PERFORMANCE IMPROVEMENT OF SPREAD SPECTRUM ADDITIVE DATA HIDING OVER CODEC-DISTORTED VOICE CHANNELS

M. Boloursaz, student member, IEEE, R. Kazemi*, F. Behnia*, and M. A. Akhaee†*

*Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

†School of Electrical and Computer Eng., College of Eng., University of Tehran, Tehran, Iran

Email: Boloursaz@ee.sharif.edu

ABSTRACT

This paper considers the problem of covert communication through dedicated voice channels by embedding secure data in the cover speech signal utilizing spread spectrum additive data hiding. The cover speech signal is modeled by a Generalized Gaussian (GGD) random variable and the Maximum A Posteriori (MAP) detector for extraction of the covert message is designed and its reliable performance is verified both analytically and by simulations. The idea of adaptive estimation of detector parameters is proposed to improve detector performance and overcome voice non-stationarity. The detector's bit error rate (BER) is investigated for both blind and semi-blind cases in which the GGD shape parameter needed for optimum detection is either estimated from the stego or cover signal respectively. The simulation results also show that the proposed method achieves acceptable robustness against the lossy compression attack by different compression rates of Adaptive Multi Rate (AMR) voice codec.

Index Terms— Data Hiding, Generalized Gaussian Distribution (GGD), Maximum A Posteriori (MAP) detector, Adaptive Multi Rate (AMR) compression attack.

1. INTRODUCTION

Nowadays, with the spread of different voice communication systems such as the Public Switched Telephone Network (PSTN), Voice Over Internet Protocol (VOIP) and cellular networks (GSM, 3G, 4G, ...), voice calls are widely available ubiquitously. Such voice communication channels can be regarded as a priceless available infrastructure for covert data communication. Apart from this, the covert data communication channel established by Data Hiding over voice benefits the advantages of wide coverage, good service availability and QoS due to the large span of cellular voice channels and higher priority of voice traffic over data traffic.

On the other hand, end to end secure message transmission through cellular networks is of high interest due to security threats present in such networks. For example, various possible cases of eavesdropping on the widely operational GSM network has been reported and it is evident that GSM network operators and third party attackers can easily eavesdrop all parlances. Some of the mentioned security threats of the GSM network are described below:

Firstly, due to the weakness of the utilized A5/1 and A5/2 encryption algorithms, online and offline cracking of the cipher is possible and there is the possibility of eavesdropping voice calls and SMS messages.

Secondly, in GSM network, data encryption is applied only to the data transmitted between BTS (Base Transceiver Station) and ME (Mobile Equipment) via air, while inside the GSM network infrastructure and between MSC (Mobile Switching Center), BSC (Base Station Controller) and BTS (Base Transceiver Station) data is not encrypted. Hence, mobile network operators can easily eavesdrop voice calls and SMS messages. In other words, end to end encryption is not provided by GSM network.

In this research, hiding the covert message in the voice signal passing through the voice channel is proposed as a way of end to end secure message transmission. Figure 1 depicts the overall block diagram of the proposed method. In this method, the covert message is compressed, encrypted, channel coded and finally embedded in voice as the cover signal. This data embedded voice (stego signal) is then transmitted to the receiver through the voice channel attracting minimal attention of the eavesdropper.

One limitation in covert communication through voice dedicated channels is the restricted available bandwidth of 4000 Hz which inherently limits the achievable embedding rate. Also, there exists the major attack of speech compression to the stego signal. The speech compression is done using low rate voice codecs that apply lossy compression algorithms based on parametric model of analysis by synthesis of the speech signal. These vocoders also provide a nonlinear channel with unmodeled memory as described in [1-3] which makes data hiding in the mentioned channel more complicated. Hence, the problem is to propose a data hiding method which is more robust to the

This paper is supported and funded by Iran's National Elites Foundation (INEF).

considered vocoder compression attack while maintaining perceptual quality of the voice signal and maximizing the data embedding capacity.

