15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP

INTEGRATING THE COCHLEA'S COMPRESSIVE NONLINEARITY IN THE BAYESIAN APPROACH FOR SPEECH ENHANCEMENT

Eric Plourde and Benoît Champagne

Department of Electrical and Computer Engineering, McGill University
Montreal, Quebec, Canada
email: eric.plourde@mail.mcgill.ca, benoit.champagne@mcgill.ca

ABSTRACT

The human ear has a great ability to isolate speech in a noisy environment and, therefore, constitutes a great source of inspiration for speech enhancement algorithms. In this work, we propose a Bayesian estimator for speech enhancement that integrates the cochlea's compressive nonlinearity in its cost function. When compared to existing Bayesian speech enhancement estimators, the proposed estimator can achieve a better compromise between speech distortion and noise reduction by favoring less speech distortion at lower frequencies, where the main formants are located, while increasing the noise reduction at higher frequencies. The proposed estimator also yields better results both in terms of objective and subjective performance measures.

1. INTRODUCTION

In speech enhancement, the general objective is to remove a certain amount of noise from a noisy speech signal while keeping the speech component as undistorted as possible. Many approaches have been proposed to achieve that goal, such as the spectral subtraction, Bayesian or subspace approaches [1]. In Bayesian approaches, an estimate of the clean speech is derived by minimizing the expectation of a defined cost function.

One possible avenue for choosing an appropriate cost function is to consider the human hearing mechanism. The ear is most sensitive to small signals and grows progressively less responsive as stimulations become stronger. This permits us to interpret sounds over a wider range of amplitudes and is also thought to play a role in the noise suppression capabilities of the auditory system [2]. This ability is due in part to the signal processing performed by the cochlea and more precisely by its basilar membrane. The basilar membrane performs a spectrum analysis and can be assimilated to an active bank of filters with non-linear gains [3]. One of the properties of the cochlea that produces nonlinearities in the gains is the cochlear amplification. In cochlear amplification, each spectral component is amplified by the active mechanism of the outer hair cells. However, as the spectral amplitude increases, the amplification saturates and, in relative terms, the higher spectral amplitudes become compressed. This behavior has been referred to as the cochlea's compressive nonlinearity [4].

In this paper we propose a Bayesian speech enhancement algorithm motivated by the cochlea's compressive nonlinearity. Results show an improvement in terms of noise reduction, Perceptual Evaluation of Speech Quality (PESQ) and informal Mean Opinion Score (MOS) when the proposed algorithm is compared to existing Bayesian algorithms such as the MMSE STSA [5] and MMSE log-STSA (LSA) [6].

This paper is organized as follows. In Section 2 we first present relevant Bayesian speech enhancement algorithms while in Section 3, we further discuss the cochlea's compressive nonlinearity. In Section 4 we derive the proposed estimator. In Section 5, we present comparative results of the proposed estimator while Section 6 concludes this work.

2. BAYESIAN STSA SPEECH ENHANCEMENT

Let the observed noisy speech be

$$y(t) = x(t) + n(t) \qquad 0 \le t \le T \tag{1}$$

where $x(t)$ is the clean speech, $n(t)$ is an additive noise and $[0,T]$ is the observation interval. Let Y_k, X_k and N_k denote the k^{th} complex spectral components of the noisy speech, clean speech and noise respectively obtained through a Fourier analysis.

In Bayesian Short Time Spectral Amplitude (STSA) estimation for speech enhancement, the goal is to obtain the estimator $\hat{\chi}_k$ of $\chi_k \triangleq |X_k|$ which minimizes $E\{C(\chi_k,\hat{\chi}_k)\}$, where $C(\chi_k,\hat{\chi}_k)$ is a chosen cost function and E denotes statistical expectation. This estimator is then combined with the phase of the noisy speech, $\angle Y_k$, to yield the estimator of the complex spectral component of the clean speech $\hat{X}_k = \hat{\chi}_k e^{j\angle Y_k}$ [5]. In MMSE STSA, $C(\chi_k,\hat{\chi}_k) = (\chi_k - \hat{\chi}_k)^2$ while in LSA, $C(\chi_k,\hat{\chi}_k) = (\log(\chi_k) - \log(\hat{\chi}_k))^2$.

