A POSTERIORI SPEECH PRESENCE PROBABILITY ESTIMATION BASED ON AVERAGED OBSERVATIONS AND A SUPER-GAUSSIAN SPEECH MODEL

Balázs Fodor

Institute for Communications Technology
Technische Universität Braunschweig
Braunschweig, Germany

Timo Gerkmann

Speech Signal Processing Group
Cluster of Excellence "Hearing4all"
Universität Oldenburg
Oldenburg, Germany

ABSTRACT

Explicit information about speech presence or absence is needed in many speech processing applications. In a Bayesian estimation framework, this information can be provided by an *a posteriori* speech presence probability (SPP) estimator. Recent improvements in SPP estimation include likelihoods of speech presence based on a super-Gaussian speech model or, alternatively, based on averaged observations. In this paper, we combine these aspects and derive a closed form solution for the likelihood of speech presence based on both averaged observations and a super-Gaussian speech model. The new approach is shown to outperform competing methods that either include averaging or super-Gaussian speech models.

1. INTRODUCTION

It was shown, e. g., in [1, 2] that minimum mean squared error (MMSE) based clean speech estimators can be enhanced by taking speech presence uncertainty (SPU) into account. As a result, MMSE estimation under SPU turns out to be a product of a common MMSE estimator for the speech and an *a posteriori* speech presence probability (SPP) estimator [1]. *A posteriori* SPP estimators are most commonly based on a Gaussian assumption for the speech [2–4], however, it was shown, e. g., in [5] that speech discrete Fourier transform (DFT) coefficients can better be modeled by super-Gaussian distributions than by a Gaussian. Accordingly, also *a posteriori* SPP estimators based on a super-Gaussian speech model have been proposed, e. g., [6, 7].

While common *a posteriori* SPP estimators are able to successfully achieve values close to one if speech is present, many estimators return only the *a priori* SPP if speech is absent [4]. This problem is mended in [3] by adapting the *a priori* SPP. In contrast to this, in [4] it was shown that SPP estimates close to zero in speech absence can also be obtained by choosing both a fixed *a priori* SPP and a fixed *a priori* signal-to-noise ratio (SNR). While in [2, 3] the *a priori* SNR reflects an SNR which is present in a local time-frequency unit, in [4] it is argued that in order to distinguish between speech presence and absence, the *a priori* SNR employed in the likelihood of speech presence should reflect an SNR which is typical for speech presence. This SNR can be mathematically optimized for typical application scenarios [4].

Outliers in *a posteriori* SPP estimation can be reduced, e. g., by driving the likelihood functions by averaged observations [4] or by employing hidden Markov models [8]. However, deriving likelihood functions for averaged observations based on super-Gaussian priors can be difficult and so far only approximate solutions have been proposed [9].

In this paper, however, we present a closed-form solution for likelihood functions based on both averaged observations and a super-Gaussian speech model. Additionally, the corresponding new *a posteriori* SPP estimator makes use of the advantage of a fixed *a priori* SNR.

This paper is organized as follows: Section 2 gives an introduction to *a posteriori* SPP estimation with averaged observations assuming a Gaussian speech model, *a posteriori* SPP estimation without averaging assuming a super-Gaussian speech model, and *a posteriori* SPP estimation using fixed prior parameters. Our proposed approach that unifies the advantages of averaged observations, a super-Gaussian speech model, and fixed *a posteriori* SPP parameters is introduced in Section 3. In Section 4 the advantage of the proposed estimators are demonstrated, while Section 5 concludes this paper.

2. MMSE ESTIMATION UNDER SPU

We assume the following short-time Fourier transform (STFT) domain signal model: We observe noisy speech $Y(\ell, k) = S(\ell, k) + N(\ell, k)$ where the clean speech coefficients $S(\ell, k)$ are disturbed by additive noise $N(\ell, k)$. Here, ℓ and k denote the frame index and the frequency bin index, respectively. Using polar notation, the speech $S(\ell, k) = A(\ell, k) \cdot e^{j\alpha(\ell,k)}$ can be described by the speech spectral amplitude $A(\ell, k)$ and the speech spectral phase $\alpha(\ell, k)$. Note that in the sequel, we will omit the indices ℓ, k for ease of readability. Employing MMSE short-time spectral amplitude (STSA) estimation [2], the absolute value of speech is estimated using the observed noisy speech Y and some *a priori* knowledge about the speech and the acoustic channel. Furthermore,