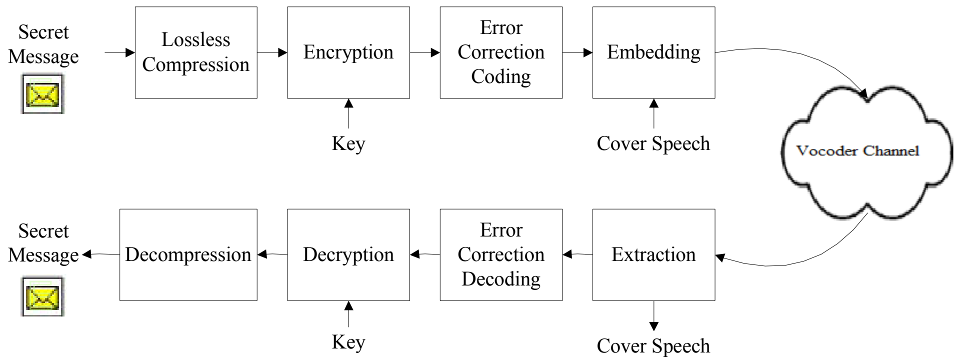

Fig. 1 The Overall Block Diagram of the Proposed Covert Communication Method through Voice Channels

A couple of prior works on data hiding over codec distorted voice channels have suggested the secret message to be hidden in less significant output bits of the vocoder [4-8].

However, this may be impossible in cases like cell phone devices where the codec is implemented by hardware on electronic chips that refuse access to the compressed bits.

Xiao et. al. suggests the Quantization Index Modulation (QIM) method for data hiding over vocoders utilizing Linear Prediction Coding (LPC) speech analysis method in [9]. Chen et. al. suggests the multiplicative Spread Spectrum (SS) data hiding in wavelet domain for covert communication through vocoder channel in [10]. Finally, [11, 12] suggests the Least Significant Bit (LSB) data hiding and its variants for data embedding in VOIP systems.

In this research the method of additive Spread Spectrum (SS) data hiding [13, 14, 15, 16] is used to embed covert data in the voice signal. An optimum detector for extracting the covert message based on speech probability distribution is proposed and its performance is evaluated both analytically and by simulation experiments on Adaptive Multi Rate (AMR) vocoder channel simulator [17, 18].

The rest of this paper is organized as follows. Section 2 illustrates the proposed data embedding method. Section 3 presents the designed optimum detector for extraction of the covert data. Section 4 presents and discusses the achieved simulation results and finally section 5 concludes the paper and considers directions for further research.

2. ADDITIVE SPREAD SPECTRUM DATA EMBEDDING

In this research Spread Spectrum (SS) additive data embedding [13, 14, 15, 16] method is proposed for covert message communication through the voice codec due to its better resistance against the vocoder compression attack. In this method data embedding is done according to (1)

$$\underline{s_k} = \underline{c_k} + b_k \alpha_k \, \underline{w} \tag{1}$$

In which $\underline{c_k}$ and $\underline{s_k}$ are vectors of length l and correspond the $k'th$ frame of cover and stego signals respectively. Also $\underline{w}$ is the vector of pseudo random watermark of length l. The embedding coefficient α_k is small enough to make the presence of the watermark completely intangible. It is adjusted adaptively based on the power of the cover signal in each speech frame such that the Signal to Watermark Ratio (SWR) remains constant. Finally b_k is set to -1 for transmitting a 0 and $+1$ for transmitting a 1.

3. OPTIMUM DETECTOR

According to [19, 20], speech probability distribution can be modeled by Generalized Gaussian Density (GGD) function given by (2)

$$f_x(x) = \frac{\theta}{2\beta\Gamma(\frac{1}{\theta})} e^{-\left(\frac{|x-\mu|}{\beta}\right)^{\theta}} \tag{2}$$

In which $\Gamma(.)$ denotes the gamma funciton and μ, β and θ are locaion, scale and shape parameters respectively. The GGD probability density function (PDF) spans a large number of symmetric PDFs keeping the Laplacian, Gaussian, Uniform and Impulse densities as special cases for $\theta = 1, 2, \infty, 0^+$ respectively.