The MMSE STSA estimator was generalized under the β-order STSA MMSE (β-SA) estimator in [7] by modifying the cost function as $C(\chi_k,\hat{\chi}_k;\beta) = (\chi_k^\beta - \hat{\chi}_k^\beta)^2$ where the exponent β is a real positive parameter. The β-SA estimator is expressible as:

$$\hat{\chi}_{\beta\text{-SA},k} = G_{\beta\text{-SA},k}\,|Y_k| \tag{2}$$

$$G_{\beta\text{-SA},k} = \frac{\sqrt{\upsilon_k}}{\gamma_k}\left[\Gamma\left(\frac{\beta}{2}+1\right) M\left(-\frac{\beta}{2},1;-\upsilon_k\right)\right]^{1/\beta} \tag{3}$$

with:

$$\upsilon_k \triangleq \frac{\xi_k}{1+\xi_k}\gamma_k, \quad \xi_k \triangleq \frac{E\{{\chi_k}^2\}}{E\{|N_k|^2\}}, \quad \gamma_k \triangleq \frac{|Y_k|^2}{E\{|N_k|^2\}}$$

and where $\Gamma(x)$ is the gamma function and $M(a,b;z)$ is the confluent hypergeometric function.

When $\beta = 1$, the β-SA estimator is identical to the MMSE STSA estimator. Furthermore, You *et al.* suggested

This work was supported by the *Fonds québécois de la recherche sur la nature et les technologies*.

in [7] that when $\beta \to 0^+$, the β-SA estimator is equivalent to the LSA estimator. Therefore, the MMSE STSA and LSA estimators are both special cases of the more general β-SA estimator.

The case $\beta > 0$ was analyzed in [7] while the analysis was extended to the case $\beta < 0$ in [8]. In the later, it was shown that the β-SA estimator introduces more speech distortion but also achieves better noise reduction as β is decreased from 1 to -2; however, serious speech distortions were reported for $\beta < -1.5$.

Some variants of the β-SA estimator with $\beta > 0$ were also proposed. In [7], You *et al.* proposed to adapt the value of β differently for each analysis frame according to the frame's Signal-to-Noise Ratio (SNR). Furthermore, they also proposed to modify β according to the masking threshold for each frequency component [9].

3. COMPRESSIVE NONLINEARITY OF THE COCHLEA

As mentioned in the introduction, the cochlea has a nonlinear compressive behavior. This so-called compressive nonlinearity has been noticed when measuring basilar membrane responses to input tones at several sound pressure levels [4]. It is thought to be caused by the active mechanism of the outer hair cells which at lower input amplitudes exhibit an amplification of the basilar membrane vibration, termed cochlear amplification. As the amplitude increases, however, this amplification saturates and, in relative terms, the larger input spectral amplitudes become compressed.

Compression rates of 0.2 dB/dB were measured at the base (i.e. for high frequencies) of the mammalian cochlea for intensities between 40 and 90 dB SPL [4]; conversational speech is at 60 dB SPL. The compression rates tended to 1 dB/dB for lower intensities, i.e. where the amplification did not saturate.

While the cochlea's compressive nonlinearities are well documented and accepted for high frequency components, there is no real consensus on the degree of compressive nonlinearity at lower frequencies (i.e. at the apex of the cochlea). In fact, some results from chinchilla show a smaller rate of compression (0.5 - 0.8 dB/dB) than at higher frequencies, when several other results from guinea pigs and squirrel monkeys fail to show any compressive nonlinearity (i.e. rate of compression of 1 dB/dB) or even show an expansion (i.e. rate of compression greater than 1 dB/dB) [4]. On the other hand, psychoacoustic experiments in humans report either a comparable rate of compression at low and high frequencies [10] or a smaller but still existent rate of compression at lower frequencies [11]. However, since those results are from psychoacoustic experiments and not from a specific physiological experiment, one cannot be sure where in the auditory processing path this compression originates and it may not occur in the cochlea but rather along the auditory neural pathway [10]. Therefore, the cochlear rate of compression at low frequencies is still an active debate [4]. For the purpose of this research, and based on the above discussion, we will assume there is no compressive nonlinearity at low frequencies.

