Modulation frequency domain adaptive gain equalizer using convex optimization

Rizwan Ishaq, Muhammad Shahid, Benny Lövström, Begona Garcıa Zapirain, Ingvar Claesson

6th International Conference on Signal Processing and Communication Systems (ICSPCS)

2012 Gold Coast, Australia

Modulation Frequency Domain Adaptive Gain Equalizer Using Convex Optimization

Rizwan Ishaq*, Muhammad Shahid**, Benny Lövström**, Begoña García Zapirain* and Ingvar Claesson**

*Dept. of Elec. Engineering, University of Deusto, Bilbao, Spain
**School. of Elec. Engineering, Blekinge Institute of Technology, Karlskrona, Sweden

***Abstract*—Adaptive gain equalizer (AGE) is a commonly used single-channel speech enhancement algorithm. AGE and its variants has been widely used for speech enhancement applications. There are two broad categories of these variants. The first deals with its improvement in time-frequency domain with readjustment of the used parameters and the second one deals with performing the main filtering operation in modulation frequency domain. This paper evaluates the working of AGE in modulation frequency domain with the use of a demodulation technique which solves the demodulation process as a convex optimization problem. The performance of the modified AGE is compared with the traditional AGE and another modulation frequency domain AGE based on demodulation using the spectral center-of-gravity. These used performance measures are Signal to Noise Ratio Improvement(SNRI), Spectral Distortion(SD) and Mean Option Score(MOS).**

***Index Terms*—Convex demodulation, Center of Gravity, filter bank, Adaptive Gain Equalizer.**

I. Introduction

Different types of background noise corrupts the otherwise clean speech signals in everyday communication. A phone call can be disturbed by a variety of noises present nearby ranging from computer fan noise to factory noise. There have been a variety of methods for reducing noise from speech signal, e.g., spectral subtraction [1] and optimum Wiener filtering [2]. The commonly used method for reducing noise is spectral subtraction but it has an inherent problem of generating musical noise due to spectral flooring [3]. There have also been some efforts to reduce this musical noise such as [4] but this improvement has the tendency of producing audible-distortion causing listening discomfort even compared to the unprocessed signal [5]. Reducing noise without generating artifacts was proposed in [6] but this method fails to address unvoiced speech.

The Adaptive gain equalizer (AGE) is a time domain speech enhancement algorithm in which the speech signal is amplified based on signal-to-noise (SNR) estimates in subbands. A signal is divided into subbands for calculation of a gain which is independent for each band. The algorithm has shown advantages over contemporary techniques because of its low complexity implementation, no requirement of voice activity detector (VAD) and has no presence of musical noise as a result of controlled gains [7]. Additionally, hardware implementations of AGE [8] indicate its importance in speech processing applications.

As an alternative to time domain processing, an implementation of AGE in the modulation domain was presented in a recent study [9]. This method was mainly inspired by the performance advantages of splitting the signal into its frequency bands. The modulation system assumes a speech signal as composed of a modulator and a carrier. Thus the signal is represented by

$$x(t) = m(t)c(t) \tag{1}$$

where $m(t)$ denotes the low frequency part of the signal, called the modulator, that modulates a high frequency carrier $c(t)$. Studies have shown that the modulators of a speech signal are more important for the intelligibility of the speech signals than their counterpart carriers [10]. Modulation systems are based on sub-band modulators and hence perfectly fit the AGE system which works on the sub-bands of the signal. Besides the fact that the study in [9] has reported improvement in performance measures in speech enhancement in comparison to time-domain AGE, the proposed center of gravity (COG) demodulation does not involve an optimization step, the need of which we state in the following.

In this work, we consider AGE in modulation domain by demodulation process as a convex optimization problem presented in [11]. The reason of adaptation of this technique for AGE in modulation domain is mainly the ambiguity associated with the demodulation process of having unlimited number of possible modulator-carrier pairs. Moreover, proven ability of this method for efficiently demodulating a variety of carriers such as harmonic, stochastic and time-varying ones further justifies its usage.

An account of related work in modulation domain and a brief introduction of AGE is provided in Section II. Section III describes a modulation system, a summary of a demodulation technique called spectral center of gravity that used in AGE implementation given in [9]. Section IV starts with an introduction of solving demodulation as an optimization problem and completes with the description of the proposed model of AGE. A comparison of performance of the proposed model is presented in Section V with its time-domain and modulation domain counterparts. Finally, some conclusive remarks about this work are drawn in Section VI with an outline of possible future works in the area.