In this research, the cover speech distribution is modeled by GGD density function. By setting $\underline{w}' = \alpha_k \times \underline{w}$, the $i'th$ sample of the $k'th$ stego frame s_{ki} has GGD distribution with a mean of $-w_i'$ or w_i' depending on the embedded bit. Also assuming Independent Identical Distribution (iid) for speech samples, a posteriori probability distributions are calculated as

$$\underline{s_k} = \underline{c_k} \underset{0}{\overset{1}{\pm}} \underline{w}'$$

$$f(\underline{s}_k|b_k = 1) = \left(\frac{\theta}{2\beta\Gamma\left(\frac{1}{\theta}\right)}\right)^l \times \prod_{i=1}^{l} e^{-\left(\frac{|s_{ki}-w_i'|}{\beta}\right)^\theta} \quad (3)$$

$$f(\underline{s}_k|b_k = -1) = \left(\frac{\theta}{2\beta\Gamma\left(\frac{1}{\theta}\right)}\right)^l \times \prod_{i=1}^{l} e^{-\left(\frac{|s_{ki}+w_i'|}{\beta}\right)^\theta}$$

Assuming equal transmission probability for 0 and 1 (P_0 and P_1), the Maximum A Posteriori (MAP) detection leads to Maximum Likelihood Detector (MLD). Applying this assumption, the MLD decision rule is derived as (4)

$$\frac{f(\underline{s}_k|b_k = 1)}{f(\underline{s}_k|b_k = -1)} \underset{0}{\overset{1}{\gtrless}} \frac{P_1}{P_0} = 1$$

$$\sum_{i=1}^{l}\left(\frac{|s_{ki} - w_i'|}{\beta}\right)^\theta \underset{1}{\overset{0}{\gtrless}} \sum_{i=1}^{l}\left(\frac{|s_{ki} + w_i'|}{\beta}\right)^\theta \quad (4)$$

$$\sum_{i=1}^{l}|s_{ki} - w_i'|^\theta \underset{1}{\overset{0}{\gtrless}} \sum_{i=1}^{l}|s_{ki} + w_i'|^\theta$$

The above decision rule considers $\underline{w}'$ (or equivalentlyα) and θ to be known at the receiver. This assumption holds if the algorithm is considered semi-blind. The algorithm's performance for both the semi-blind and blind cases are reported in section IV.

According to [20], the shape parameter θ is estimated about 1.5 for speech frame lengths of about 50ms (equal to 400 speech samples at a sampling frequency of 8 KHz). Hence, if little performance degradation is tolerated, the speech probability distribution can be assumed Gaussian $\theta = 2$ to simplify the proceeding calculations. For $\theta = 2$ the decision rule derived in Equ. 4 simplifies to the matched filter receiver given by (5)

$$\underline{s}_k \cdot \underline{w} \underset{0}{\overset{1}{\gtrless}} 0 \quad (5)$$

Calculating the exact detection error probability is complicated for the case of general θ, but placing $\theta = 2$, an estimate on the error probability is obtained as follows. A transmitted 0 is detected in error if and only if (6)

$$\sum_{i=1}^{l}(c_{ki} - w_i')w_i > 0$$

$$\sum_{i=1}^{l} c_{ki} \times w_i > \alpha_k \sum_{i=1}^{l} w_i^2 \quad (6)$$

As l is large enough and mostly larger than 100 in this application, the Central Limit Theorem (CLT) can be applied and $z = \sum_{i=1}^{l} c_{ki} \times w_i$ can be approximated by a Gaussian distribution $N(0, \sigma_c^2 \sum_{i=1}^{l} w_i^2)$, in which σ_c^2 is the variance of the cover voice signal. Hence, the Gaussian upper bound on detection error probability is given by (7)

$$P_e = P_{e0} \le Q\left(\frac{\alpha_k \sum_{i=1}^{l} w_i^2}{\sigma_c \times \sqrt{\sum_{i=1}^{l} w_i^2}}\right) = Q(\frac{\sqrt{l}}{\sqrt{SWR}}) \quad (7)$$

$$SWR = \frac{l \times \sigma_c^2}{\alpha_k^2 \sum_{i=1}^{l} w_i^2}$$

It should be stressed that as stated in [19], the actual value of the shape parameter θ varies for different recorded speech samples according to the recording SNR, and the percentage of the silence periods present in that voice sample. It also varies during a single observation as the cover speech changes between voiced, unvoiced and silence periods and causes the voice PDF to change between Gaussian, Laplacian and Gamma. The fact of cover non-stationarity seems to be overlooked by previous researches on spread spectrum Data Hiding [13, 14, 15]. In this research, we propose the idea of adaptive estimation of the shape parameter to cope with this phenomenon. This adaptive estimation may be carried out from the cover voice at the transmitter (semi-blind case) or from the stego signal at the receiver (blind case). These two cases are explained more accurately in section IV.

