The 7th International Conference on Signal Processing Applications & Technology, Boston MA, pp. 476-480, 7-10 October 1996.

Encoding a Hidden Digital Signature onto an Audio Signal Using Psychoacoustic Masking

John F. Tilki and A. A. (Louis) Beex
The Bradley Department of Electrical Engineering,
VIRGINIA TECH
Blacksburg, VA 24061-0111

We report on the development of a system for use in an interactive television application. A method of encoding a hidden digital signature onto the audio component of a television signal is presented. This digital signature is about 35 bits in length and is encoded utilizing psychoacoustic principles such that it is nearly inaudible to human observers yet detectable by an inexpensive hand-held decoder. The encoding scheme is robust against most extraneous room noise as well as the wow and flutter of video tape machines.

1. Introduction

During the development of an interactive television application, a need arose for a method of encoding a digital signature onto the audio component of a television signal. Several design constraints prevented the use of most conventional coding schemes. The digital signature had to be about 35 bits in length and should be nearly inaudible to human observers yet detectable by an inexpensive hand-held decoder. The decoder had to be battery operated and physical connections to the television equipment were not allowed. Furthermore, the encoding scheme must be robust against most extraneous room noise as well as the wow and flutter of video tape machines.

2. System Description

After considering various approaches including options based on audio frequency spread spectrum and differential phase shift keying, we decided to use a hybrid technique similar to amplitude-shift keying (ASK) and frequency-shift keying (FSK) [1]. The digital signature is encoded using 167 sinusoids added to a filtered version of the audio component of the television signal.

To make the signature as inaudible as possible psychoacoustic masking properties were employed. The signature is of short time duration and has a low amplitude relative to the local audio. Furthermore, the sinusoidal frequencies were chosen to be in the range from 2.4 to 6.4 kHz, where human sensitivity declines compared to its peak around 1 kHz. This frequency range also allows the signature to be placed where strong low frequency content is present in the audio signal to help mask the weaker high frequency sinusoids. Since humans are much more sensitive to the lower frequencies, this masking can be quite effective. Using frequencies above 2.4 kHz also provides some resistance to human voice interference at the receiver. Although the audio component of a television signal is not bandlimited to 6.4 kHz, and frequencies above this could have been used to take advantage of further reduction in human sensitivity, the sampling rate of the decoder had to be kept as low as possible because of computational requirements.

Once the target location within the audio signal has been chosen, a zero-phase lowpass filter is used to locally remove any frequency content above 2.4 kHz. The sinusoids are then added to the signal. Figure 1 below shows the time-averaged power spectral density (PSD) of a typical window of audio signal. Figure 2 shows the PSD of the same window after lowpass filtering and addition of the sinusoids.

