MINIMUM MEAN-SQUARE ERROR AMPLITUDE ESTIMATORS FOR SPEECH ENHANCEMENT UNDER THE GENERALIZED GAMMA DISTRIBUTION

R.C. Hendriks, J.S. Erkelens, J. Jensen, and R. Heusdens

{R.C.Hendriks, J.S.Erkelens, J.Jensen, R.Heusdens}@tudelft.nl
Delft Univeristy of Technology, Department of Mediamatics
Information and Communication Theory Group
Mekelweg 4, 2628 CD Delft, The Netherlands

ABSTRACT

In this paper we derive minimum mean-square error (MMSE) amplitude estimators for DFT-based noise suppression. The optimal estimators are found under a generalized Gamma distribution, which takes as special cases (different parameter settings) all priors used in noise suppression schemes so far. Deriving the MMSE estimators involves integration of (weighted) Bessel functions. In order to end up with analytical solutions, for some parameter settings we have to approximate the Bessel functions. In this paper we combine two types of approximations by using a simple binary decision between the two. We show by computer simulations that the estimators thus obtained are very close to the exact MMSE estimators for all SNR conditions. The presented estimators lead to improved performance compared to the suppression rule proposed by Ephraim and Malah. Furthermore, the maximum performance is the same as compared to state of the art amplitude estimators.

1. INTRODUCTION

The increasing number of speech processing applications has resulted in more interest for ambient noise reduction methods. Among those methods is the class of single-channel speech enhancement methods. Many of these methods are based on the discrete Fourier transform (DFT) where speech DFT coefficients are estimated on a frame-by-frame basis by processing the noisy DFT coefficients. Existing methods estimate either the complex-valued speech DFT coefficients or the DFT amplitudes. It has often been assumed that DFT coefficients have a Gaussian distribution, e.g. [1]. Both complex DFT and amplitude estimators can be derived by minimizing the mean-square error (MMSE) or finding the maximum *a posteriori* (MAP) estimate, under this assumption. More recently there has been increased interest for estimators under supergaussian distributions, because they give a better approximation of the observed distribution of speech DFT coefficients. In [2] complex DFT MMSE estimators under the Gamma and Laplace distributions were derived.
In this paper we investigate MMSE speech amplitude estimators under the generalized Gamma distribution of the following form

$$f_A(a) = \frac{\gamma\beta^{\nu}}{\Gamma(\nu)} a^{\gamma\nu-1}\exp(-\beta a^{\gamma}), \qquad (1)$$

The research is supported by Philips Research, MultimediaN, the Technology Foundation STW, applied science division of NWO, and the technology programme of the ministry of Economics Affairs.

where $\beta > 0,\ \gamma > 0,\ \nu > 0,\ a \geq 0$, and the random variable A represents the DFT magnitude. In [3] a MAP amplitude estimator was proposed for the case $\gamma = 1$. In this paper, we derive MMSE amplitude estimators for the distribution classes $\gamma = 1$ and $\gamma = 2$. For $\gamma = 2$ the derivation is exact. A special case of this class appears for $\nu = 1$ when the density in Eq. (1) equals the Rayleigh prior used in [1]. For the case $\gamma = 1$ we cannot derive exact MMSE estimators analytically, and therefore two approximations are applied, one of which is most accurate at low SNRs and the other at high SNRs. Analytical expressions can be found under these approximations. Furthermore, it is shown that a simple binary strategy can be used to choose between the resulting amplitude estimators.