4. SIMULATION RESULTS

To access the Bit Error Rate (BER) performance of the proposed data embedding method against the AMR codec's compression attack, statistically efficient amount of random data was embedded in cover signals taken from TIMIT [21] which is a database of high quality recorded human speech. The achieved stego signal is then passed through different codec data rates of the AMR channel simulator [17, 18] and finally, the data is extracted from the channel output signal.

The empirical Bit Error Rate (BER) curves against the SWR parameter achieved by this method are depicted in Fig. 2 for different frame lengths. It should be noted that the covert data transmission rate for each frame length is $R = \frac{F_s}{l} = \frac{8000}{l}\ bps$. In this figure, the embedded data are extracted from the original stego signal and the reported BERs are not subject to vocoder's compression attack. The BERs are reported for 3 cases of

a) Utilizing the optimum detection rule (4) in the semi-blind case.

b) Utilizing the optimum detection rule (4) in the blind case.

c) Utilizing the suboptimum matched filter detector (Correlation demodulator for $\theta = 2$ assumption).

It should be noted that the Moment Matching Estimator (MME) proposed in [20, 22] is used to estimate the θ parameter for the cases *a* and *b*. In case *a* (the semi-blind case) this parameter estimation method is applied to the cover signal at the transmitter and the estimated θ parameter is passed to the receiver, however in case b (the blind case), the receiver itself estimates θ by application of this method on the received stego signal.

It should also be noted that in the semi-blind case, the cover signal's variance σ_c^2 is transmitted to the receiver and it will be able to calculate the α_k parameter needed for optimum detection by utilizing (7), however in the blind case the receiver itself estimates σ_c^2 from the received stego signal.

Similarly, Fig. 3 shows the same results for AMR codec data rate of 12.2 Kbps.

It is observed from figures 2 and 3 that larger frame lengths lead to more significant performance improvements achieved by blind and semi-blind detectors in comparison with the conventional matched filter. This is due to the fact that larger l leads to more accurate estimates of the θ parameter which in turn improves system performance. Another observation is that the performance of the blind detector approaches that of the semi-blind detector for larger SWRs and codec data rates. This can similarly be described as the effect of a more accurate estimate of the GGD shape parameter. As known, in the blind case this parameter is estimated from codec distorted stego signal instead of the original cover voice. Hence less power of the added watermark (larger SWR) and less distorted voice signal (larger codec data rate) leads to more accurate estimates of the original shape parameter and improves the BER for blind detection.

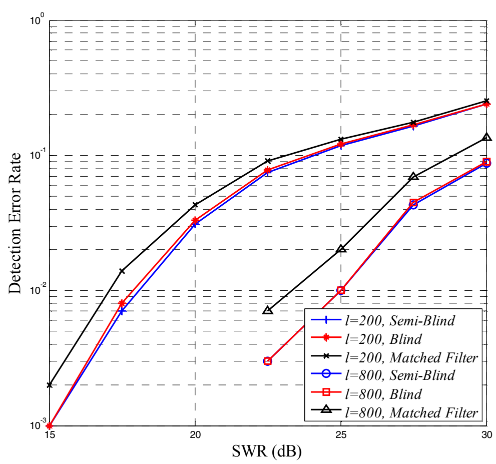

Fig. 2 BER Performance of the Proposed Method for Different Embedding Rates (Without Vocoder Compression Attack)

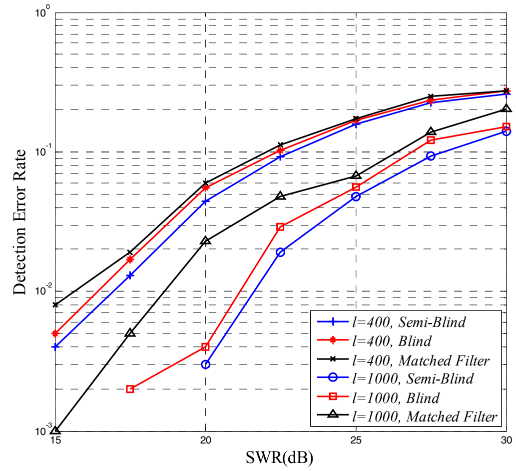

Fig. 3 BER Performance of the Proposed Method for Different Embedding Rates (AMR 12.2 Compression Attack)

