DATA HIDING IN SPEECH USING PHASE CODING

Yasemin Yardimci[†], A. Enis Çetin[†] and Rashid Ansari[††]

[†]*Department of Electrical and Electronic Engineering, Bilkent University, Ankara* 06533, *Turkey.*
[††]*Department of Electrical Engineering and Computer Science University of Illinois, Chicago, Illinois, USA.*

ABSTRACT

The human auditory system is insensitive to phase information of the speech signal. By taking advantage of this fact data such as the transcript, some keywords, and copyright information can be embedded into the speech signal by altering the phase in a predefined manner. In this paper, an all-pass filtering based data embedding scheme is developed for speech signals. Since all-pass filters modify only the phase without effecting the magnitude response they are employed to diffuse data into the speech signal by filtering different portions of the speech signal with different all-pass filters. The embedded data can be retrieved by tracking the zeros of the all-pass filters.

1. Introduction

The task of embedding data into image, speech, and audio signals is called data hiding [1]-[6]. In a typical application the embedded data should be diffused into the original signal rather than into a header so that it can not be altered intentionally or unintentionally. For example, the transcript of the speech or other binary data such as copyright information can be embedded into the speech signal and stored together with the original speech waveform. Degradation of the host signal should be also minimal, in other words the embedded data should be inaudible. It should be noted that data hiding is neither encryption in which the aim is to restrict access nor speech coding.

It is well known that human auditory system is insensitive to phase information of the speech signal. By altering the phase in a predefined manner binary data can be embedded into the speech signal. Phase coding is recently used to embed data into speech by Bender et.al [1] in which the phase of the Short-time Fourier Transform of the speech frames are modified in a prescribed manner.

In this paper, data embedding is carried out using all-pass IIR filters. Consider the following two first order all-pass sections,

$$H_i(z) = \frac{za_i + 1}{z + a_i}, \quad i = 0, 1 \qquad (1)$$

The filters H_o and H_1 both have unity gain [7,8] but they modify the phase of the input signal in a different manner. In our algorithm the speech signal is first divided into frames. Each frame is filtered by one of the above IIR all-pass filter sections depending on the bit to be embedded in that frame. If the bi-

nary number $'0'$ ('1') is to be embedded to the current speech frame then it is filtered by the all-pass filter H_o (H_1).

Ternary and higher order data can be also embedded to the speech signal by using three or more all-pass filters.

2. Detection of the Embedded Data

The embedded data can be extracted by tracking the zero locations of the speech frames. Let $y[n]$ be the received signal. This signal is divided into frames and processed frame by frame. If the current speech frame has a zero at $z_o = -1/a_0$ ($z_1 = -1/a_1$) then this frame corresponds to the binary number '0' ('1'). In other words, whether the current speech frame was filtered by the all-pass filter H_o or H_1 is determined by an inner product computation.

Let the current speech frame be

$$y_k[n] = y[n]w[n - kN_o] \tag{2}$$

where the window

$$w[n - kN_o] = \begin{cases} 1, & \text{for } n = kN_o, kN_o + 1, \\ & \quad ..., kN_o + N_o - 1 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The speech frame $y_k[n]$ has a zero at $z = z_o$, if the inner product

$$Y_k(z_o) = \sum_{i=0}^{N-1} y_k[kN_o + i]z_o^{-i}, \quad N < N_o \tag{4}$$

is very close to zero. Ideally the above sum must be an infinite sum and it should be equal to zero. In practice, the frame size N_o is greater than 600 in order to avoid any audible distortion and $N << N_o$.

The embedded data detection problem can be posed as a hypothesis testing problem as

$$\mathcal{H}_o : \ z = z_o$$

$$\mathcal{H}_1 : \ z = z_1$$

for which the test statistic $|Y_k(z)|$ is evaluated for $z = z_O$ and $z = z_1$. The decision is made in favor of the hypothesis yielding a smaller value for the test statistic, $|Y_k(z)|$. If more than two all-pass filters are used for data embedding then one has to consider a multiple hypothesis testing problem.

