isvr Institute of Sound and Vibration Research

Bayesian Algorithms for Speech Enhancement

I. Andrianakis and P. R. White

ISVR Technical Report No 305

January 2006

University of Southampton

SCIENTIFIC PUBLICATIONS BY THE ISVR

Technical Reports are published to promote timely dissemination of research results by ISVR personnel. This medium permits more detailed presentation than is usually acceptable for scientific journals. Responsibility for both the content and any opinions expressed rests entirely with the author(s).

Technical Memoranda are produced to enable the early or preliminary release of information by ISVR personnel where such release is deemed to the appropriate. Information contained in these memoranda may be incomplete, or form part of a continuing programme; this should be borne in mind when using or quoting from these documents.

Contract Reports are produced to record the results of scientific work carried out for sponsors, under contract. The ISVR treats these reports as confidential to sponsors and does not make them available for general circulation. Individual sponsors may, however, authorize subsequent release of the material.

UNIVERSITY OF SOUTHAMPTON

INSTITUTE OF SOUND AND VIBRATION RESEARCH

SIGNAL PROCESSING AND CONTROL GROUP

Bayesian Algorithms for Speech Enhancement

by

I Andrianakis and P R White

ISVR Technical Report No.305

January 2006

Authorised for issue by

Prof. R Allen

Group Chairman

CONTENTS

ABSTRACT

Bayesian algorithms have been proven very successful in enhancing speech from background noise. A large number of such algorithms can be found in the scientific literature of the past 25 years. In this report a number of frequency domain Bayesian algorithms for speech enhancement is examined and evaluated. The algorithms can be grouped according to the feature of the Short Time Fourier Transform (STFT) they act upon, the estimator applied and the speech prior density function used. The STFT features considered are the DFT coefficients of the STFT (real and imaginary part) and their amplitude. The estimators applied are the Minimum Mean Square Error (MMSE) and the Maximum A Posteriori (MAP). Finally, the priors used are the one and two sided Chi and Gamma probability density functions (*pdfs*). Of particular interest is the value of the parameter a, which greatly influences the shape of the *pdfs* and subsequently the performance of the respective algorithms.

Results from extensive simulations performed with all the examined algorithms are also presented. The algorithms are evaluated according to two objective measures, the Segmental SNR (SegSNR) and the Perceptual Evaluation of Speech Quality (PESQ). An informal subjective evaluation of the enhanced speech is also given.

CHAPTER 1 INTRODUCTION

The continuous evolution of computers and digital systems and the widespread use of mobile phones has given rise to numerous man-machine voice interfaces, with applications such as voice portals, v-commerce etc. The portability of such devices enables them to be deployed in environments where background noise conditions can be adverse. Background noise poses a serious problem for both voice-based communication and automated services. Speech quality and intelligibility can be seriously hindered and automatic speech recognition systems are far less robust to noise than humans.

Speech enhancing algorithms can restore, to some extent, the noise corrupted speech, increasing its quality and potentially its intelligibility. The success rate of speech recognition engines can also be improved. Most modern speech enhancing algorithms work in the frequency domain. The transformation to the frequency domain is usually achieved with the Short Time Fourier Transform (STFT), because speech is a non stationary signal. To obtain the STFT, the time signal is divided into segments of $\sim$ 30ms overlapped by $\sim$ 20ms, windowed by a tapered window, and transformed with a DFT. The resulting DFT segments are then placed as columns in a matrix, which is called the STFT matrix (or STFT). The columns of this matrix are called time frames, while its rows are called frequency bins.

Speech enhancement algorithms typically try to restore the clean speech signal STFT given the STFT of the noisy signal. Some algorithms restore the DFT coefficients (real and imaginary parts) whilst others only modify the amplitude, which is combined with the phase of the noisy signal to produce the enhanced speech. Processing is usually performed in each frequency bin independently.

The speech and noise data in each frequency bin are modelled as random variables (r.v.), which are assumed to be distributed according to a probability density function (*pdf*). These modelling assumptions make techniques from Bayesian theory powerful tools for the estimation of the clean speech coefficients. The most widely used Bayesian estimators are the Minimum Mean Square Error (MMSE) and the Maximum A Posteriori (MAP),

both of which have been successfully employed in the construction of speech enhancement algorithms. A number of density functions have also been used to model speech (also known as priors) such as Gaussian, Laplacian, Gamma etc. The most common assumption used for the noise is that is distributed according to a Gaussian distribution.

The scientific literature has produced a number of Bayesian speech enhancement algorithms. Beginning with those that enhance the DFT coefficients, we have the Wiener filter (Gaussian speech priors, MMSE or MAP estimators [1]), and the method proposed by Martin in [1] employing Gamma priors with $a = 0.5$ (see eq. 3.19) and the MMSE estimator. Algorithms that enhance the amplitude of the STFT include the well known Ephraim and Malah MMSE STSA [2] (Rayleigh speech priors, MMSE estimator), while the MAP estimator with the same speech priors was proposed by Wolfe and Godsill in [3]. Finally, MAP estimators with Gamma and Chi priors were proposed by Dat et al in [4] and Lotter and Vary in [5, 6].

We can therefore see that a framework of Bayesian speech enhancement algorithms is forming. The algorithms that belong to it can be divided into the following categories: Firstly, according to the feature of the STFT they are enhancing, which can be either the DFT coefficients (real and imaginary parts) or their amplitude. Secondly, according to the Bayesian estimator that is applied, which is either the MMSE or the MAP. Finally, they can be grouped according to the prior density which is used for the speech. The speech priors mentioned above belong to two general families of priors. These families are the Chi (eqs. 3.7, 4.3) and the Gamma (eqs. 3.19, 4.15) density functions. The algorithms of the above framework are the issue of investigation in this report.

Although some of the algorithms already exist in the literature, some are new in this report. These include the MAP and MMSE estimators in the DFT domain with the Chi and Gamma priors (note that the Wiener filter and the algorithm proposed by Martin in [1] are only special cases) and the MMSE estimator in the amplitude domain with the Chi priors (again the algorithm proposed by Ephraim and Malah in [2] is a special case).

Due to the large number of algorithms examined in this report a 'code name' is given to each for easy reference. The names are of the form 'Estimator-Prior', where estimator is either the MAP or the MMSE and prior is the Chi or Gamma. For example the name 'MMSE-Chi' refers to the MMSE estimator with the Chi priors. The feature of the spec-

trogram the estimator is applied on (DFT or Amplitude) should be clear from the context. Otherwise the name will be extended i.e. MAP-Gamma-DFT or MAP-Gamma-AMP.

The organisation of this report is as follows: In chapter 2 we lay down the theoretical background of Bayesian estimation and introduce key concepts such as the prior and posterior density functions and the MMSE and MAP estimators. In chapter 3 we present the algorithms that work with the DFT coefficients (real and imaginary part), and in chapter 4 those who work with their amplitude. In chapter 5 we give the results from simulations preformed with all the presented algorithms and chapter 6 concludes this report.

CHAPTER 2 BAYESIAN ESTIMATION

2.1 Introduction

Most modern speech enhancing algorithms work in the Short Time Fourier Transform (STFT) domain, where each frequency bin is processed as a separate time series. The frequency bin representations most often used in the processing are either the complex DFT coefficients, or their amplitude. These features are assumed to be corrupted by noise and some estimation rule has to be employed for their recovery.

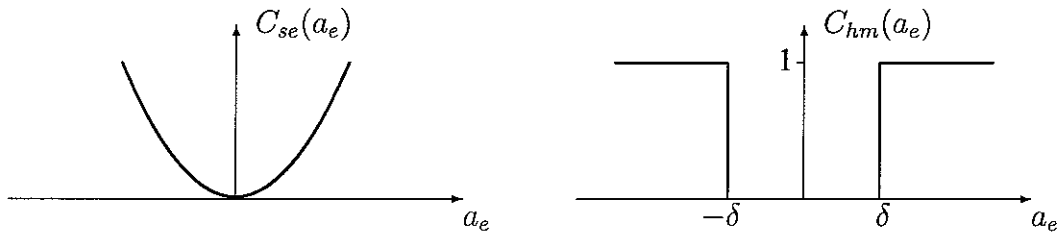
Both representations are successfully modelled as non-stationary stochastic processes, about whose probability distributions some knowledge is usually available. This modelling assumption and the knowledge about the form of the distributions make Bayesian estimation a promising method for a successful speech enhancing scheme. Indeed, Bayesian estimators are highly applicable when the parameter we need to estimate is a random variable (r.v.) itself and there is also some prior knowledge about its distribution.

A central concept in Bayesian estimation is the *cost* function $C\left(a, \hat{a}(b)\right)$, where a is the parameter we are trying to estimate, b is the observation and $\hat{a}(b)$ is an estimate of a once b is observed. The cost function defines the cost of observing b and saying that the estimate for a is $\hat{a}(b)$. It is often possible to express the cost as a function of a single variable $a_e(b)$, which is called the error and is defined as:

$$a_e(b) = \hat{a}(b) - a \tag{2.1}$$

Typical cost functions include the square error (eq. 2.2) and the 'hit-or-miss' cost function (eq. 2.3), which assigns a uniform cost for absolute error values above a threshold δ. These cost functions are also shown in figure 2.1.

$$C_{se}(a_e) = a_e^2 \tag{2.2}$$



Square error Hit-or-miss

Figure 2.1: Typical cost functions.

$$C_{hm}(a_e) = \begin{cases} 0 & \text{if } |a_e| < \delta \\ 1 & \text{if } |a_e| > \delta \end{cases} \tag{2.3}$$

Once a cost function is chosen, the objective is to minimise its expected value. The expectation (average) is with respect to all the possible values of the parameters a and b and is often referred to as *risk*, which is defined in equation 2.4. $p_{a,b}(a,b)$ is the joint probability density function (joint *pdf*) of a and b.

$$\mathcal{R} \triangleq E[C\left(a, \hat{a}(b)\right)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C\left(a, \hat{a}(b)\right) p_{a,b}(a,b)\, da\, db \tag{2.4}$$

Minimisation of the risk for different cost functions leads to different estimators. The estimators that are derived when the square error and hit-or-miss cost functions are used are the Minimum Mean Square Error (MMSE) and Maximum A Posteriori (MAP) estimators respectively. These estimators are the principal ones used in practice and will be examined in the following sections.

2.2 Minimum Mean Square Error Estimator

The MMSE estimator is obtained by minimising the risk function (eq. 2.4) with respect to $\hat{a}(b)$, using the square error cost function (eq. 2.2). The risk function can be written as:

$$\mathcal{R} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(a - \hat{a}(b)\right)^2 p_{a,b}(a,b)\, da\, db \tag{2.5}$$

Application of Bayes' theorem transforms the above equation to:

$$\mathcal{R} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(a - \hat{a}(b)\right)^2 p_{a|b}(a|b)\, da\, p_b(b)\, db \tag{2.6}$$

As $p_b(b)$ and the inner integral are non-negative, minimising the latter with respect to $\hat{a}$ also minimises the risk. Differentiation of the inner integral w.r.t. $\hat{a}$ yields:

$$\frac{d}{d\hat{a}}\left[\int_{-\infty}^{\infty}(a-\hat{a}(b))^2\,p(a|b)\,da\right] = -2\int_{-\infty}^{\infty}(a-\hat{a}(b))\,p(a|b)\,da \tag{2.7}$$

For simplicity we have dropped the subscript of the *pdfs*, as their argument should make clear which random variables they refer to. Setting equation 2.7 to zero and considering that the integral of $p(a|b)$ from $-\infty$ to ∞ is one, we see that the estimate that minimizes the mean square error is:

$$\hat{a}(b) = E[a|b] = \int_{-\infty}^{\infty} a\,p(a|b)\,da \tag{2.8}$$

It is interesting to note that the MMSE estimate is always the mean of the a posteriori density $p(a|b)$ (see figure 2.2). Further application of the Bayes theorem on eq. 2.8 can yield the expression in eq. 2.9, where the MMSE estimate is expressed in terms of the likelihood function $p(b|a)$ (evidence) and the prior $p(a)$.

$$\hat{a}(b) = \frac{\int_{-\infty}^{\infty} a\,p(a,b)\,da}{p(b)} = \frac{\int_{-\infty}^{\infty} a\,p(a,b)\,da}{\int_{-\infty}^{\infty} p(a,b)\,da} = \frac{\int_{-\infty}^{\infty} a\,p(b|a)p(a)\,da}{\int_{-\infty}^{\infty} p(b|a)p(a)\,da} \tag{2.9}$$

2.3 Maximum A Posteriori Estimator

The maximum a posteriori estimator can be found by substituting the hit-or-miss error function (eq. 2.3) in the expression for the risk (eq. 2.4), which then reads:

$$\mathcal{R} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_{hm}(a_e(b))p(a,b)\,da\,db \tag{2.10}$$

Applying the Bayes rule and following the same argument as in equations 2.6 and 2.7 we see that for the minimisation of the risk it suffices to minimise:

$$\hat{\mathcal{R}} = \int_{-\infty}^{\infty} C_{hm}(a_e(b))p(a|b)\,da \tag{2.11}$$

Considering that the cost function $C_{hm}(a_e(b))$ is 1 only for $a_e(b) > |\delta|$ or equivalently for $a > \hat{a}(b) + \delta$ and $a < \hat{a}(b) - \delta$ while it is zero everywhere else, eq. 2.11 can be written

as:

$$\hat{\mathcal{R}} = \int_{-\infty}^{\hat{a}(b)-\delta} 1 \cdot p(a|b)\, da + \int_{\hat{a}(b)+\delta}^{\infty} 1 \cdot p(a|b)\, da \tag{2.12}$$

or as:

$$\hat{\mathcal{R}} = 1 - \int_{\hat{a}(b)-\delta}^{\hat{a}(b)+\delta} p(a|b)\, da \tag{2.13}$$

if we recall that $\int_{-\infty}^{\infty} p(a|b)\, da = 1$. As δ approaches zero, the value of $\hat{a}(b)$ that minimises $\hat{\mathcal{R}}$ is the value of a for which $p(a|b)$ has its maximum. In other words, the risk is minimized for the hit-or-miss cost function when the estimate is the maximum (*mode*) of the posterior density function (see fig. 2.2); hence the Maximum A Posteriori name for the estimator.

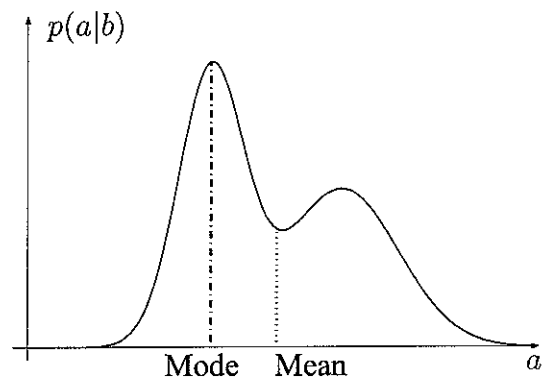


Figure 2.2: Posterior density with the mode and the mean.

CHAPTER 3 DFT DOMAIN ALGORITHMS

3.1 Introduction

The Bayesian methods presented previously can be applied for enhancing speech in the DFT domain or more precisely, one of the estimation methods presented in chapter 2 can be applied to enhance the real and the imaginary parts of the noisy speech STFT coefficients in each frequency bin separately. This is in juxtaposition with methods that act upon the amplitude of the STFT coefficients which are presented in the next chapter.

Suppose that we observe a signal $x(t)$ which is the sum of a speech signal $s(t)$ and a noise signal $n(t)$, i.e.:

$$x(t) = s(t) + n(t) \tag{3.1}$$

Also denote by $\mathbf{X}_k$, $\mathbf{S}_k$, and $\mathbf{N}_k$ the k^{th} frequency bin of $x(t)$, $s(t)$, and $n(t)$ respectively, obtained with the STFT. Because the estimation rules are applied independently to the real and the imaginary part of each frequency bin, to simplify the analysis we shall only consider the real part. The results for the imaginary parts are identical. X, S and N in the following, denote the real part of an STFT sample in the k^{th} frequency bin. Because of the linearity of the Fourier Transform the following equation will also hold:

$$X = S + N \tag{3.2}$$

The estimation problem can be formulated as follows: we observe a sample of X and we want to estimate S given the noise and the speech statistics. The estimators that will be employed are the MMSE and the MAP, the derivation of which requires first the calculation of the posterior probability density function $p(S|X)$. According to Bayes theorem the posterior density can be written as:

$$p(S|X) = \frac{p(X|S)p(S)}{p(X)} \tag{3.3}$$

We first state a fundamental result that if $X = S + N$ then:

$$p(X|S) = p_N(X - S) \tag{3.4}$$

where p_N is the *pdf* of N. Assuming that N is a zero mean Gaussian r.v. with variance σ_N^2, the likelihood $p(X|S)$ can be written as:

$$p(X|S) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \tag{3.5}$$

The prior $p(S)$ is a density function that reflects our knowledge about the distribution of S. We will see in the following that the form of the prior strongly affects the performance of the resulting algorithm; hence an appropriate selection of a prior is of critical importance. The prior densities considered here are the 2-sided *Chi* and *Gamma pdfs* that will be presented shortly.

The probability of the data $p(X)$ is a normalising factor that does not depend on S and ensures that the integral of the posterior density with respect to S equals 1. It can be calculated according to Bayes rule as:

$$p(X) = \int_{-\infty}^{\infty} p(X|S)p(S)\, dS \tag{3.6}$$

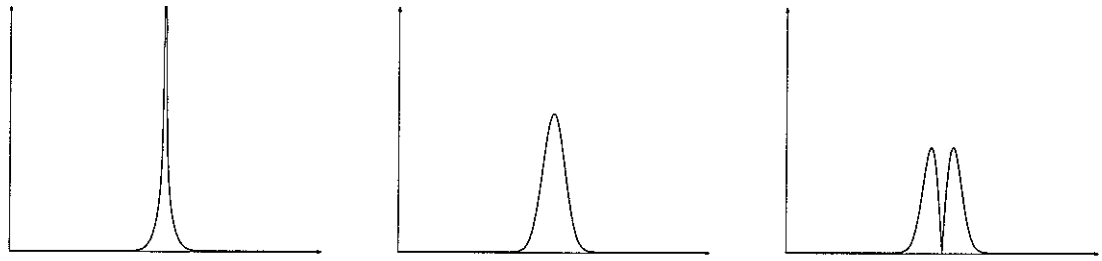
Note the similarity of the numerator and the denominator of 3.3 if $p(X)$ is replaced by equation 3.6.

3.2 2-Sided Chi Speech Priors

The 2-sided Chi *pdf* is given by:

$$p(S) = \frac{1}{\theta^a \Gamma(a)} |S|^{2a-1} \exp\left[-\frac{S^2}{\theta}\right] \tag{3.7}$$

where $\Gamma(.)$ is the gamma function. Special cases of this distribution occur when $a = 0.5$ (Gaussian) and when $a = 1$ (2-sided Rayleigh). Figure 3.1 shows some instances of the 2-sided Chi *pdf* for some characteristic values of a.



(a) $a = 0.1$ (b) $a = 0.5$ (c) $a = 1$

Figure 3.1: 2-Sided Chi *pdfs* for different values of a.