978-1-4799-6808-4/14/$31.00 ©2014 IEEE

taking SPU into account, i. e., introducing the hypothesis of speech presence H_1 and speech absence H_0, the speech spectral amplitude $A = |S|$ estimate results in [1]

$$\widehat{A} = P(H_1|Y) \cdot E\{A|Y, H_1\} \tag{1}$$

with the *a posteriori* SPP estimator $P(H_1|Y)$, the statistical expectation $E\{\cdot\}$, and the MMSE estimator for the speech amplitude under speech presence $E\{A|Y, H_1\}$. In this paper we make the common assumptions that the speech S and the noise N are statistically independent, the speech spectral phase α is uniformly distributed and statistically independent of the speech spectral amplitude A [10, 11], and that the complex-valued noise DFT coefficients N are bivariate Gaussian distributed [2]. Accordingly, the *a posteriori* SNR $\gamma = |Y|^2/\sigma_N^2$ with σ_N^2 being the noise power spectral density (PSD) turns out to be independent of the speech spectral phase. Therefore, the *a posteriori* SPP can also be written as a function of γ, i. e., $P(H_1|Y) = P(H_1|\gamma)$ [12].

2.1. SPP Estimation Assuming Averaged Observations

Random fluctuations in the noisy observation may cause estimation outliers in the *a posteriori* SPP that may result in annoying musical noise in a speech enhancement task. To improve estimation robustness, these outliers can be reduced by averaging *a posteriori* SNRs in adjacent time-frequency bins as [4]

$$\bar{\gamma}(\ell, k) = \frac{1}{\nu} \sum_{\substack{\lambda \in \mathbb{L} \\ \kappa \in \mathbb{K}}} \gamma(\lambda, \kappa) \tag{2}$$

with $\mathbb{L}$ and $\mathbb{K}$ being a set of frames and a set of frequency bins within the averaging window, respectively. Furthermore, $\nu = |\mathbb{L}| \cdot |\mathbb{K}|$ is the total number of *a posteriori* SNR values within the averaging window. Driven by the averaged *a posteriori* SNRs $\bar{\gamma}$, the *a posteriori* SPP can be written as a function of the generalized likelihood ratio (GLR) $\bar{\Lambda}$

$$P(H_1|\bar{\gamma}) = \frac{\bar{\Lambda}}{1 + \bar{\Lambda}} \tag{3}$$

with

$$\bar{\Lambda} = \frac{P(H_1)}{P(H_0)} \cdot \frac{p(\bar{\gamma}|H_1)}{p(\bar{\gamma}|H_0)}. \tag{4}$$

Here, $P(H_1)$ and $P(H_0) = 1 - P(H_1)$ are the *a priori* SPP and speech absence probability, while $p(\bar{\gamma}|H_1)$ and $p(\bar{\gamma}|H_0)$ are the likelihood of speech presence and the likelihood of speech absence, respectively, both based on averaged observations.

Assuming a bivariate Gaussian distribution for the speech S and the noise N, the *a posteriori* SNR values $\gamma = |S + N|^2/\sigma_N^2$ follow the exponential distribution, which is a special case of the chi-squared distribution with shape parameter $\nu_{\chi^2} = 1$ [4]. Furthermore, averaged *a posteriori* SNR values $\bar{\gamma}$ can be modeled by a chi-squared distribution [4] with an increased shape parameter [13]. Accordingly, the GLR for averaged observations (4) turns out to be [4]

$$\bar{\Lambda}_{[4]} = \frac{P(H_1)}{P(H_0)} \cdot \left(\frac{1}{1+\xi}\right)^{\nu_{\chi^2}} \cdot \exp\left(\nu_{\chi^2} \frac{\xi}{1+\xi} \bar{\gamma}\right) \tag{5}$$

with the *a priori* SNR $\xi = \sigma_S^2/\sigma_N^2$ and the speech PSD σ_S^2.

2.2. SPP Estimation for Super-Gaussian Speech

While (5) is based on a Gaussian assumption for the speech, it was shown in [5] that speech can better be modeled by super-Gaussian distributions. As in [6, 14], the super-Gaussian character of speech is modeled by assuming that the speech spectral amplitudes A follow a chi-distribution with a shape parameter $\mu < 1$. Without averaging, the likelihood of speech presence can be obtained by [6, Eq. (20)]

$$p(\gamma|H_1) = p(\gamma|H_0) \cdot \left(\frac{\mu}{\mu+\xi}\right)^{\mu} \cdot {}_1F_1\left(\mu; 1; \frac{\xi}{\mu+\xi}\gamma\right) \tag{6}$$

where ${}_1F_1(\cdot)$ is the confluent hypergeometric function [15]. The novelty of this paper is that in Section 3 we will derive a closed form solution for the likelihood of speech presence for averaged observations (2) assuming a super-Gaussian speech model as in (6).