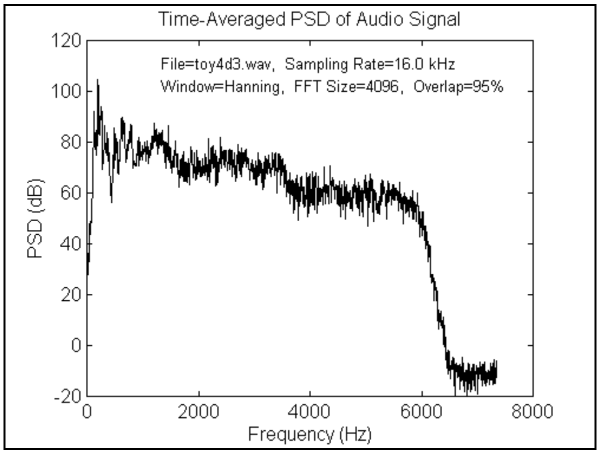

Figure 1: Time-Averaged Power Spectral Density of Audio Signal

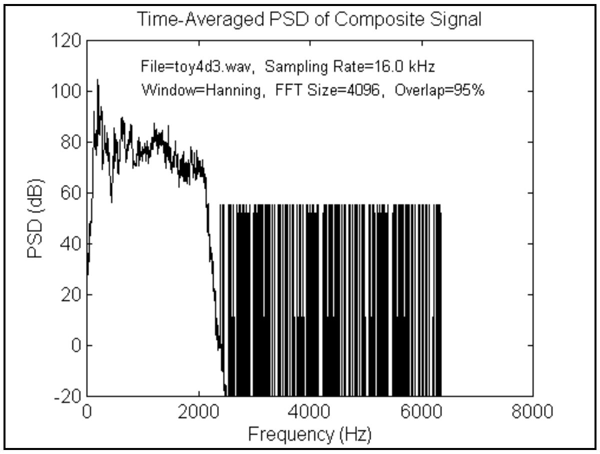

Figure 2: Time-Averaged Power Spectral Density of Composite Signal

The frequencies of the 167 sinusoids are chosen to correspond with bin frequencies of a 4096 point FFT performed on data sampled at a rate of 16.0 kHz. Thus the decoder can use a simple FFT to detect the presence of the sinusoids. The magnitude estimate at each candidate FFT bin is compared with neighboring bins two away on each side to determine whether a sinusoid is present. If the neighbor FFT values are below the center value by at least 4 dB, then a sinusoid is assumed to be present and a digital ‘1’ is indicated. If this condition is not satisfied, a digital ‘0’ is indicated. Figure 3 depicts this detection process. The asterisks mark the candidate FFT bins, and the circles mark the bins two away on either side of the centers. The bit sequence [1 0 1 0 0 1 1 1] is represented in the example.

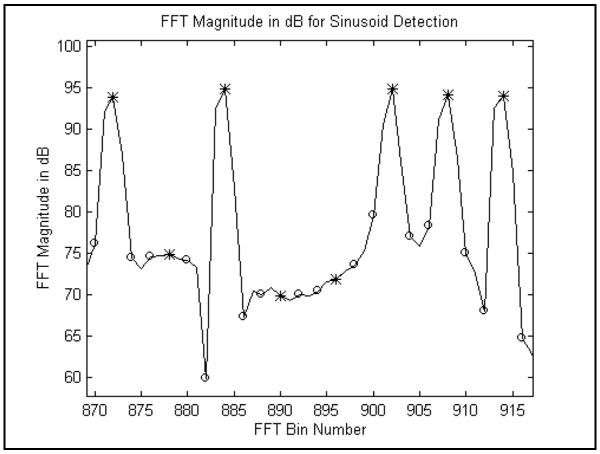

Figure 3: FFT Magnitude in dB for Sinusoidal Detection

Because interference due to sidelobe leakage can be a problem with closely packed sinusoids, a Hanning window is applied before the FFT is computed. The mainlobe width of the Hanning window dictates that the sinusoids be spaced at least six FFT bins apart, and also requires that we examine the neighbor bins two away on each side rather than the immediate neighbors a single bin away.

When detecting the presence of the sinusoids, calculating the true magnitude spectrum is not possible due to the computational burden imposed by the square-root operation. We also found that the alternative of using the magnitude squared spectrum created dynamic range problems when implemented with 16-bit wordlength. A solution was found by computing an estimate of the magnitude spectrum by calculating the maximum of the absolute values of the real and imaginary parts for each FFT bin. Simulations have shown that performance is only slightly degraded by substituting this magnitude estimate.

The 167 sinusoids used in coding the digital signature perform several functions. Since many sinusoids can be attenuated due to transmission losses, multipath effects, and noise interference, redundancy and error correction techniques are necessary. The digital signature itself is 35 bits in length. For error detection purposes, a cyclic redundancy check of 12 bits was added [2]. These 47 bits are then repeated in a triplication code to provide double redundancy, bringing the total to 141 bits. The three blocks in the triplication code occur in distinct frequency regions between 2.4 and 6.4 kHz. Thus, if attenuation occurs for any of the above-stated reasons in a single frequency band, the data bit patterns are still detectable due to the double redundancy present in the other bands.

Five sinusoids are used for frequency shift detection, to be described later in the paper. The final subset of 21 sinusoids is used solely for self-synchronization. If a two-thirds majority of these “control” sinusoids is detected, valid data is considered to be present. Figure 4 below demonstrates self-synchronization with the control function. As the FFTs are performed on blocks of signal, the 21 control sinusoid locations are examined. If sinusoids are detected at 14 or more of these locations, valid data is assumed to be present in that FFT block, and the data sinusoid locations can be polled. In the figure below the asterisks indicate when a two thirds majority of control sinusoids is present, and hence when the digital data is available.