3.2.1 MMSE Estimator

As shown in chapter 2 the MMSE estimator is the mean of the posterior density. Therefore, the MMSE estimator of S will be:

$$\hat{S} = \mathrm{E}\left[S|X\right] = \frac{\int_{-\infty}^{\infty} S\, p(X|S)\, p(S)\, dS}{\int_{-\infty}^{\infty} p(X|S)\, p(S)\, dS} \tag{3.8}$$

where $p(S)$ and $p(X|S)$ are given by eqs. 3.7 and 3.5 respectively. Calculation of the integrals in 3.8 yields (see Appendix A.2):

$$\hat{S} = 2a\sigma_N^2\zeta\, \frac{D_{-2a-1}(-\zeta X) - D_{-2a-1}(\zeta X)}{D_{-2a}(-\zeta X) + D_{-2a}(\zeta X)} \qquad \text{where} \quad \zeta = \sqrt{\frac{\theta/\sigma_N^2}{\theta + 2\sigma_N^2}} \tag{3.9}$$

where $D_v(x)$ is the *Parabolic Cylinder Function* (eq. 9.240, [7])

It is also possible to express the above estimator as a gain of the noisy coefficients X, which is a function of the a priori and a posteriori SNRs. The a priori SNR is given by the relationship $\xi = \mathrm{E}[S^2]/\mathrm{E}[N^2]$, which for this case can be written as $\xi = \theta a/\sigma_N^2$. The a posteriori SNR is defined as $\gamma = X^2/\mathrm{E}[N^2]$ and in this case can be written as $\gamma = X^2/\sigma_N^2$. Substituting the expressions for ξ and γ in equation 3.9 we get:

$$\hat{S} = X\left[\frac{2a\eta}{\gamma}\frac{D_{-2a-1}(-\eta) - D_{-2a-1}(\eta)}{D_{-2a}(-\eta) + D_{-2a}(\eta)}\right] \qquad \text{where} \quad \eta = \mathrm{sgn}(X)\sqrt{\frac{\xi\gamma}{\xi + 2a}} \tag{3.10}$$

3.2.2 MAP Estimator

The MAP estimator $\hat{S}$ is the value of S for which the posterior density has its maximum. The probability of the data $p(X)$ is not a function of S so it suffices to find the maximum of $p(X|S)p(S)$, which are respectively defined by 3.5 and 3.7. The algebraic manipulations are substantially simplified if $\ln(p(X|S)p(S))$ is maximised instead. The resulting estimator is given by (see Appendix A.3):

$$\hat{S} = \zeta\frac{X}{2} + \text{sgn}(X)\left[\left(\zeta\frac{X}{2}\right)^2 + (a-0.5)\,2\sigma_N^2\,\zeta\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \tag{3.11}$$

where $\text{sgn}(.)$ is the signum function. It is also possible to express the above estimator as a gain of the noisy coefficients, which is a function of the a priori and a posteriori SNR, defined in section 3.2.1. The resulting expression is:

$$\hat{S} = X\left[\frac{\eta}{2} + \left[\left(\frac{\eta}{2}\right)^2 + (a-0.5)\frac{2\eta}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{\xi}{\xi + 2a} \tag{3.12}$$

It is easy to see that when $a < 0.5$ the posterior density has a pole at zero. The strategy we follow in this case is to take the maximum provided by equation 3.11 when it exists and when it does not (or when the argument of the square root is negative) we suppress the noisy sample by a fixed amount i.e. 25 dB.

Although it is not evident at first sight (especially in the case of the MMSE), both the MAP and the MMSE estimators give the well-known Wiener solution for $a = 0.5$:

$$\hat{S} = \frac{X\sigma_S^2}{\sigma_S^2 + \sigma_N^2} \tag{3.13}$$

where σ_S^2 is the variance of the speech prior, which for a general 2-sided Chi *pdf* is equal to θa and in this particular case is $\theta/2$.

3.2.3 On Line Estimation of the a Parameter

A method for the estimation of the a parameter can be found using the method of *moment matching*. A method for finding Maximum Likelihood estimates of a is described in [8], but requires a significantly greater amount of computation and the availability of clean

speech samples. The moment matching method is quite simple in its implementation, can be applied directly on the noisy samples and the accuracy of the estimates is found to be satisfactory for our purposes.

Given the model $X = S + N$ the fourth moment of the noisy speech can be shown to be:

$$\mathrm{E}[X^4] = \mathrm{E}[S^4] + 6\mathrm{E}[S^2]\mathrm{E}[N^2] + \mathrm{E}[N^4] \tag{3.14}$$

Given the Gaussian noise model we have:

$$\mathrm{E}[N^2] = \sigma_N^2, \qquad \text{and} \qquad \mathrm{E}[N^4] = 3\sigma_N^4 \tag{3.15}$$

The corresponding moments of the 2-sided Chi *pdf* are:

$$\mathrm{E}[S^2] = \theta a, \qquad \text{and} \qquad \mathrm{E}[S^4] = \theta^2 a(a+1) \tag{3.16}$$

Finally, the second moment of S can be also found from the a priori SNR ξ as $\mathrm{E}[S^2] \triangleq \sigma_S^2 = \xi\sigma_N^2$.

Substituting the above equations in 3.14 we obtain:

$$\mathrm{E}[X^4] = \sigma_S^4 \frac{a+1}{a} + 6\sigma_S^2\sigma_N^2 + 3\sigma_N^4$$

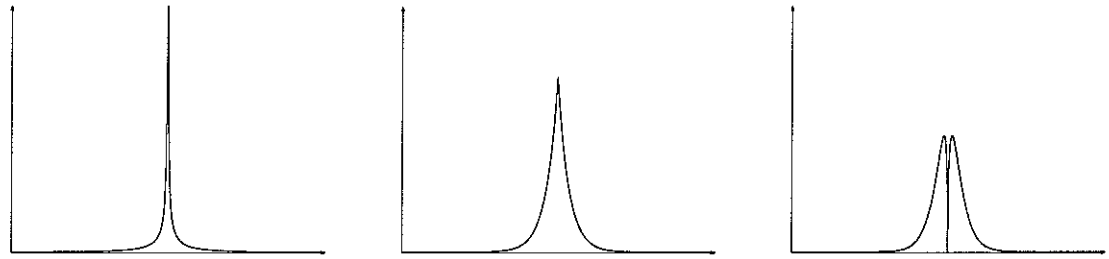
or:

$$\frac{a+1}{a} = \frac{\mathrm{E}[X^4] - 6\sigma_S^2\sigma_N^2 - 3\sigma_N^4}{\sigma_S^4} = \kappa \tag{3.17}$$

Therefore:

$$a = \frac{1}{\kappa - 1} \tag{3.18}$$

In equation 3.17 κ can be recognised as the *kurtosis*, defined as $\kappa \triangleq \mathrm{E}[S^4]/\mathrm{E}[S^2]^2$. Note that as the kurtosis tends to infinity a tends to zero and the priors are getting narrower with longer tails. If the kurtosis is below 1, a has a negative value, which is not acceptable.



(a) $a = 0.1$ (b) $a = 1$ (c) $a = 1.5$

Figure 3.2: 2-Sided Gamma *pdfs* for different values of a.

3.3 2-Sided Gamma Speech Priors

The 2-sided Gamma density function is a generalisation of the Laplacian *pdf* and is given by:

$$p(S) = \frac{1}{2\theta^a \Gamma(a)} |S|^{a-1} \exp\left[-\frac{|S|}{\theta}\right] \tag{3.19}$$

The 2-sided Gamma *pdf* is more leptokurtic (higher value at zero, longer tails) than the 2-sided Chi *pdf* for the same value of a, while for $a = 1$ yields the Laplacian *pdf*. Some plots for characteristic values of a are shown in figure 3.2.

3.3.1 MMSE Estimator

To obtain the MMSE estimator we need to substitute in equation 3.8 the expression for the likelihood (eq. 3.5) and the Gamma prior, which is given by equation 3.19. The resulting estimator is given by (see Appendix A.4):

$$\hat{S} = a\sigma_N \frac{\exp\left[\frac{\zeta_1^2}{4}\right] D_{-a-1}(\zeta_1) - \exp\left[\frac{\zeta_2^2}{4}\right] D_{-a-1}(\zeta_2)}{\exp\left[\frac{\zeta_1^2}{4}\right] D_{-a}(\zeta_1) + \exp\left[\frac{\zeta_2^2}{4}\right] D_{-a}(\zeta_2)} \tag{3.20}$$

where $\zeta_1 = \frac{\sigma_N}{\theta} - \frac{X}{\sigma_N}, \zeta_2 = \frac{\sigma_N}{\theta} + \frac{X}{\sigma_N}$

To express the above estimator as a gain of the noisy coefficients we must bear in mind that the expression for the a priori SNR is now $\xi = \theta^2 a(a+1)/\sigma_N^2$ and the a posteriori

SNR is again $\gamma = X^2/\sigma_N^2$. The resulting expression is:

$$\hat{S} = X \left[\frac{a\,\mathrm{sgn}(X)}{\sqrt{\gamma}} \frac{\exp\left[\frac{\eta_1^2}{4}\right] D_{-a-1}(\eta_1) - \exp\left[\frac{\eta_2^2}{4}\right] D_{-a-1}(\eta_2)}{\exp\left[\frac{\eta_1^2}{4}\right] D_{-a}(\eta_1) + \exp\left[\frac{\eta_2^2}{4}\right] D_{-a}(\eta_2)} \right] \tag{3.21}$$

where $\eta_1 = \frac{\sqrt{a(a+1)}}{\sqrt{\xi}} - \mathrm{sgn}(X)\sqrt{\gamma}, \quad \eta_2 = \frac{\sqrt{a(a+1)}}{\sqrt{\xi}} + \mathrm{sgn}(X)\sqrt{\gamma}$

3.3.2 MAP Estimator

The MAP estimator for the 2-sided Gamma priors can be obtained in the same way the corresponding estimator for the 2-sided Chi priors was found. It therefore suffices to find the maximum of $\ln(p(X|S)p(S))$ where $p(X|S)$ is again given by eq. 3.5 and $p(S)$ by eq. 3.19. The resulting estimator is (see Appendix A.5):

$$\hat{S} = \zeta + \mathrm{sgn}(X)\left[\zeta^2 + (a-1)\sigma_N^2\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{X}{2} - \mathrm{sgn}(X)\frac{\sigma_N^2}{2\theta} \tag{3.22}$$

The expression of the above estimator as a gain of the noisy coefficients is given below. The expressions for the a priori and the a posteriori SNRs are the same as in section 3.3.1.

$$\hat{S} = X\left[\eta + \mathrm{sgn}(X)\left[\eta^2 + \frac{a-1}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{a(a+1)}{\xi\gamma}} \tag{3.23}$$

When $a < 1$ the posterior density has a pole. In this case we take the solution provided by eq. 3.22 when it exists (or when the argument of the square root is positive) and suppress the noisy sample by a fixed amount (25 dB) when it does not. If we also observe the form of the posterior density (eq. A.22) we can see that the value of S where its maximum occurs must have the same sign with X. It is possible however, that the expression in eq. 3.22 can yield negative solutions for positive X and vice versa. This is not acceptable and in these cases the noisy samples are again suppressed by the same fixed amount.

3.3.3 On Line Estimation of the a Parameter

An estimate for the a parameter can be obtained in a similar way as in section 3.2.3. The corresponding moments for the Gamma prior are now:

$$\mathrm{E}[S^2] = \theta^2 a(a+1), \qquad \text{and} \qquad \mathrm{E}[S^4] = \theta^4 a(a+1)(a+2)(a+3) \tag{3.24}$$

Following the same procedure as in section 3.2.3 we have:

$$\frac{(a+2)(a+3)}{a(a+1)} = \frac{\mathrm{E}[X^4] - 6\sigma_S^2\sigma_N^2 - 3\sigma_N^4}{\sigma_S^4} = \kappa \tag{3.25}$$

Solving the quadratic equation we consecutively have:

$$a = \frac{5 - \kappa \pm \sqrt{(5-k)^2 - 24(1-k)}}{2(\kappa - 1)} \tag{3.26}$$

$$a = \frac{5 - \kappa + \sqrt{\kappa^2 + 14\kappa + 1}}{2\kappa - 2} \tag{3.27}$$

The root with the $(+)$ is selected because for $\kappa > 1$ the root with the $(-)$ is negative as it can be seen from equation 3.26. If the kurtosis κ is less than 1 the both roots are negative, which is not acceptable.

3.4 Estimation of the Remaining Parameters

Although a number of density functions that model speech were discussed previously, the noise coefficients were assumed to have a Gaussian distribution. This might need some justification. If noise is stationary the Gaussian assumption for the DFT coefficients is justified by the central limit theorem. If on the other hand, noise is non-stationary its DFT coefficients can still be modelled as Gaussian but with a time varying variance σ_N^2, which can track the noise changes. This model is viable because noise usually changes more slowly than speech so we do not need to resort to other assumptions about its distribution.

It is therefore apparent, that some algorithm has to be employed to track the temporal changes in the noise variance. A simple method involves the use of a Voice Activity Detector (VAD), which identifies the time frames that contain noise only and then estimates its variance in each frequency bin. A more sophisticated method called noise estimation

based on minimum statistics was presented in [9] and estimates the variance in each frequency bin by searching for power minima. To be precise, the above algorithm estimates the variance of the amplitude of the DFT coefficients. However, if the DFT coefficients are independent, Gaussian, zero mean and of variance σ_N^2 then the amplitude has a Rayleigh distribution with variance $2\sigma_N^2$. Hence, an estimate of the DFT coefficients variance can be obtained from an estimate of the variance of their amplitude by a simple division by 2.

All the estimators presented previously were given in two forms: one involving only the parameters of the density functions (a,θ and σ_N^2) and one where the two latter parameters were replaced by two popular quantities in the Bayesian speech enhancement literature, the a priori and a posteriori SNR (ξ and γ respectively). From all the quantities and parameters involved in the estimators the only ones whose calculation we have either not discussed yet or is not straight forward are the a priori SNR ξ and the parameter θ of the prior density functions. These two quantities are related by simple expressions, which depend on the prior density used, so calculation of one of them is sufficient to define all the involved quantities.

It is possible to estimate a fixed value for θ from clean data samples or some other adaptive scheme. It has been proven highly successful however, to estimate the a priori SNR instead, with a method such as the Decision Directed Approach, presented in [2]. The success of this approach is that it aids the suppression of the musical noise phenomenon (narrow band noise with time changing frequency center that gives the impression of musical tones) [10]. The rule for the decision directed estimation of the a priori SNR is given by:

$$\xi_k = \alpha \frac{\hat{S}_{k-1}^2}{\sigma_{N|k-1}^2} + (1-\alpha)\max(\gamma_k - 1) \tag{3.28}$$

The subscripts k and $k-1$ denote the current and the previous time frames, while α is a smoothing parameter, which is typically set to 0.99. $\hat{S}_{k-1}^2$ is the estimated clean speech sample in the previous time frame. Other methods for estimating the a priori SNR also exist [11], but they are more involved computationally.

CHAPTER 4 AMPLITUDE DOMAIN ALGORITHMS

4.1 Introduction

In the previous chapter we presented methods for estimating the clean speech DFT coefficients (real and imaginary parts) in every frequency bin given the noisy observations. An alternative option is to estimate the *amplitude* and the *phase* of the clean speech frequency bins instead, which generates a whole new family of algorithms. In practice it is sufficient to estimate the amplitude only and then combine it with the noisy speech phase to create the enhanced speech waveform. That is because it has been widely argued that the perception of speech is phase insensitive [12], [13] and moreover, Ephraim in [2] has shown that the optimal estimate for the clean speech phase is the noisy speech phase itself. This property gives the amplitude estimation methods an advantage compared to their DFT coefficients counterparts, which is that the number of data points that need to be estimated is halved.

For every frequency bin k we can express the DFT coefficients of the clean and noisy speech in terms of their amplitude and phase as $\mathbf{X}_k \triangleq R_k e^{j\psi_k}$, and $\mathbf{S}_k \triangleq A_k e^{j\phi_k}$, where R_k, ψ_k, A_k and ϕ_k are the amplitude and phase of the noisy and clean speech respectively. As the estimation procedure has to be applied in every frequency bin independently, the subscript k will be dropped in the following for notational simplicity.

The estimation problem can be formulated as follows: we are trying to find an estimate of the clean speech amplitude A given the noisy speech amplitude R and phase ψ. Recall from chapter 3 that in order to apply both the MMSE and the MAP estimators, the

calculation of the posterior density $p(A|R,\psi)$ is first necessary. This can be written as:

$$\begin{aligned} p(A|R,\psi) &= \frac{p(R,\psi|A)p(A)}{\int_0^\infty p(R,\psi|A)p(A)\,dA} \\ &= \frac{\int_0^{2\pi} p(R,\psi|A,\phi)p(A)p(\phi)\,d\phi}{\int_0^\infty \int_0^{2\pi} p(R,\psi|A,\phi)p(A)p(\phi)\,dA\,d\phi} \end{aligned} \tag{4.1}$$

In the above equation we note that $p(A)$ and $p(\phi)$ are factorised, which implies that A and ϕ are independent. This is indeed supported by simulation results, which also show that the distribution of the clean speech phase is uniform; hence we can replace $p(\phi)$ with $1/2\pi$.

The density function of R and ψ conditioned on A and ϕ is given by (see Appendix A.1):

$$p(R,\psi|A,\phi) = \frac{R}{2\pi\sigma_N^2}\,\exp\left[-\frac{R^2+A^2-2RA\cos(\psi-\phi)}{2\sigma_N^2}\right] \tag{4.2}$$

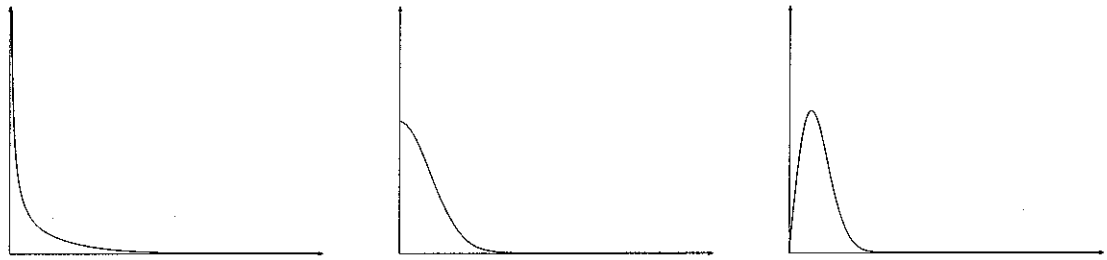
We now proceed to derive the MMSE and MAP estimators for different families of speech amplitude priors.

4.2 Chi Speech Priors

The Chi density function is the 1-sided version of the *pdf* described in section 3.2 and its functional form is given by:

$$p(A) = \frac{2}{\theta^a\Gamma(a)}\,A^{2a-1}\,\exp\left[-\frac{A^2}{\theta}\right],\ \text{with } A \geq 0 \tag{4.3}$$

It is easy to see that for $a = 1$ yields the Rayleigh *pdf*, while for $a = 0.5$ the half Gaussian. Some of its characteristic instances can be seen in figure 4.1. Let us now present the expressions for the MMSE and the MAP estimators.



