ENCODING A HIDDEN AUXILIARY CHANNEL ONTO A DIGITAL AUDIO SIGNAL USING PSYCHOACOUSTIC MASKING

John F. Tilki and A. A. (Louis) Beex
DSP Research Laboratory
The Bradley Department of Electrical Engineering
VIRGINIA TECH
Blacksburg, VA 24061-0111, USA
voice: (540) 231-4877 fax: (540) 231-3362
e-mail: tilki@vt.edu and beex@vt.edu

Abstract **- We report on the development of a method of encoding an auxiliary channel onto a digital audio signal such that it is imperceptible to human observers. The encoding is accomplished by imposing slight and controlled changes on the phase spectrum of short-time signal windows. By employing principles of psychoacoustics the auxiliary channel is masked from human perception. Furthermore, the encoding and decoding are achieved via fast algorithms which allow real time processing. The method is applicable for any digitized audio signal containing voice, music, or other acoustic signals to be heard by humans. Possible signal sources include compact discs, digital television, digital radio, digital telephony, and any other source where an audio signal is in a digital format.**

INTRODUCTION

The ability to add extra hidden information to a signal could be useful in many applications. Audio compact discs could be modified to contain artist information, song titles, song lyrics, karaoke information, still images, and perhaps even video clips. In other contexts an extra channel could be used to transmit control information. The want and need for additional bandwidth is ever-increasing, and the capability to add extra information to digital audio can be quite useful even for archival and storage purposes.

The coding method presented in this paper allows the addition of such an auxiliary channel, and it does so in a manner that does not significantly degrade the perceived quality of the primary audio channel. Furthermore, the method does not require changes in the underlying digital format, and so it completely maintains backward compatibility with existing data formats. One can imagine the possibilities for the audio CD example. Existing CD players could still play CDs encoded with the auxiliary channel. With additional signal processing capability however, new CD players could provide the enhanced functionality described above.

Other researchers have aimed at the same objective using methods such as subband coding coupled with adaptive quantization, and subtractively dithered noise-shaped quantization [1-3]. However, a direct comparison between the schemes in terms of auxiliary channel data rate versus perceived quality of the primary signal could not be performed. Since the other approaches seem to require more computations, our method could be an excellent alternative especially at high sampling rates and where computational capability is a limiting factor.

DESCRIPTION OF CODING METHOD

Encoding

It is well known that humans are relatively insensitive to certain types of phase distortions. We use this fact to encode additional information onto the phase of selected bins of an FFT. The signal bins are selected according to human sensitivity in the corresponding frequency bands, and are spaced to facilitate masking by unmodified neighboring bins in the same spectral region. In particular the lower frequencies (below say 2 kHz), which typically contain the most energy in common audio signals, are left untouched. The stronger low frequency components help to mask the changes made in the higher frequency region. Additionally, the unmodified FFT bins in the higher frequency region help to mask their modified neighbors.

The FFT bins immediately preceding the signal bins are used as references. While keeping the digital audio signal's FFT amplitude the same, the phase of the FFT at the signal bins is discarded, and a new phase is assigned based on the neighboring reference bins, a differential phase change, and the digital information to be encoded. For example, a small clockwise rotation of phase relative to the reference can represent a digital "1" while a counter-clockwise rotation of phase represents a digital "0".

Suppose that the FFT has a complex value of X_r at a reference bin. In polar form this value is represented as

$$X_r = R_r e^{j\theta_r} \tag{1}$$

where R_r is the magnitude and θ_r is the phase. Similarly the value of the FFT at the neighboring signal bin is

$$X_s = R_s e^{j\theta_s} \tag{2}$$

Then if φ is the differential phase change, the new FFT value for the signal bin is assigned as

$$\hat{X}_s = R_s e^{j(\theta_r \pm \phi)} \quad (3)$$

Note that it is the phase of the *reference* bin that is used along with the differential phase to replace the phase of the signal bin. Addition of the differential phase represents a digital one and subtraction a digital zero.

Once the phases of the signal bins have been modified, the conjugate phase is assigned to the matching negative frequency bins. An inverse FFT then yields the modified signal block containing both the primary audio and the secondary channel. The process is then repeated for subsequent signal blocks.

Figure 1 shows a small region of an example phase spectrum and its modified version after encoding. The phase at the reference bins is denoted by asterisks, the original phase at the signal bins is denoted by 'o', and the modified phase at the signal bins is denoted by 'x'. This example shows a digital pattern of [0 1 0 0 1].