2.3. SPP Estimation with Fixed Parameters

Generally, the likelihood of speech presence is a function of the *a priori* SNR ξ (cf., e. g., (6)). While in MMSE spectral estimation ξ represents the local SNR in each time frequency unit, in *a posteriori* SPP estimation the situation is different. As the likelihood of speech presence (e. g., (6)) is a model for speech presence, also its parameter ξ can be interpreted as a model parameter for speech presence. However, if ξ represents the local SNR, in speech absence we have $\xi = 0$ and, thus, $p(\gamma|H_1) = p(\gamma|H_0)$ results in (6). As a consequence, the distinction between speech presence and absence using the likelihoods is not possible and the *a posteriori* SPP simply outputs the *a priori* SPP (cf. (3) and (4) with $p(\gamma|H_1) = p(\gamma|H_0)$), instead of a value close to zero. To overcome this issue, instead of adapting ξ to follow the local SNR in each time-frequency bin, in [4] it was proposed to find a fixed *a priori* SNR that reflects a typical SNR which can be expected in speech presence. With this fixed *a priori* SNR ξ, the *a posteriori* SPP estimate is capable of yielding values close to zero in speech absence. The fixed *a priori* SNR ξ is obtained by interpreting the *a posteriori* SPP estimator as a detector and minimizing the total probability of false detections for a typical range of input SNRs [4].

3. PROPOSED SPP ESTIMATOR

In this section, we extend (6) to averaged *a posteriori* SNRs obtained by (2). Assuming statistically independent observations, the probability density function (PDF) of the sum of two random *a posteriori* SNR values can be obtained by convolving (6) with itself. If the resulting PDF is again convolved with (6) we obtain the PDF for the sum of three random variables and so forth. Accordingly, using one-sided convolution [16, Eq. (15.93)] by [15, Eq. (7.613.4)] and induction, we obtain the likelihood of speech presence for averaged observations and a super-Gaussian speech model

$$p(\bar{\gamma}|H_1) = p(\bar{\gamma}|H_0) \cdot \left(\frac{\mu}{\mu+\xi}\right)^{\nu\mu} \cdot {}_1F_1\left(\nu\mu; \mu; \frac{\nu \cdot \xi}{\mu+\xi}\bar{\gamma}\right) \quad (7)$$

where, as in Section 2.1, $p(\bar{\gamma}|H_0)$ is a chi-squared distribution with shape parameter ν. As can be seen, this likelihood is a function of both the shape parameter of the chi distribution μ (cf. Section 2.2) and the number of averaged SNR values ν (cf. (2)). Inserting (7) into (4) results in the GLR for averaged *a posteriori* SNRs assuming chi-distributed speech spectral amplitudes

$$\boxed{\bar{\Lambda}_{[\text{new}]} = \frac{P(H_1)}{P(H_0)} \cdot \left(\frac{\mu}{\mu+\xi}\right)^{\nu\mu} \cdot {}_1F_1\left(\nu\mu; \mu; \frac{\nu \cdot \xi}{\mu+\xi}\bar{\gamma}\right).} \quad (8)$$

Note that (8) generalizes both the GLR for averaged observations under a Gaussian speech model and the GLR for non-averaged observations but a super-Gaussian speech model: For $\mu = 1$ we obtain the GLR for averaged Gaussian speech (5) derived in [4], while for $\nu = 1$ we obtain the GLR for non-averaged super-Gaussian speech [6] with the likelihood of speech presence (6).