(a) $a = 0.1$ (b) $a = 0.5$ (c) $a = 1$

Figure 4.1: 1-Sided Chi *pdfs* for different values of a.

4.2.1 MMSE Estimator

The MMSE estimator of the clean speech amplitude A given the noisy speech amplitude R and phase ψ is given by:

$$\begin{aligned} \hat{A} = \mathrm{E}[A|R,\psi] &= \int_0^\infty A p(A|R,\psi)\, dA \\ &= \frac{\int_0^\infty \int_0^{2\pi} A p(R,\psi|A,\phi) p(A) p(\phi)\, d\phi\, dA}{\int_0^\infty \int_0^{2\pi} p(R,\psi|A,\phi) p(A) p(\phi)\, d\phi\, dA} \end{aligned} \tag{4.4}$$

Substitution of $p(R,\psi|A,\phi)$ and $p(A)$ from 4.2 and 4.3 respectively and the assumption of a uniform phase distribution ($p(\phi) = \frac{1}{2\pi}$) yields (see Appendix A.6):

$$\hat{A} = \sqrt{2\sigma_N^2 \zeta}\, \frac{\Gamma(a+0.5)}{\Gamma(a)}\, \frac{{}_1F_1(a+0.5;1;\frac{R^2}{2\sigma_N^2}\zeta)}{{}_1F_1(a;1;\frac{R^2}{2\sigma_N^2}\zeta)} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \tag{4.5}$$

${}_1F_1(\alpha;\beta;\gamma)$ is the Confluent Hypergeometric Function (eq. 9.210.1, [7]). The above estimator with $a = 1$ (Rayleigh speech prior) is the one found in the well known Ephraim-Malah algorithm [2].

The estimator in equation 4.5 can be expressed as a gain of the noisy coefficients, which is a function of the a priori and the a posteriori SNRs. The expression for the a priori SNR is $\xi = \mathrm{E}[A^2]/\mathrm{E}[N^2]$ or $\xi = \theta a/2\sigma_N^2$. The a posteriori SNR is given by $\gamma = R^2/\mathrm{E}[N^2]$ or $\gamma = R^2/2\sigma_N^2$. The estimator can be written as:

$$\hat{A} = R\left[\sqrt{\frac{\eta}{\gamma}}\, \frac{\Gamma(a+0.5)}{\Gamma(a)}\, \frac{{}_1F_1(a+0.5;1;\gamma\eta)}{{}_1F_1(a;1;\gamma\eta)}\right] \qquad \text{where} \quad \eta = \frac{\xi}{\xi + a} \tag{4.6}$$

4.2.2 MAP Estimator

The MAP estimator can be found by maximising with respect to A the posterior density $p(A|R,\psi)$. Once again since the denominator in the expression for the posterior density in equation 4.1 is not a function of A it suffices to maximise the numerator only, or its logarithm as the calculations are simplified significantly; thus:

$$\hat{A} = \arg\max_{A} \ln\left(\int_0^{2\pi} p(R,\psi|A,\phi)p(A)p(\phi)\,d\phi\right) \tag{4.7}$$

Substituting $p(R,\psi|A,\phi)$ and $p(A)$ from 4.2 and 4.3 and $p(\phi) = \frac{1}{2\pi}$ yields (see Appendix A.7):

$$\hat{A} = \zeta\frac{R}{2} + \left[\left(\zeta\frac{R}{2}\right)^2 + (a-0.75)\,2\sigma_N^2\,\zeta\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \tag{4.8}$$

The above expression as a gain of the noisy coefficients is:

$$\hat{A} = R\left[\frac{\eta}{2} + \left[\left(\frac{\eta}{2}\right)^2 + (a-0.75)\frac{\eta}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{\xi}{\xi + a} \tag{4.9}$$

By examining equations 4.8, 4.9 we can see that if $a < 0.75$ the square root argument can be negative, which reflects the fact that the posterior density has no finite maximum. (It actually has a pole at zero). In this case the noisy samples are suppressed by a fixed amount (i.e. 25 dB).

4.2.3 On Line Estimation of the a Parameter

A method for estimating a from the noisy speech samples, similar to that developed in section 3.2.3, is developed below. Given the model for the DFT coefficients $\mathbf{X} = \mathbf{S} + \mathbf{N}$ the fourth moment of the noisy speech spectral amplitude can be written as:

$$\mathrm{E}[R^4] = \mathrm{E}[A^4] + 4\mathrm{E}[A^2]\mathrm{E}[B^2] + \mathrm{E}[B^4] \tag{4.10}$$

where B denotes the amplitude of the noise DFT coefficients, which are modelled as independent, zero mean Gaussian r.vs. with variance σ_N^2. Given this model, the 2^{nd} and

4^{th} moments of the noise spectral amplitude are given by:

$$\mathrm{E}[B^2] = 2\sigma_N^2, \qquad \text{and} \qquad \mathrm{E}[B^4] = 8\sigma_N^4 \tag{4.11}$$

The corresponding moments for the speech spectral amplitude given the Chi prior are:

$$\mathrm{E}[A^2] = \theta a, \qquad \text{and} \qquad \mathrm{E}[A^4] = \theta^2 a(a+1) \tag{4.12}$$

The second moment of A can again be found from the a priori SNR ξ as $\mathrm{E}[A^2] \triangleq \sigma_A^2 = 2\xi\sigma_N^2$. Substituting the above in equation 4.10 we have:

$$\frac{(a+1)}{a} = \frac{\mathrm{E}[R^4] - 8\sigma_A^2\sigma_N^2 - 8\sigma_N^2}{\sigma_A^4} = \kappa \tag{4.13}$$

or finally

$$a = \frac{1}{\kappa - 1} \tag{4.14}$$

where κ is the kurtosis of the clean speech amplitude, defined as $\kappa \triangleq \mathrm{E}[A^4]/\mathrm{E}[A^2]^2$. Note the similarity of equation 4.14 with equation 3.18, which is the result of the second and fourth raw moments being the same for the 1-sided and the 2-sided Chi *pdfs*. Note however, that the definition of κ changes as in section 3.2.3 it is the kurtosis of the DFT coefficients, while here it represents the kurtosis of the amplitude coefficients.

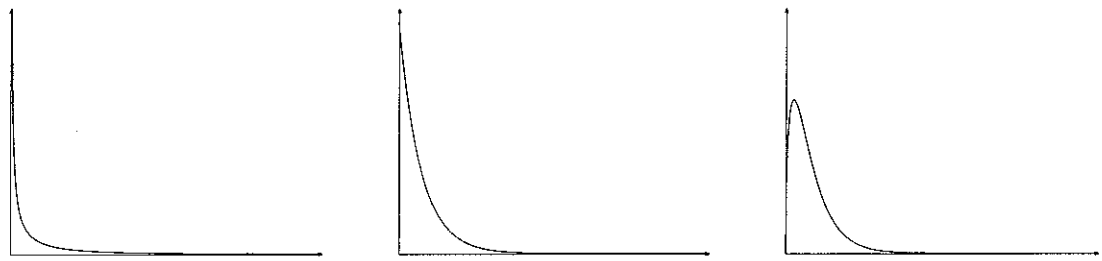
4.3 Gamma Speech Priors

Another family of speech priors is be given by the Gamma density function, described by equation:

$$p(A) = \frac{1}{\theta^a \Gamma(a)} A^{a-1} \exp\left[-\frac{A}{\theta}\right], \text{ with } A \geq 0 \tag{4.15}$$

The Gamma density function is the 1-sided variant of the *pdf* described in section 3.3. Some of its characteristic instances for various values of the parameter a are shown in figure 4.2.

One peculiarity of modelling speech amplitude with Gamma priors is that a closed form solution for the MMSE estimator cannot be found. It can be shown that the resulting integrals are of the form $\int_0^\infty x^\mu \exp(-\alpha x^2 - \beta x)\, I_0(\gamma x)\, dx$, where $I_0(\gamma x)$ is the zeroth order



(a) $a = 0.1$ (b) $a = 1$ (c) $a = 1.5$

Figure 4.2: 1-Sided Gamma *pdfs* for different values of a.

modified Bessel function of the first kind. The above integral has no analytic solution known to the author and its value can be approximated only with numerical techniques. For this reason the above algorithm is not considered here. Thus, for the Gamma speech priors only the MAP estimator will be given below.

4.3.1 MAP Estimator

The MAP estimator can be found by maximising the expression in (4.7), where the likelihood is again given in (4.2), the phase density is $p(\phi) = \frac{1}{2\pi}$ and the Gamma speech prior is given in (4.15). The resulting estimator is (see Appendix A.8):

$$\hat{A} = \zeta + \left[\zeta^2 + (a - 1.5)\sigma_N^2\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{R}{2} - \frac{\sigma_N^2}{2\theta} \tag{4.16}$$

The a priori SNR is given by $\xi = \theta^2 a(a+1)/2\sigma_N^2$ and the a posteriori SNR by $\gamma = R^2/2\sigma_N^2$. The estimator in equation 4.16 can be written as:

$$\hat{A} = R\left[\eta + \left[\eta^2 + \frac{a - 1.5}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{1}{2} - \frac{1}{4}\sqrt{\frac{a(a+1)}{\xi\gamma}} \tag{4.17}$$

If $a < 1.5$ the argument of the square root can be negative, which indicates that the posterior density has no finite maximum. In this case the noisy samples are suppressed by a fixed amount (i.e. 25 dB). The estimator in the above equations can sometimes yield negative estimates, which are not acceptable, as the parameter we are estimating is amplitude. In these cases the noisy samples are again suppressed by the same fixed amount.

4.3.2 On Line Estimation of the a Parameter

A value for the parameter a can be obtained in a similar fashion as in section 4.2.3. The procedure is identical apart from the different expressions for the speech moments, which have changed due to the new prior. These are:

$$\mathrm{E}[A^2] = \theta^2 a(a+1), \qquad \text{and} \qquad \mathrm{E}[A^4] = \theta^4 a(a+1)(a+2)(a+3) \tag{4.18}$$

Following the same steps as in section 4.2.3 we have:

$$\frac{(a+2)(a+3)}{a(a+1)} = \frac{\mathrm{E}[R^4] - 8\sigma_A^2\sigma_N^2 - 8\sigma_N^2}{\sigma_A^4} = \kappa \tag{4.19}$$

Or finally, solving the quadratic equation w.r.t a:

$$a = \frac{5 - \kappa + \sqrt{\kappa^2 + 14\kappa + 1}}{2\kappa - 2} \tag{4.20}$$

The valid root from the solution of the quadratic equation is the one with the $(+)$ for the same reasons stated in section 3.3.3. Note again the similarity with equation 3.27, which is the result of the second and fourth raw moments of the 1-sided and 2-sided Gamma *pdfs* being identical.

4.4 Estimation of the Remaining Parameters

The estimation of the remaining parameters is quite similar to that developed in section 3.4. Here, we will just point out some minor differences.

The variance of the amplitude of the noise coefficients can again be estimated with either the use of a VAD or with the minimum statistics algorithm. Note however that the noise variance in this section is $2\sigma_N^2$ while in section 3 was σ_N^2.

The a priori SNR can be estimated with the decision directed approach which is given in equation 4.21.

$$\xi_k = \alpha \frac{\hat{A}_{k-1}^2}{2\sigma_{N|k-1}^2} + (1-\alpha)\max(\gamma_k - 1) \tag{4.21}$$

where $\hat{A}_{k-1}^2$ is the estimated speech amplitude sample of the $(k-1)^{\text{th}}$ time frame (previ-

ous), $2\sigma^2_{N|k-1}$ is the noise estimate of the $(k-1)^{\text{th}}$ time frame and γ_k is the a posteriori SNR of the k^{th} time frame (current).

CHAPTER 5 RESULTS

In this chapter we present the results from simulations performed with the family of Bayesian algorithms described in this report. The simulations were performed with a number of clean speech phrases, artificially corrupted by white Gaussian noise, and then enhanced with the algorithms described here. The performance of the algorithms is evaluated using a number of objective measures, while an attempt has also been made to subjectively assess the quality of the resulting speech. Of particular interest in this evaluation, is the effect of the form of the speech prior densities.
Some details about the specifics of the simulation and the evaluation measures will be given first, before presenting the simulation results.

5.1 Simulation Setup

The clean speech database comprised of 16 phrases, half of which were spoken by men and half by women. The sampling frequency was 8 KHz and the total duration 27 seconds. The speech phrases were corrupted by white Gaussian noise at 0, 10 and 20 dB input Segmental SNR. For these input Segmental SNR levels the corresponding PESQ values were 1.536, 2.239 and 2.914 respectively[1]. The transformation to the frequency domain was performed with Hamming windows of 256 samples length and using a 192 sample overlap. The windows were also normalised so that their amplitude when overlapped and added was 1.

The speech variance in all the algorithms was estimated from the a priori SNR which was in turn estimated with the decision directed approach of Ephraim and Malah [2]. The smoothing parameter α was set to 0.99. A lower limit was also set for the a priori SNR at -25 dB as this was reported to help reducing the amount of musical noise [10].

[1]For a definition of Segmental SNR and PESQ see section 5.2

5.2 Evaluation of the Algorithms

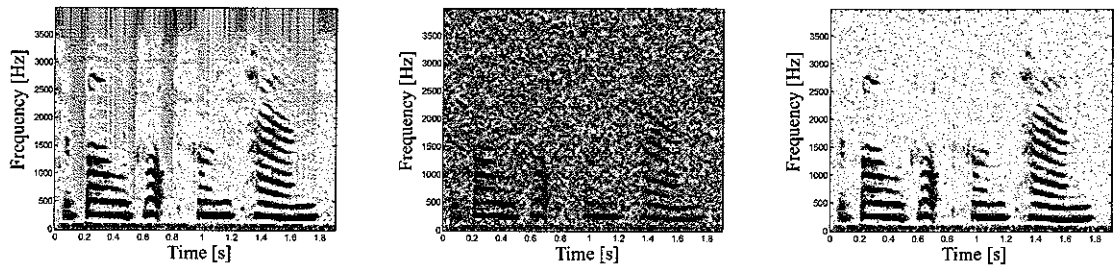
Evaluating the performance of speech enhancing algorithms is not a trivial task. Although a number of evaluation measures has been proposed over the years, no objective measure has been accepted as the 'gold standard'. There is a number of aspects that affect the overall enhanced speech quality, such as intelligibility, naturalness of speech, level and type of residual noise, the assessment of which is difficult to be incorporated into a single measure. Moreover, the subjective character of the above aspects complicates their assessment even further.

Evaluation measures can be divided into two categories: subjective and objective. Subjective measures are based on comparison of the clean and enhanced signals by a group of listeners, who then subjectively rank the quality of the enhanced signals. Objective measures on the other hand, are based on a mathematical model, which may or may not try to emulate a subjective measure. Two objective measures are used in this report: the Segmental Signal to Noise Ratio (SegSNR) and the Perceptual Evaluation of Speech Quality (PESQ).

The SegSNR is an extension of the traditional (or total) SNR and is believed to be more suitable for the evaluation of speech enhancement algorithms. Segmental SNR is calculated by finding the SNR in each speech analysis frame in dB and then averaging across the frames. Analytically it is given by [14]:

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{1}{M}\sum_{j=1}^{M} 10\log_{10}\left[\sum_{n=m_j}^{m_j+N-1}\frac{s^2(n)}{[s(n)-\hat{s}(n)]^2}\right] \tag{5.1}$$

where M is the number of speech frames, m_j is the starting sample of the j^{th} frame, s is the clean and $\hat{s}$ the cleaned speech signal. The motivation for this measure is to emphasize the effect of noise in the low energy speech segments, which are more sensitive to noise than the high energy ones. Indeed, a segment with a very low SNR will weight much more toward the final result in equation 5.1, because of the addition of the logarithms, compared to the total SNR where the square errors would be summed across the entire waveform. A problem that arises often when the Segmental SNR is used, is that the existence of silent frames in the signal can produce large negative SNRs, which are not representative of the enhanced speech quality. This problem however, is sidestepped if the silent frames are



(a) Clean Speech (b) Input SegSNR = 0 dB (c) Input SegSNR = 20 dB

Figure 5.1: Spectrograms of clean and noisy speech at different input SegSNRs.

identified in the clean speech and excluded from the calculation of the Segmental SNR.

PESQ is an perceptual quality measurement for voice quality in telecommunications and has been approved as the International Telecommunication Union (ITU) recommendation P.862. It is designed to predict the Mean Opinion Scores (MOS) from a subjective listening test, returning a score for the degraded audio sample between 1.0 (worst) and 4.5 (best). It has been reported in [15] that the correlation between PESQ scores and those obtained from subjective listening tests was indeed very high.

Throughout the presentation of the results, along with the above two objective measures we will give an informal subjective evaluation of the degraded audio samples. The evaluation will be mainly focused on aspects such as the nature of the residual noise (musical vs. broadband) and the amount of speech distortion, which are not always illustrated in the numerical results of the objective measures. A visual supplement will be provided by spectrograms of a segment of the enhanced speech. To facilitate comparison between different algorithms, all the spectrograms will correspond to the same phrase (*the bowl dropped from his hands*) and will be normalised so that same colors indicate same levels. The spectrograms of the above phrase prior to noise corruption and mixed with two different noise levels are given in figure 5.1 for reference purposes.

A popular visualisation of an algorithm's properties is given by its suppression curves. These are plots of the suppression the algorithm applies (in dB) as a function of some of its input parameters. A number of suppression curves plots will be given for each algorithm and for some key values of the prior density function parameter a, so that their shape for the whole range of a values should be easily inferred. The suppression curves will be shown as a function of the a priori and the a posteriori SNR.

5.3 DFT Domain Results

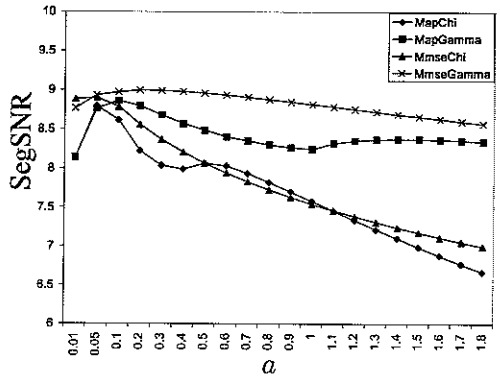
In this section we present the results of the algorithms that work in the DFT domain, i.e. with the real and the imaginary part of the STFT coefficients separately. We will first present the results for a range of different values of a to hopefully gain some insight about the algorithms' behaviour. The results obtained with the fixed optimum values of a and values that are adaptively estimated will be presented in the following.