2. MMSE ESTIMATION OF AMPLITUDES

We assume that the speech and the noise process are uncorrelated and that the noise process is additive, i.e.,

$$X(k,m) = S(k,m) + W(k,m), \qquad (2)$$

where $X(k,m),\ S(k,m),\ W(k,m) \in \mathbb{C}$ are complex-valued random variables representing the DFT coefficients obtained in signal frame m at frequency index k from the noisy speech, clean speech and noise process, respectively. Since DFT coefficients from different time frames and frequency indices are assumed to be independent, the indices m and k will be omitted for simplicity. We can write $S = Ae^{j\Phi}$ and $X = Re^{j\Theta}$, where random variables A and R represent the clean and noisy amplitude, and Φ and Θ the corresponding phase values. In this work we focus on MMSE estimation of the clean amplitudes A. MMSE estimators for complex DFT coefficients under similar distribution assumptions are derived in [4]. The MMSE estimate of A is the expectation of the clean amplitude conditional on the noisy amplitude r, i.e., $E\{A|r\}$. With Bayes' formula we can express the MMSE estimate $\hat{A}$ as

$$\hat{A} = E\{A|r\} = \frac{\int_0^{\infty} a f_{R|A}(r|a) f_A(a) da}{\int_0^{\infty} f_{R|A}(r|a) f_A(a) da}. \qquad (3)$$

Assuming that the noise DFT coefficients have a zero-mean Gaussian density, $f_{R|A}(r|a)$ can be written as [5]

$$f_{R|A}(r|a) = \frac{2r}{\sigma_W^2}\exp\left(-\frac{r^2+a^2}{\sigma_W^2}\right) I_0\left(\frac{2ar}{\sigma_W^2}\right), \qquad (4)$$

where I_0 is the zero'th order modified Bessel function of the first kind, and $\sigma_W^2 = E\{|W|^2\}$ is the noise spectral variance. In the following we derive MMSE amplitude estimators for the cases $\gamma = 1$ and $\gamma = 2$.

2.1. MMSE estimators for $\gamma = 2$

Inserting Eq. (1) with $\gamma = 2$ and Eq. (4) into Eq. (3) gives

$$\hat{A}^{(2)} = \frac{\int_0^\infty a^{2\nu} \exp\left(-\frac{a^2}{\sigma_W^2} - \beta a^2\right) I_0\left(\frac{2ar}{\sigma_W^2}\right) da}{\int_0^\infty a^{2\nu-1} \exp\left(-\frac{a^2}{\sigma_W^2} - \beta a^2\right) I_0\left(\frac{2ar}{\sigma_W^2}\right) da}, \quad (5)$$

where the superscript (2) indicates that $\gamma = 2$. Using [6, Thm. 6.643.2], we can solve the integrals for $\nu > 0$. After inserting the relation between β and the second moment $E\{A^2\}$, which for this case is $\beta = \nu/\sigma_S^2$, with $\sigma_S^2 = E\{|S|^2\}$, we obtain [4]

$$\hat{A}^{(2)} = \frac{\Gamma(\nu + 1/2)}{\Gamma(\nu)} \frac{\sqrt{Q}}{\zeta} \frac{M_{(-\nu,0)}(Q)}{M_{(-\nu+1/2,0)}(Q)} r. \quad (6)$$

where $Q \triangleq \zeta\xi/(\nu+\xi)$, $M_{\nu,\mu}$ is known as the Whittaker function and $\zeta = |r|^2/\sigma_W^2$ and $\xi = \sigma_S^2/\sigma_W^2$ are known as the *a posteriori* and *a priori* SNR, respectively. The special case $\nu = 1$ is the estimator derived in [1].

2.2. MMSE estimators for $\gamma = 1$

For $\gamma = 1$, Eq. (3) becomes:

$$\hat{A}^{(1)} = \frac{\int_0^\infty a^{\nu} \exp\left(-\frac{a^2}{\sigma_W^2} - \beta a\right) I_0\left(\frac{2ar}{\sigma_W^2}\right) da}{\int_0^\infty a^{\nu-1} \exp\left(-\frac{a^2}{\sigma_W^2} - \beta a\right) I_0\left(\frac{2ar}{\sigma_W^2}\right) da}, \quad (7)$$

where β is now related to the speech spectral variance as $\beta^2 = \nu(\nu+1)/\sigma_S^2$. Analytical solutions to these integrals are unknown to the authors, but introducing approximations of the Bessel functions allows us to solve the integrals analytically.