Figure 4 depicts the BER performance of the proposed semi-blind detector against SWR for different compression rates of AMR codec and a frame length of 800 which is equivalent to a data embedding rate of 10 bps. This figure shows that a $SWR = 20\,dB$ maintains an acceptable performance ($BER < 0.1$) as a feasible solution for all codec data rates. To gain an insight of the stego signal's subjective quality for this feasible case, the PESQ measure (Perceptual Evaluation of Speech Quality) introduced by [23] is calculated as 4.0, 4.1, 4.1 and 3.8 for the four cases of "No Vocoder", "AMR 12.2", "AMR 7.4" and "AMR 4.75" respectively. Finally it should be noted that the watermark $\underline{w}$ is a Gaussian noise filtered in [500-3700] Hz to lie in the codec's natural voice transmission frequency range. If the lower limit of 500 Hz is decreased, the watermark's spectrum will overlap significant frequency components of the original voice and degrade its perceptual quality (latency). On the other hand, if it is increased, the method becomes less robust against a simple low pass filtering attack.

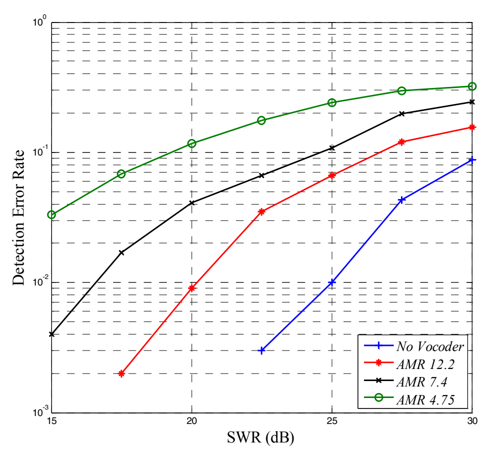

Fig. 4 BER Performance of the Proposed Method for Semi-Blind Detection and 10 bps Data Rate

It also seems obvious that by increasing the number of watermarks used ($\underline{w}_1, \underline{w}_2, \ldots$), the data embedding capacity can be increased, however at the cost of increased detector complexity.

5. CONCLUSION AND FUTURE WORKS

A covert communication channel through codec distorted voice channels was proposed by embedding secure data in the cover speech. The method of spread spectrum additive data hiding was utilized and its optimum detector was derived by modeling the speech signal as a GGD random variable. A major contribution of the authors is to propose the idea of adaptive estimation of the detector parameters to cope with cover voice non-stationarity. The detector's bit error rate (BER) performance was investigated for both blind and semi-blind detections. Simulation results also show that the proposed data hiding method achieves acceptable robustness against the lossy compression attack by different compression rates of AMR voice codec. Finally, this research can be extended by deriving the optimum detector for the case of multiple watermarks used to boost embedding capacity and calculating theoretical bounds on its detection error probability.

8. REFFERENCES

[1] V. Sapozhnykov, and S. Fienberg, “A Low-rate Data Transfer Technique for Compressed Voice Channels,” *Journal of Signal Processing Systems*, vol. 68, no. 2, pp. 151-170, 2012.

[2] M. Boloursaz, A. H. Hadavi, R. Kazemi and F. Behnia, “Secure Data Communication through GSM Adaptive Multi Rate Voice Channel,” *6th International Symposium on Telecommunicaions (IST 2012),* pp. 1021-1026, November 2012.

[3] M. Boloursaz, A. H. Hadavi, R. Kazemi, F. Behnia, "A data modem for GSM Adaptive Multi Rate voice channel," *East-West Design & Test Symposium,* pp. 1-4, Sept. 2013.