5.3.1 Results for Different Values of a

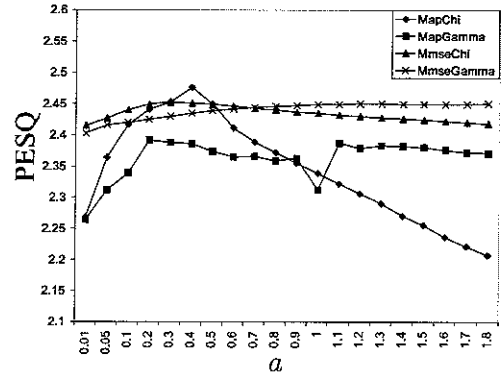
Figure 5.2 shows the SegSNR and PESQ scores for all four algorithms, three different input SegSNRs and a range of a values. Examination of the SegSNR plots reveals that the Gamma speech prior algorithms yield better SegSNR scores. The PESQ plots show that MMSE algorithms perform better but it is not clear whether Chi or Gamma priors are preferable. We proceed with a closer examination of the results.

Comparison of MMSE and MAP Estimators for Gamma Speech Priors The MMSE estimator yields almost uniformly better results for both the SegSNR and PESQ measures. The residual noise is musical for small values of a and becomes slightly more broadband as the value of a increases, although it retains its musical character. The MAP estimator produces musical noise for small values of a, which is concentrated in few frequency bands. It has however, a relatively high amplitude which makes it much more perceivable and possibly more annoying than the musical noise produced by the MMSE estimator. As the value of a approaches 1, the musical noise is reduced but the speech is suppressed significantly and only its strong components are retained. For values of a greater than one, more residual noise is present but its character is now broadband rather than musical, a rather uncommon feature of the algorithms that operate in the DFT domain. This change is also reflected in the MAP-Gamma curves in figure 5.2, which have a local minimum at $a = 1$. The above comments are illustrated in figure 5.3 where the phrase *The bowl dropped from his hands* is corrupted by white Gaussian noise at 0 dB SegSNR and then enhanced with the MAP and MMSE-Gamma algorithms for different values of a.

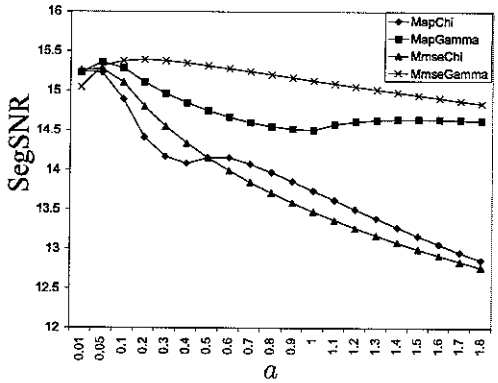
Figure 5.4 shows the suppression curves of the two algorithms for different values of a. For $a = 0.1$ note that as the a posteriori SNR drops the MMSE algorithm applies more



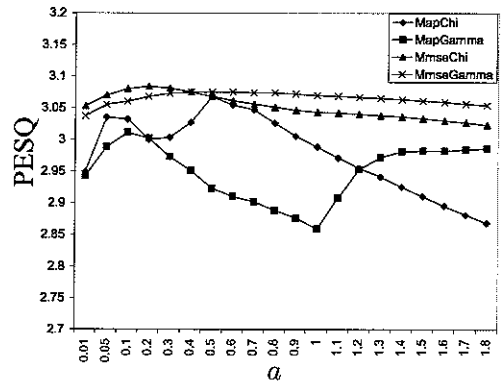
(a) Input SegSNR = 0 dB



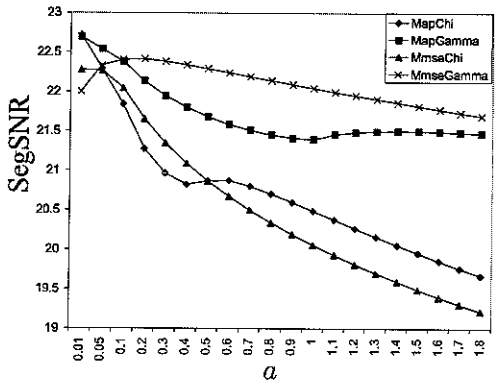
(b) Input SegSNR = 0 dB



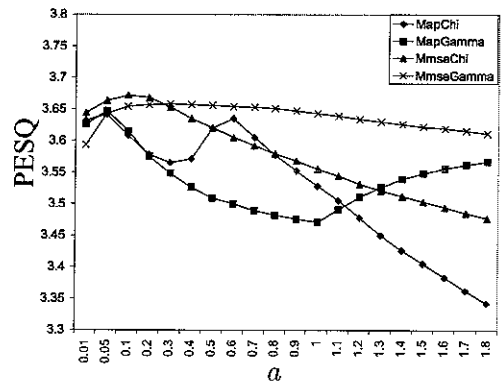
(c) Input SegSNR = 10 dB



(d) Input SegSNR = 10 dB

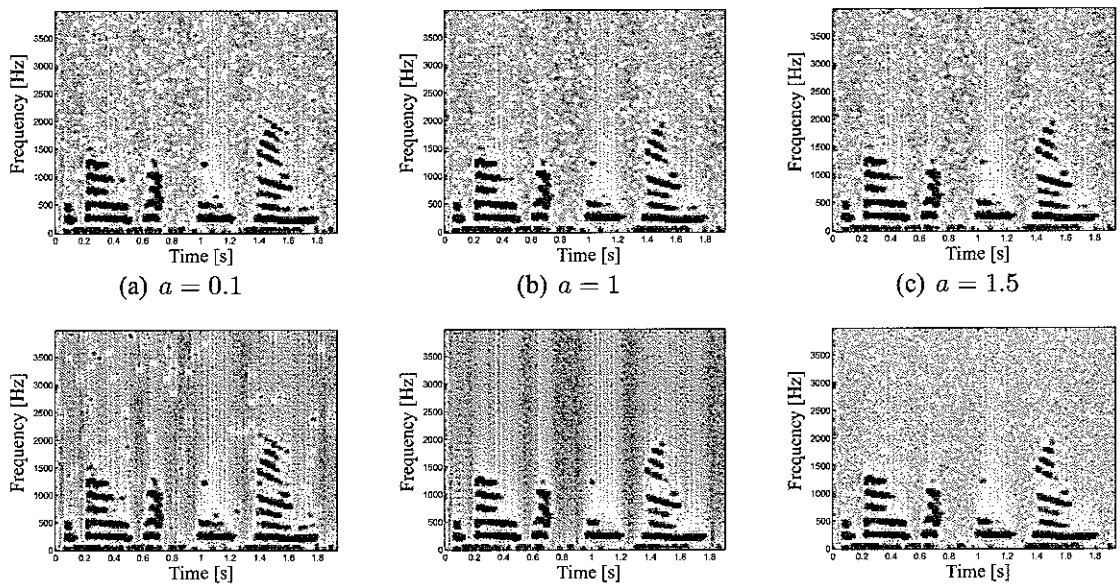


(e) Input SegSNR = 20 dB



(f) Input SegSNR = 20 dB

Figure 5.2: SegSNR and PESQ scores for different values of a, Input SegSNRs and algorithms.

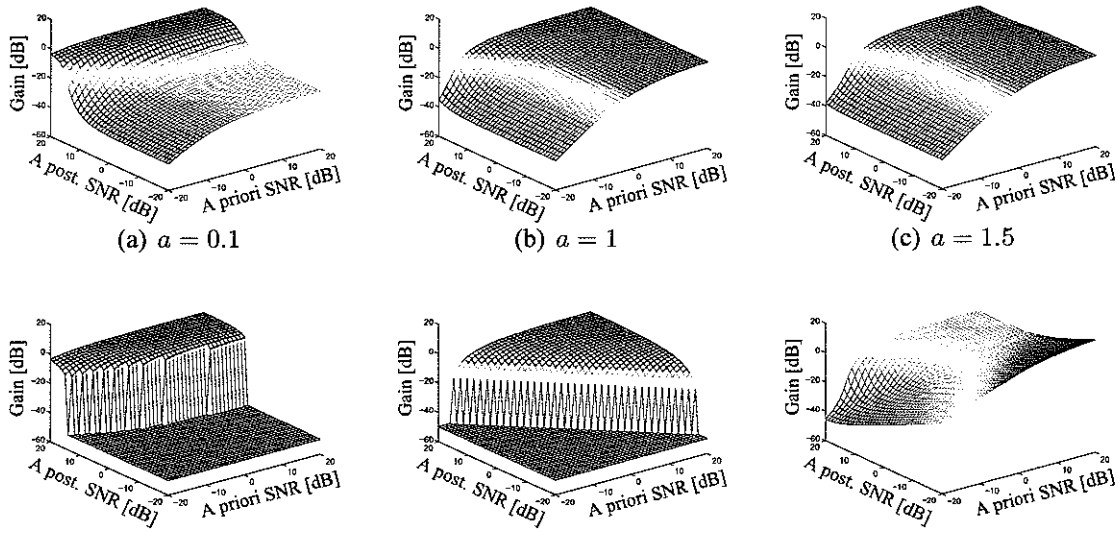


(a) $a = 0.1$ (b) $a = 1$ (c) $a = 1.5$

Figure 5.3: Spectrograms of enhanced speech with the MMSE-Gamma (upper row) and MAP-Gamma (lower row) algorithms for different values of a.

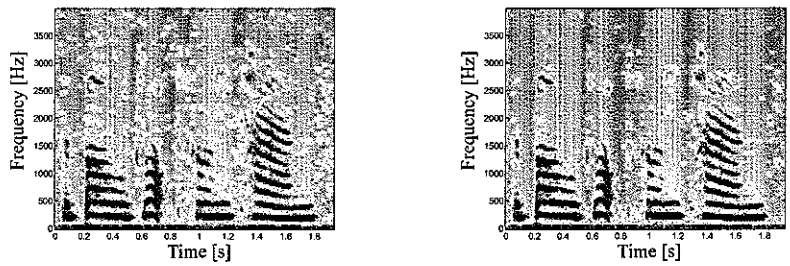
suppression before the 'cut-off' of the MAP algorithm is reached. This results in the MAP enhanced speech exhibiting a few but high amplitude musical noise peaks, while spectral components below the 'cut-off' threshold are heavily suppressed. Another point that needs mentioning is that for high values of a, the MAP estimator presents the 'counter-intuitive' behaviour of increasing the suppression for larger a posteriori SNR values. This property is believed to generate broadband noise instead of musical as it was reported in [10].

From figure 5.2 one might also notice that for small values of a, $(a < 0.1)$ and higher input SegSNRs the MAP algorithm scores are better than those of the MMSE. For an explanation of this phenomenon we must keep in mind that for small values of a the MMSE estimator applies more suppression for relatively high a posteriori SNR values. The application of less suppression by the MAP estimator makes the resulting speech look spectrally closer to the clean speech than the MMSE enhanced speech does. Observe for instance the two spectrograms in figure 5.5 between 0.8 and 1.4 secs in the frequency band between 1.5 and 3 KHz and after 1.8 secs ('s' from hands) and compare them with the original signal in figure 5.1. The MAP algorithm has retained more energy in these regions, which is believed to be the reason for the better results in the objective measures. This excess of energy however is perceived as noise rather than as the original weak spectral components. On the other hand, its elimination by the MMSE algorithm does not



(a) $a = 0.1$ (b) $a = 1$ (c) $a = 1.5$

Figure 5.4: Suppression Curves for MMSE-Gamma (upper row) and MAP-Gamma (lower row) algorithms for different values of a as a function of the a priori and a posteriori SNRs.



(a) MAP Enh. Speech (b) MMSE Enh. Speech

Figure 5.5: Spectrograms of clean and enhanced speech with the MMSE-Gamma and MAP-Gamma algorithms. Input SegSNR was 20 dB and $a = 0.01$.

introduce any significant distortion in speech, while it noticeably reduces the background noise level. Therefore, despite the slightly inferior scores in the objective measures, the MMSE algorithm still delivers a signal with lower background noise, at the expense of a rather negligible distortion.

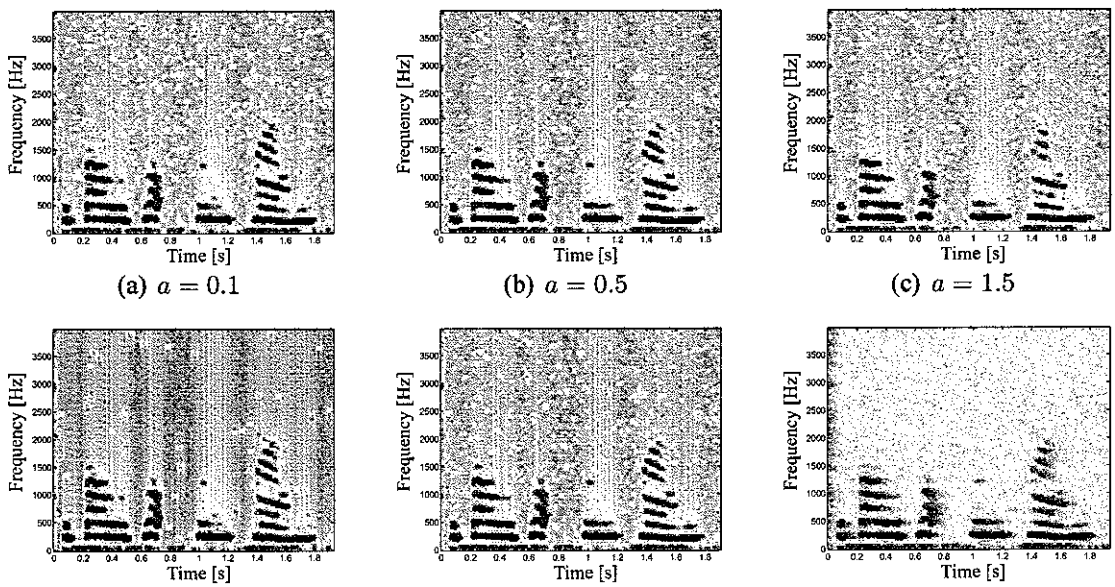
We make a final point, which also holds for all the MAP and MMSE algorithms found in this report, and has to do with the efficiency of the algorithms rather than the quality of the resulting speech. Recall from the theory chapters that the calculation of the MMSE estimators involves the integration of the posterior density, which results in expressions containing special functions (Parabolic Cylinder and Confluent Hypergeometric Function). If the special functions are chosen to be evaluated for each spectrogram sample

separately, MMSE algorithms can turn to be much slower than their MAP counterparts. Also care has to be taken with extreme inputs because it is easy for these functions to create an over- or underflow. The use of asymptotic forms is recommended for such cases. An alternative option could be the use of look-up tables which should increase the speed at the expense of some memory requirements and potentially some approximations errors.

Comparison of MMSE and MAP Estimators for Chi Speech Priors As one might expect, a number of parallels can be drawn between the comparison of the MAP and MMSE estimators when Chi and Gamma speech priors are used. Again the MMSE algorithm delivers almost constantly better PESQ scores, while the SegSNR scores are better for $a < 0.5$. The change in the quality of the enhanced speech with changing values of a is also quite similar to the case of Gamma priors. For small a the MAP algorithm produces a residual musical noise concentrated in few frequency bands, while noise is heavily suppressed elsewhere. As a increases the residual noise becomes broadband but its level also increases. The turning point in the behaviour of the algorithm is now at $a = 0.5$, and the peak in the PESQ score in the vicinity of that value is found because both the musical noise is suppressed and the overall noise level is low. The MMSE algorithm on the other hand produces musical noise for all the examined values of a which turns only slightly more broadband for increasing a. The above comments are illustrated in the spectrograms of figure 5.6.

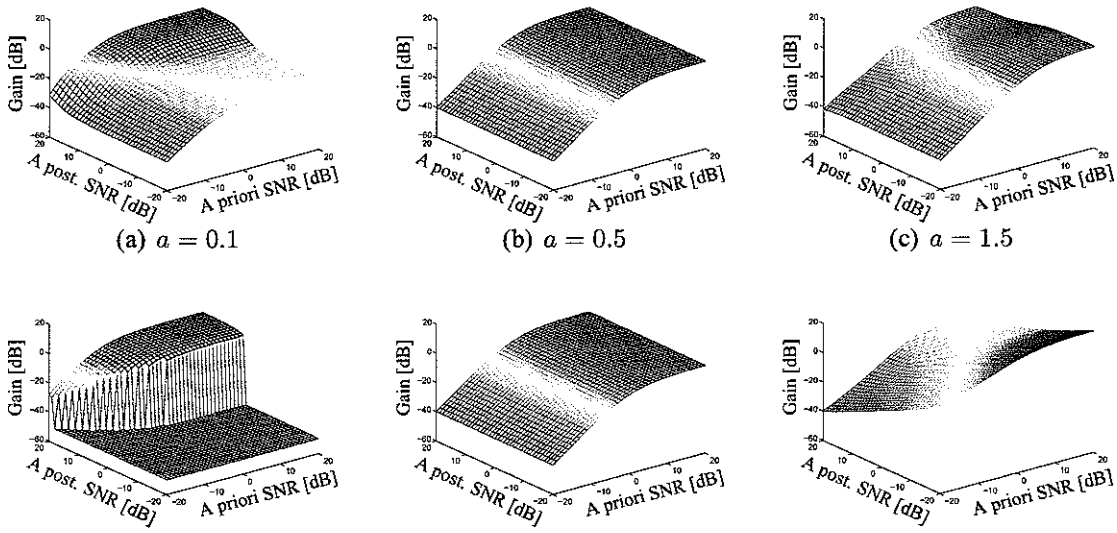
A value of a which is of particular interest is $a = 0.5$. For this value the Chi function yields the Gaussian *pdf* and the MMSE and MAP estimators produce the same estimate, which is the Wiener filter. The suppression curves for $a = 0.5$ are constant with respect to the a posteriori SNR.

Finally, let us mention that the reason why the MAP algorithm yields higher SegSNR scores for high values of a and high input SegSNR is similar to the one developed around figure 5.5. The MAP algorithm applies less suppression and the resulting waveform is spectrally closer to the original speech signal. This energy excess however, is again perceived as noise, making the MMSE enhanced signal sound less noisy. Note also that the PESQ results clearly favor the MMSE enhanced signal.



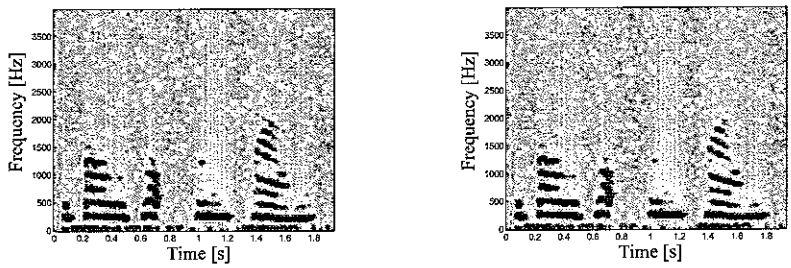
(a) $a = 0.1$ (b) $a = 0.5$ (c) $a = 1.5$

Figure 5.6: Spectrograms of enhanced speech with the MMSE-Chi (upper row) and MAP-Chi (lower row) algorithms for different values of a.



(a) $a = 0.1$ (b) $a = 0.5$ (c) $a = 1.5$

Figure 5.7: Suppression Curves for MMSE-Chi (upper row) and MAP-Chi (lower row) algorithms for different values of a as a function of the a priori and a posteriori SNRs.