If the data is corrupted by noise during transmission or the receiver is not synchronized with the transmitter then the test statistic, $|Y_k(z)|$, can be significantly higher than zero for $z = z_o$ and z_1, i.e., both

$$|Y_k(z_j)| > T_j, \quad j = 0, 1 \tag{5}$$

where T_j, $j = 0, 1$, are experimentally determined thresholds close to zero. In this case, one may prefer not to assign any value to the transmitted bit.

In practice, the receiver may have to detect the time instance at which the all-pass filtering starts. The receiver computes the inner products over sliding windows until N_o is detected. At this point, either $|Y(z_o)|$ or $|Y(z_1)|$ is less than the thresholds, T_o or T_1, respectively. Once this location is detected the inner products are computed in a periodic manner at time instants, $2N_o$, $3N_o$, ...

If the number of all-pass filters are more than two, say K, then K inner products have to be computed to detect the embedded data. In this case, the computational cost of detection is basically equivalent to K inner product computations.

The all-pass IIR filters are also characterized by their pole locations along with their zero locations. Therefore, the detection can be based on only zeros, only poles, and both poles and zeros. In the case of poles the test statistic

is $|Y_k(1/z_j)|$ which becomes a very large number at pole locations. The use of both poles and zeros for detection is expected to improve the robustness of the detection scheme against noise.

3. Experimental Studies and Conclusion

In order to test the effectiveness of the all-pass filtering based data embedding algorithm various simulation studies are carried out.

The speech signal is divided into 0.1 second long frames which contain 800 samples and each frame is filtered by an IIR all-pass filter corresponding to the binary data. To achieve the 30 bits/sec embedding rate eight IIR all-pass filters are used. Thus each filter represents 3 bits. Filters have poles at ± 0.2, ± 0.4, ± 0.6, and ± 0.8. Consequently, they have zeros at $\pm 1/0.2$, $\pm 1/0.4$, $\pm 1/0.6$, and $\pm 1/0.8$, respectively. At the receiver all of the above zero locations are tested for a given speech frame and the all-pass filter used at the transmitter is determined. In order to determine the correct all-pass filter eight inner products have to be computed. In Figure 1, a speech signal corresponding to the sentence "Do you have any advice for college ?" is shown and the same signal after embedding 30 bits is shown in Figure 2.

Higher data embedding bit rates can be achieved using higher order all-pass filters. Assume that the speech signal is filtered by second order systems in the form of:

$$H(z) = \frac{zb+1}{z+b} \times \frac{zc+1}{z+c} \qquad (6)$$

This system has two poles at $z = -b$ and $z = -c$, or equivalently two zeros at $z = -1/b$ and $z = -1/c$. If b and c take the values ± 0.2, ± 0.4, ± 0.6, and ± 0.8 then there are $C(8:2) = 28$ possible combinations which

