SPEECH ENHANCEMENT BASED ON RAYLEIGH MIXTURE MODELING OF SPEECH SPECTRAL AMPLITUDE DISTRIBUTIONS

J.S. Erkelens, J. Jensen†, and R. Heusdens**

*Delft University of Technology, Department of Mediamatics, Mekelweg 4, 2628 CD Delft, The Netherlands
phone: +(31) 152782188, fax: +(31) 152781843, email: {j.s.erkelens, r.heusdens}@tudelft.nl, web: www-ict.ewi.tudelft.nl
†Oticon A/S, Kongebakken 9, 2765 Smørum, Denmark, email: jsj@oticon.dk

ABSTRACT

DFT-based speech enhancement algorithms typically rely on a statistical model of the spectral amplitudes of the noise-free speech signal. It has been shown in the literature recently that the speech spectral amplitude distributions, conditional on estimated a priori SNR, may differ significantly from the traditional Gaussian model and are better described by super-Gaussian probability density functions. We show that these conditional distributions can be accurately approximated by a mixture of Rayleigh distributions. The MMSE amplitude estimators based on Rayleigh Mixture Models perform at least as well as the estimators based on super-Gaussian models. Furthermore, the proposed Rayleigh Mixture Models allow for derivation of closed-form estimators minimizing other perceptually relevant distortion measures, which may be difficult for other models.

1. INTRODUCTION

The traditional assumption for speech enhancement in the DFT domain is that the distribution of the complex speech DFT coefficients is Gaussian [1]–[3]. Consequently, the spectral amplitude distribution is modeled by a Rayleigh distribution. Recently, super-Gaussian models of the DFT coefficients have received quite some attention, because they lead to estimators with better performance than those based on a Gaussian model. Martin [4] derived complex-DFT estimators for Laplacian and Gamma speech priors, and Lotter and Vary [5] proposed a Maximum A Posteriori (MAP) amplitude estimator for a generalized Gamma amplitude distribution. MMSE estimators for the amplitudes, assuming a one-sided generalized Gamma distribution, were treated in [6] and [7].

In this paper we propose to model the distributions of speech DFT amplitudes by Rayleigh Mixture Models (RMMs). RMMs have some important advantages over existing speech models. They offer more accurate fits to the amplitude distributions, and can also adapt better to the *a priori* SNR estimator used. Furthermore, analytical derivation of estimators for relevant distortion measures is relatively simple.

The paper is organized as follows. Section 2 recapitulates MMSE speech spectral estimation and introduces RMMs. Section 3 motivates the use of RMMs. Estimators under an RMM speech prior are derived in Section 4. The amplitude estimator is evaluated in Section 5 and compared with existing estimators. Section 6 concludes the paper.

The research is supported by MultimediaN, the Technology Foundation STW (applied science division of NWO), and the technology programme of the ministry of Economic Affairs.

2. MMSE SPECTRAL ESTIMATION

2.1 Signal model and assumptions

We consider an additive-noise signal model of the form

$$X(k,m) = S(k,m) + D(k,m),$$

where $X(k,m)$, $S(k,m)$, and $D(k,m)$ are complex-valued random variables representing the short-time DFT coefficients obtained at frequency index k in signal frame m from the noisy speech, clean speech and noise process, respectively. Applying the standard assumption that $S(k,m)$ and $D(k,m)$ are statistically independent across time and frequency as well as from each other, leads to expressions for the resulting estimators that are independent of time and frequency. For ease of notation we therefore drop the time and/or frequency index when this does not cause confusion. We use capitals for random variables and the corresponding lower-case letters for their realizations. The speech amplitude is $A = |S|$, and the noisy amplitude is $R = |X|$. The noise DFT coefficients D are assumed to follow a complex Gaussian distribution with variance λ_D.