II. Background

AGE can attenuate noise in speech signals in real time with low computational complexity [12]. It uses an FIR filter bank to divide a speech signal into subbands where speech in each subband is amplified independently. It was also shown that the system can adopt itself for different types of noise. The proposed AGE method using the mixed analog and digital

hybrid approach yield around 13 dB speech enhancement [13]. The AGE was originally intended for the digital domain, but [13] provides an analog implementation which does not use quantization and digitization and is best suited for battery powered applications. A hybrid solution to overcome problems related to a digital and an analog implementation of the AGE is found in [14].

Zadeh [15] introduced the modulation domain as a two dimensional bi-frequency system, where time variation of the ordinary frequency is the second dimension. Since then, there have been reasonably large interest in this field for various tasks related to speech processing. Atlas et al. used the concept of coherent modulation for the target talker enhancement in speech enhancement [16]. They proved that working in modulation domain can increase the speech intelligibility. Coherent modulation using the frequency reassignment has been used for speech enhancement and for demodulation of a signal into modulator and carrier [17]. Speech polluted by wind noise has been enhanced by using coherent modulation comb filtering as reported in [18]. Although the modulation filtering has mostly been used for the purpose of speech enhancement, we find some of its applications in audio compression as well [19]. It was showed that a 32 kb/s/channel outperformed MPEG-1 coded at 56 kb/s/channel (both at 44.1 kHz), using the modulation technique.

III. Modulation Domain and AGE

An acoustic spectrum is transformed by short-time Fourier transform into the modulation domain spectrum at a particular acoustic frequency. It has been observed that speech intelligibility can be altered by operating on modulator part of the signal. Shamma [20] reported that auditory cortex neurons possibly decompose the acoustic contents into spectro-temporal modulation contents. It has been found that if the modulators of the speech signal are replaced by constant amplitude modulators, while carriers are preserved, speech does not remain intelligible anymore. However, when the modulators are preserved but carriers are altered, the speech is intelligible [10]. A modulation frequency system is described by the following steps:

- Filter bank to get sub-band signals
- Demodulation i.e., decomposition of each sub-band signal into a modulator and a carrier.
- Analysis of the modulators of the sub-band signals by discrete Fourier transform of each modulators
- Modification of the modulators (e.g. linear filtering)
- Re-modulation (recombination of modified modulators with original carriers)
- Synthesis of signals

The modulation system's filter bank divides the wide-band signal into K narrow-band sub-bands. The signal $x(t)$ is passed through the filter bank set of band-pass filters h_k, which renders the sub-band signals $x_k(t)$.

$$x_k(t) = h_k(t) * x(t) \tag{2}$$

s where $*$ is convolution operator. The demodulation process decomposes the sub-band signal into its envelope and carrier. It is efficient to decimate the sub-band signals so that the redundant samples may be removed. Modification of the modulators is done by the modulation filtering $g(t)$, i.e., $\tilde{m_k}(t) = m_k(t)g(t)$. A modulation spectrogram and modulation analysis can be done by computing the Fourier transform along the time-axis of the spectrogram (magnitude) or by utilizing the spectrum of the envelop signals, which gives the modulation frequency along the horizontal axis and acoustic frequency along the vertical axis. Re-modulation is the process in which modified modulators $\tilde{m_k}(t)$ are combined with the original carriers, obtained in the process of demodulation, to get the modified sub-band signals $\tilde{x}_k(t)$. The synthesis process reconstructs the modified signal $\tilde{x}(t)$ using the modified sub-band signals $\tilde{x}_k(t)$, according to the following equation. Interpolation must be performed prior to this stage if decimation was done before.

$$\tilde{x}(t) = \sum_{k=1}^{K} \tilde{x}_k(t) \tag{3}$$

Following is a brief description on one of the methods used for coherent carrier detection which is also used in this work, apart from convex optimization demodulation process.