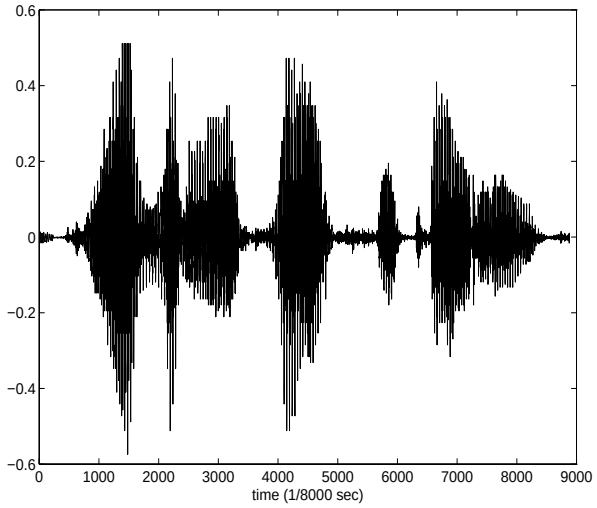

Figure 1: *The original speech waveform, 'Do you have any advice for college ?'*

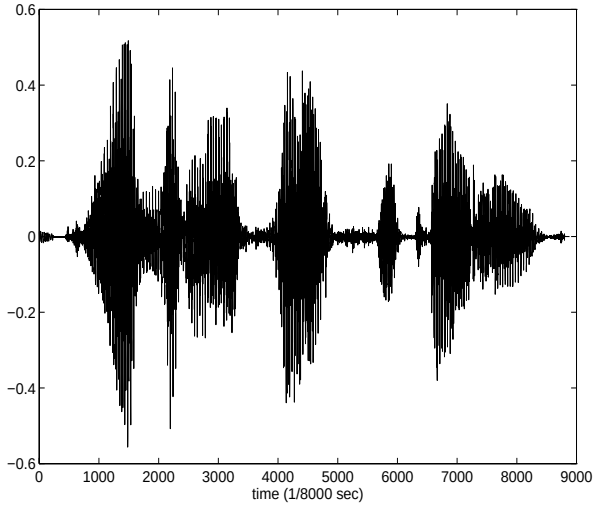

Figure 2: *The speech waveform after data embedding.*

corresponds to 4.8074 bits per frame. If one uses both first order and second order all-pass filters then the number of possible combinations increases to $C(8:1) + C(8:2) = 36$ or 5.17 bits per frame. This improvement is achieved without increasing the computational complexity of the decoding process.

By using third order all-pass filters we are able to embed more than 100 bits/sec data into the speech signal without introducing any distortion. Our data embedding rates are higher than those reported in [1]. For both methods, the limiting effect of noise on embedding rates has not been fully investigated.

The use of shorter frames also increases the embedding rate. However, distortion due to filter changes become noticeable below 0.08 second frames which corresponds to $N_o = 640$ sample frames with the 8 Khz sampling frequency. A smooth transition from filter to filter may increase the embedding rate and is currently under investigation.

Our data embedding algorithm is not affected by PCM coding. The performance of the algorithm under CELP-type coders is also under investigation.

Acknowledgment

The authors would like to thank Ahmed Tewfik, Bin Zhu, and Mitch Swanson for useful discussions. This work is supported by Turkish Scientific and Technical Research Council (TUBITAK). The project number is COST-249.

References

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, 'Techniques for data hiding,' MIT Media Lab Report, also appeared in *IBM Technical Journal,* Vol. 35, No. 3-4, 1996.

[2] M. D. Swanson, Bin Zhu , and A. H. Tewfik, "Transparent Robust Image Watermarking," *Proceedings of IEEE Int. Conf. on Image Processing (ICIP 96),* vol. III, pp 211-214, Lausanne, Switzerland, September, 1996.

[3] M. D. Swanson and A. H. Tewfik, "Embedded Object Dictionaries for Image Database Browsing and Searching," *Proceedings of IEEE Int. Conf. on Image Processing (ICIP 96),* vol. III, pp. 875-878, Lausanne, Switzerland, September, 1996.

[4] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust Data Hiding for Images," *Proceedings of 7th IEEE Digital Signal Processing Workshop (DSP 96),* Loen, Norway, pp. 37-40, September, 1996.

[5] A. Bors and I. Pitas, "Image Watermarking Using DCT Domain Constraints," *Proceedings of IEEE International Conference on Image Processing (ICIP'96)*, Lausanne, Switzerland, vol. III, pp. 231-234, 16-19, September 1996.

[6] F. Hartung, B. Girod, "Digital watermarking of MPEG-2 Coded video in the bitstream domain," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97)*, Munich, April 1997.

[7] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, 1993.

[8] R. Ansari and B. Liu, "A class of low-noise computationally efficient recursive digital filters with applications to sampling rate alterations," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol-33, pp. 90-97, 1985.