2.2 p-th Order amplitude estimators

For given noise spectral variance λ_D and given speech spectral variance λ_S, the MMSE estimator of some power p of the speech amplitude is (see, e.g., [1]):

$$\widehat{A^p} = E\{A^p|R\} = \frac{\int_0^\infty a^p \exp(-\frac{a^2}{\lambda_D}) I_0(\frac{2aR}{\lambda_D}) f_A(a)da}{\int_0^\infty \exp(-\frac{a^2}{\lambda_D}) I_0(\frac{2aR}{\lambda_D}) f_A(a)da}, \quad (1)$$

where $f_A(a)$ is the probability density function of A, which depends on λ_S, and $I_0(\cdot)$ is the zeroth order Bessel function of the fist kind. $\widehat{A^p}$ is called the *p-th order* amplitude estimator. In practice, the spectral variances λ_S and λ_D are unknown and have to be estimated. This will affect the optimality of the estimators. We will take into account, to some extent, the influence of the speech spectral variance estimation, by matching our model of $f_A(a)$ to measured histograms that are conditional on a high value of estimated λ_S. In the following, we assume that the noise spectral variance can be estimated accurately during speech pauses for stationary noise or by using approaches based on minimum-statistics [8][9], for example, for non-stationary noise.

2.2.1 Generalized Gamma distribution

Recently, the clean-amplitude distribution $f_A(a)$ in (1) has been modeled using the generalized Gamma distribution [5, 6, 7]. This distribution is given by

$$f_A(a) = \frac{\gamma\beta^\nu}{\Gamma(\nu)} a^{\gamma\nu-1} \exp(-\beta a^\gamma),\ a \geq 0, \tag{2}$$

with the constraints on the parameters $\gamma > 0,\ \nu > 0$. We will consider the cases $\gamma = 1$ and $\gamma = 2$. Because $E\{A^2\}$ equals λ_S by definition, β is related to γ, ν and λ_S. For $\gamma = 1$ we have $\beta^2 = \nu(\nu+1)/\lambda_S$, and $\beta = \nu/\lambda_S$ for $\gamma = 2$. For $\gamma = 2$, $\nu = 1$, the Rayleigh distribution appears as a special case.

2.2.2 *Rayleigh Mixture Model*

As an alternative to the generalized Gamma model, we propose a Rayleigh Mixture Model. If the complex speech DFT coefficients are modeled by a Gaussian Mixture Model, then the amplitude distribution is a sum of Rayleigh distributions:

$$f_A(a) = \sum_{j=1}^{J} c_j \frac{2a}{\lambda_j} \exp\left\{-\frac{a^2}{\lambda_j}\right\}, \tag{3}$$

where J is the number of components and the c_j are positive weighting factors that satisfy $\sum c_j = 1$. The λ_j are the variances of the individual components; they satisfy $\sum c_j \lambda_j = \lambda_S$.

2.3 *A priori* SNR estimation

Speech amplitude estimators are usually written in terms of gain functions, e.g., $\hat{A} = G(\xi, \zeta)R$. These gain functions depend on *a priori* SNR ξ, defined as $\xi = \lambda_S/\lambda_D$, and *a posteriori* SNR ζ, defined as $\zeta = R^2/\lambda_D$. We will use the decision-directed approach [1] to estimate *a priori* SNR, with a bias correction [10]:

$$\hat{\xi}_k(m) = \max\left[\alpha \frac{\widehat{A_k^2}(m-1)}{\lambda_D(k, m-1)} + (1-\alpha)[\zeta_k(m) - 1], \xi_{min}\right]. \tag{4}$$

Note that in the first term, the second order amplitude estimator is used, instead of the square of the first order amplitude estimator, which was the original definition [1]. The second order amplitude estimator used in (4) will be based on the generalized Gamma distribution (2), with either $\gamma = 1$ or $\gamma = 2$. We have observed that this new *a priori* SNR estimator (4) leads to less musicality than the old definition, for parameter settings (ν, α) with the same speech quality versus noise reduction trade-off [11].

3. RAYLEIGH MIXTURE MODELING OF CONDITIONAL SPEECH AMPLITUDE DISTRIBUTIONS

It has been shown in several papers [4]–[7] that better noise suppression performance can be achieved by abandoning the Gaussian speech model. There may be several reasons for the suboptimality of the Gaussian model. Often, the normal distribution of DFT coefficients is motivated by the central limit theorem. For speech DFT coefficients, the central limit theorem may not be applicable, because of the long span of correlation which can be larger than the frame lengths [4][5]. Speech is also non-stationary, causing many time frames to contain non-identically distributed samples [6]. Furthermore, gain functions are derived for *known a priori* SNR. In practice, *a priori* SNR has to be estimated. This means that the optimal statistical model for enhancement may differ from the true underlying speech distribution, and should be adapted to the *a priori* SNR estimator used [10]–[13].