A. Spectral Center of Gravity Carrier Estimation

In the Center-of-Gravity(CoG) approach, instantaneous frequency $\omega_k(n)$ is defined as instantaneous spectrum average frequency of $x_k(t)$ at time t [21]. An instantaneous spectrum with short-time Fourier transform is computed as,

$$S_k(\omega, t) = \sum_p g(p)x_k(t+p)e^{-j\omega p} \tag{4}$$

where $g(p)$ is a short spectral-estimation window. The instantaneous frequency $\omega_k(t)$ of the sub-band signal $x_k(t)$ is estimated as,

$$\omega_k(t) = \frac{\int_{-\pi}^{\pi} \omega |S_k(\omega, t)|^2 d\omega}{\int_{-\pi}^{\pi} |S_k(\omega, t)|^2 d\omega} \tag{5}$$

The phase $\phi_k(t)$ of the carrier is computed as follows

$$\phi_k(t) = \sum_{p=0}^{t} \omega_k(p) \tag{6}$$

The carrier c_k is

$$c_k(t) = e^{j\phi_k(t)} \tag{7}$$

and the complex valued modulator $m_k(t)$ is given by

$$m_k(t) = x_k(t)c_k^*(t) \tag{8}$$

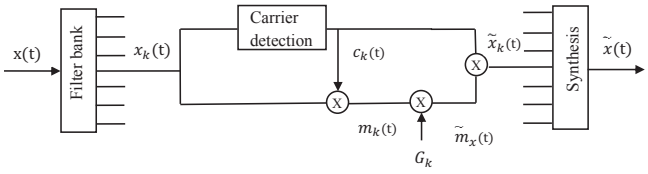

Fig. 1. Adaptive gain equalizer in modulation domain

B. Adaptive Gain Equalizer System

The AGE consists of a filter bank and each sub-band is weighted by a gain function which amplifies the signal when speech is present and keeps the noisy part of the signal, where no speech is present, to unity [7]. A filter bank of K bandpass filters divides the input signal $x(n)$ into K sub-bands $x_k(n)$.

$$x_k(n) = h_k(n) * x(n) \tag{9}$$

Here h_k is the impulse response of the filter bank sub-band k and $*$ denotes the convolution. The output signal $\tilde{x}(n)$, with the amplified speech signal, is computed as

$$\tilde{x}(n) = \sum_{k=1}^{K} G_k(n) x_k(n) \tag{10}$$

where $G_k(n)$ is the AGE weighting function which amplifies the signal when speech is active and is given by

$$G_k(n) = min\left\{ \left(\frac{A_k(n)}{L_{opt}.B_k(n)} \right)^{p_k}, L_k \right\} \tag{11}$$

where L_{opt} is the optimized suppression level for gain function and p_k gain rise exponent constant. L_k is a limiting threshold limiting gain function value. Fast average $A_k(n)$ and slow average $B_k(n)$ of sub-band k calculated according to:

$$A_k(n) = \alpha_k A_k(n-1) + (1-\alpha_k)|x_k(n)| \tag{12}$$

where $\alpha_k = \frac{1}{f_s T_a}$ is forgetting factor constant and f_s is sampling frequency.

$$B_k(n) = \begin{cases} A_k(n) & \text{if } A_k(n) \leq B_k(n-1) \\ (1+\beta_k)(B_k(n-1) & \text{otherwise} \end{cases} \tag{13}$$

where $\beta_k = \frac{1}{f_s T_b}$ is a positive constant control the noise level. Based on the above mentioned principle of AGE, a speech signal modulator can also be enhanced by the equalizer. Modulation domain separates each sub-band signal into a carrier and a modulator. While only modulators are considered here, the AGE is implemented on each modulator to enhance the speech. The system is shown in figure 1. The mathematics for AGE in the modulation domain is the same as for AGE in the sub-band domain, the long term average and the short term average are calculated for each sub-band modulator, instead of the sub-band itself. The gain function is multiplied with the modulator of the sub-band to yield a modified modulator $\tilde{m}_k(n)$ which is then used with the carrier in the reconstruction stage of the modulation system.

$$\tilde{m}_k(n) = m_k(n) G_k \tag{14}$$

$$\tilde{x}_k(n) = c_k(n) \tilde{m}_k(n) \tag{15}$$

The synthesized signal $y(n)$ is finally calculated by adding up all the components.