2.2.1. *Bessel function approximation for small arguments*

For small arguments of the Bessel function I_0, we approximate it using a Taylor series expansion around $x = 0$. The Taylor series of I_0, truncated after K terms, is given by [7]

$$I_0(x; K) \triangleq \sum_{k=0}^{K-1} \left(\frac{x}{2}\right)^{2k} \frac{1}{(k!)^2}. \quad (8)$$

Substituting this expression for $I_0(\,\cdot\,; K)$ in Eq. (7) gives

$$\hat{A}_K^{(1)} \triangleq \frac{\int_0^\infty a^{\nu} \exp\left(-\frac{a^2}{\sigma_W^2} - \beta a\right) \sum_{k=0}^{K-1} \left(\frac{ar}{\sigma_W^2}\right)^{2k} \left(\frac{1}{k!}\right)^2 da}{\int_0^\infty a^{\nu-1} \exp\left(-\frac{a^2}{\sigma_W^2} - \beta a\right) \sum_{k=0}^{K-1} \left(\frac{ar}{\sigma_W^2}\right)^{2k} \left(\frac{1}{k!}\right)^2 da},$$

which, for $\nu > 0$, using [6, Thm. 3.462.1] leads to [4]

$$\hat{A}_K^{(1)} = \frac{\sum_{k=0}^{K-1} \left(\sqrt{\frac{\zeta}{2}}\right)^{2k} \left(\frac{1}{k!}\right)^2 \Gamma(\nu + 1 + 2k) D_{-(\nu+1+2k)}(T)}{\sqrt{2\zeta} \sum_{k=0}^{K-1} \left(\sqrt{\frac{\zeta}{2}}\right)^{2k} \left(\frac{1}{k!}\right)^2 \Gamma(\nu + 2k) D_{-(\nu+2k)}(T)} r,$$

with $T \triangleq \sqrt{(\nu+1)\nu/2\xi}$. For $K \to \infty$, $\hat{A}_K^{(1)} \to \hat{A}^{(1)}$ [4].

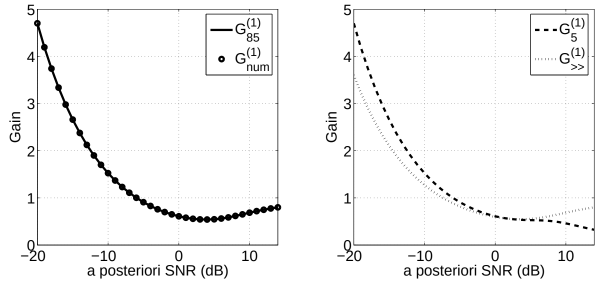

Figure 1: *Approximated gain functions for $\xi = 0$ dB and $\nu = 0$. A) $G_{85}^{(1)}$ with $K = 85$ terms, and numerically evaluated gain function, $G_{num}^{(1)}$ B) $G_5^{(1)}$ and high-SNR approximation, $G_{\gg}^{(1)}$.*

2.2.2. *Bessel function approximation for large arguments*

For large input arguments, however, a large value of K is needed to accurately approximate I_0, which may result in numerical problems and demanding computations. Therefore, for this case the following well-known approximation of I_0 is applied [7]:

$$I_0(x) \sim \frac{1}{\sqrt{2\pi x}} \exp(x). \quad (9)$$

Substituting this approximation in Eq. (7) and using [6, Thm. 3.462.1] we find for $\nu > 0.5$ [4]

$$\hat{A}_{\gg}^{(1)} \triangleq (\nu - 1/2)\sqrt{\frac{1}{2\zeta}} \frac{D_{-(\nu+1/2)}(P)}{D_{-(\nu-1/2)}(P)} r, \quad (10)$$

where $P \triangleq \sqrt{\nu(\nu+1)}/\sqrt{2\xi} - \sqrt{2\zeta}$, and D_ν is a parabolic cylinder function of order ν.