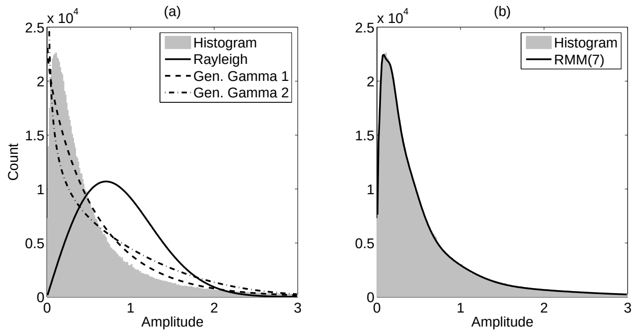

Figure 1: Normalized histogram of clean spectral amplitudes from frequency bins with an estimated *a priori* SNR in the range 19-21 dB, (a) maximum-likelihood fits of a Rayleigh distribution and generalized Gamma distributions with $\gamma = 1$ and $\gamma = 2$, and (b) maximum-likelihood fit of a Rayleigh Mixture Model with J=7 components.

3.1 Measured amplitude distributions

Following an idea of Martin [4], Lotter and Vary [5] have attempted to measure the distribution of amplitudes of speech DFT coefficients. For this purpose, a speech database is processed in a standard DFT-based enhancement algorithm, and coefficients are collected from those frequency bins for which the estimated *a priori* SNR is within a narrow range of high values. We performed a similar experiment. Figure 1(a) shows a histogram of one million of such amplitudes from TIMIT, normalized such that the second moment equals one, i.e., $\overline{A^2} = 1$, where the overbar indicates the sample mean. Also shown are maximum-likelihood fits of a Rayleigh distribution and generalized Gamma distributions. Clearly the measured amplitude distribution does not follow the Rayleigh model, while the generalized Gamma models fit better. Amplitude estimators based on generalized Gamma distributions improve speech enhancement performance over those based on a Gaussian speech model [5]–[7]. Figure 1(b) shows a maximum-likelihood fit (see section 4.2.1) of the proposed RMM (3) to the histogram, using $J = 7$ components. Clearly, the RMM model offers a much better fit to the histogram. The experiments of Section 5 show that the resulting estimators also perform very well in a speech enhancement context. However, online adaptation to speech characteristics would be easier for the generalized Gamma models, because of the smaller number of parameters.

3.2 Discussion

Ephraim and Cohen [14] have shown that the Gaussian speech model and other models are not necessarily contradictory. If the spectral variance λ_S is treated as a random variable with pdf $f(\lambda_S)$, then the joint distribution of real and imaginary parts of the corresponding DFT coefficient is given by

$$f(s_R, s_I) = \int_0^\infty f(s_R, s_I|\lambda_S) f(\lambda_S) d\lambda_S,$$

where s_R and s_I are the real and imaginary parts of a clean speech DFT coefficient, respectively. If $f(s_R, s_I|\lambda_S)$ is a Gaussian distribution, then $f(s_R, s_I)$ is a continuous mixture of Gaussian distributions, which can take many different forms depending on $f(\lambda_S)$. For example, if $f(\lambda_S)$ is assumed to be exponential, then the pdf of the real and imaginary parts each follow a Laplace pdf, as it was assumed in [4].

The distributions faced in practice are conditional on *estimated a priori* SNR and are given by

$$f(s_R, s_I|\hat{\lambda}_S) = \int_0^\infty f(s_R, s_I|\lambda_S, \hat{\lambda}_S) f(\lambda_S|\hat{\lambda}_S) d\lambda_S,$$

where $\hat{\lambda}_S$ is given by $\hat{\xi}\lambda_D$. *Given* λ_S, we expect s_R and s_I to be only weakly dependent on $\hat{\lambda}_S$ for the following reasons. Since $\hat{\lambda}_S$ is only an estimate of the true spectral variance λ_S, it may contain less information about s_R and s_I than λ_S itself. The second term in the decision directed estimator (4), $(1-\alpha)[r^2(m)/\lambda_D - 1]$, depends on the noisy amplitude and therefore contains some information about the clean s_R and s_I. However, the weighting factor $(1-\alpha)$ of this term is generally small (0.02 is a typical value). We therefore expect the following approximation to be reasonable:

$$f(s_R, s_I|\hat{\lambda}_S) \approx \int_0^\infty f(s_R, s_I|\lambda_S) f(\lambda_S|\hat{\lambda}_S) d\lambda_S. \tag{5}$$