For a given source sample rate the FFT size is chosen to be short enough in time to maintain the imperceptibility of the introduced phase distortions. However, the FFT size should conversely be chosen as large as possible under the above constraint such that the FFT bin frequencies are as dense as possible. In our limited simulations, FFT sizes corresponding to block lengths of 10 to 50 msec worked well. Smaller blocks often yielded "granular" noise, while larger blocks usually created noticeable distortion of the primary signal.

Decoding

The signal is decoded by comparing the phase at the (known) signal bins to the phase at the neighboring reference bins, and assigning bits based on the direction of relative phase rotation.

Synchronization

The decoding process requires perfect synchronization with the encoded signal blocks. For a storage medium such as a compact disc synchronization is simple and only requires knowledge of the offset of the first signal block. For real-time data streams however, synchronization is more difficult. Several options are available, including decoding candidate blocks until an error detection criterion is satisfied. Error detection can be accomplished with cyclic redundancy checks (CRC) or other efficient error detection schemes [4]. The candidate block alignments are changed one sample at a time until a valid data transmission is detected. The offset at the point of valid data detection provides synchronization for all subsequent signal blocks. Figure 2 shows how zero bit errors result when perfect synchronization is achieved. When synchronization is incorrect by even a single sample delay or advance several bit errors result (≥38 in the case shown).

Another possibility for synchronization is the inclusion of known bit patterns in the data stream. This situation also requires shifts by a single sample at a time until the embedded bit pattern is detected. Such a "control" function for synchronization has been used in a similar context in the analog channel case [5].

SIMULATION RESULTS

The coding method has been tested and works well. The example provided in the paper is based on four seconds of the audio component of a TV commercial, sampled monophonically at a rate of 44.1 kHz, consistent with CD quality audio. The lowpass nature of this audio signal is shown in Figure 3. (Note that the sinusoid at 15.734 kHz is interference from the horizontal synchronization pulse for the video electron beam.) The decoded phase errors would be zero were it not for the quantization back to 16-bits that occurs after the inverse FFT. The phase errors due to quantization are shown in Figure 4. A differential phase of $\pi/8$ was used to modify the phase of every fourth FFT bin above 2 kHz in 2048 point FFT blocks. A secondary channel resulted with a capacity of over 5000 bits per second with little or no degradation in the perceived quality of the primary audio signal.

Higher data rates can be achieved by expanding the frequency region used for coding and using smaller spacing between signal bins in the FFT. However, as would be expected, the perceived quality of the primary signal suffers as the data rate of the auxiliary channel increases.

FUTURE IMPROVEMENTS

Future improvements include using several increments of the differential phase for the forward and reverse rotations of phase, allowing for the coding of multiple bits per FFT bin. Also, pre-quantization of encoded FFT blocks should allow pre-compensation for quantization effects during the encoding process. In addition, the audio signal can be decomposed into subbands, with each subband independently used for coding. This would allow more bits to be encoded in selected subbands based on the relative influence the individual subbands have on overall perception.

Finally, adaptive encoding based on local power levels and frequency content in the primary audio signal can perhaps increase the effective capacity of the auxiliary channel. A related variation would be adaptation of the differential phase change based on the local frequency content and on some measure of phase deviation in the region. This could be used to compensate for the fact that frequency regions with lower energy are more sensitive to

quantization effects than regions with more energy. Note that the phase errors in Figure 4 are largest in the regions of lowest energy, as seen from comparison with Figure 3.

CONCLUSION

In conclusion, we have developed a method of successfully encoding an auxiliary channel onto a digital audio signal. This channel is encoded using psychoacoustic principles such that it is imperceptible to human observers. Both the encoding and decoding are accomplished with the FFT, allowing real time implementations. The coding process is appropriate for any application containing a digitized audio signal.

REFERENCES

[1] W.R.Th. ten Date, L.M. van de Kerkhof, and F.F.M. Zijderveld, "Digital Audio Carrying Extra Information," ICASSP 1990, pp. 1097-1100.

[2] Michael A. Gerzon and Peter G. Craven, "A High-Rate Buried-Data Channel for Audio CD," J. Audio Eng. Soc., Vol. 43, No. 1/2, Jan./Feb. 1995, pp. 3-22.