$$\tilde{x}(n) = \sum_{k=1}^{K} \tilde{x_k}(n). \tag{16}$$

IV. Convex Optimization and the Proposed Model

One inherent problem with the demodulation technique is the unfortunate presence of unlimited number of possible yet valid modulator-carrier pairs. This predicament can be understood by taking example of a sinusoidal signal that is composed of multiple frequency sinusoids. Such a signal can be decomposed into more than one legitimate modulator and carrier pairs, that are equally correct mathematically. Similar is the case with speech signals when the problem of demodulating it into modulator and carrier is dealt. Thus there is need to add some conditions to the problem which can make the algorithm result into the desired solution. A general optimization problem minimizes a given objective function while fulfilling a set of equality and unequally constraints. If the objective function and inequality constraints are all convex and the equality constraints are all affine, the problem is called a convex optimization problem [22]. Although the modulation problem of equation 1 is not convex as it is, two methods have been suggested in [11] for constraining modulation into convex restrictions. One solution is to work in logarithm domain where the optimization variables can be defined simply as the logarithm of the squared linear optimization variables $m(t)$ and $c(t)$. A convex relation is then obtained by just summing the two logarithmic domain variables. The other method of making the problem convex is to work in linear domain where the process involves eliminating the carrier $c(t)$ and minimization of only the modulator signal is done. The final expression obtained in linear domain convex optimization is given by the following:

Minimize $C_m(m(t)) + C_c(m(t)^{-1}x(t))$

where the modulator cost function C_m can be any convex function but the carrier cost function C_c must be both convex and non-decreasing as a requirement of making the problem a convex one. We have followed the linear domain convex optimization method in our work. The interested reader is referred to [11] for detailed analysis of these methods.

V. Comparative Performance Analysis

A. Mean Opinion Score(MOS)

The Mean Opinion Score (MOS) calculated by observing the clean speech signal processed by a system to check how much it degrades the clean speech signal. Fig. 2 shows a female speech signal processed by a system where SNR has been set -10dB for both Engine Noise (EN) and Factory Noise (FN). The system with convex demodulation has MOS value around 3.5 for EN and 3.8 for FN which provides

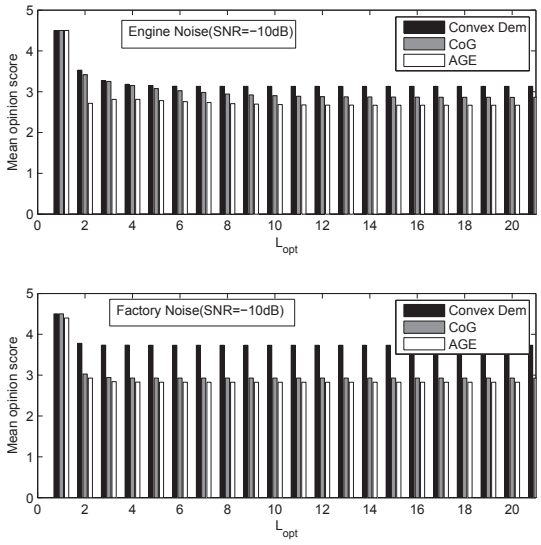

Fig. 2. Mean Opinion Scores(MOS) for all systems with SNR=-10dB

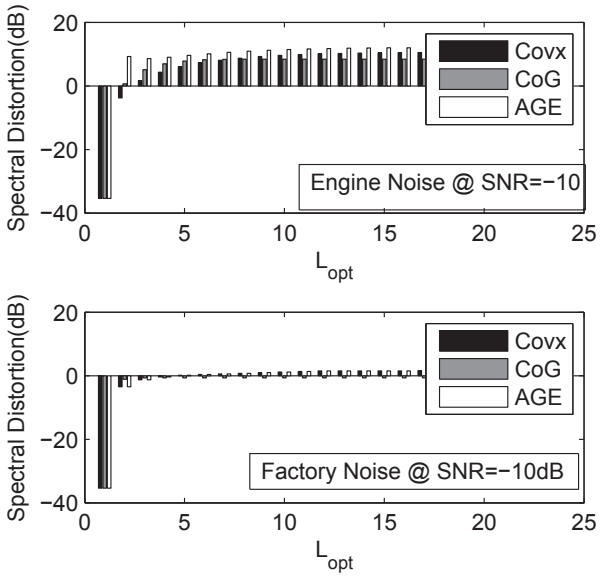

Fig. 3. Spectral Distortion with SNR=-10dB

less degradation as compare to CoG modulation and AGE system where is average MOS observed 3, and less than 3, respectively.