(a) MMSE-Gamma (b) MMSE-Chi

Figure 5.8: Spectrograms of enhanced speech with the MMSE Gamma and Chi algorithms. Input SegSNR was 0 dB and $a = 0.2$.

Comparison of the Gamma and Chi Speech Priors Observation of the SegSNR plots in figure 5.2 reveals that the algorithms that use Gamma priors produce speech with the larger SegSNR scores for every value of a. The MMSE-Gamma algorithm also gets PESQ scores which are consistently among the highest. One could therefore argue, that Gamma density functions are more appropriate for modelling the speech DFT coefficients.

One point we have to make however, is that the MMSE-Chi algorithm gets better PESQ scores for small values of a ($a \leq 0.2$). Figure 5.8 shows a sample of speech corrupted by Gaussian White noise at 0 dB input SegSNR and enhanced by the MMSE Gamma and Chi algorithms with $a = 0.2$. Note that the peaks of the musical noise are slightly smoother when the Chi prior is used. This is believed to increase the PESQ score. The speech harmonics however, are somewhat better restored with the Gamma priors, which probably contributes to the higher SegSNR. Let us add here, that the trade off between sharply restored harmonics and smoothness of the residual noise seems to encompass all the algorithms that are examined in this report.

Conclusion As a general remark about the MMSE algorithms we can say that they both produce musical residual noise which becomes slightly more broadband for increasing a. The MAP algorithms yield musical noise for small a, which is concentrated in very few frequency bands, and it is therefore quite distinctive. As a increases the residual noise turns to broadband. This can be useful in some applications where musical noise is unwanted.

As for the suitability of priors, Gamma densities seem to be more appropriate as the resulting speech scores better in the objective measures. The MMSE-Chi algorithm with

small a should also be mentioned here as it produces somewhat better PESQ results, at the expense of a small decrease in SegSNR.

5.3.2 Results for Fixed Optimal Values of a

In this section we present the results for fixed values of a, which are optimal for the given clean speech data. The optimal values of a were found by minimising the Kullback-Leibler (KL) divergence between the data and each speech prior. The definition of the KL divergence (for the discrete case) is:

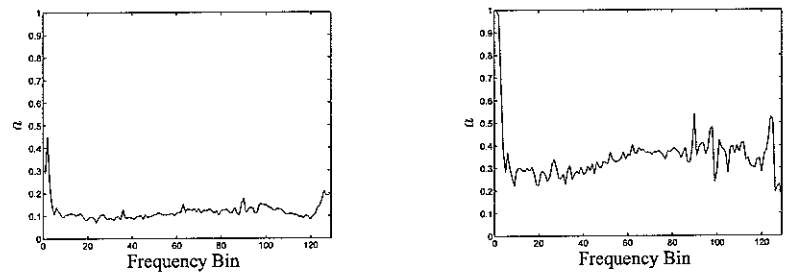
$$\mathrm{KL} = \sum_{n=1}^{N} (p_d(n) - p_s(n)) \ln \left(\frac{p_d(n)}{p_s(n)} \right) \tag{5.2}$$

where $p_d(n)$ is the *pdf* of the data, calculated by their histogram and evaluated on N bins, and $p_s(n)$ is the speech prior evaluated on the same N points.

Two 'optimal' sets of a were found: the first comprised of a separate a for each frequency bin, while the second had a single value of a for all frequency bins, estimated from all the available data. Separate sets of a were found for the real and the imaginary parts of the DFT coefficients, although the results were quite similar. Especially in the case when a single value of a was estimated from the data in all the frequency bins, the results were virtually identical. We proceed by showing the results for each prior, together with the results for the values of a that yielded the best SegSNR and PESQ scores from the previous section for comparison.

Figure 5.9 shows the optimal values of a for each frequency bin and for the two different priors (Chi and Gamma). The results shown were obtained by fitting the real part of the DFT coefficients to the priors. The results from the imaginary parts were quite similar and for this reason are not shown. Notice the similarity of the overall shape of the curves. The values of a for the Gamma priors are constantly larger ($\sim$ 3 times), because for the same value of a the Gamma priors have much longer tails than the Chi.

Tables 5.1 - 5.4 show the results for the four DFT algorithms examined with fixed values of a. The first column indicates the values of a that were used. '$\mathrm{opt}_{\mathrm{ind}}$' indicates the values estimated independently for each frequency bin, and '$\mathrm{opt}_{\mathrm{tot}}$' refers to the case when the value of a was estimated from all the data available in the STFT. This value is



(a) Chi (b) Gamma

Figure 5.9: Optimal values of a for each frequency bin and different speech priors.

shown in the column under the header 'a'. 'Best SNR' refers to the value of a that yielded the best SegSNR score for the respective algorithm and Input SegSNR and similarly for the 'Best PESQ'.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	8.43	2.43	-	14.72	3.03	-	21.74	3.60	-
opt_{tot}	8.73	2.37	0.07	15.08	3.04	0.07	22.09	3.62	0.07
Best SNR	8.79	2.36	0.05	15.25	2.94	0.01	22.71	3.63	0.01
Best PESQ	7.98	2.47	0.40	14.15	3.06	0.50	22.29	3.64	0.05

Table 5.1: Results from the MAP-Chi Algorithm.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	8.67	2.44	-	15.00	3.08	-	21.98	3.67	-
opt_{tot}	8.85	2.43	0.07	15.20	3.07	0.07	22.18	3.66	0.07
Best SNR	8.89	2.42	0.05	15.27	3.07	0.05	22.27	3.64	0.01
Best PESQ	8.36	2.45	0.30	14.79	3.08	0.30	22.04	3.67	0.10

Table 5.2: Results from the MMSE-Chi Algorithm.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	8.54	2.38	-	14.87	2.96	-	21.89	3.54	-
opt_{tot}	8.75	2.39	0.23	15.05	3.00	0.23	22.06	3.56	0.23
Best SNR	8.85	2.33	0.10	15.35	2.98	0.05	22.69	3.62	0.01
Best PESQ	8.79	2.39	0.20	15.28	3.01	0.10	22.54	3.64	0.05

Table 5.3: Results from the MAP-Gamma Algorithm.

Examination of the tables reveals that the 'optimal' values do not produce the highest scores, although they are reasonably close. Possible reasons for this discrepancy could be estimation errors due to the finite number of samples, or the ability of the priors to model the data accurately.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	8.93	2.43	-	15.35	3.07	-	22.36	3.65	-
opt_{tot}	8.99	2.43	0.23	15.38	3.07	0.23	22.40	3.65	0.23
Best SNR	8.99	2.43	0.30	15.39	3.06	0.20	22.41	3.65	0.20
Best PESQ	8.75	2.45	1.20	15.31	3.07	0.50	22.41	3.65	0.20

Table 5.4: Results from the MMSE-Gamma Algorithm.

An important observation that can be also made is that estimation of a for each frequency bin independently does not necessarily give better results than when a is estimated for all frequency bins from all the available data. This could be attributed to the fact that much more data are used in the second case, which may allow for a more accurate estimation. A degree of similarity of the distributions across the frequency spectrum might also be of some help towards this end. Note finally, that the values of a estimated from all the frequency bins is lower than the average of values estimated for each bin separately.

Examination of the best scores in the above tables, shows that MMSE-Gamma algorithm yields the best SegSNR scores, and its PESQ scores are almost identical to those of the MMSE-Chi which are the best overall. The MAP algorithms are somewhat inferior to the respective MMSE, but this should also be balanced with their computational and implementation complexity which is clearly lower.

5.3.3 Results for Adaptively Estimated Values of a

The value of a can be adaptively estimated using equation 3.18 from section 3.2.3 for the Chi priors and equation 3.27 from section 3.3.3 for the Gamma priors. The fourth moment for the noisy speech DFT coefficients $\mathrm{E}[X^4]$ can be found with a first order recursive averaging. If we define an estimator e for $\mathrm{E}[X^4]$ an estimate for the fourth moment can be obtained as:

$$e_n = (1 - \lambda)e_{n-1} + \lambda X_n^4 \tag{5.3}$$

where the subscripts n and $n-1$ indicate the current and previous time frames respectively. λ is a smoothing parameter whose best value was found through simulations. σ_S^2 in equations 3.18 and 3.27 was found by the a priori SNR ξ as $\sigma_S^2 = \xi\,\sigma_N^2$ and by a subsequent smoothing according to equation 5.3, as this was found to improve the performance. Applying lower and higher limits to the allowable values of a was found to be

highly beneficial. A reason for that was that despite their simplicity in implementation, moment methods are sensitive to outliers, and prone to produce erroneously high results in their presence. Bounding the values of a within certain limits circumvents this problem. To sum up, three parameters were found to affect the adaptive estimation of a; these were λ, the lower a limit a_{min} and the upper limit a_{max}. Their values that maximised the algorithms' performance were calculated through simulations and will be presented together with the results.

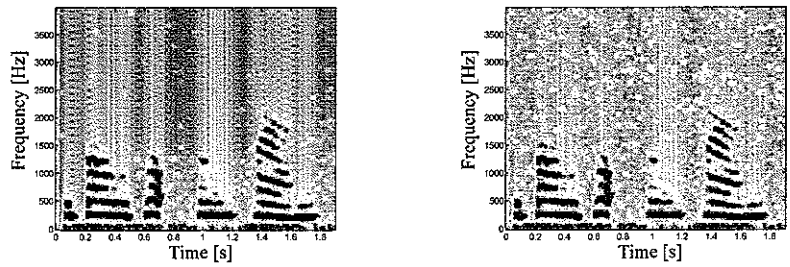
The scheme described for the adaptive estimation of a was found to have different effects when combined with an MMSE or a MAP estimator. For this reason we will present the results for each estimator separately. The parameters that were found to give the best results for the MMSE were $\lambda = 0.005$, $a_{min} = 0.0001$ and $a_{max} = 0.5$. To demonstrate the sensitivity of the algorithm to the above parameters we also state the following observations: Values of λ in the region of [0.001,0.01] were also producing good results. Outside these limits a was fluctuating either very rapidly or very slowly, with a negative effect on the performance. Decreasing a_{min} below 0.0001 did not seem to have any effect while increasing it above 0.001 resulted in higher background noise levels. Finally, the value of a_{max} did not have a major impact, and any value in the [0.1,2] range produced reasonable results.

Table 5.5 shows the results from the above adaptive scheme for the MMSE-Chi and MMSE-Gamma algorithms. Comparison with the results in tables 5.2, 5.4 shows that for high input SegSNR the adaptive scheme produces results as good as the best that could be obtained with fixed values of a. For low input SegSNR however, the results of the adaptive scheme are better.

Input SegSNR	0 dB		10 dB		20 dB	
	SegSNR	PESQ	SegSNR	PESQ	SegSNR	PESQ
MMSE-Chi	8.96	2.5	15.22	3.08	22.26	3.66
MMSE-Gamma	9.07	2.5	15.38	3.08	22.40	3.65

Table 5.5: Results from the MMSE-Chi and MMSE-Gamma Algorithms with the adaptive scheme.

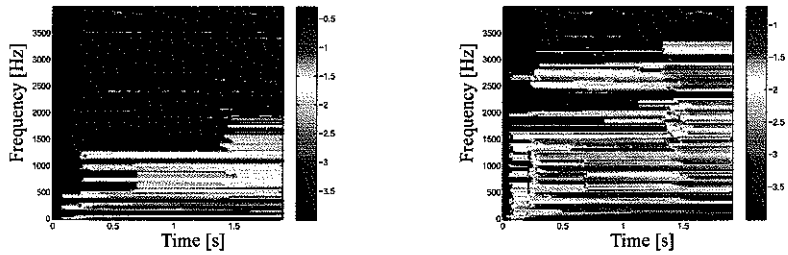
Figure 5.10 shows spectrograms of speech enhanced with the MMSE-Chi algorithm for different input SegSNRs. The value of a was either fixed ($a = 0.07$) or adaptively estimated with the above parameters. Although the spectrograms that correspond to the high input SegSNR are rather similar, those who correspond to the low input SegSNR have



(a) Adaptive, 0 dB (b) Fixed, 0 dB

(c) Adaptive, 20 dB (d) Fixed, 20 dB

Figure 5.10: Spectrograms of enhanced speech with the MMSE-Chi algorithm with adaptive and fixed values of a and different input SegSNRs.

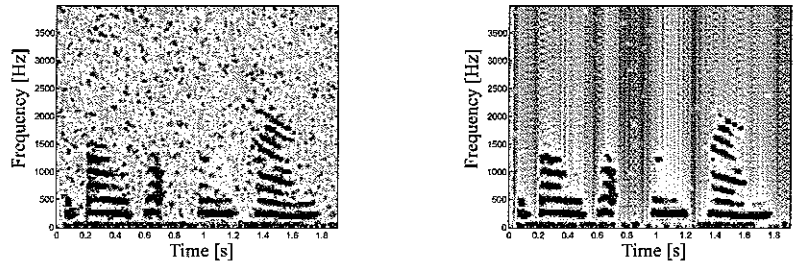


(a) 0 dB (b) 20 dB

Figure 5.11: Values of a corresponding to the spectrograms in figure 5.10.

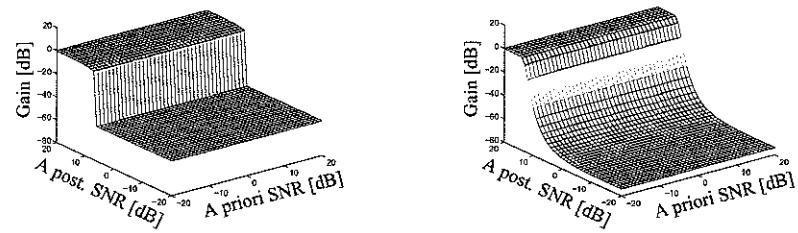
significant differences. The reason is that the adaptive scheme detected that the noise was dominant in the higher frequencies and reduced the value of a significantly, making the prior very narrow. This resulted in the suppression of the background noise in the higher frequencies while the speech components in the lower frequencies were preserved. Had there been some extended silent periods in the processed speech sample, the noise suppression would have occurred across the frequency spectrum. Figure 5.11 shows the values of $\log_{10}(a)$ for every sample of the spectrogram and for the two different input SegSNRs.

The above strategy for estimating a has quite different effects when used with the MAP estimator and the same parameters. The main difference is that when very narrow priors



(a) MAP-Chi (b) MMSE-Chi

Figure 5.12: Output of the MAP-Chi and MMSE-Chi for $a = 0.0001$. Input SegSNR = 0 dB.



(a) MAP-Chi (b) MMSE-Chi

Figure 5.13: Suppression curves for the MAP-Chi and MMSE-Chi for $a = 0.0001$.

are combined with the MAP estimator ($a << 0.1$) a significant amount of musical noise is produced. On the contrary, when narrow priors are used with the MMSE estimator both the speech and noise spectral components are heavily suppressed. See for example the two spectrograms in figure 5.12, where noisy speech with 0 dB input SegSNR is enhanced with the MAP-Chi and MMSE-Chi algorithms. The value of a was 0.0001. Figure 5.13 shows the suppression curves of the respective algorithms and priors. Note that as the a posteriori SNR drops the MMSE algorithm applies more suppression.

On the whole, the simulation results showed that MAP algorithms did not benefit by the use of the adaptive scheme. Numerous simulations were performed with the MAP algorithms and the adaptive scheme for a grid of values of the smoothing and limiting parameters λ, a_{min} and a_{max}. The results did not exceed the best results that were obtained with the fixed values of a. Tables 5.6, 5.7 show the results of the MAP-Gamma algorithm with the adaptive scheme. A number of different λ and a_{min} values were used while a_{max} was kept to 2, as varying its value did not seem to affect the results. Input SegSNR was 0 and 20 dB.

Examination of the above tables shows that the best results occur for $\lambda = 0.0001$. Note

	SegSNR					PESQ				
a_{min}	0.001	0.01	0.1	0.2	1.1	0.001	0.01	0.1	0.2	1.1
λ										
0.0001	7.55	8.15	8.85	8.79	8.31	2.15	2.24	2.34	2.39	2.38
0.001	7.76	8.20	8.70	8.75	8.31	2.13	2.20	2.33	2.38	2.38
0.01	7.93	8.14	8.57	8.64	8.31	2.12	2.18	2.31	2.38	2.38
0.1	7.60	8.03	8.78	8.82	8.37	2.13	2.19	2.33	2.39	2.39

Table 5.6: Results from the MAP-Gamma Algorithm for different parameter values of the adaptive scheme. Input SegSNR was 0 dB.

	SegSNR					PESQ				
a_{min}	0.001	0.01	0.05	0.1	1.1	0.001	0.01	0.05	0.1	1.1
λ										
0.0001	22.58	22.60	22.52	22.37	21.45	3.63	3.64	3.64	3.61	3.49
0.001	22.27	22.28	22.29	22.26	21.45	3.59	3.60	3.60	3.59	3.49
0.01	22.08	22.09	22.10	22.08	21.45	3.57	3.57	3.57	3.57	3.49
0.1	22.17	22.19	22.20	22.16	21.49	3.55	3.56	3.57	3.57	3.48

Table 5.7: Results from the MAP-Gamma Algorithm for different parameter values of the adaptive scheme. Input SegSNR was 20 dB.

also that the values of a_{min} for which the adaptive scheme produces the best results coincide with the fixed values of a that give the best results in table 5.3. Examination of the values of a returned by the adaptive scheme for every sample of the STFT showed that they were fairly constant and close to a_{min}. Thus, the adaptive scheme produces the best results when the estimated values of a are the same as the fixed values from table 5.3. The adaptive scheme however, was forced to converge to these values through the choice of a_{min}, rather than freely selecting them. We argue therefore, that the adaptive scheme does not offer a benefit to the MAP-Gamma algorithm. The MAP-Chi algorithm showed the same behaviour when combined with the adaptive scheme.

5.4 Amplitude Domain Results

In this section we present the results from the algorithms that operate on the amplitude of the DFT coefficients in the STFT domain. We develop our presentation again by presenting the algorithms' behaviour for different fixed values of a and then the results obtained with fixed optimal values and the adaptive estimation scheme.

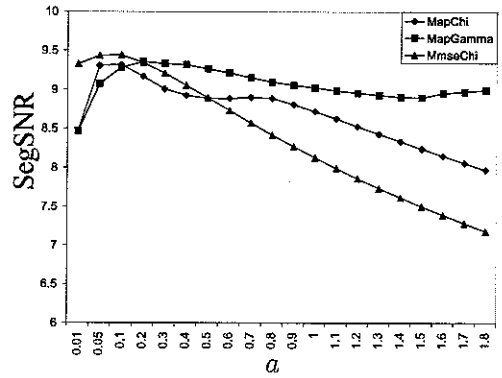
5.4.1 Results for Different Values of a

Figure 5.14 shows the SegSNR and PESQ scores for the different algorithms for a range of input SegSNR and a values. If we observe the SegSNR plots we note that the MAP-Gamma algorithm delivers high scores for the whole range of a, which are constantly higher than those of the MAP-Chi algorithm. The MMSE-Chi algorithm produces some of the best SegSNR scores for small values of a, but its performance deteriorates rapidly as a increases. Examination of the PESQ plots reveals that the MAP-Gamma scores are among the lowest. The scores for the MAP-Chi algorithm vary with the value of a but they seem to reach a peak when a is in the range of 0.6-0.8. Finally, the MMSE-Chi scores drop rapidly as a increases, but are again among the best for small values of a. Let us now examine the behaviour of each algorithm analytically.