[3] A.W.J. Oomen, M.E. Groenewegen, R.G. Van Der Waal, and R.N.J. Veldhuis, "A Variable-Bit-Rate Buried-Data Channel for Compact Disc," J. Audio Eng. Soc., Vol. 43, No. 1/2, Jan./Feb. 1995, pp. 23-28.

[4] Stephen B. Wicker, Error Control Systems for Digital Communications and Storage, Englewood Cliffs, New Jersey: Prentice Hall, 1995.

[5] John F. Tilki and A. A. (Louis) Beex, "Encoding a Hidden Digital Signature onto an Audio Signal Using Psychoacoustic Masking," International Conference on Signal Processing Applications & Technology, October 7-10, 1996, Boston, MA, pp. 476-480.

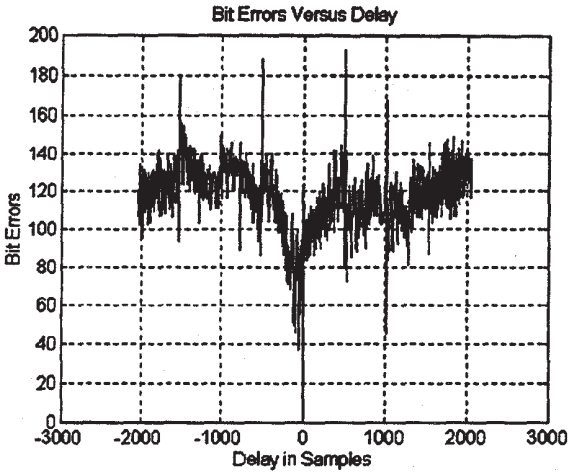

Figure 2. Block Synchronization by Error Detection.

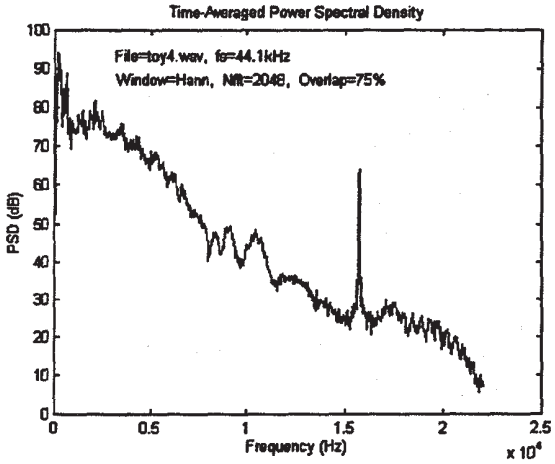

Figure 3. Time-Averaged PSD of Audio Signal.

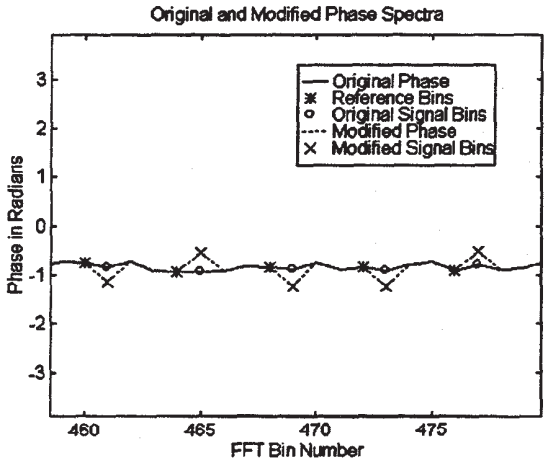

Figure 1. Example Phase Spectra.

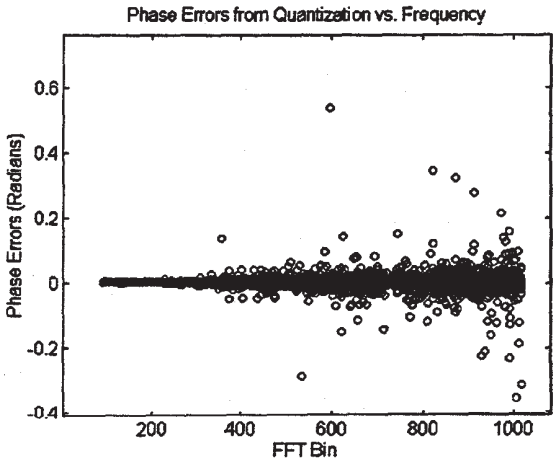

Figure 4. Phase Errors from Quantization vs. Frequency.