2.2.3. *Combining $A_K^{(1)}$ and $A_{\gg}^{(1)}$*

We now describe a simple and efficient strategy for combining $\hat{A}_K^{(1)}$ and $\hat{A}_{\gg}^{(1)}$ which leads to a combined estimator that is a good approximation of the true MMSE estimator under all conditions of interest. Which of the approximations considered is closest to the true MMSE estimator depends on the *a priori* and *a posteriori* SNRs, and the value of ν. Consider the following change of variable: $x = 2ar/\sigma_W^2$. This transformation makes it easier to see under which conditions approximations are expected to be accurate. The expression for $\hat{A}^{(1)}$ now becomes

$$\hat{A}^{(1)} = \frac{\sigma_W^2}{2r} \frac{\int_0^\infty x^{\nu} \exp\left[-\frac{x^2}{4\zeta} - \frac{\mu x}{2\sqrt{\zeta\xi}}\right] I_0(x)\, dx}{\int_0^\infty x^{\nu-1} \exp\left[-\frac{x^2}{4\zeta} - \frac{\mu x}{2\sqrt{\zeta\xi}}\right] I_0(x)\, dx}, \quad (11)$$

with $\mu = \sqrt{\nu(\nu+1)}$. The function $x^\nu \exp[-\frac{x^2}{4\zeta} - \frac{\mu x}{2\sqrt{\zeta\xi}}]$ attains its maximum for small x when the exponentials decay fast and x^ν rises slowly. In this case it is especially important to approximate the Bessel function well at small arguments. This happens when ζ or $\sqrt{\zeta\xi}$ is small and ν is small. For these conditions we may expect $\hat{A}_K^{(1)}$ to be more accurate than $\hat{A}_{\gg}^{(1)}$, while $\hat{A}_{\gg}^{(1)}$ is more accurate for large SNRs. Note that ζ is the more dominant parameter of ζ and ξ, because ξ is not present in the quadratic term in the exponentials.

A gain function $G(\zeta, \xi)$ for a certain amplitude estimator is defined as the estimate divided by the noisy amplitude, for example $G_{MMSE}^{(1)}(\zeta, \xi) = \hat{A}^{(1)}/r$ is the MMSE gain function for

$\gamma = 1$. As an example, Fig. 1 shows several gain functions for an *a priori* SNR ξ of 0 dB and $\nu = 1.6$ as a function of *a posteriori* SNR ζ. Fig. 1A shows the gain function $G_{85}^{(1)}$, which uses $K = 85$ terms, and $G_{num}^{(1)}$ which evaluates $G_{MMSE}^{(1)}$ using numerical integration of (7). The algorithms in [8] have been used to evaluate the parabolic cylinder functions. We see that $G_{85}^{(1)}$ and $G_{num}^{(1)}$ lie virtually on top of each other, showing that $G_{85}^{(1)}$ is a very good approximation of the true MMSE gain function for these combinations of *a priori* and *a posteriori* SNRs. These gain functions could not be evaluated accurately for larger *a posteriori* SNRs, mainly because of overflow problems. Fig. 1B shows $G_5^{(1)}$ and $G_{\gg}^{(1)}$ from Eq. (10). Comparison with Fig. 1A shows that $G_5^{(1)}$ is too small for high ζ, while $G_{\gg}^{(1)}$ is too small for low ζ. It can easily be proven that $G_K^{(1)}$ is always less than $G_{MMSE}^{(1)}$ for all K [4], but $G_{\gg}^{(1)}$ can sometimes be slightly larger than $G_{MMSE}^{(1)}$. Numerical calculations, however, have shown that $G_{\gg}^{(1)}$ does not exceed $G_{MMSE}^{(1)}$ by more than 0.10 dB, for the parameter range of interest. These results, and the difference in behavior of the two approximations as illustrated in Fig. 1B, suggests a simple binary strategy: take the *maximum* of the two approximations $G_K^{(1)}$ and $G_{\gg}^{(1)}$ as the gain function.