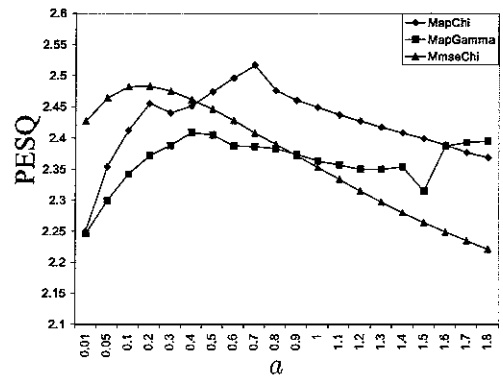
Comparison of the MMSE and MAP Estimators for Chi Speech Priors For small values of a the MMSE-Chi algorithm leaves some residual noise, which although it has some musical character it is much more broadband than the residual noise of the MAP-Chi algorithm for the same value of a. The noise level is also higher, as it can be verified by inspection of the spectrograms in figure 5.15, although this might contribute to the higher PESQ values as the speech sounds a little more natural. As a increases the residual noise becomes more broadband, although its level is relatively high even for moderate values of a, which explains the rapid drop of SegSNR and PESQ values. A special case of this algorithm is for $a = 1$, when the Chi function simplifies to the Rayleigh *pdf* and the resulting algorithm is the well-known Ephraim-Malah MMSE-STSA [2].

The MAP-Chi algorithm's behaviour for different values of a resembles the behaviour of the MAP algorithms examined in section 5.3. For small values of a the residual noise has musical character and is concentrated at few frequency bands, which makes it quite distinctive. As a increases, the musical noise is suppressed but some weak speech spectral components are suppressed as well. For values of a larger than 0.75, the residual noise becomes again broadband with the amount of suppression dropping for increasing a. The peak in the PESQ measure in the vicinity of $a = 0.75$ is found because the musical noise peaks are suppressed and the background noise level is low.

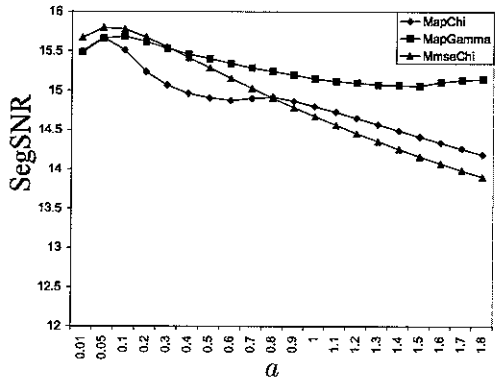
Figure 5.16 shows the suppression curves for some characteristic values of a. A signifi-



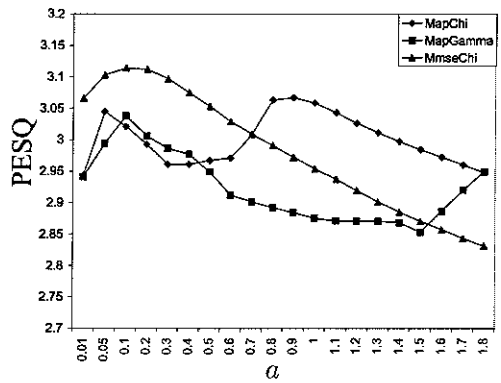
(a) Input SegSNR = 0 dB



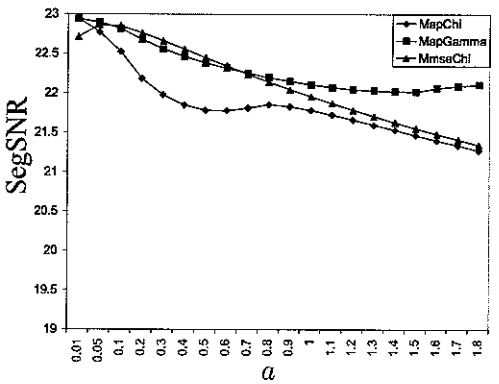
(b) Input SegSNR = 0 dB



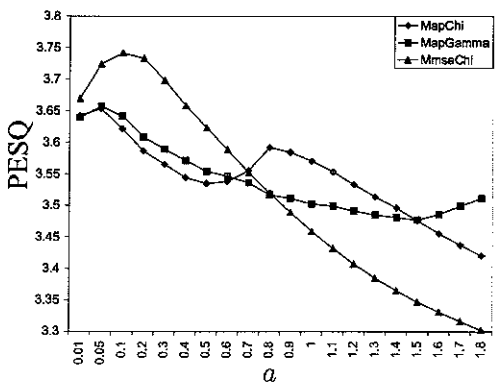
(c) Input SegSNR = 10 dB



(d) Input SegSNR = 10 dB

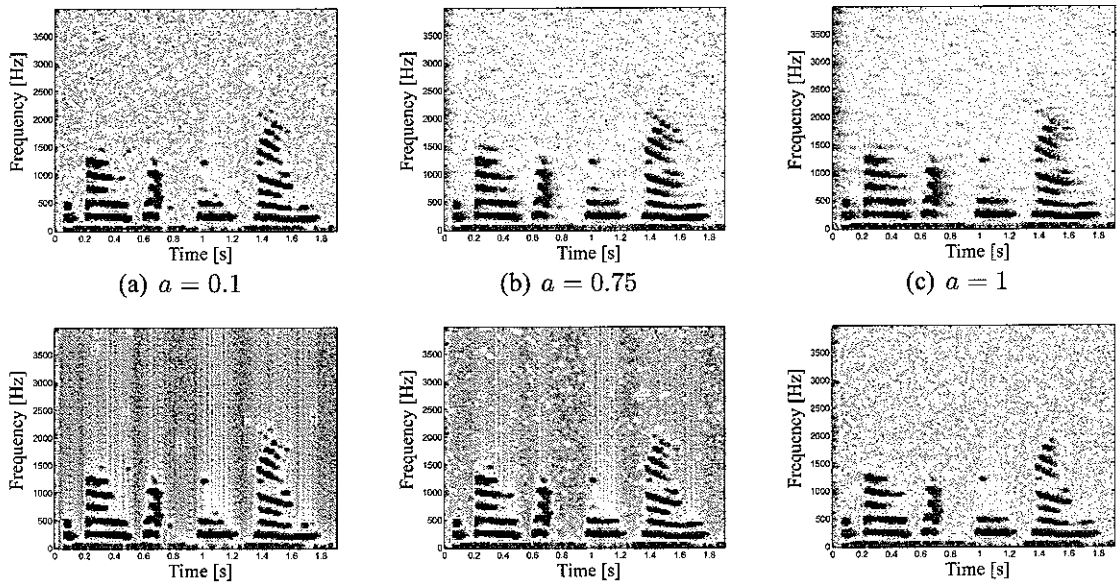


(e) Input SegSNR = 20 dB



(f) Input SegSNR = 20 dB

Figure 5.14: SegSNR and PESQ scores for different values of a, Input SegSNRs and algorithms.



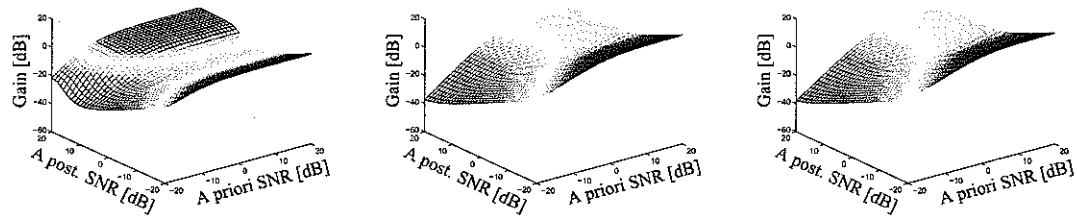
(a) $a = 0.1$ (b) $a = 0.75$ (c) $a = 1$

Figure 5.15: Spectrograms of enhanced speech with the MMSE-Chi (upper row) and MAP-Chi (lower row) algorithms for different values of a. Input SegSNR was 0 dB.

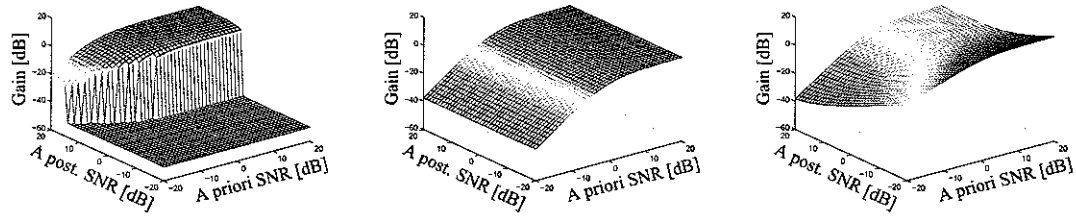
cant point is found for $a = 0.75$ where the curves are non decreasing as the a posteriori SNR drops.

The MAP-Gamma Algorithm The MAP-Gamma algorithm bears again a number of similarities with the MAP-Chi. For very small values of a the residual noise has a strong musical nature. Inspection of figures 5.15(a) and 5.17(a) reveals that for $a = 0.1$ the Gamma priors produce musical noise of higher amplitude and in more frequency bands. This should be attributed to the spikier form of the Gamma prior compared to the Chi for the same value of a. As a increases the musical noise is suppressed but only the strong speech spectral components are retained. The value of a after which the residual noise becomes broadband is now 1.5, while for the MAP-Chi algorithm was 0.75. This also interprets the change of the SegSNR and PESQ curves at $a = 1.5$.

Figure 5.18 shows the suppression curves for some characteristic values of a. Note the similarity of these curves with those of the MAP-Gamma-DFT algorithm in figure 5.4. Compare also the curves in figure 5.18 with those in 5.16 and note that the MAP-Gamma algorithm applies less suppression for high a posteriori and low a priori SNR. This contributes to the higher levels of musical noise, especially for small a.

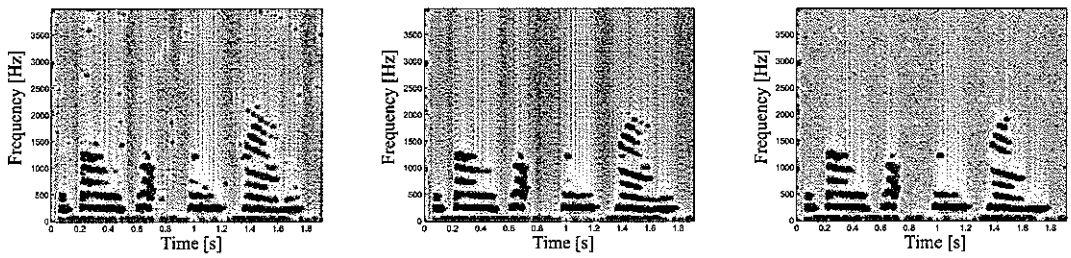


(a) $a = 0.1$ (b) $a = 0.75$ (c) $a = 1$



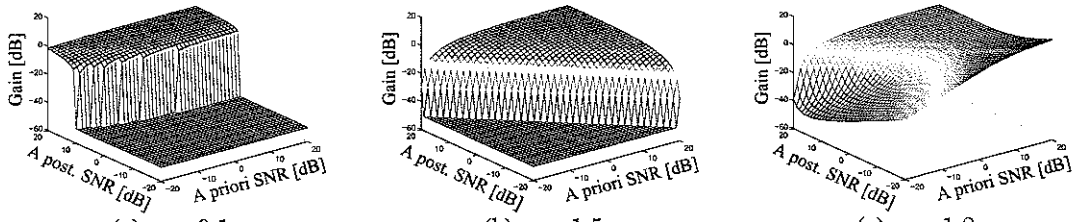
(d) $a = 0.1$ (e) $a = 0.75$ (f) $a = 1$

Figure 5.16: Suppression Curves for MMSE-Chi (upper row) and MAP-Chi (lower row) algorithms for different values of a as a function of the a priori and a posteriori SNRs.



(a) $a = 0.1$ (b) $a = 1.5$ (c) $a = 1.8$

Figure 5.17: Spectrograms of enhanced speech with the MAP-Gamma algorithm for different a. Input SegSNR was 0 dB.



(a) $a = 0.1$ (b) $a = 1.5$ (c) $a = 1.8$

Figure 5.18: Suppression Curves of the MAP-Gamma algorithm for different a.

Conclusion For small values of a both MAP algorithms produce musical noise, whose concentration in few frequency bands make it quite distinctive. Its levels are higher for the MAP-Gamma algorithm. The residual noise of the MMSE-Chi algorithm for small a is again of musical nature but much more broadband compared with that of the MAP algorithms. Its overall level is also much higher.

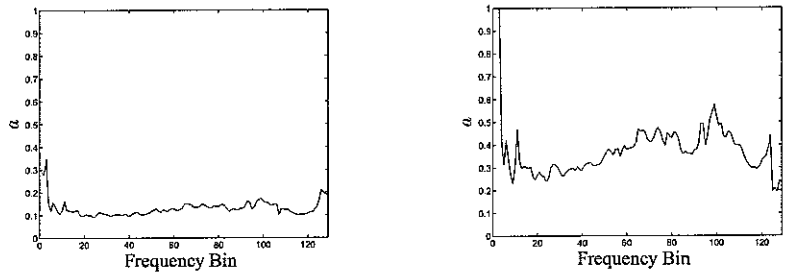
As a increases, the residual noise of the MMSE-Chi algorithm becomes even more broadband, but its overall level increases as well, causing a rapid drop in the objective measures. The musical noise peaks of the MAP-Chi algorithm are suppressed with increasing a, but so are the weak speech spectral components. This causes the PESQ drop until $a \sim 0.7$. For $a > 0.7$ the residual noise becomes broadband, which shows in the change of the SegSNR and PESQ curves. A similar behaviour is encountered by the MAP-Gamma algorithm, although the turning point is now at $a = 1.5$.

5.4.2 Results for Fixed Optimal Values of a

Optimal values for the a parameter of the prior densities were calculated for the amplitude of the DFT coefficients. The optimal values for a were found by minimising the KL distance between the amplitude of the coefficients and the prior densities. Two sets were found again, one with values of a for each frequency bin separately and one with a single value for the whole STFT using the data from all frequency bins. The first set of values is shown in figure 5.19. Note the similarity of the values compared to those shown in figure 5.9.

Results for the three amplitude algorithms are summarised in tables 5.8 - 5.10, which have the same format as those in section 5.3.2. We remind the reader that 'opt_{ind}' indicates the results obtained with values of a estimated independently for each frequency bin, and 'opt_{tot}' refers to the case when the a value was estimated from all the data available in the spectrogram. 'Best SNR' refers to the value of a that yielded the best SegSNR score for the respective algorithm and Input SegSNR and similarly for the 'Best PESQ'.

The conclusions that can be drawn by examining the above tables are similar to those of section 5.3.2. Firstly, the optimal set of a values does not necessarily produce the best results, although they are definitely close. Secondly, estimating a value of a for each frequency bin separately seems to produce inferior results compared to using a single



(a) Chi (b) Gamma

Figure 5.19: Optimal values of a for each frequency bin and different speech priors.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	9.19	2.42	-	15.40	3.01	-	22.42	3.61	-
opt_{tot}	9.34	2.39	0.07	15.61	3.04	0.07	22.66	3.63	0.07
Best SNR	9.31	2.41	0.10	15.66	3.04	0.05	22.94	3.64	0.01
Best PESQ	8.89	2.51	0.70	14.86	3.06	0.90	22.76	3.65	0.05

Table 5.8: Results from the MAP-Chi Algorithm.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	9.44	2.47	-	15.76	3.11	-	22.82	3.74	-
opt_{tot}	9.44	2.47	0.07	15.79	3.11	0.07	22.86	3.73	0.07
Best SNR	9.43	2.48	0.10	15.79	3.10	0.05	22.86	3.72	0.05
Best PESQ	9.43	2.48	0.10	15.77	3.11	0.10	22.84	3.74	0.10

Table 5.9: Results from the MMSE-Chi Algorithm.

Input SegSNR	0 dB			10 dB			20 dB		
	SegSNR	PESQ	a	SegSNR	PESQ	a	SegSNR	PESQ	a
opt_{ind}	9.2	2.41	-	15.45	2.98	-	22.48	3.57	-
opt_{tot}	9.34	2.37	0.23	15.58	2.99	0.23	22.63	3.60	0.23
Best SNR	9.35	2.37	0.20	15.68	3.03	0.10	22.93	3.63	0.01
Best PESQ	9.31	2.40	0.40	15.68	3.03	0.10	22.89	3.65	0.05

Table 5.10: Results from the MAP-Gamma Algorithm.

value of a for all the STFT, calculated by the data available in all the frequency bins.

A comparison between the three algorithms reveals that MMSE-Chi algorithm yields better SegSNR and PESQ scores. Note also that the scores obtained with the optimal a values (case 'opt_{tot}') are either the best or a compromise between the 'Best SNR' and the 'Best PESQ'. The MAP algorithms seem to yield similar results, with the MAP-Gamma being marginally better in the SegSNR and the MAP-Chi in PESQ results, especially for low input SegSNRs.

5.4.3 Results for Adaptively Estimated Values of a

The parameter a can be adaptively estimated using equations 4.14 and 4.20 in sections 4.2.3 and 4.2.3 respectively. The fourth moment $\mathrm{E}[R^4]$ was estimated via an estimator e and according to equation

$$e_n = (1 - \lambda)e_{n-1} + \lambda R_n^4 \tag{5.4}$$

where n and $n - 1$ denote the current and the previous time frames respectively. The speech variance σ_A^2 was estimated by the a priori SNR, and was also smoothed with the parameter λ, according to the above equation. The permissible values of a had upper and lower limits given by a_{min} and a_{max} respectively.

The adaptive scheme had different effects when combined with the MMSE or the MAP estimator, much like in section 5.3.3. MAP algorithms did not seem to benefit, producing a lot of musical noise for very narrow priors ($a << 0.1$). The best results from the combination of the adaptive scheme and the MAP algorithms occurred when the adaptive scheme was forced to pick the values of a that yielded the best results, without being able to select them on its own. On the other hand a combination of the adaptive scheme with the MMSE-Chi algorithm produced results that were as good as the best found by fixed and preselected a values, and for low input SegSNR were even better. The parameters that produced the best results for the MMSE-Chi algorithm were $a_{min} = 0.0001$, $a_{max} = 0.1$ and $\lambda = 0.005$. The results obtained with the above values are shown in table 5.11.

Input SegSNR	0 dB		10 dB		20 dB	
	SegSNR	PESQ	SegSNR	PESQ	SegSNR	PESQ
MMSE-Chi	9.54	2.52	15.77	3.12	22.84	3.71

Table 5.11: Results from the MMSE-Chi Algorithm with the adaptive scheme.