Note that any dependency that may exist between the real and imaginary parts of the current time frame, $s_R(m)$ or $s_I(m)$, and the estimated (second order) amplitude estimate of the previous time frame, $\widehat{A^2}(m-1)$, is also neglected given λ_S, as is usually done in the derivation of estimators. If $f(s_R, s_I|\lambda_S)$ is Gaussian, then (5) expresses $f(s_R, s_I|\hat{\lambda}_S)$ as a continuous mixture of Gaussians. The corresponding amplitude distribution is a continuous mixture of Rayleigh distributions. We propose to model such amplitude distributions by RMMs (3). That model is used in (1) to obtain estimators that take into account statistics of the speech *and* the particular *a priori* SNR estimator used. Note that we do not really rely on (5), because the RMM model can accurately match the histograms with a sufficiently large number of components, regardless of whether (5) is accurate or not. As was illustrated in Figure 1(b), only a small number of components suffices in practice.

4. AMPLITUDE ESTIMATORS

4.1 Generalized Gamma distribution

A MAP amplitude estimator for the model (2) for $\gamma = 1$ was derived in [5], while MMSE amplitude estimators for the classes $\gamma = 1$ and $\gamma = 2$ have been studied in [6] and [7]. For $\gamma = 2$, the expressions are exact, while approximations have to be made for $\gamma = 1$. The maximum achievable performance for both classes is about the same. Because of lack of space, we show only the expressions for the estimators of the $\gamma = 2$ class, which contain the Gaussian speech model as a special case for $\nu = 1$. The MMSE amplitude estimator is given by

$$\hat{A}_\nu^{(2)} = \frac{\Gamma(\nu+0.5)}{\Gamma(\nu)} \sqrt{\frac{\xi}{\zeta(\nu+\xi)}} \frac{{}_1F_1\left(\nu+0.5; 1; \frac{\zeta\xi}{\nu+\xi}\right)}{{}_1F_1\left(\nu; 1; \frac{\zeta\xi}{\nu+\xi}\right)} R,$$

where ${}_1F_1(a;b;x)$ is a confluent hypergeometric function [15, 13.1.2]. The superscript $^{(2)}$ indicates that $\gamma = 2$. The corresponding second order amplitude estimator is given by

$$\widehat{A_\nu^2}^{(2)} = \frac{\nu\xi}{\zeta(\nu+\xi)} \frac{{}_1F_1\left(\nu+1; 1; \frac{\zeta\xi}{\nu+\xi}\right)}{{}_1F_1\left(\nu; 1; \frac{\zeta\xi}{\nu+\xi}\right)} R^2. \tag{6}$$

It can be shown that for $\nu \to \infty$, the first and second order amplitude estimates approach $\hat{\lambda}_S^{0.5}$ and $\hat{\lambda}_S$, respectively. The reason for this behavior is that for a given λ_S, (2) tends to a delta-function centered around $\lambda_S^{0.5}$ when ν goes to infinity. This is also true for the $\gamma = 1$ case. Consequently, the decision-directed *a priori* SNR estimator behaves like an ordinary exponential smoother for $\nu \to \infty$.

4.2 Rayleigh Mixture Model

The derivation of the MMSE estimator for the RMM speech amplitude priors goes much along the lines of the derivations in [1]. If we define ξ_j as λ_j/λ_D, and v_j, g_j and q_j as

$$\mathrm{v}_j = \frac{\xi_j}{1+\xi_j}\zeta, \quad g_j = \frac{c_j}{1+\xi_j} e^{\mathrm{v}_j}, \quad q_j = \frac{g_j}{\sum_{i=1}^{J} g_i},$$

respectively, then the amplitude estimator is

$$\hat{A}_{RMM} = \frac{\sqrt{\pi}}{2\zeta} \sum_{j=1}^{J} q_j \sqrt{\mathrm{v}_j}\, {}_1F_1(-0.5; 1; -\mathrm{v}_j) R. \tag{7}$$

Existing estimators under a Gaussian speech model that minimize other perceptually relevant distortion measures, such as those in [2][3], may also be generalized to the RMM case. For example, the estimator that minimizes the log distortion measure $E\{(\log A - \log \hat{A})^2\}$, called $\hat{A}_{RMM}^{log}$, is given by

$$\hat{A}_{RMM}^{log} = \exp\left\{ \sum_{j=1}^{J} q_j \left\{ \log \mathrm{v}_j - \log \zeta + \frac{1}{2} \int_{\mathrm{v}_j}^{\infty} \frac{e^{-t}}{t} dt \right\} \right\} R.$$

We are unaware of closed-form log-amplitude estimators under the generalized Gamma model.