2.2.4. Error analysis

For the range $1 \leqslant \nu \leqslant 3.2$, $-20\ dB \leqslant \xi \leqslant +20\ dB$, $-20\ dB \leqslant \zeta \leqslant +14\ dB$ the true MMSE gain function $G_{MMSE}^{(1)}$ could be evaluated without numerical problems. This range is sufficient, because for larger ζ, the accuracy of the high-SNR approximation $G_{\gg}^{(1)}$ only increases, so the maximum error will not increase. For the binary decision $\max[\,G_5^{(1)}, G_{\gg}^{(1)}]$, the maximum positive error was $+0.35$ dB, and the maximum negative error was -0.10 dB. A positive error means that $G_{MMSE}^{(1)}$ was larger than the approximation. When $G_{10}^{(1)}$ is used in a binary decision with $G_{\gg}^{(1)}$, i.e., $\max[\,G_{10}^{(1)}, G_{\gg}^{(1)}]$, the maximum positive and negative errors are $+0.12$ dB and -0.10 dB, respectively. However, the maximum positive error increases with decreasing ν. For $\nu = 0.51$, about 20 terms are needed in $\hat{A}_K^{(1)}$. Although $G_{\gg}^{(1)}$ can be larger than $G_{MMSE}^{(1)}$, it will not exceed it by more than 0.10 dB.

2.3. Input-output characteristics

In Fig. 2 we show input-output (IO) characteristics of the derived estimators. In Fig. 2A we consider the case $\gamma = 1$ where $\hat{A}_5^{(1)}$, and $\hat{A}_{\gg}^{(1)}$ are combined into one estimate $\hat{A}_C^{(1)}$ by means of the binary decision, that is, $\hat{A}_C^{(1)} = \max[\,\hat{A}_5^{(1)}, \hat{A}_{\gg}^{(1)}]$. The values $\nu \in \{0.8, 1, 1.5\}$, the constraint $\sigma_S^2 + \sigma_W^2 = 2$, and $\xi = -5$ dB and $\xi = 5$ dB are used. The IO characteristics are fairly insensitive to ν. In Fig. 2B we consider $\nu \in \{0.5, 1, 1.5\}$ for the case $\gamma = 2$. The IO characteristics are more sensitive to ν values here and a smaller ν value clearly leads to less suppression at higher input values and to more suppression for lower input values.

3. SIMULATION RESULTS

In this section we present experimental results for $\hat{A}^{(2)}$ and two approximations of $\hat{A}^{(1)}$, namely $\hat{A}_C^{(1)} = \max[\,\hat{A}_5^{(1)}, \hat{A}_{\gg}^{(1)}]$ and $\hat{A}_{\gg}^{(1)}$. Further, we make comparisons with a modification of the MAP amplitude estimator as presented in [3], which is a MAP estimator under a generalized Gamma distribution with $\gamma = 1$. The MAP estimator proposed originally in [3] is

$$\hat{A}_{MAP}^{(1)} = \max_a \log f_A(a) f_{R|a}(r|a) \tag{12}$$

using the approximation for the Bessel function Eq. (9). This approximation is made *before* taking the derivative with respect to a to find the maximum. This leads to the gain function

$$G_{MAP}^{(1)} = u + \sqrt{u^2 + \frac{\nu' - 0.5}{2\zeta}} \quad , \quad u = 1/2 - \frac{\mu}{4\sqrt{\zeta\xi}}, \tag{13}$$

where $\nu' = \nu - 1$ and which is valid for $\nu' > 0.5$ only. A joint amplitude and phase MAP estimator was proposed as well, which can be derived without approximations. The gain function G_{JMAP} of the joint MAP estimator is given by

$$G_{JMAP}^{(1)} = u + \sqrt{u^2 + \frac{\nu'}{2\zeta}} \quad , \quad u = 1/2 - \frac{\mu}{4\sqrt{\zeta\xi}}. \tag{14}$$