CHAPTER 6 CONCLUSION

A family of Bayesian algorithms for speech enhancement was presented. The algorithms that comprised it can be divided according to the following criteria: the STFT feature estimated, the Bayesian estimator applied and the speech prior used. The STFT features were the DFT coefficients and their amplitude. The Bayesian estimators applied were the MMSE and the MAP and the speech priors considered were the Gamma and the Chi density functions. A key point in the selection of the prior was the parameter a, whose value influenced the shape of the prior and the performance of the algorithm significantly. We conclude this report by summing up the major findings from the algorithms' evaluation.

Focusing on the algorithms that process the real and imaginary parts independently we note that a common feature is that they produce musical noise for small values of a, which turns more broadband as a increases. The musical noise is quite distinctive for the MAP algorithms and small a, while for large a it loses its musical character almost completely. This effect is less dramatic for the MMSE algorithms. Small values of a also preserve better the weaker speech spectral components.

The MMSE algorithms generally produce better results, according to our objective measures, than their MAP counterparts. Using them in combination with the adaptive scheme for the estimation of a yields results which are as good as the best that can be obtained with fixed a. This value of a might be unknown so the adaptive scheme offers a way of estimating it from the noisy samples. For low input SegSNR the adaptive scheme might even produce results better than those obtained with any fixed value of a for the whole STFT.

The algorithm with the best overall performance is the MMSE-Gamma, because it produces some of the best SegSNR and PESQ scores for a large range of a values. The MMSE-Chi algorithm yields very good PESQ results for $a \sim 0.1$ at the expense of a smaller SegSNR score. Compared to the MMSE-Gamma acoustically, it produces a little less annoying musical noise but some weak speech components are not recovered. The MAP algorithms are easier to implement alternatives to their MMSE counterparts but the

resulting speech is somewhat inferior. Good SegSNR scores are obtained for $a \sim 0.1$, although a significant amount of musical noise is present. For more broadband residual noise and better PESQ results $a \sim 1.8$ and $a \sim 0.5$ can be selected for the MAP-Gamma and MAP-Chi respectively. For these values however, the weakest speech components are not preserved.

As far as the algorithms that work with the amplitude of the STFT coefficients are concerned, again small values of a create musical noise which turns to broadband as a increases. This effect is more pronounced in the MAP algorithms.

The MAP-Gamma algorithm produces better SegSNR results compared to the MAP-Chi, while the latter produces better PESQ results. The MAP-Gamma produces more musical noise while preserving better the weakest speech components, while MAP-Chi although it suppresses some weak speech components, produces more broadband background noise, which possibly explains the increase in the PESQ scores.

The MMSE-Chi algorithm generally produces better results in the objective measures than the two MAP algorithms. The value of a that gives the best results is around 0.1. It preserves the weak speech components while the background noise, although it has some musical character it is much more broadband than that of the MAP algorithms for the same values of a. Its overall level is higher compared to that of the MAP algorithms, but the use of the adaptive scheme for the estimation of a can significantly reduce it in the frequency bands where it is dominant.

A comparison between the algorithms that work with the real and imaginary parts and those who work with the amplitude reveals that the latter produce better results (see figures 5.2 and 5.14). Additionally, the algorithms that work with the amplitude have only half the data to process and are therefore faster. A theoretical justification for halving the data is that the optimal estimate for the phase is the noisy phase itself [2]. Hence, it suffices to estimate the amplitude only.

APPENDIX A DERIVATION OF THE ESTIMATORS

A.1 Derivation of the Amplitude Posterior Density

If we denote by $\mathbf{S}_k$, and $\mathbf{X}_k$ the clean and noisy speech DFT coefficients in the k^{th} frequency bin we then have:

$$p(\mathbf{X}_k|\mathbf{S}_k) = p_N(\mathbf{X}_k - \mathbf{S}_k) \tag{A.1}$$

where p_N is the *pdf* of the noise DFT coefficients. Assuming that these are Gaussian and independent random variables with zero mean and variance σ_N^2 eq. A.1 can be written as:

$$p(\mathbf{X}|\mathbf{S}) \triangleq p(X_r, X_i|S_r, S_i) = \frac{1}{2\pi\sigma_N^2} \exp\left[-\frac{(X_r - S_r)^2 + (X_i - S_i)^2}{2\sigma_N^2}\right] \tag{A.2}$$

where X_r and X_i denote the real and the imaginary part of $\mathbf{X}$ and similarly for $\mathbf{S}$. The subscripts k were also dropped as the procedure holds for all frequency bins independently. Our goal is to find $p(R, \psi|A, \phi)$ when we know $p(\mathbf{X}|\mathbf{S})$ or equivalently $p(X_r, X_i|S_r, S_i)$. If we define by $D_{R\psi}$ the slice of a circle of radius R_0 and angle ψ_0 centered at zero on the plane X_r, X_i, the probability mass that it encloses can be written as:

$$P_{R,\psi|A,\phi}(R_0, \psi_0|A, \phi) = \iint_{D_{R\psi}} p(X_r, X_i|S_r, S_i)\, dX_r\, dX_i \tag{A.3}$$

where $P_{R,\psi|A,\phi}(R_0, \psi_0|A, \phi)$ is the probability *distribution* function of R and ψ given A and ϕ, or in other words, the probability that $R \leq R_0$ and $\psi \leq \psi_0$ given A and ϕ. If we change the cartesian to polar coordinates in the integral in eq. A.3 (i.e. $X_r = r\cos\omega$, $X_i = r\sin\omega$ and $dX_r\, dX_i = r dr d\omega$) and express S_r, S_i in their polar form A, ϕ we get:

$$P_{R,\psi|A,\phi}(R_0, \psi_0|A, \phi) = \int_0^{R_0} \int_0^{\psi_0} p(r, \omega|A, \phi) r\, dr\, d\omega \tag{A.4}$$

Substituting the expression for $p(X_r, X_i|S_r, S_i)$ from eq. A.2 we have:

$$P_{R,\psi|A,\phi}(R_0, \psi_0|A, \phi) =$$

$$\frac{1}{2\pi\sigma_N^2} \int_0^{R_0} \int_0^{\psi_0} \exp\left[-\frac{(r\cos\omega - A\cos\phi)^2 + (r\sin\omega - r\sin\phi)^2}{2\sigma_N^2}\right] r\, dr\, d\omega =$$

$$\frac{1}{2\pi\sigma_N^2} \int_0^{R_0} \int_0^{\psi_0} \exp\left[-\frac{r^2 + A^2 - 2rA\cos(\omega - \phi)}{2\sigma_N^2}\right] r\, dr\, d\omega \tag{A.5}$$

The probability density function of R and ψ given A and ϕ is easily obtained by differentiating the distribution function with respect to R_0 and ψ_0.

$$p_{R,\psi|A,\phi}(R_0, \psi_0|A, \phi) = \frac{\partial^2}{\partial R_0\, \partial\psi_0} P_{R,\psi|A,\phi}(R_0, \psi_0|A, \phi) =$$

$$\frac{R_0}{2\pi\sigma_N^2} \exp\left[-\frac{R_0^2 + A^2 - 2R_0A\cos(\psi_0 - \phi)}{2\sigma_N^2}\right] \tag{A.6}$$

Finally, by denoting R_0 and ψ_0 with R and ψ we have:

$$p(R, \psi|A, \phi) = \frac{R}{2\pi\sigma_N^2} \exp\left[-\frac{R^2 + A^2 - 2RA\cos(\psi - \phi)}{2\sigma_N^2}\right] \tag{A.7}$$

About the Notation of Distribution and Density Functions The formal notation of the probability distribution and density functions requires a subscript and an argument i.e. $P_x(x_0)$ or $p_x(x_0)$. The subscript denotes the random variable the function refers to, while the argument is a mere number (i.e. $x_0 = 5$). $P_x(x_0)$ for example, denotes then the probability that the r.v. x is less or equal than x_0. However, when there is no fear of ambiguity the subscript is dropped and the argument defines both the independent variable of the function (input) and the random variable.

A.2 Derivation of the MMSE-Chi-DFT Estimator

Substitution of eqs. 3.7 and 3.5 in eq. 3.8 yields:

$$\hat{S} = \frac{\displaystyle\int_{-\infty}^{\infty} S \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{|S|^{2a-1}}{\theta^a \Gamma(a)} \exp\left[-\frac{S^2}{\theta}\right] dS}{\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{|S|^{2a-1}}{\theta^a \Gamma(a)} \exp\left[-\frac{S^2}{\theta}\right] dS} \tag{A.8}$$

The numerator can be written as:

$$\begin{aligned} \text{num} &= \int_{-\infty}^{0} S \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{(-S)^{2a-1}}{\theta^a \Gamma(a)} \exp\left[-\frac{S^2}{\theta}\right] dS \\ &+ \int_{0}^{\infty} S \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{S^{2a-1}}{\theta^a \Gamma(a)} \exp\left[-\frac{S^2}{\theta}\right] dS \end{aligned}$$

Making the substitution $S = -S$ in the first integral we have:

$$\begin{aligned} \text{num} &= \int_{0}^{\infty} -S \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X+S)^2}{2\sigma_N^2}\right] \frac{S^{2a-1}}{\theta^a \Gamma(a)} \exp\left[-\frac{S^2}{\theta}\right] dS \\ &+ \int_{0}^{\infty} S \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{S^{2a-1}}{\theta^a \Gamma(a)} \exp\left[-\frac{S^2}{\theta}\right] dS \end{aligned}$$

Expanding the exponentials and taking common factors:

$$\begin{aligned} \text{num} &= \frac{\exp\left[-\frac{X^2}{2\sigma_N^2}\right]}{\sqrt{2\pi\sigma_N^2}\,\theta^a\,\Gamma(a)} \left[-\int_{0}^{\infty} S^{2a} \exp\left[-S^2\left(\frac{1}{2\sigma_N^2}+\frac{1}{\theta}\right) - S\frac{X}{\sigma_N^2}\right] dS \right. \\ &\left. + \int_{0}^{\infty} S^{2a} \exp\left[-S^2\left(\frac{1}{2\sigma_N^2}+\frac{1}{\theta}\right) - S\frac{X}{\sigma_N^2}\right] dS \right] \end{aligned} \tag{A.9}$$

The above integrals can be solved with equation 3.462.1 found in [7], which is stated below.

$$\int_{0}^{\infty} x^{\nu-1} \exp[-\beta x^2 - \gamma x]\, dx = (2\beta)^{-\nu/2}\, \Gamma(\nu) \exp\left[\frac{\gamma^2}{8\beta}\right] D_{-\nu}\left(\frac{\gamma}{\sqrt{2\beta}}\right) \tag{A.10}$$

where $D_\nu(z)$ is the Parabolic Cylinder Function (eq. 9.240, [7]).

Solving the integrals in eq. A.9 according to eq. A.10 we have:

$$\text{num} = \frac{\exp\left[-\frac{X^2}{2\sigma_N^2}\right]}{\sqrt{2\pi\sigma_N^2}\,\theta^a\,\Gamma(a)}\left(\frac{1}{\sigma_N^2}+\frac{2}{\theta}\right)^{-(2a+1)/2}\Gamma(2a+1)\,\exp\left[\frac{\left(\frac{X}{\sigma_N^2}\right)^2}{8\left(\frac{1}{2\sigma_N^2}+\frac{1}{\theta}\right)}\right]$$
$$\left[-D_{-2a-1},\left(\frac{\frac{X}{\sigma_N^2}}{\sqrt{\frac{1}{\sigma_N^2}+\frac{2}{\theta}}}\right)+D_{-2a-1},\left(\frac{-\frac{X}{\sigma_N^2}}{\sqrt{\frac{1}{\sigma_N^2}+\frac{2}{\theta}}}\right)\right] \qquad \text{(A.11)}$$

Performing the same steps on the denominator of eq. A.8 we get:

$$\text{den} = \frac{\exp\left[-\frac{X^2}{2\sigma_N^2}\right]}{\sqrt{2\pi\sigma_N^2}\,\theta^a\,\Gamma(a)}\left(\frac{1}{\sigma_N^2}+\frac{2}{\theta}\right)^{-2a/2}\Gamma(2a)\,\exp\left[\frac{\left(\frac{X}{\sigma_N^2}\right)^2}{8\left(\frac{1}{2\sigma_N^2}+\frac{1}{\theta}\right)}\right]$$
$$\left[D_{-2a},\left(\frac{\frac{X}{\sigma_N^2}}{\sqrt{\frac{1}{\sigma_N^2}+\frac{2}{\theta}}}\right)+D_{-2a},\left(\frac{-\frac{X}{\sigma_N^2}}{\sqrt{\frac{1}{\sigma_N^2}+\frac{2}{\theta}}}\right)\right] \qquad \text{(A.12)}$$

Dividing the two above equations we get:

$$\hat{S} = \left(\frac{1}{\sigma_N^2}+\frac{2}{\theta}\right)^{-1/2}\frac{\Gamma(2a+1)}{\Gamma(2a)}\frac{D_{-2a-1}(-\zeta X)-D_{-2a-1}(\zeta X)}{D_{-2a}(-\zeta X)+D_{-2a}(\zeta X)} \qquad \text{(A.13)}$$

where

$$\zeta = \frac{1/\sigma_N^2}{\sqrt{1/\sigma_N^2+2/\theta}} = \sqrt{\frac{\theta/\sigma_N^2}{\theta+2\sigma_N^2}}$$

Considering that $\Gamma(2a+1)/\Gamma(2a) = 2a$ and expressing the first square root of eq. A.13 in terms of ζ we have:

$$\hat{S} = 2a\sigma_N^2\zeta\,\frac{D_{-2a-1}(-\zeta X)-D_{-2a-1}(\zeta X)}{D_{-2a}(-\zeta X)+D_{-2a}(\zeta X)} \qquad \text{where} \quad \zeta = \sqrt{\frac{\theta/\sigma_N^2}{\theta+2\sigma_N^2}} \qquad \text{(A.14)}$$

A.3 Derivation of the MAP-Chi-DFT Estimator

The MAP estimate is the value of S which maximises $\ln(p(X|S)p(S))$, where $p(X|S)$ and $p(S)$ are given by 3.5 and 3.7 respectively. We therefore have:

$$\ln(p(X|S)p(S)) = \ln\left[\frac{1}{\sqrt{2\pi\sigma_N^2}}\exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right]\frac{|S|^{2a-1}}{\theta^a\,\Gamma(a)}\exp\left[-\frac{S^2}{\theta}\right]\right]$$

Taking the derivative w.r.t. S we get:

$$\frac{d(\ln(p(X|S)p(S)))}{dS} = \frac{X-S}{\sigma_N^2} + \frac{2a-1}{S} - \frac{2S}{\theta} \tag{A.15}$$

Setting the above equation to zero and solving w.r.t S we get:

$$\hat{S} = \zeta\frac{X}{2} + \text{sgn}(X)\left[\left(\zeta\frac{X}{2}\right)^2 + (a-0.5)\,2\sigma_N^2\,\zeta\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta+2\sigma_N^2} \tag{A.16}$$

The above estimator comes from solving a quadratic equation, which can have two solutions. We briefly describe which one is chosen and how the sgn(.) appears in the above equation. The value of S for which the posterior density has its maximum has the same sign as X as it can be seen from the form of $p(X|S)P(S)$. For $a > 0.5$ the two solutions have different signs, so we chose the one that has the same sign as X. For $a < 0.5$ both of the solutions have the same sign but one only is a maximum, which is what we are looking for. Following these rules, it turns that the correct sign from the $\pm$ is the one that matches the sign of X.

A.4 Derivation of the MMSE-Gamma-DFT Estimator

Substituting in equation 3.8 the expression for the likelihood (eq. 3.5) and the Gamma prior, which is given by equation 3.19 we have:

$$\hat{S} = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{S}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{|S|^{a-1}}{2\,\theta^a\,\Gamma(a)} \exp\left[-\frac{|S|}{\theta}\right] dS}{\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{|S|^{a-1}}{2\,\theta^a\,\Gamma(a)} \exp\left[-\frac{|S|}{\theta}\right] dS} \tag{A.17}$$

The numerator can be written as:

$$\begin{aligned} \text{num} &= \int_{-\infty}^{0} \frac{S}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{(-S)^{a-1}}{2\,\theta^a\,\Gamma(a)} \exp\left[\frac{S}{\theta}\right] dS \\ &+ \int_{0}^{\infty} \frac{S}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{S^{a-1}}{2\,\theta^a\,\Gamma(a)} \exp\left[-\frac{S}{\theta}\right] dS \end{aligned}$$

Making the substitution $S = -S$ in the first integral we have:

$$\begin{aligned} \text{num} &= \int_{0}^{\infty} -S\frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X+S)^2}{2\sigma_N^2}\right] \frac{(S)^{a-1}}{2\,\theta^a\,\Gamma(a)} \exp\left[-\frac{S}{\theta}\right] dS \\ &+ \int_{0}^{\infty} S\frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right] \frac{S^{a-1}}{2\,\theta^a\,\Gamma(a)} \exp\left[-\frac{S}{\theta}\right] dS \end{aligned}$$

Expanding the exponentials and taking common factors:

$$\begin{aligned} \text{num} = \frac{\exp\left[-\frac{X^2}{2\sigma_N^2}\right]}{\sqrt{2\pi\sigma_N^2}\,2\theta^a\,\Gamma(a)} \cdot \Bigg[&- \int_0^{\infty} S^a \exp\left[-\frac{S^2}{2\sigma_N^2} - S\left(\frac{X}{\sigma_N^2} + \frac{1}{\theta}\right)\right] dS \\ &+ \int_0^{\infty} S^a \exp\left[-\frac{S^2}{2\sigma_N^2} - S\left(-\frac{X}{\sigma_N^2} + \frac{1}{\theta}\right)\right] dS \Bigg] \end{aligned} \tag{A.18}$$

Solving the above integrals with A.10 we get:

$$\text{num} = \frac{1}{2\sqrt{2\pi\sigma_N^2}} \frac{\sigma_N^{a+1}}{\theta^a} \frac{\Gamma(a+1)}{\Gamma(a)} \exp\left[-\frac{X^2}{2\sigma_N^2}\right]$$
$$\left[\exp\left[\left(\frac{\zeta_1}{2}\right)^2\right] D_{-a-1}(\zeta_1) - \exp\left[\left(\frac{\zeta_2}{2}\right)^2\right] D_{-a-1}(\zeta_2)\right] \tag{A.19}$$

where $\zeta_1 = \frac{\sigma_N}{\theta} - \frac{X}{\sigma_N}, \quad \zeta_2 = \frac{\sigma_N}{\theta} + \frac{X}{\sigma_N}$

if we perform the same operations on the denominator of eq. A.17 we have:

$$\text{den} = \frac{1}{2\sqrt{2\pi\sigma_N^2}} \frac{\sigma_N^{a}}{\theta^a} \exp\left[-\frac{X^2}{2\sigma_N^2}\right]$$
$$\left[\exp\left[\left(\frac{\zeta_1}{2}\right)^2\right] D_{-a}(\zeta_1) + \exp\left[\left(\frac{\zeta_2}{2}\right)^2\right] D_{-a}(\zeta_2)\right] \tag{A.20}$$

Dividing the numerator and the denominator we get:

$$\hat{S} = a\sigma_N \frac{\exp\left[\frac{\zeta_1^2}{4}\right] D_{-a-1}(\zeta_1) - \exp\left[\frac{\zeta_2^2}{4}\right] D_{-a-1}(\zeta_2)}{\exp\left[\frac{\zeta_1^2}{4}\right] D_{-a}(\zeta_1) + \exp\left[\frac{\zeta_2^2}{4}\right] D_{-a}(\zeta_2)} \tag{A.21}$$

A.5 Derivation of the MAP-Gamma-DFT Estimator

The estimate of this algorithm is the value of S that maximises $\ln(p(X|S)p(S))$ where $p(X|S)$ is again given by eq. 3.5 and $p(S)$ by eq. 3.19. we consecutively have:

$$\ln(p(X|S)p(S)) = \ln\left[\frac{1}{\sqrt{2\pi\sigma_N^2}}\exp\left[-\frac{(X-S)^2}{2\sigma_N^2}\right]\frac{|S|^{a-1}}{2\theta^a\,\Gamma(a)}\exp\left[-\frac{|S|}{\theta}\right]\right]$$

Taking the derivative w.r.t. S we get:

$$\frac{d(\ln(p(X|S)p(S)))}{dS} = \frac{X-S}{\sigma_N^2} + \frac{a-1}{S} - \frac{\mathrm{sgn}(S)}{\theta} \tag{A.22}$$

Setting the above equation to zero and solving w.r.t S we get:

$$\hat{S} = \zeta + \mathrm{sgn}(X)\left[\zeta^2 + (a-1)\sigma_N^2\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{X}{2} - \mathrm{sgn}(X)\frac{\sigma_N^2}{2\theta} \tag{A.23}$$

The $\mathrm{sgn}(.)$ in the definition of ζ comes from the fact that the maximum of the posterior density occurs at an S which has the same sign with X. The $\mathrm{sgn}(.)$ before the square root appears because one of the two solutions of $d(\ln(p(X|S)p(S)))/dS = 0$ is chosen according to the rules stated in appendix A.3.