3.1. Parameter Choice

Just as in [1], we employ $P(H_1) = 0.5$. Furthermore, we model the speech spectral amplitudes with a shape parameter $\mu = 0.5$ which has been shown to be a good compromise between musical noise and speech preservation in speech enhancement [6]. Moreover, the proposed *a posteriori* SPP estimator reduces estimation outliers using averaging. It was shown in [3] that a good tradeoff between speech distortion and the amount of estimation outliers can be achieved by using two averaging processes, instead of one. Accordingly, we used (2) to obtain averaged *a posteriori* SNRs within a local and a global window, resulting in the averaged *a posteriori* SNR $\bar{\gamma}_\Theta$ with $\Theta \in \{\text{local}, \text{global}\}$ [4]. Each averaging window covers the current frame ℓ and the previous $\Delta\ell_\Theta$ frames and, therefore, has the width $|\mathbb{L}_\Theta| = \Delta\ell_\Theta + 1$. The height of each averaging window is $|\mathbb{K}_\Theta| = 2 \cdot \Delta k_\Theta + 1$, i. e., besides the current frequency bin k, Δk_Θ frequency bins below it and Δk_Θ frequency bins above it are employed for averaging. Therefore, each averaging window contains $\nu_\Theta = |\mathbb{K}_\Theta| \cdot |\mathbb{L}_\Theta|$ *a posteriori* SNR values. In this paper, we use the same window sizes as in [4] (cf. Table 1).

Then, applying the resulting averaged *a posteriori* SNRs to the new GLR (8) results in two likelihood ratios, denoted by $\bar{\Lambda}_\Theta$ with $\Theta \in \{\text{local}, \text{global}\}$. However, as in Section 2.3, we employ a fixed *a priori* SNR ξ_Θ for calculating the GLR $\bar{\Lambda}_\Theta$, instead of an adapted one. For the training of ξ_Θ, the same training steps were employed as in [4] and the resulting ξ_Θ values are given in Table 1. Accordingly, $\bar{\Lambda}_\Theta$ is calculated by (8) using the averaged SNR $\bar{\gamma}_\Theta$ from (2) as well as $\xi = \xi_\Theta$, $\nu = \nu_\Theta$, and μ from Table 1. Assuming that the time-frequency units in each averaging window are uncorrelated, $\nu_\Theta = |\mathbb{K}_\Theta| \cdot |\mathbb{L}_\Theta|$ corresponds to the number of time-frequency points in the respective averaging window.

Θ	Δk_Θ	$\Delta\ell_\Theta$	ν_Θ	μ	ξ_Θ
local	1	2	9	0.5	12.6 dB
global	8	2	51		10.4 dB

Table 1. Parameters of the averaging framework

The local and global *a posteriori* SPP can be obtained by applying $\bar{\Lambda}_\Theta$ to (3), resulting in $P(H_1|\bar{\gamma}_\Theta)$ with $\Theta \in \{\text{local}, \text{global}\}$. Similar to the proposals in [3, 4] we obtain the final *a posteriori* SPP estimate by $P(H_1|\bar{\gamma}) = P(H_1|\bar{\gamma}_{\text{local}}) \cdot P(H_1|\bar{\gamma}_{\text{global}})$ which can directly be used in (1) [4].

4. PERFORMANCE EVALUATION

The proposed approach and the reference approaches were evaluated by the following simulations: As speech data, we employed the English subset of the NTT Multi-Lingual Speech Database [17]. As noise data, we used car, factory, and babble noise from the NTT Ambient Noise Database [18]. First, the database signals were downsampled to 8 kHz sampling rate. The input SNR was adjusted between -5 dB and 20 dB in 5 dB steps by scaling the speech and noise components separately by means of the active speech level and the root mean square (RMS) noise level, respectively, as recommended by ITU-T P.56 [19]. The noisy speech signal was then transformed into the STFT domain using frames with a length of $L = 256$ samples, a frame shift of 50 %, a square-root Hann window for both spectral analysis and synthesis, and a DFT. Using the resulting noisy speech STFT coefficients, the noise power was estimated by the minimum statistics approach [20]. Subsequently, the *a priori* SNR was obtained by the well-known decision-directed approach [2] with the smoothing factor 0.98. The speech DFT coefficients were estimated by (1) and then transformed back into the time domain using an inverse DFT, synthesis windowing, and an overlap-add step.