4.2.1 Parameter estimation and estimator implementation

The parameters of the RMM in (3) are found by fitting to measured amplitude data from TIMIT, as in Figure 1(b). First, the amplitude data is normalized such that $\overline{A^2} = 1$. Next, a least-squares fit of (3) to the histogram is made under the constraints $\sum c_j = \sum c_j \lambda_j$ and all c_j and λ_j positive. The c_j thus found are normalized with $\sum c_j$, such that the pdf (3) integrates to 1. Finally, the parameters are used as initial conditions for the EM-algorithm. It can be shown that, under the constraint $\sum c_j = 1$, the resulting maximum-likelihood estimates of the parameters satisfy $\sum c_j \lambda_j = \overline{A^2}$. To apply the estimators of Section 4.2, the variance parameters λ_j have to be scaled since they are found from normalized data. In every frequency bin of every time frame, the parameters to be used in (7) are found by multiplying each of the fitted λ_j by $\hat{\lambda}_S(k,m) = \hat{\xi}_k(m)\lambda_D(k,m)$. For *a priori* SNR estimation, (4) is used with the second order amplitude estimator from a generalized Gamma model (which for $\gamma = 2$ is given by (6)).

To gain speed, we tabulated all gain functions in the experiments, for the range -19 dB $< \xi <$ 40 dB and -30 dB $< \zeta <$ 40 dB, both in steps of 1 dB.

5. EXPERIMENTAL RESULTS

5.1 Experimental set-up

In the enhancement system, we use 50%-overlapping frames of 32 ms. The data window used was a cosine-squared window. The smoothing parameter α is set to 0.98 and ξ_{min} to

−19 dB. We use all 30 clean sentences of the NOIZEUS database [16]. Noisy signals were generated by adding white and nearly stationary car noise from the Noisex-92 database [17] to the clean signals, at 5 and 15 dB overall SNR. The noise and speech are limited to telephone bandwidth (300-3400 Hz). The noise variance was estimated from 0.64 seconds of noise only preceding speech activity. Objective quality was measured in two different ways. We measure mean-square error (*MSE*), because it is what MMSE estimators should minimize on the average. We compute *MSE* as

$$MSE = \frac{1}{M}\sum_{m=1}^{M}\sum_{k}\{a(k,m)-\hat{a}(k,m)\}^2,$$

where $a(k,m)$ and $\hat{a}(k,m)$ are the clean speech spectral amplitude and the estimated amplitude of frequency bin k and time frame m, respectively, and M is the number of frames containing speech in a sentence. To exclude silence intervals, frames with a clean energy more than 40 dB below the maximum clean frame energy of a speech sentence are not taken into account. All results at a given SNR are averages over all test sentences. Furthermore, to quantify the speech distortion versus noise reduction trade-off, we also measure separately segmental Speech Quality (*SQ*) and Noise Reduction (*NR*) as in [5], and plot these quantities against each other while varying ν. The enhanced speech $\hat{s}(n)$ can be written as

$$\hat{s}(n) = \tilde{s}(n) + \tilde{d}(n),$$

where $\tilde{s}(n)$ and $\tilde{d}(n)$ result from applying the gain functions to the clean speech and noise DFT coefficients separately, and transforming back to the time domain. We define segmental Speech Quality as

$$SQ = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\left(\frac{\|\mathbf{s}_m\|^2}{\|\mathbf{s}_m-\tilde{\mathbf{s}}_m\|^2}\right),$$

where $\mathbf{s}_m$ and $\tilde{\mathbf{s}}_m$ denote time frame m of the signals $s(n)$ and $\tilde{s}(n)$, respectively. The operator $\|\cdot\|^2$ computes the energy of a time frame. Segmental Noise Reduction is defined as

$$NR = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\left(\frac{\|\mathbf{d}_m\|^2}{\|\tilde{\mathbf{d}}_m\|^2}\right).$$

Strong suppression leads to low *SQ* and high *NR*, while the opposite happens for weak suppression.