4. INTEGRATING THE COMPRESSIVE NONLINEARITY IN THE BAYESIAN COST FUNCTION

In this section we will show how the cochlea's compressive nonlinearities can be incorporated in the cost function of a Bayesian STSA estimator for speech enhancement.

4.1 β as the compression rate

We wish to modify the STSA to integrate the cochlear compressive nonlinearity. One way to achieve our goal is to apply a relevant exponent to the STSA. In fact, consider two different input spectral amplitudes say $|W_k| < |V_k|$ to which we apply an exponent β. We can compute the compression rate m in dB/dB as:

$$m = \frac{20\log\left(\frac{|V_k|^\beta}{|W_k|^\beta}\right)}{20\log\left(\frac{|V_k|}{|W_k|}\right)} = \beta$$

Therefore, β can be directly interpreted as the compression rate of the input spectral amplitudes and thus set to physiological values identified in the previous section. Note however that by doing so, we will apply the compression rate on the entire range of possible input intensities rather than on the 40 to 90 dB SPL range.

It is interesting to note that power laws have been used in the past to model cochlear nonlinearities [12] as well as in speech processing to perform an intensity-to-loudness conversion (e.g. in Perceptual Linear Predictive (PLP) analysis [13]).

4.2 The proposed cost function and estimator

As mentioned in Section 3, we will assume there is no compressive nonlinearity in the cochlea at low frequencies. Since a high rate of compression is known to occur at high frequencies, β will therefore become frequency dependent. The proposed cost function is thus:

$$C(\chi_k, \hat{\chi}_k; \beta_k) = (\chi_k^{\beta_k} - \hat{\chi}_k^{\beta_k})^2 \tag{4}$$

where β_k accounts for the compression rate at frequency k. The corresponding estimator is:

$$\hat{\chi}_{\mathrm{CNSA},k} = G_{\mathrm{CNSA},k}\,|Y_k| \tag{5}$$

$$G_{\mathrm{CNSA},k} = \frac{\sqrt{\upsilon_k}}{\gamma_k}\left[\Gamma\left(\frac{\beta_k}{2}+1\right) M\left(-\frac{\beta_k}{2}, 1; -\upsilon_k\right)\right]^{1/\beta_k} \tag{6}$$

which will be identified as the CNSA (Compressive Nonlinear transformation of the Spectral Amplitude) estimator.

This estimator is closely related to the β-SA estimator, however, β_k is now interpreted as a physiological parameter accounting for the rate of compression and varying as a function of the frequency.

4.3 Deriving the appropriate β_k values

We need to adequately define the cochlea's rate of compression, β_k, for every frequency k. To do so, we will choose relevant values of β_k for low and high frequencies as well as means of interpolation between those two values to obtain intermediate rate of compressions.

Since for low frequency we consider the absence of compressive nonlinearity, we will therefore choose $\beta_{\text{low}} = 1$. As indicated in Section 3, the compressive nonlinearity at high frequencies is thought to have a rate of compression of approximately 0.2 dB/dB. For high frequencies, it therefore seems plausible to set $\beta_{\text{high}} = 0.2$ as an initial value.

As shown in [8], when β is decreased, more noise reduction is achieved by the β-SA estimator while more speech distortion is simultaneously introduced. Choosing the previous values for β_{high} and β_{low} will therefore imply less speech distortion at lower frequencies, where the main speech formants are present and may mask the noise, and more noise reduction at higher frequencies. While the value of $\beta_{\text{high}} = 0.2$ is based on physiological observations, it would be relevant to include another value of β_{high} to our study that may not be based on such observations but that would imply further noise reduction at higher frequencies, while keeping $\beta_{\text{low}} = 1$ to limit speech distortion at lower frequencies. We will therefore consider also the value $\beta_{\text{high}} = -1.5$.