This estimator allows for a broader range of ν'-values, namely $\nu' > 0$. Our first modification concerns the order in which an approximation is made and the derivative taken in Eq. (12). More specifically, we compute the amplitude MAP estimator by *first* taking the derivative in Eq. (12) and *then* using the large-argument approximation $I_1/I_0 \approx 1$, where I_1 is the first-order modified Bessel function of the first kind. Interestingly, the resulting MAP estimator is identical to the joint MAP estimator in Eq. (14). Our second modification concerns the parameter μ. In [3] the estimators were derived as a function of two free parameters, while there is in fact only one free parameter. The amplitude MAP estimator we use in our experiments is modified accordingly and is equal to Eq. (14) with μ set to $\sqrt{\nu(\nu+1)}$ [4].
For the experiments, the Noizeus database [9] was used which consists of 30 IRS-filtered speech signals sampled at 8 kHz, contaminated by various additive noise sources. We added computer-generated telephone-bandwidth white Gaussian noise as an extra noise source, since it is not present in the data base. The frame size is 256 samples, with an overlap of 50 %. The decision-directed approach with a smoothing factor $\alpha = 0.98$ was used to estimate ξ [1]. The noise variance was estimated with the minimum statistics approach [10]. Further, in all experiments the maximum suppression was limited to 0.1.
To express the performance of the estimators in terms of speech distortion and noise reduction separately, we follow the approach of [3] and define segmental speech SNR as

$$\text{SNR-S} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} 10 \log_{10} \left(\frac{\|\mathbf{s}_p\|_2^2}{\|\mathbf{s}_p - \tilde{\mathbf{s}}_p\|_2^2} \right), \tag{15}$$

where the vector $\mathbf{s}_p$ represents the p'th clean speech (time-domain) frame and $\tilde{\mathbf{s}}_p$ is the result of applying the gain functions to the clean speech in the frequency domain and transforming back to the time domain. To discard non-speech frames, an index set $\mathcal{P}$ is used of clean signal frames with energy no less than 30 dB of the maximum frame energy in a particular speech signal. $|\mathcal{P}|$ denotes the cardinality of $\mathcal{P}$. Similarly, noise reduction is measured as

$$\text{SNR-N} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} 10 \log_{10} \left(\frac{\|\mathbf{w}_p\|_2^2}{\|\tilde{\mathbf{w}}_p\|_2^2} \right), \tag{16}$$

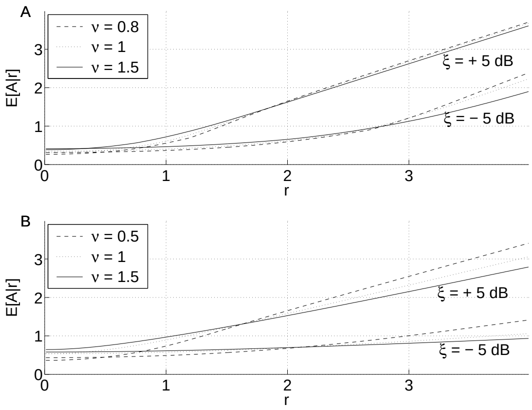

Figure 2: *Input-output characteristics with* $\sigma_S^2 + \sigma_W^2 = 2$ *for A)* $\gamma = 1$ *B)* $\gamma = 2$*.*

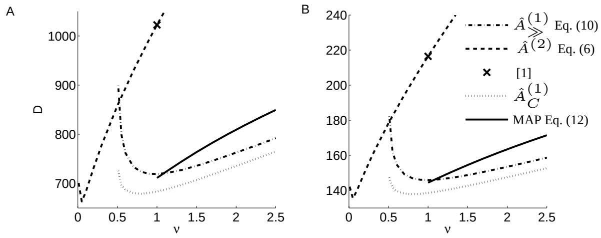

Figure 3: *Measured squared error for white noise with A) SNR = 0 dB. B) SNR = 10 dB.*