5.2 Performance evaluation

We will compare amplitude estimators for the generalized Gamma model with those of RMM models, while varying the parameter ν. *A priori* SNR estimation with (4) was always based on the generalized Gamma model. The parameters of the RMM models are found from the corresponding histograms, as was outlined in Sections 3.1 and 4.2.1. Figure 2 shows the results for the $\gamma = 1$ case. The dash-dotted lines result for white noise when the generalized-Gamma amplitude estimators are used for reconstruction, while the solid curves are for RMM amplitude estimators with $J = 7$ components. The dotted and dashed lines are the corresponding results for car noise. The value of ν is limited to values larger than 0.5 for $\gamma = 1$, because of an approximation that is used in the derivation of the estimators [7]. The crosses and pluses

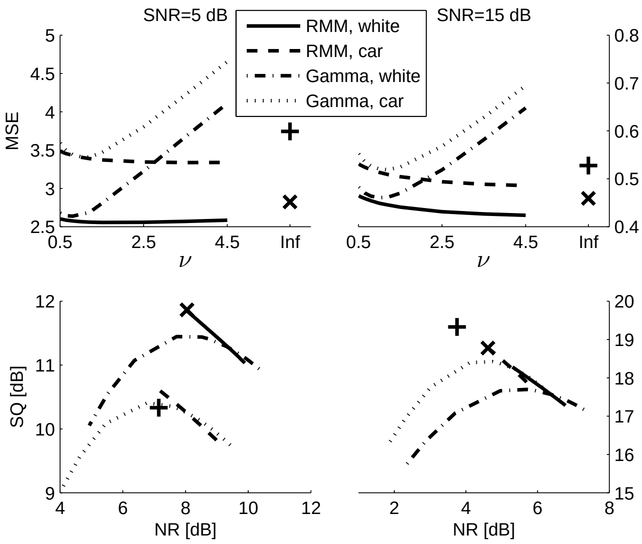

Figure 2: MSE versus ν and SQ versus NR for $\gamma = 1$. White and car noise have been used at overall SNRs of 5 and 15 dB.

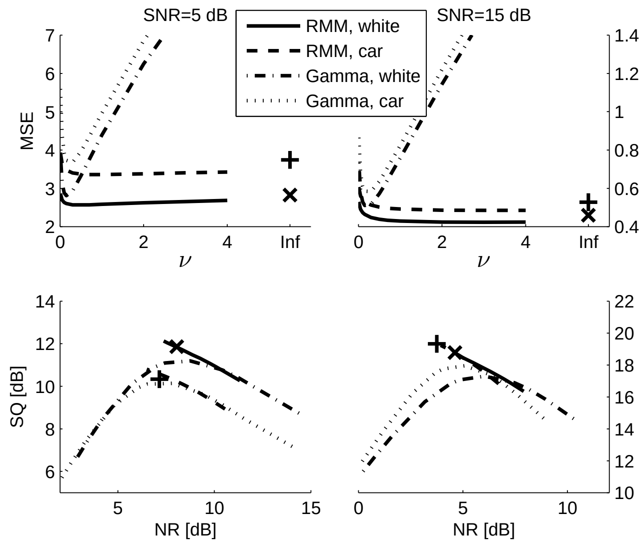

Figure 3: MSE versus ν and SQ versus NR for $\gamma = 2$. White and car noise have been used at overall SNRs of 5 and 15 dB.

result for the RMM amplitude estimator, for white and car noise respectively, when an exponential smoother is used for *a priori* SNR estimation (corresponding to $\nu \to \infty$). The upper two panels show that minimum achievable *MSE* is lower for the RMM amplitude estimators. Furthermore, the RMM estimators are much less sensitive to the value of ν. The main reason for this behavior is that the RMM models have been adapted to some extent to the *a priori* SNR estimator used, because the parameters are found from measured data that depend on it (see sections 3.1 and 4.2.1). It is clear that using an exponential smoother for *a priori* SNR estimation is not optimal. The lower two panels show *SQ* versus *NR* when ν is varied over the same range as for the upper two panels. The value of ν increases when going from the right to the left along the curves. More noise reduction is possible for the generalized Gamma amplitude estimators, but the maximum achievable speech quality is higher for the RMM amplitude estimators. Similar trends are seen for car noise.