B. Spectral Distortion

Fig.3 shows the Spectral Distortion(SD) for female speech signal contaminated by EN and FN at the SNR of -10dB. The increasing value of L_{opt} increases SD up to 10dB for EN when the system uses AGE while for convex demodulation average SD around 7dB and for CoG demodulation its around 9dB, but for FN, SD for all the system observed around 3dB average.

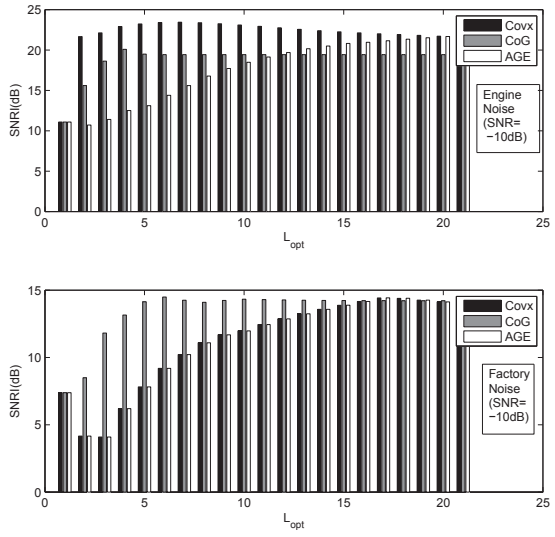

Fig. 4. Signal to Noise Ratio Improvement(SNRI) with SNR=-10dB

C. Signal to Noise Ratio Improvement(SNRI)

Fig. 4 shows the Signal to Noise Ratio Improvement (SNRI) for AGE, MAGE (CoG and Convex demodulation) for a female speech signal distorted by EN and FN having SNR of -10dB. The MAGE methods with convex demodulation has the highest SNRI for all the values of L_{opt} and around 5dB and 8dB improvement over the AGE and MAGE (CoG) methods for EN. But for FN system show improvement after $L_{opt} = 12$. The MAGE (CoG) in start improved significantly but with increasing value of L_{opt} MAGE (Convex demodulation) has better improvement.

D. Spectrogram Analysis

Fig. 5 and 6 shows spectrogram of original signal with processed signal with AGE, MAGE (convex and CoG demodulation) for FN and EN respectively. The MAGE (convex demodulation) improvement can be observed in term of speech formants being not effected, as visible in spectrogram for both EN and FN.

VI. Conclusion

An alternative method of demodulation has been proposed for AGE in the modulation frequency domain. The presented method solves the demodulation process as a convex optimization problem, thereby avoiding the inherent problem of multiple solutions of a demodulation algorithm. We have tested the proposed method for various conditions and magnitudes of noise injected in a clean speech signal. The performance of our method has been validated by mean opinion score, spectral distortion, signal to noise ratio improvement and spectrogram analysis in comparison to two other techniques.

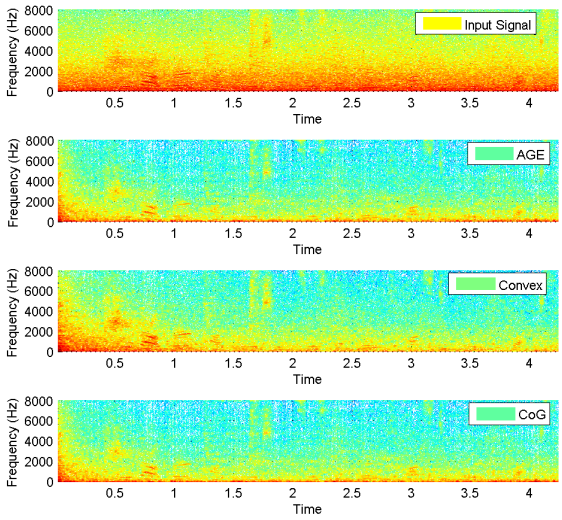

Fig. 5. Spectrogram with Factory Noise(FN) (SNR=-10dB)