The proposed and the reference approaches were evaluated with respect to the speech component quality, the amount of noise attenuation, and the amount of musical noise in a signal-component-wise manner: Applying the spectral weights to the signal components of the noisy speech, namely the speech component and the noise component, results in the processed speech and the processed noise component, respectively. Accordingly, the speech component quality was measured by the segmental speech-to-speech distortion ratio SSDR_{seg} using the speech component and its processed replica [21]. The larger the SSDR_{seg}, the less distortion was introduced to the speech component through processing. The amount of noise attenuation was measured by the segmental noise attenuation measure NA_{seg} based on the noise component and its processed noise replica [21]. The larger the NA_{seg}, the more is the noise component attenuated through processing. The amount of musical noise was measured by

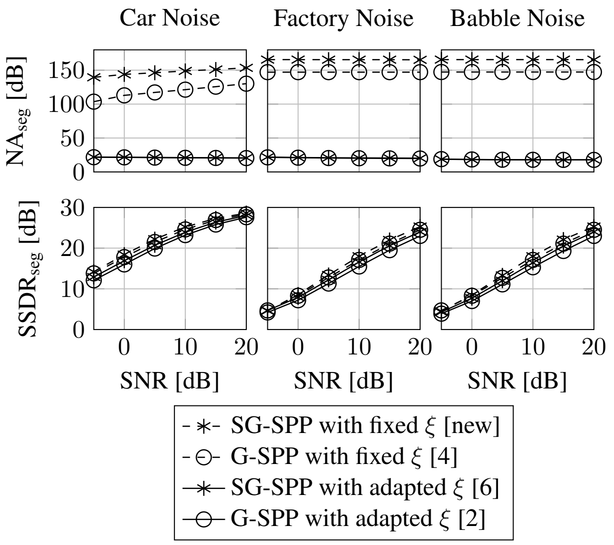

Fig. 1. Evaluation results in terms of segmental speech-to-speech distortion ratio ($SSDR_{seg}$, the larger the better) and segmental noise attenuation (NA_{seg}, the larger the better) [21]

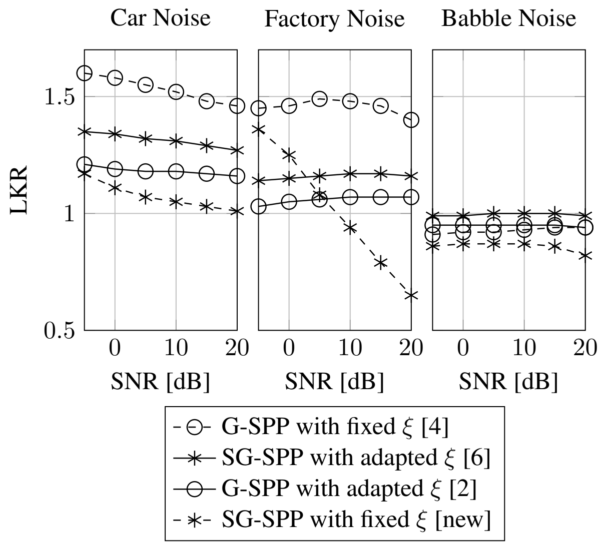

Fig. 2. Evaluation results in terms of the amount of musical noise based on the weighted log-kurtosis ratio (LKR, the smaller the better) [22]

the weighted log-kurtosis ratio (LKR) [22]. This measure compares the kurtosis of the *noise component* of the noisy speech signal before and after processing in speech pauses, resulting in the LKR [22]. Large values of the LKR indicate a large amount of processing outliers that may be perceived as annoying musical noise.

The proposed and the reference approaches were of the form (1), i. e., they consist of a common MMSE estimator for the speech and an *a posteriori* SPP estimator. The first reference approach (denoted as "G-SPP with adapted ξ [2]") consists of the MMSE estimator [2, Eq. (7)] and the *a posteriori* SPP with the GLR in [2, Eq. (27)] both based on Gaussian speech and noise models. The SPP parameter ξ was adapted using the decision-directed approach and thus follows the local SNR in the time-frequency plane. The second reference approach (denoted as "SG-SPP with adapted ξ [6]") consists of the MMSE estimator [14, Eq. (6)] and the corresponding *a posteriori* SPP estimator with the GLR [6, Eq. (20)]. Both the MMSE and the SPP estimator assume a super-Gaussian speech model, i. e., chi-distributed speech amplitudes with the shape parameter $\mu = 0.5$. Again, the model parameter ξ of the GLR is adapted and follows the local SNR in the time-frequency plane. The third reference approach (denoted as "G-SPP with fixed ξ [4]") consists of the MMSE estimator [2, Eq. (7)] and the *a posteriori* SPP estimator with the GLR (5). This GLR is driven by averaged *a posteriori* SNRs (2) and has fixed parameters, such as a fixed *a priori* SNR [4]. Both estimators are based on a Gaussian speech model. The proposed approach (denoted as "SG-SPP with fixed ξ [new]") consists of the MMSE estimator [14, Eq. (6)] and the *a posteriori* SPP estimator with the new GLR (8). This GLR is driven by averaged *a posteriori* SNRs and its fixed parameters are from Table 1. Both estimators are based on a super-Gaussian speech model (chi-distributed speech amplitudes with the shape parameter $\mu = 0.5$).