Figure 3 shows the results for the $\gamma = 2$ case. The maximum achievable performance is about the same as for the $\gamma = 1$ case, although the results are much more sensitive to the value of the ν-parameter. The RMM amplitude estimators perform about the same on the *a priori* SNR estimators of both cases.

5.2.1 Informal listening

For increasing values of ν the enhanced speech sounds more reverberant but the musicality decreases, especially for the amplitude estimators of the $\gamma = 2$ class. For the RMM amplitude estimators the ν-dependency is the weakest, although these effects are clearly noticeable for $\nu \to \infty$. For the lowest values of ν, all estimators sound very similar.

5.3 Discussion

Amplitude and *a priori* SNR estimation for the generalized Gamma models is based on one and the same prior speech distribution (i.e., (2) with the same values of γ and ν). This does not necessarily lead to optimal results. To a given *a priori* SNR estimator corresponds a certain measured histogram of spectral amplitudes. This histogram depends on the unknown dynamical and statistical properties of this particular *a priori* SNR estimator. There is no reason why the parametric amplitude distribution used in calculating the *a priori* SNR estimates should fit accurately to its corresponding measured amplitude histogram. In fact, we have seen that an RMM model can fit much better to histograms found with the generalized gamma *a priori* SNR estimator. We have not investigated whether using different γ and/or ν values for the amplitude and *a priori* SNR estimation tasks leads to significant improvements for the generalized Gamma models.

Many estimators found in literature that are based on parametric models of the speech prior distribution, including the ones presented here, are implicitly assuming that the conditional distribution $f_A(a|\hat{\lambda}_S)$ has the same shape (except for a variance scaling) for all values of $\hat{\lambda}_S$. This may not be an accurate assumption for all SNRs, because the properties of the *a priori* SNR estimator depend on the SNR. A data-driven approach has been proposed in [12] to deal with this problem.

6. SUMMARY AND CONCLUSIONS

In this paper we have proposed Rayleigh Mixture Models to describe measured speech amplitude distributions in the context of speech enhancement. We have shown that the resulting amplitude estimators can compete with state-of-the-art estimators. Furthermore, analytical derivation of estimators for meaningful distortion measures is relatively simple.

REFERENCES

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[2] ———, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[3] P. C. Loizou, "Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 857–869, Sept. 2005.

[4] R. Martin, "Speech enhancement based on Minimum Mean-Square Error Estimation and supergaussian priors", *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 845–856, Sept. 2005.

[5] T. Lotter, and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model," *EURASIP Journ. Appl. Signal Proc.*, vol. 7, pp. 1110–1126, 2005.

[6] I. Andrianakis, and P. R. White, "MMSE speech spectral amplitude estimators with Chi and Gamma speech priors", *Proc. Int. Conf. Acoust., Speech and Signal Proc.*, vol. III, pp. 1068–1071, May 2006.

[7] 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 Trans. Audio, Speech and Language Proc.*, vol. 15, July 2007.

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

[9] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging", *IEEE Trans. Speech Audio Proc.*, vol. 11, no. 5, pp. 466–475, Sept. 2003.

[10] J. S. Erkelens, J. Jensen, and R. Heusdens, "A data-driven approach to optimizing spectral speech enhancement methods for various error criteria", *Speech Communication, Special Issue on Speech Enhancement*, 2007.

[11] J. S. Erkelens, J. Jensen, and R. Heusdens, "Improved speech spectral variance estimation under the generalized Gamma distribution", *Proc. IEEE BENELUX/DSP Valley Sign. Proc. Symp.*, pp. 43-46, March 2007.

[12] J. Jensen and R. Heusdens, "A numerical approach for estimating optimal gain functions in single-channel DFT based speech enhancement", *Proc. EUSIPCO*, Sept. 2006.

[13] I. Cohen, "Supergaussian GARCH models for speech signals", *Proc. Interspeech*, pp. 2053-2056, Sept. 2005.

[14] Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement", in *The Electrical Engineering Handbook*, CRC Press, 2006. ece.gmu.edu/ ∼ yephraim/Papers/crc 2004 enhan ee.pdf

[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1964.

[16] Y. Hu and P. Loizou, "Subjective comparison of speech enhancement algorithms", *Proc. Int. Conf. Acoust., Speech and Signal Proc.*, vol. I, pp. 153–156, May 2006. www.utdallas.edu/ ∼ loizou/speech/noizeus/

[17] A. Varga and H. J. M. Steeneken, "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. 247–253, 1993.