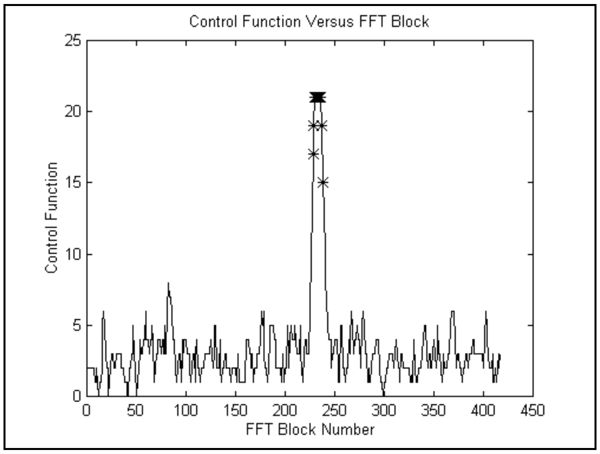

Figure 4: Control Function Versus FFT Block Number

The control sinusoids are uniformly interspersed with the data sinusoids throughout the entire 4 kHz band being used. Thus the control sinusoids not only serve a synchronization purpose, but they also provide an indication of the quality of the received data in that particular FFT block. Since a triplication code is being used for the data, a two-thirds majority for the control function is appropriate. Furthermore, when the data is being tabulated over successive FFT blocks, the results of each block are weighted according to the value of the control function in that block. For example, we have more confidence in the data when the control function is 21/21 versus when it is 14/21. The data associated with such blocks should be weighted accordingly. Thus digital ones are represented by positive control function values and digital zeros by negative values.

Once the detection process is initiated by the control function, valid data received in the current and subsequent FFT blocks are tabulated. The end of data transmission is detected by the level of the control function dropping below 14/21 and remaining low for a specified period of time. When it is determined in this manner that the data transmission is complete, final decisions are made regarding the individual data bits. Since the bit votes from each FFT block (weighted by the control function from each block) have been summed over time, a final decision can be made regarding each bit's status by a simple threshold test. If a bit's value is positive it is considered to be a digital one. If it is negative it is considered to be a zero. The triplication code is then decoded by a two-thirds majority vote among the three frequency subbands. Finally the CRC is checked to verify error-free reception. The 35 bit digital signature results.

Table 1 below demonstrates the data decoding process (without the CRC) through an example. Suppose we desire to transmit a digital signature of two bits [1 0], and on the decoding end the control function is detected as shown in the table. When the control function is below 14 no data is present. When the control function is 14 or larger the data bit locations are analyzed to see if sinusoids are present. When a sinusoid is present, the value of the control function is added to the corresponding data bit location. Likewise the lack of a sinusoid represents a digital zero, and the value of the control function is subtracted from the corresponding data bit location. Once the control function drops below 14 and stays there, the data collection process terminates. Any bit locations containing positive values are considered to be digital ones, and any negative locations are zeros.

Note that Bit 1-1 contains an error during FFT block 7. However, the correct value was received often and strong enough in other FFT blocks to produce the correct bit decision at the end. Bit 1-3, however, has been corrupted several times (in FFT blocks 3, 4, and 7). Multipath interference can cause a null in the frequency domain resulting in such a repeating bit error. In this case the bit decision is incorrectly made as a zero. However, the proper digital signature will still be extracted due to the redundancy of the triplication code. The values for Bit 1 are [1 1 0] yielding a 1. Similarly the values for Bit 2 are [0 0 0] yielding a 0.

Table 1: Example of the Data Decoding Process.

FFT Block	Control Function	Bit 1-1	Bit 2-1	Bit 1-2	Bit 2-2	Bit 1-3	Bit 2-3
1	4	0	0	0	0	0	0
2	3	0	0	0	0	0	0
3	14	14	-14	14	-14	-14	-14
4	18	18	-18	18	-18	-18	-18
5	21	21	-21	21	-21	21	-21
6	20	20	-20	20	-20	20	-20
7	16	-16	-16	16	-16	-16	-16
8	13	0	0	0	0	0	0
9	7	0	0	0	0	0	0
10	2	0	0	0	0	0	0
Final Values		57	-89	89	-89	-7	-89
Bit Decisions		1	0	1	0	0	0

Since the playback speed of video tape machines is not perfectly constant, the time scale of the audio signal can expand and contract locally. This phenomenon, called wow, has the effect of shifting all sinusoids in the frequency domain. A frequency locking mechanism was developed to detect, quantify, and compensate for the resulting spectral shifting. This locking mechanism consists of five non-uniformly spaced sinusoids placed at the upper end of the frequency region, and the amount of spectral shift is determined with a frequency domain matched filter. Figure 5 shows the frequency domain matched filter. This matched filter is applied to the FFT bins in the region of the locking sinusoids. When the matched filter overlaps with the locations of the five sinusoids in the FFT a large correlation results. The output of the matched filter during zero frequency shift is shown in Figure 6.