Physiological experiments on the cochlear rate of compression at intermediate frequencies (i.e. between the apex and the base of the cochlea) are extremely scarce. Therefore to interpolate β_k for intermediate frequencies we propose two approaches.

First we propose to interpolate the β_k values linearly with respect to the frequency therefore implying a linear relation between the rate of compression and its associated frequency. We will refer to this approach as the frequency interpolation approach.

In the second approach, we consider the fact that each frequency corresponds to a position on the basilar membrane following the so-called tonotopic mapping [4]. One such tonotopic mapping, proposed in [14], is given by:

$$p = \frac{1}{\alpha}\log_{10}\left(\frac{f(k)}{A} + l\right) \tag{7}$$

where p is the position on the basilar membrane in mm, $\alpha = 0.06$ mm^{-1}, $A = 165.4$ Hz, $l = 1$ are parameters set as per [14] and $f(k)$ is the frequency in Hz corresponding to spectral component k. In this approach, we will therefore consider the compression rate to vary linearly not with respect to the frequency but to the position on the basilar membrane corresponding to that frequency following the given tonotopic mapping. In fact, the compressive nonlinearity is thought to be caused by the active process of the outer hair cells and it is known that the hair cells follow a tonotopic organization where they are optimally sensitive to a particular frequency according to their position on the basilar membrane [15]. Interestingly, some of the outer hair cell properties, such as their lengths, have been shown to have a linear relation with respect to their position on the basilar membrane [16]. For this second approach, the values of β_k are thus derived by linearly interpolating between β_{low} and β_{high} according to the position on the basilar membrane corresponding to a given intermediate frequency as given by the tonotopic mapping. Therefore:

$$\beta_k = \frac{1}{\alpha}\log_{10}\left(\frac{f(k)}{A} + l\right)\frac{(\beta_{\text{high}} - \beta_{\text{low}})}{\left(\frac{1}{\alpha}\log_{10}\left(\frac{F_s}{2A} + l\right)\right)} + \beta_{\text{low}} \tag{8}$$

where F_s is the sampling frequency set to 8 kHz in this study. We will denote this second approach as the tonotopic interpolation approach. Figure 1 represents the different values of β_k as a function of the frequency for $\beta_{\text{high}} = 0.2$ and $\beta_{\text{high}} = -1.5$ using the frequency and tonotopic interpolation approaches.

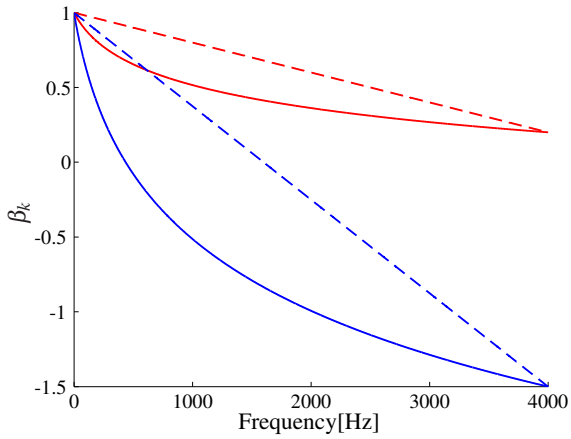

Figure 1: β_k versus frequency [Hz] for $\beta_{\text{high}} = 0.2$ and $\beta_{\text{high}} = -1.5$ (Legend: – – – frequency interpolation; —— tonotopic interpolation).

In summary, the new estimator will therefore be given by (5) and (6) where β_k will be frequency dependent and chosen to reflect the cochlear rate of compression for every frequency k as shown in Figure 1. The estimators will be referred to as CNSA-*f* and CNSA-*t* to indicate a frequency or tonotopic interpolation respectively.