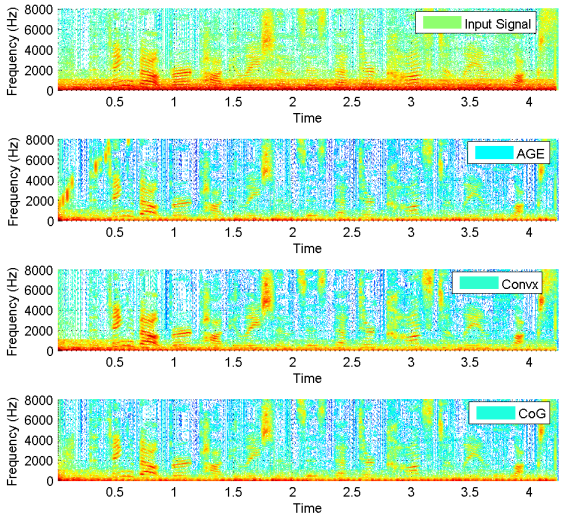

Fig. 6. Spectrogram with Engine Noise(FN) (SNR=-10dB)

REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE trans. Accoust. Speech and Sig. Proc.*, vol. 27, no. 2, pp. 113–120, 1979.

[2] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1996.

[3] Z. Goh, K.-C. Tan, and T. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," *Speech and Audio Processing, IEEE Transactions on*, vol. 6, no. 3, pp. 287 –292, may 1998.

[4] 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, dec 1984.

[5] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Musical noise generation analysis for noise reduction methods based on spectral subtraction and mmse stsa estimation," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, april 2009, pp. 4433 –4436.

[6] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 6, pp. 2098 –2108, nov. 2006.

[7] N. Westerlund, M. Dahl, and I. Claesson, "Speech enhancement for personal communication using an adaptive gain equalizer," *Elsevier Signal Processing.*, vol. 85, pp. 1089–1101, 2005.

[8] B. Sällberg, N. Grbic, and I. Claesson, "Implementation aspects of the adaptive gain equalizer," 2006.

[9] M. Shahid, R. Ishaq, B. Sällberg, N. Grbic, B. Lövström, and I. Claesson, "Modulation domain adaptive gain equalizer for speech enhancement," in *Signal and Image Processing Application 2011, by IASTED*, 2011.

[10] S. Schimmel, "Theory of modulation frequency analysis with applications to hearing devices," *Ph.D. dissertation*, 2007.

[11] G. Sell and M. Slaney, "Solving demodulation as an optimization problem," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 8, pp. 2051 –2066, nov. 2010.

[12] N. Westerlund, M. Dahl, and I. Claesson, "Real-time implementation of an adaptive gain equalizer for speech enhancement purposes," *WSEAS.*, 2003.

[13] M. Dahl, I. Claesson, B. Sällberg, and H. Akesson, "A mixed analog -digital hybrid for speech enhancement purposes," *ISCAS.*, 2005.

[14] M. Dahl and B. Sällberg, "Speech enhancement implementations in the digital, analog and hybrid domain," *Swedish System on Chip Conference*, 2005.

[15] L. Zadeh, "Frequency analysis of variable networks," in *Proc. IRE*, vol. 38, no. 3, Mar. 1950, pp. 291–299.

[16] S. Schimmel and L. Atlas, "Target talker enhancement in hearing devices," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008-april 4 2008, pp. 4201 –4204.

[17] S. M. Schimmel, K. R. Fitz, and L. Atlas, "Frequency reassignment for coherent modulation filtering," *IEEE, Acoustics, Speech and Signal Processing, ICASSP*, vol. 5, pp. 261–264, 2006.

[18] B. King and L. Atlas, "Coherent modulation comb filtering for enhancing speech in wind noise," *International Workshop on Acoustice Echo and Noise Control*, Sep 2008.

[19] M. S. Vinton and L. Atlas, "A scalable and progressive audio codec," *IEEE, Acoustics, Speech and Signal Processing, ICASSP*, vol. 5, pp. 3277–3280, 2001.

[20] S. Shamma, "Encoding sound timbre in the auditory system," *IETE J. Res.*, vol. 49, no. 2, pp. 193–205, 2003.

[21] P. Clark and L. E. Atlas, "Time-frequency coherent modulation filtering of non-stationary signals," *IEEE transaction on Signal Processing*, vol. 45, no. 57, pp. 4323–4332, 2009.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge, UK: Cambridge University Press, 2004.