A.6 Derivation of the MMSE-Chi-AMP Estimator

The estimator for this algorithm can be obtained by substituting eqs. 4.2 and 4.3 into 4.4. The numerator of the last equation will then read:

$$num = \int_0^{\infty} \int_0^{2\pi} \frac{AR}{2\pi\sigma_N^2} \exp\left[-\frac{R^2 + A^2 - 2RA\cos(\psi - \phi)}{2\sigma_N^2}\right] \frac{2A^{2a-1}\exp\left[-\frac{A^2}{\theta}\right]}{2\pi\,\theta^a\,\Gamma(a)}\, d\phi dA \tag{A.24}$$

which after some algebraic manipulations can be written as:

$$\text{num} = K \int_0^{\infty} A^{2a} \exp\left[-A^2\left(\frac{\theta + 2\sigma_N^2}{\theta\, 2\sigma_N^2}\right)\right] J_0\left(i\frac{RA}{\sigma_N^2}\right) dA \tag{A.25}$$

where

$$J_0\left(i\frac{RA}{\sigma_N^2}\right) = \frac{1}{2\pi}\int_0^{2\pi} \exp\left[\frac{RA\cos(\psi - \phi)}{\sigma_N^2}\right] d\phi \tag{A.26}$$

and $J_\nu(z)$ is the Bessel function of the first kind (see [7] eqs. 8.406.3, 8.431.5). K is:

$$K = \frac{2R}{2\pi\,\sigma_N^2\,\theta^a\,\Gamma(a)} \exp\left[-\frac{R^2}{2\sigma_N^2}\right] \tag{A.27}$$

The integral in eq. A.25 can be solved with formula 6.631.1 from [7] which is stated below.

$$\int_0^{\infty} x^{\mu} \mathrm{e}^{-\delta x^2} J_\nu(\beta x)\, dx = \frac{\beta^\nu\, \Gamma\left(\frac{\nu+\mu+1}{2}\right)}{2^{\nu+1}\,\delta^{(\mu+\nu+1)/2}\,\Gamma(\nu+1)}\, {}_1F_1\left(\frac{\nu+\mu+1}{2}; \nu+1; -\frac{\beta^2}{4\delta}\right) \tag{A.28}$$

Solving the integral we get:

$$\text{num} = K\left(\frac{2\sigma_N^2\theta}{\theta + 2\sigma_N^2}\right)^{a+0.5} \frac{\Gamma(a+0.5)}{2}\, {}_1F_1\left(a+0.5; 1; \frac{R^2\theta}{2\sigma_N^2(\theta + 2\sigma_N^2)}\right) \tag{A.29}$$

Performing the same operations on the denominator we get:

$$\text{den} = K\left(\frac{2\sigma_N^2\theta}{\theta + 2\sigma_N^2}\right)^{a} \frac{\Gamma(a)}{2}\, {}_1F_1\left(a; 1; \frac{R^2\theta}{2\sigma_N^2(\theta + 2\sigma_N^2)}\right) \tag{A.30}$$

Dividing the the numerator (num) with the denominator (den) we get:

$$\hat{A} = \sqrt{2\sigma_N^2 \zeta}\, \frac{\Gamma(a+0.5)}{\Gamma(a)}\, \frac{{}_1F_1(a+0.5;1;\frac{R^2}{2\sigma_N^2}\zeta)}{{}_1F_1(a;1;\frac{R^2}{2\sigma_N^2}\zeta)} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \qquad \text{(A.31)}$$

A.7 Derivation of the MAP-Chi-AMP Estimator

Substituting $p(R,\psi|A,\phi)$ and $p(A)$ from 4.2 and 4.3 and $p(\phi)=\frac{1}{2\pi}$ in eq. 4.7 yields:

$$\hat{A} = \arg\max_{A} \ln\left[\int_0^{2\pi} \frac{R}{2\pi\sigma_N^2}\exp\left[-\frac{R^2+A^2-2RA\cos(\psi-\phi)}{2\sigma_N^2}\right] \cdot \frac{2A^{2a-1}}{2\pi\,\theta^a\,\Gamma(a)}\exp\left[-\frac{A^2}{\theta}\right] d\phi\right] \quad \text{(A.32)}$$

After some rearrangement the logarithm can be written as:

$$\ln\left[\frac{2R}{2\pi\sigma_N^2\,\theta^a\,\Gamma(a)}A^{2a-1}\exp\left[-\frac{R^2+A^2}{2\sigma_N^2}-\frac{A^2}{\theta}\right]\frac{1}{2\pi}\int_0^{2\pi}\exp\left[\frac{RA\cos(\psi-\phi)}{\sigma_N^2}\right]d\phi\right] \quad \text{(A.33)}$$

Using eq. 8.431.5 from [7] we have:

$$I_0\left(\frac{RA}{\sigma_N^2}\right)=\frac{1}{2\pi}\int_0^{2\pi}\exp\left[\frac{RA\cos(\psi-\phi)}{\sigma_N^2}\right]d\phi \quad \text{(A.34)}$$

where $I_0(z)$ is the modified bessel function of the first kind. Using also the approximation $I_0(z)\sim \mathrm{e}^z/\sqrt{2\pi z}$ the logarithm in eq. A.33 can be written as:

$$\ln\left[\frac{2R}{2\pi\sigma_N^2\,\theta^a\,\Gamma(a)}A^{2a-1}\exp\left[-\frac{R^2+A^2}{2\sigma_N^2}-\frac{A^2}{\theta}\right]\frac{\exp\left[\frac{RA}{\sigma_N^2}\right]}{\sqrt{2\pi\frac{RA}{\sigma_N^2}}}\right] \quad \text{(A.35)}$$

Taking the derivative of the above expression w.r.t. A, setting to zero and solving w.r.t A we get:

$$\hat{A}=\zeta\frac{R}{2}\pm\left[\left(\zeta\frac{R}{2}\right)^2+(a-0.75)\,2\sigma_N^2\,\zeta\right]^{1/2} \quad \text{where} \quad \zeta=\frac{\theta}{\theta+2\sigma_N^2} \quad \text{(A.36)}$$

From the above two solutions the valid is the one which is a maximum and positive. Some further analysis shows that this is always the one with the $(+)$.

A.8 Derivation of the MAP-Gamma-AMP Estimator

Substituting $p(R,\psi|A,\phi)$ and $p(A)$ from 4.2 and 4.15 and $p(\phi)=\frac{1}{2\pi}$ in eq. 4.7 yields:

$$\hat{A} = \arg\max_{A} \ln\left[\int_0^{2\pi} \frac{R}{2\pi\sigma_N^2}\exp\left[-\frac{R^2+A^2-2RA\cos(\psi-\phi)}{2\sigma_N^2}\right] \cdot \frac{A^{a-1}}{2\pi\,\theta^a\,\Gamma(a)}\exp\left[-\frac{A}{\theta}\right] d\phi\right] \quad \text{(A.37)}$$

After some rearrangement the logarithm can be written as:

$$\ln\left[\frac{R}{2\pi\sigma_N^2\,\theta^a\,\Gamma(a)}A^{a-1}\exp\left[-\frac{R^2+A^2}{2\sigma_N^2}-\frac{A}{\theta}\right]\frac{1}{2\pi}\int_0^{2\pi}\exp\left[\frac{RA\cos(\psi-\phi)}{\sigma_N^2}\right]d\phi\right] \quad \text{(A.38)}$$

Transforming the integral as in Appendix A.7 the above expression becomes:

$$\ln\left[\frac{R}{2\pi\sigma_N^2\,\theta^a\,\Gamma(a)}A^{a-1}\exp\left[-\frac{R^2+A^2}{2\sigma_N^2}-\frac{A}{\theta}\right]\frac{\exp\left[\frac{RA}{\sigma_N^2}\right]}{\sqrt{2\pi\frac{RA}{\sigma_N^2}}}\right] \quad \text{(A.39)}$$

Taking the derivative of the above expression w.r.t. A, setting to zero and solving w.r.t A we get:

$$\hat{A}=\zeta\pm\left[\zeta^2+(a-1.5)\sigma_N^2\right]^{1/2}\text{where}\quad \zeta=\frac{R}{2}-\frac{\sigma_N^2}{2\theta} \quad \text{(A.40)}$$

From the above two solutions the valid is the one with the $(+)$ because it is always positive and a maximum.

APPENDIX B SUMMARY OF THE ESTIMATORS

MMSE-Chi-DFT

$$\hat{S} = 2a\sigma_N^2\zeta \frac{D_{-2a-1}(-\zeta X) - D_{-2a-1}(\zeta X)}{D_{-2a}(-\zeta X) + D_{-2a}(\zeta X)} \qquad \text{where} \quad \zeta = \sqrt{\frac{\theta/\sigma_N^2}{\theta + 2\sigma_N^2}} \tag{B.1}$$

$$\hat{S} = X\left[\frac{2a\eta}{\gamma}\frac{D_{-2a-1}(-\eta) - D_{-2a-1}(\eta)}{D_{-2a}(-\eta) + D_{-2a}(\eta)}\right] \qquad \text{where} \quad \eta = \text{sgn}(X)\sqrt{\frac{\xi\gamma}{\xi + 2a}} \tag{B.2}$$

MAP-Chi-DFT

$$\hat{S} = \zeta\frac{X}{2} + \text{sgn}(X)\left[\left(\zeta\frac{X}{2}\right)^2 + (a - 0.5)\,2\sigma_N^2\,\zeta\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \tag{B.3}$$

$$\hat{S} = X\left[\frac{\eta}{2} + \left[\left(\frac{\eta}{2}\right)^2 + (a - 0.5)\frac{2\eta}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{\xi}{\xi + 2a} \tag{B.4}$$

MMSE-Gamma-DFT

$$\hat{S} = a\sigma_N \frac{\exp\left[\frac{\zeta_1^2}{4}\right] D_{-a-1}(\zeta_1) - \exp\left[\frac{\zeta_2^2}{4}\right] D_{-a-1}(\zeta_2)}{\exp\left[\frac{\zeta_1^2}{4}\right] D_{-a}(\zeta_1) + \exp\left[\frac{\zeta_2^2}{4}\right] D_{-a}(\zeta_2)} \tag{B.5}$$

where $\zeta_1 = \frac{\sigma_N}{\theta} - \frac{X}{\sigma_N}, \zeta_2 = \frac{\sigma_N}{\theta} + \frac{X}{\sigma_N}$

$$\hat{S} = X\left[\frac{a\,\text{sgn}(X)}{\sqrt{\gamma}} \frac{\exp\left[\frac{\eta_1^2}{4}\right] D_{-a-1}(\eta_1) - \exp\left[\frac{\eta_2^2}{4}\right] D_{-a-1}(\eta_2)}{\exp\left[\frac{\eta_1^2}{4}\right] D_{-a}(\eta_1) + \exp\left[\frac{\eta_2^2}{4}\right] D_{-a}(\eta_2)}\right] \tag{B.6}$$

where $\eta_1 = \dfrac{\sqrt{a(a+1)}}{\sqrt{\xi}} - \mathrm{sgn}(X)\sqrt{\gamma}, \quad \eta_2 = \dfrac{\sqrt{a(a+1)}}{\sqrt{\xi}} + \mathrm{sgn}(X)\sqrt{\gamma}$

MAP-Gamma-DFT

$$\hat{S} = \zeta + \mathrm{sgn}(X)\left[\zeta^2 + (a-1)\sigma_N^2\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{X}{2} - \mathrm{sgn}(X)\frac{\sigma_N^2}{2\theta} \tag{B.7}$$

$$\hat{S} = X\left[\eta + \mathrm{sgn}(X)\left[\eta^2 + \frac{a-1}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{a(a+1)}{\xi\gamma}} \tag{B.8}$$

MMSE-Chi-AMP

$$\hat{A} = \sqrt{2\sigma_N^2\,\zeta}\,\frac{\Gamma(a+0.5)}{\Gamma(a)}\,\frac{{}_1F_1(a+0.5;1;\frac{R^2}{2\sigma_N^2}\zeta)}{{}_1F_1(a;1;\frac{R^2}{2\sigma_N^2}\zeta)} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \tag{B.9}$$

$$\hat{A} = R\left[\sqrt{\frac{\eta}{\gamma}}\,\frac{\Gamma(a+0.5)}{\Gamma(a)}\,\frac{{}_1F_1(a+0.5;1;\gamma\eta)}{{}_1F_1(a;1;\gamma\eta)}\right] \qquad \text{where} \quad \eta = \frac{\xi}{\xi + a} \tag{B.10}$$

MAP-Chi-AMP

$$\hat{A} = \zeta\frac{R}{2} + \left[\left(\zeta\frac{R}{2}\right)^2 + (a-0.75)\,2\sigma_N^2\,\zeta\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{\theta}{\theta + 2\sigma_N^2} \tag{B.11}$$

$$\hat{A} = R\left[\frac{\eta}{2} + \left[\left(\frac{\eta}{2}\right)^2 + (a-0.75)\frac{\eta}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{\xi}{\xi + a} \tag{B.12}$$

MAP-Gamma-AMP

$$\hat{A} = \zeta + \left[\zeta^2 + (a-1.5)\sigma_N^2\right]^{1/2} \qquad \text{where} \quad \zeta = \frac{R}{2} - \frac{\sigma_N^2}{2\theta} \tag{B.13}$$

$$\hat{A} = R\left[\eta + \left[\eta^2 + \frac{a-1.5}{\gamma}\right]^{1/2}\right] \qquad \text{where} \quad \eta = \frac{1}{2} - \frac{1}{4}\sqrt{\frac{a(a+1)}{\xi\gamma}} \tag{B.14}$$

APPENDIX C RAW MOMENTS OF THE PRIOR DENSITY FUNCTIONS

$f(s)$	Raw k^{th} moment
2-sided Chi	
$\frac{1}{\theta^a \Gamma(a)} \lvert S\rvert^{2a-1} \exp\left[-\frac{S^2}{\theta}\right]$	$M_k = \begin{cases} \frac{\theta^{k/2}\Gamma(a+k/2)}{\Gamma(a)} & \text{if } k \text{ even} \\ 0 & \text{if } k \text{ odd} \end{cases}$
2-sided Gamma	
$\frac{1}{2\theta^a \Gamma(a)} \lvert S\rvert^{a-1} \exp\left[-\frac{\lvert S\rvert}{\theta}\right]$	$M_k = \begin{cases} \frac{\theta^{k}\Gamma(a+k)}{\Gamma(a)} & \text{if } k \text{ even} \\ 0 & \text{if } k \text{ odd} \end{cases}$
1-sided Chi	
$\frac{2}{\theta^a \Gamma(a)} S^{2a-1} \exp\left[-\frac{S^2}{\theta}\right]$, with $S \geq 0$	$M_k = \frac{\theta^{k/2}\Gamma(a+k/2)}{\Gamma(a)}$
1-sided Gamma	
$\frac{1}{\theta^a \Gamma(a)} S^{a-1} \exp\left[-\frac{S}{\theta}\right]$, with $S \geq 0$	$M_k = \frac{\theta^{k}\Gamma(a+k)}{\Gamma(a)}$

Table C.1: Prior Density functions and Raw Moments.

REFERENCES

[1] R. Martin, "Speech enhancement using mmse short time spectral estimation with gamma distributed speech priors," in *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on,*, vol. 1, pp. 253–256, 2002.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, pp. 1109–1121, 1984.

[3] P. J. Wolfe and S. J. Godsill, "Efficient alternatives to the ephraim malah suppression rule for audio signal enhancement," *EURASIP Journal on Applied Signal Processing 2003*, vol. 10, pp. 1043–1051, 2003.

[4] T. H. Dat, K. Takeda, and F. Itakura, "Generalized gamma modeling of speech and its online estimation for speech enhancement," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on,*, vol. 4, pp. 181–184, 2005.

[5] T. Lotter and P. Vary, "Noise reduction by joint maximum a posteriori spectral amplitude and phase estimation with super-gaussian speech modelling," in *in Proc. of EUSIPCO-04 (Vienna, Austria),*, pp. 1457–1460, 2004.

[6] T. Lotter and P. Vary, "Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modeling," in *International Workshop on Acoustic Echo and Noise Control (IWAENC2003),*, pp. 83–86, Sep 2003.

[7] I. S. Gradshteyn and I. W. Ryzik, *Tables of Integrals Series and Products*. New York: Academic Press, 1965.

[8] J. W. Shin, J.-H. Chang, and N. S. Kim, "Statistical modeling of speech signals based on generalized gamma distribution," *Signal Processing Letters, IEEE,*, vol. 12, no. 3, pp. 258–261, 2005.

[9] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing,*, vol. 9, pp. 504–512, 2001.

[10] O. Cappe, "Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor," *Speech and Audio Processing, IEEE Transactions on,*, vol. 2, no. 2, pp. 345–349, 1994.

[11] I. Cohen, "On the decision-directed estimation approach of ephraim and malah," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on,*, vol. 1, pp. I–293–6, 2004.

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

[13] D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on,*, vol. 30, no. 4, pp. 679–681, 1982.

[14] J. R. J. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. New York, NY: Macmillan Publishing Company, 1993.

[15] E. Zavarehei and S. Vaseghi, "Speech enhancement in temporal dft trajectories using kalman filters," in *Interspeech*, 2005.