5. RESULTS

In this section, we will present a speech distortion and noise reduction analysis of the new estimator which will be followed by PESQ and MOS results.

5.1 Speech distortion versus noise reduction

In order to study the speech distortion and noise reduction properties of the estimator, we used the following speech distortion and noise reduction metrics in the frequency domain:

$$\Upsilon(G_k) \triangleq E\{[\chi_k - G_k\chi_k]^2\} \tag{9}$$

$$\Psi(G_k) \triangleq \frac{1}{E\{[G_k|N_k|]^2\}} \tag{10}$$

In (9), $\Upsilon(G_k)$ reflects the clean speech distortion energy and, therefore, its value increases for increasing speech distortions. In (10), $\Psi(G_k)$ reflects the inverse of the noise energy remaining in the enhanced speech and increases for increasing noise reduction.

Figure 2 plots $\Upsilon(G_k)$ and $\Psi(G_k)$ versus the frequency for different gains G_k as given by the MMSE STSA (β-SA with $\beta = 1$), LSA (β-SA with $\beta \to 0$) and CNSA-*t* algorithms (average of 30 sentences, white noise, SNR = 0 dB). The frequency interpolation approach, CNSA-*f*, has been left out in Figure 2 for clarity purposes, however, the observations made for CNSA-*t* would also apply to CNSA-*f*.

As can be observed, the speech distortion is greater for LSA than for MMSE STSA however LSA produces more

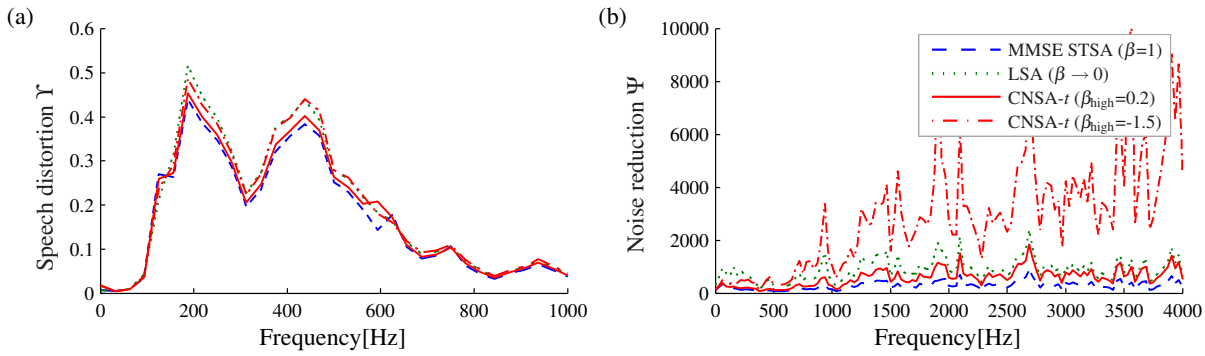

Figure 2: (a) Speech distortion metric versus frequency (0 - 1000 Hz) (b) Noise reduction metric versus frequency (0 - 4000 Hz) (average of 30 sentences, white noise, SNR = 0 dB).

noise reduction. The results obtained by the CNSA-*t* with $\beta_{\text{high}} = 0.2$ are in between MMSE STSA and LSA for both speech distortion and noise reduction. This could be expected since the β values of that estimator are interpolated from 1 to 0.2 while MMSE STSA corresponds to $\beta = 1$ and LSA corresponds to $\beta \to 0$. The CNSA-*t* estimator with $\beta_{\text{high}} = -1.5$, on the other hand, achieves increasing noise reduction while limiting the speech distortion. In fact, it keeps the speech distortion close to the LSA level for lower frequencies ($\approx$ 0-1000 Hz) where the main formants are located and may mask the noise. However, it maximizes the noise reduction at higher frequencies ($\approx$ 1000-4000 Hz) where the speech distortion energies are lower.