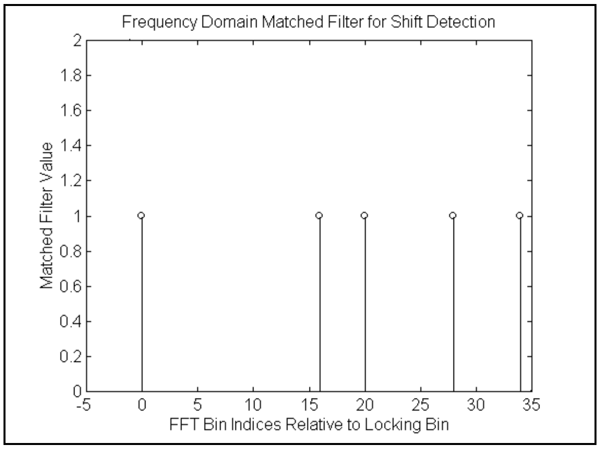

Figure 5: Frequency Domain Matched Filter for Shift Detection

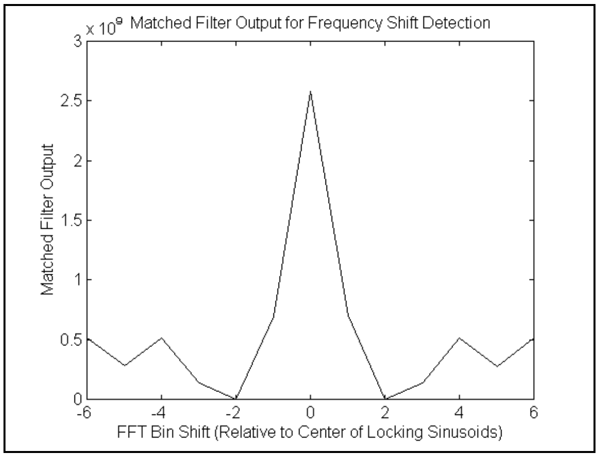

Figure 6: Matched Filter Output for Frequency Shift Detection

The spacing of the locking sinusoids was chosen to maximize the correlation during lock and to minimize the correlation for all other shifts. Once the amount of shift is determined from the peak of the matched filter output, the new data and control bin locations can be read from lookup tables. The data extraction then proceeds as discussed earlier.

3. Hardware Implementation

The hardware subsystem consists of an electret microphone, three stages of active filters, an Analog Devices AD1847 codec, and an Analog Devices ADSP2181 digital signal processor. The active filters amplify the frequency band of interest and attenuate all other frequencies, especially the lows. Since human voice contains mostly low frequency energy, these filters provide some robustness against voice interference.

The sampling rate for the codec is 16 kHz, and all decoding calculations are performed every 205 samples (every 12.8125 msec). A large order FIR filter pre-emphasizes the sampled audio signal by compensating for imperfections in the analog filters and providing further gain in the passband. The FIR filter was designed via the Parks-McClellan algorithm [3].

4. Conclusion

We have successfully developed a method of encoding a hidden digital signature onto an audio signal. By taking advantage of psychoacoustic properties, this signature is inaudible to most human observers yet detectable by a hand-held decoder. Furthermore the encoding scheme is robust against most extraneous room noise as well as the wow and flutter of video tape machines. The hardware implementation has been successfully tested, and is now a part of an interactive television application.

REFERENCES

[1] B. P. Lathi, Modern Digital and Analog Communication Systems, pp. 179-182, (New York: Holt, Rinehart and Winston, 1983).

[2] Stephen B. Wicker, Error Control Systems for Digital Communications and Storage, pp. 68-127, (Englewood Cliffs, New Jersey: Prentice Hall, 1995).

[3] Alan V. Oppenheim and Ronald W. Schafer, Discrete-Time Signal Processing, pp. 464-468, (Englewood Cliffs, New Jersey: Prentice Hall, 1989).