[4] Hui Tian, Ke Zhou, Hong Jiang, Jin Liu, Yongfeng Huang, Dan Feng, "An M-Sequence Based Steganography Model for Voice over IP," *IEEE International Conference on Communications (ICC '09)*, pp.1-5, June 2009.

[5] C. Kratzer, J. Dittmann, T. Vogel, R. Hillert, "Design and evaluation of steganography for voice over IP," *Proceedings IEEE International Symposium on Circuits and Systems (ISCAS '06)*, May 2006.

[6] Chungyi Wang, Quincy Wu, "Information Hiding in Real-Time VoIP Streams," *Ninth IEEE International Symposium on Multimedia (ISM '07),* pp. 255-262, Dec. 2007.

[7] Takahashi Takehiro, Lee Wenke, "An assessment of VoIP covert channel threats," *Third International Conference on Security and Privacy in Communications Networks and the Workshops*, pp. 371-380, Sept. 2007.

[8] Yong Feng Huang, Shanyu Tang, Jian Yuan, "Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 296-306, June 2011.

[9] Bo Xiao, Yong Feng Huang, Shanyu Tang, "An Approach to Information Hiding in Low Bit-Rate Speech Stream," *IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 1-5, 2008.

[10] Yun-feng Chen, Tao Li, Dong-ni Gao, Xiao-qin Hu, Xiao-ping Zhang, Jin Liu, "A secure mobile communication approach based on information hiding," *2nd International Conference on Mobile Technology, Applications and Systems,* Nov. 2005.

[11] Hui Tian, Ke Zhou, Yongfeng Huang, Dan Feng, Jin Liu, "A Covert Communication Model Based on Least Significant Bits Steganography in Voice over IP," *The 9th International Conference for Young Computer Scientists (ICYCS '08),* pp. 647-652, Nov. 2008.

[12] Hui Tian, Ke Zhou, Hong Jiang, Yongfeng Huang, Jin Liu, Dan Feng, "An adaptive steganography scheme for voice over IP," *IEEE International Symposium on Circuits and Systems (ISCAS 2009)*, pp. 2922-2925, May 2009.

[13] I. J. Cox, J. Kilian, F. T. Leighton and T. Shanon, "Secure spread spectrum watermaing for multimedia," *IEEE Transaction on Image processing,* vol. 6, no. 12, pp. 1673-1678, Dec 1997.

[14] L. M. Marvel, C. G. Boncelet, C. T. Retter, "Spread spectrum image steganography," *IEEE Transactions on Image Processing*, vol. 8, no. 8, pp. 1075-1083, Aug 1999.

[15] A. Valizadeh, J. Wang, "Correlation-and-Bit-Aware Spread Spectrum Embedding for Data Hiding," *IEEE Transactions on Information Forensics and Security,* vol. 6, no. 2, pp. 267-282, June 2011.

[16] D. Kirovski, H. S. Malvar, "Spread-Spectrum Watermarking of Audio Signals," *IEEE Transactions on Signal Processing,*, vol. 51, no. 4, pp. 1020-1033, Apr 2003.

[17] ETSI, Digital Cellular Telecommunications Systems (Phase 2+), Adaptive Multi-Rate (AMR) speech transcoding, (GSM 06.90 version 7.2.1 Release 1998).

[18] ETSI, Digital Cellular Telecommunications Systems (Phase 2+), Adaptive Multi-Rate (AMR) speech transcoding, ANSI-C code for AMR speech codec, (GSM 06.90 version 7.2.1 Release 1998).

[19] S. Gazor and Z. Wei, "Speech Probability Distribution," *IEEE Signal Processing Letters,* vol. 10, no. 7, pp. 204-207, July 2003.

[20] Kostas Kokkinakis, and Asoke K. Nandi, “Speech Modeling Based On Generalized Gaussian Probability Density Functions”, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 1, pp. 381-384, March 2005.

[21] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren and V. Zue, "TIMIT Acoustic-Phonetic Continuous Speech Corpus," Philadelphia, 1993.

[22] T. Wang, H. Li, Z. Li, Z. Wang, "A Fast Parameter Estimation of Generalized Gaussian Distribution," *8th International Conference on Signal Processing,* vol. 1, pp. 16-20, 2006.

[23] Yi Hu, P.C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 16, no. 1, pp. 229-238, Jan. 2008.