5.2 PESQ and MOS results

We present comparative PESQ results for MMSE STSA, LSA and CNSA-*f,t* with $\beta_{\text{high}} = 0.2$ and $\beta_{\text{high}} = -1.5$. Thirty sentences from the TIMIT database, each sampled at 8 kHz, were used where 3 men and 3 women each spoke 5 sentences. Three types of noise were used from the NOISEX database (white, pink and f16 which is mainly composed of low frequency noise along with a peak around 2700 Hz) [17]. The observation frames were of 32ms and a 50% overlap was used between all frames in the overlap-add method for the reconstruction of the enhanced speech. All algorithms used the *decision-directed* approach for the estimation of ξ_k [5] and a voice activity detector proposed in [18] was used to evaluate the noise spectral amplitude variance.

Table 1 presents the PESQ results on a scale from 1 to 4.5. First we observe that the CNSA-*t* estimator outperforms the CNSA-*f* estimator for all cases. Secondly, the CNSA-*f,t* estimators with $\beta_{\text{high}} = 0.2$ both show inferior results when compared to LSA. Therefore, the CNSA-*f,t* estimators with $\beta_{\text{high}} = 0.2$, while they perform better than MMSE STSA, do not show advantages over LSA. Considering the CNSA estimator with $\beta_{\text{high}} = -1.5$, we observe that the CNSA-*f* estimator performs better than the LSA for almost all cases (except pink and f16 noises at 10dB) while the CNSA-*t* yields better results for all cases. The improvements in both CNSA-*f* and CNSA-*t* are, however, more important for the white noise case than for pink and f16 noises. This is mainly due to the fact that the CNSA-*f,t* estimators with $\beta_{\text{high}} = -1.5$ produce more noise reduction for higher frequencies and are therefore more advantageous when the noise has more high frequency components.

In order to support the results obtained with PESQ, we performed informal MOS subjective listening tests on 6 subjects using a subset of 4 sentences (2 men, 2 women) from the initial 30 considering white noise. The listening test involved the MMSE STSA, LSA and CNSA-*t* estimator with $\beta_{\text{high}} = -1.5$. As suggested by ITU-T P.835 [19], MOS tests included an assessment of the speech distortion where the subjects concentrated only on the speech (5 = Not distorted, 1 = Very distorted), background noise where the subjects concentrated only on the noise (5 = Not noticeable, 1 = Very intrusive) and overall speech quality (5 = Excellent, 1 = Bad). Tests were performed in an isolated acoustic room using *beyerdynamic DT880* headphones.

Table 2: MOS results for MMSE STSA, LSA and CNSA-*t* ($\beta_{\text{high}} = -1.5$) estimators (white noise, SNR = 0 dB).

	Noisy speech	**MMSE STSA**	**LSA**	**CNSA-*t*** ($\beta_{\text{high}} = -1.5$)
Speech	3.9	2.4	2.9	3.0
Noise	1.2	2.2	2.5	2.9
Overall	1.7	2.1	2.5	2.7

As can be observed in Table 2, the CNSA-*t* estimator with $\beta_{\text{high}} = -1.5$ performs better than MMSE STSA and LSA in terms of the speech distortion, background noise reduction and overall appreciation for the white noise case.

Comparing speech distortion and background noise reduction MOS results with those of Figure 2, we notice that the big advantage in noise reduction of the CNSA-*t* estimator with $\beta_{\text{high}} = -1.5$ observed in Figure 2 is confirmed by the MOS results. On the other hand, although the speech distortions of LSA and CNSA-*t* should have been greater than those of MMSE STSA, it was not perceived as such in the MOS tests. In fact this may be due to a small perceivable echo produced by MMSE STSA with a 50% overlap which is greatly reduced in LSA and CNSA-*t* and does not seem to be well taken into account by (9).

6. CONCLUSION

In this paper, we presented a speech enhancement algorithm motivated by cochlear compressive nonlinearity accompanied by some speech distortion and noise reduction analysis as well as PESQ and MOS test results. The CNSA-*f,t* estima-

Table 1: PESQ results for MMSE STSA, LSA and CNSA estimators.