The results are depicted in Figures 1 and 2. As can be seen in Figure 1, all approaches achieve approximately the same speech component quality $SSDR_{seg}$ with a slight advantage for the proposed approach "SG-SPP with fixed ξ [new]". Since *a posteriori* SPP estimation approaches with fixed parameters are able to gain values close to zero, corresponding approaches ("G-SPP with fixed ξ [4]" and "SG-SPP with fixed ξ [new]") achieve significantly larger noise attenuation levels compared to usual SPP approaches with an adapted ξ ("G-SPP with adapted ξ [2]" and "SG-SPP with adapted ξ [6]"). As can be seen in Figure 2, the proposed approach nicely achieves the lowest musical noise levels, reflected by small LKR values, for car noise and babble noise at all input SNR levels. In case of factory noise, the proposed approach achieves the best LKR values above 5 dB input SNR.

5. CONCLUSIONS

This paper presents an *a posteriori* SPP estimation approach which unifies the advantages of a super-Gaussian speech model, averaged observations, and a fixed *a priori* SNR. Accordingly, a new generalized likelihood ratio for averaged *a posteriori* SNRs is derived taking a super-Gaussian speech model into account. To obtain SPP estimates close to zero in speech absence, fixed SNR and SPP priors are employed. The resulting new approach is shown to outperform reference approaches that consider averaged observations but a Gaussian speech model, a super-Gaussian model but no averaging, and also approaches that are based on a Gaussian model without averaging.

6. REFERENCES

[1] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 2, pp. 137–145, Apr. 1980.

[2] 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.

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

[4] T. Gerkmann, C. Breithaupt, and R. Martin, "Improved a posteriori speech presence probability estimation based on a likelihood ratio with fixed priors," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 910–919, Jul. 2008.

[5] R. Martin, "Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, USA, May 2002, vol. 1, pp. 253–256.

[6] C. Breithaupt and R. Martin, "Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 277–289, Feb. 2011.

[7] B. Fodor and T. Fingscheidt, "MMSE speech enhancement under speech presence uncertainty assuming (generalized) gamma speech priors throughout," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 4033–4036.

[8] D. H. T. Vu and R. Haeb-Umbach, "Using the turbo principle for exploiting temporal and spectral correlations in speech presence probability estimation," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, BC, Canada, Sept. 2013, pp. 863–867.

[9] B. Fodor and T. Gerkmann, "A speech presence uncertainty estimator based on fixed priors and a heavy-tailed speech model," in *Proc. of EURASIP European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, Sept. 2014, accepted for publication.

[10] T. Lotter 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, Jan. 2005.

[11] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete fourier coefficients with generalized Gamma priors," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[12] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State-of-the-art*, Morgan & Claypool, Feb. 2013.

[13] R. Martin and T. Lotter, "Optimal recursive smoothing of non-stationary periodograms," in *Proc. of International Workshop on Acoustic Signal Enhancement (IWAENC)*, Darmstadt, Germany, Sept. 2001, pp. 167–170.

[14] I. Andrianakis and P. R. White, "MMSE speech spectral amplitude estimators with chi and gamma speech priors," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, May 2006, vol. III, pp. 1068–1071.

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 4th edition, 1965.

[16] I.N. Bronshtein, K.A. Semendyayev, G. Musiol, and H. Mühlig, *Handbook of Mathematics*, Springer, 2007.

[17] NTT, "Multi-lingual speech database for telephonometry," NTT Advanced Technology Corporation, 1994.

[18] NTT, "Ambient noise database for telephonometry," NTT Advanced Technology Corporation, 1996.

[19] ITU, "Objective measurement of active speech level," International Telecommunication Union, ITU-T Recommendation P.56, Mar. 1993.

[20] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[21] T. Fingscheidt, S. Suhadi, and S. Stan, "Environment-optimized speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 825–834, May 2008.

[22] H. Yu and T. Fingscheidt, "Instrumental musical tones measurement of arbitrary noise reduction systems," in *Proc. of German Annual Congress on Acoustics (DAGA)*, Darmstadt, Germany, Mar. 2012, pp. 255–256.