where $\mathbf{w}_p$ is the p'th noise frame, and $\tilde{\mathbf{w}}_p$ is the residual noise frame resulting from applying the noise suppression filter to $\mathbf{w}_p$. Fig. 4 shows performance in terms of SNR-S versus SNR-N for several values of ν and signals degraded by white noise. For a fixed SNR-N, $\hat{A}^{(1)}_C$ often leads to the best speech quality. This is also audible, in the sense that weak speech components are preserved slightly better than with Eq. (14). The distortion measure $D = \sum_{m,k}(A(k,m) - \hat{A}(k,m))^2$ is also considered, which is an estimate of the quantity minimized by the estimators derived in this paper. Fig. 3 plots D versus ν. We see that $\hat{A}^{(1)}_C$ improves over $\hat{A}^{(1)}_{\gg}$ and the MAP estimator, and that $\hat{A}^{(2)}$ scores very well for $\nu \approx 0.1$. Fig. 5 shows performance in terms of PESQ [11] versus ν for input SNRs of 5 and 15 dB and speech signals degraded by street noise and white noise. The maximum attainable PESQ scores are about the same for all estimators.

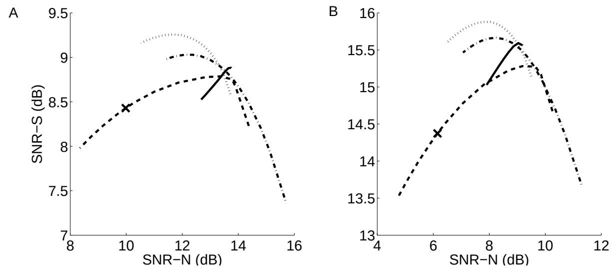

Figure 4: $SNR\!-\!S$ *versus* $SNR\!-\!N$ *for white noise, evaluated for different* ν *values. The legend of Fig. 3 applies. A) SNR = 0 dB. B) SNR = 10 dB.*

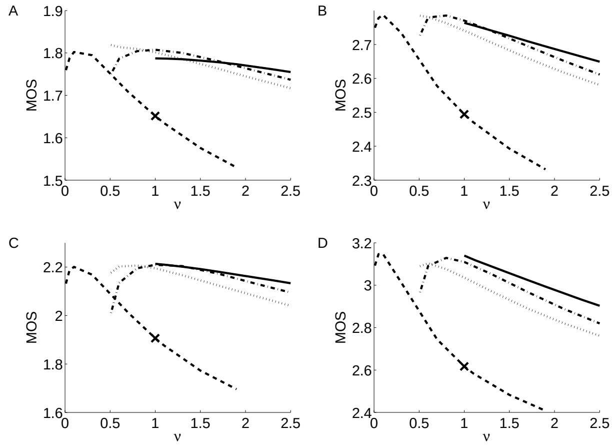

Figure 5: *PESQ performance versus* ν *for: A) Street noise at SNR = 5 dB. B) Street noise at SNR = 15 dB. C) White noise at SNR = 5 dB. D) White noise at SNR = 15 dB. The legend of Fig. 3 applies.*

4. 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 Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, December 1984.

[2] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 845–856, Sept. 2005.

[3] T. Lotter and 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, May 2005.

[4] J. Jensen, R. C. Hendriks, and J. S. Erkelens, "MMSE Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors," Tech. Rep., Delft Univ. Techn., 2006.

[5] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 2, pp. 137–145, April 1980.

[6] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, New York: Academic, 6th edition, 2000.

[7] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New-York, 9th Dover printing.

[8] Shanjie Zhang and Jianming Jin, *Computation of Special Functions*, John Wiley, New-York, 1996.

[9] "Noizeus: A noisy speech corpus for evaluation of speech enhancement algorithms," `http://www.utdallas.edu/~loizou/speech/noizeus/`.

[10] 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.

[11] J. G. Beerends, "Extending p.862 PESQ for assessing speech intelligibility," *White contribution COM 12-C2 to ITU-T Study Group 12*, October 2004.