		Noisy	MMSE	LSA	CNSA-*f*		CNSA-*t*	
		speech	STSA		$\beta_{\text{high}} = 0.2$	$\beta_{\text{high}} = -1.5$	$\beta_{\text{high}} = 0.2$	$\beta_{\text{high}} = -1.5$
white	0 dB	1.29	1.39	1.44	1.43	1.53	1.44	1.55
	5 dB	1.37	1.60	1.70	1.67	1.79	1.69	1.82
	10 dB	1.58	1.83	1.95	1.91	2.03	1.93	2.05
pink	0 dB	1.35	1.54	1.64	1.60	1.70	1.63	1.74
	5 dB	1.50	1.78	1.91	1.85	1.95	1.88	1.99
	10 dB	1.79	2.00	2.14	2.06	2.12	2.09	2.16
f16	0 dB	1.35	1.54	1.64	1.59	1.69	1.62	1.73
	5 dB	1.54	1.78	1.90	1.84	1.94	1.88	1.97
	10 dB	1.83	2.00	2.13	2.06	2.11	2.08	2.14

tors when $\beta_{\text{high}} = 0.2$ were not found to yield better results than LSA, however, when β_{high} was set to -1.5, CNSA-*f,t* yielded an advantage over both MMSE STSA and LSA by favoring less speech distortion in lower frequency regions where the speech energy is usually greater and may therefore mask the noise and, at the same time, performing more noise reduction at higher frequencies.

REFERENCES

[1] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*, Springer, 2005.

[2] X. Yang, K. Wang, and S. A. Shamma, "Auditory representations of acoustic signals," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 824–839, March 1992.

[3] R. Nobili, F. Mammano, and J. Ashmore, "How well do we understand the cochlea?," *Trends in Neurosciences (TINS)*, vol. 21, no. 4, pp. 159–167, 1998.

[4] L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," *Physiological Reviews*, vol. 81, no. 3, pp. 1305–1352, July 2001.

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

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

[7] C. H. You, S. N. Koh, and S. Rahardja, "β-order MMSE spectral amplitude estimation for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 4, pp. 475–486, July 2005.

[8] E. Plourde and B. Champagne, "Further analysis of the β-order MMSE STSA estimator for speech enhancement," in *Proc. 20th IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE)*, Vancouver, Canada, April 2007.

[9] C. H. You, S. N. Koh, and S. Rahardja, "An MMSE speech enhancement approach incorporating masking properties," in *Proc. ICASSP '04*, May 17-21 2004, vol. 1, pp. 725–728.

[10] E. A. Lopez-Poveda, C. J. Plack, and R. Meddis, "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," *J. Acoust. Soc. Am.*, vol. 113, no. 2, pp. 951–960, Feb. 2003.

[11] P. S. Rosengard, A. J. Oxenham, and L. D. Braida, "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," *J. Acoust. Soc. Am.*, vol. 117, pp. 3028–3041, May 2005.

[12] R. Meddis and L. P. O'Mard, "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.*, vol. 109, no. 6, pp. 2852–2861, June 2001.

[13] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Am.*, vol. 87, no. 4, pp. 1738–1752, 1990.

[14] D. D. Greenwood, "A cochlear frequency-position function for several species - 29 years later," *J. Acoust. Soc. Am.*, vol. 87, no. 6, pp. 2592–2605, June 1990.

[15] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, chapter Hearing, pp. 591–613, McGraw-Hill, Fourth edition, 2000.

[16] L. Brundin, A. Flock, and B. Canlon, "Sound-induced motility of isolated cochlear outer hair cells is frequency-specific," *Nature*, vol. 342, pp. 814–816, 1989.

[17] Rice University, "Signal processing information base: Noise data," [Online] Available http://spib.rice.edu/spib/select_noise.html, Accessed December 20, 2006.

[18] J. Sohn and N. S. Kim, "A statistical model-based voice activity detection," *IEEE Signal Processing Lett.*, vol. 6, no. 1, pp. 1–3, Jan. 1999.

[19] "ITU-T P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," 2003.