, instead of STFT bins, reduces the frequency components of the KF algorithm in [7]. Performing temporal dynamics tracking in the log Bark power spectral domain is beneficial for enhancement and the Bark-band-based KF algorithm shows better speech quality results compared to KF baselines that are implemented in STFT bins. For a perceptual comparison, the reader is referred to [26] where some recordings processed by the BSNT algorithm are available.

IV. Conclusion

In this paper, we present a phase-sensitive enhancement algorithm based on modulation-domain Kalman filtering in the log Bark power spectral domain. The algorithm jointly tracks speech and noise using a KF estimator that operates in the low-dimensional log Bark power spectral domain. By combining STFT bins into Bark bands, we reduce the number of frequency components. The use of Bark bands, instead of STFT bins, speeds up the algorithm. The nonlinear KF update step models the effect of noise on the speech spectral log-power using the distribution of the phase factor in Bark bands. Experimental results show that tracking speech in the log Bark power spectral domain, taking into account the temporal dynamics of each Bark subband envelope, is beneficial.

References

[1] S. So and K. K. Paliwal, "Modulation-domain Kalman filtering for single-channel speech enhancement," *Speech Communication, vol. 53, no. 6, pp. 818-829*, July 2011.

[2] Y. Wang and M. Brookes, "Model-based speech enhancement in the modulation domain," *IEEE Trans. on Audio, Speech and Language Process., vol. 26, no. 3, pp. 580-594*, March 2018.

[3] Y. Wang, "Speech enhancement in the modulation domain, Ch. 5: Model-based speech enhancement in the modulation domain," Ph.D. dissertation, Imperial College London, 2015.

[4] T. Esch and P. Vary, "Speech enhancement using a modified Kalman filter based on complex linear prediction and supergaussian priors," *in Proc. IEEE Int. Conf. Audio and Speech Signal Process., pp. 4877-4880, Las Vegas*, April 2008.

[5] T. Esch, "Model-based speech enhancement exploiting temporal and spectral dependencies, Ch. 3: Speech enhancement incorporating temporal correlation," Ph.D. dissertation, Aachen University, 2012.

[6] M. S. Kavalekalam, M. G. Christensen and J. B. Boldt, "Model based binaural enhancement of voiced and unvoiced speech," *in Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, pp. 666-670*, March 2017.

[7] N. Dionelis and M. Brookes, "Modulation-domain speech enhancement using a Kalman filter with a Bayesian update of speech and noise in the log-spectral domain," *in Proc. IEEE Int. Work. Hands-free Speech Communication and Microphone Arrays, San Francisco*, March 2017.

[8] N. Dionelis and M. Brookes, "Phase-aware single-channel speech enhancement with modulation-domain Kalman filtering," *IEEE Trans. on Audio, Speech and Language Process., vol. 26, no. 5, pp. 937-950*, May 2018.

[9] N. Dionelis and M. Brookes, "Speech enhancement using modulation-domain Kalman filtering with active speech level normalized log-spectrum global priors," *in Proc. European Signal Process. Conf., Kos*, Aug. 2017.

[10] V. Leutnant and R. Haeb-Umbach, "An analytic derivation of a phase-sensitive observation model for noise-robust speech recognition," *in Proc. Conf. Int. Speech Communication Association, pp. 2395-2398, Brighton*, Sept. 2009.

[11] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continously Spoken Sentences," *IEEE Trans. on Acoustics, Speech and Signal Process.*, vol. 28, no. 4, pp. 357–366, Aug. 1980.

[12] C. S. J. Doire, M. Brookes et al., "Single-channel online enhancement of speech corrupted by reverberation and noise," *IEEE Trans. on Audio, Speech and Language Process., vol. 25, no. 3, pp. 572-587*, March 2017.

[13] M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," Imperial College London, Software Library, 2018, [Online]. Available: http://www.ee.imperial.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[14] M. Kendall and A. Stuart, *Distribution Theory*, 4th ed., ser. The Advanced Theory of Statistics. Charles Griffin, 1977, vol. 1.

[15] L. Deng, J. Droppo, and A. Acero, "Enhancement of log Mel power spectra of speech using a phase-sensitive model of the acoustic environment and sequential estimation of the corrupting noise," *IEEE Trans. on Speech and Audio Process., vol. 12, no. 2, pp. 133-143*, April 2004.

[16] J. Li, L. Deng, R. Haeb-Umbach and Y. Gong, *Robust automatic speech recognition, Ch. 3: Background of robust speech recognition, ISBN: 978-0-12-802398-3*. Elsevier, 2016.

[17] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. on Acoustics, Speech and Signal Process.*, vol. 33, no. 2, pp. 443–445, April 1985.

[18] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. on Audio, Speech, and Language Process., vol. 20, no. 4, pp. 1383-1393*, 2012.

[19] J. Garofolo, L. Lamel, W. Fisher et al, "TIMIT accoustic-phonetic continuous speech corpus," *Corpus LDC93S1, Linguistic Data Consortium, Philadelphia*, 1993.

[20] H. Steeneken and F. Geurtsen, "Description of the RSG-10 noise database," *TNO Institute for perception, Tech. Rep. IZF 1988-3*, 1988.

[21] ITU-T, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU-T Rec P.862, Feb. 2001.

[22] C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. on Audio, Speech and Language Process., vol. 19, no. 7, pp. 2125-2136*, Sept. 2011.

[23] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403–2418, 2001.

[24] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, pp. 12–15, 2002.

[25] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech Audio Process., vol. 13, no. 5, pp. 857-869*, 2005.

[26] N. Dionelis and M. Brookes, "Single-channel speech enhancement using modulation-domain Kalman filtering," 2017, [Online]. Available: https://www.commsp.ee.ic.ac.uk/~sap/speech-enhancement-using-modulation-domain-kalman-filtering/.

ISBN 978-90-827970-1-5 © EURASIP 2018