ABSTRACT

We extensively evaluated a data hiding algorithm for stereo audio signals which embeds data using the polarity of the echoes added to the high-frequency channels, which we have previously proposed. Its performance was also compared to conventional data hiding using spread spectrum, and those using echoes with different delays. Raw embedded data was detected with little or no errors for added noise at 20 dB SNR and above, or with MP3 coders, although the spread spectrum method showed almost no errors at all. However, sample rate conversion and random bit cropping were shown not to affect the embedded data with the proposed method, while other methods, including spread spectrum, showed significant amount of errors. The embedded audio quality test using the MUSHRA standard method resulted in little noticeable degradation, far better quality compared to other hiding methods.

Index Terms— Audio signal, stereo, data hiding, inter-channel coherence, robustness.

1. INTRODUCTION

Recent advances in high-speed digital communication networks, most notably the Internet, has made it quite practical to use these networks to access digital media on the fly. This situation has called for protection of intellectual ownership and prevention of unauthorized tampering of digital media content. One solution is to digitally hide copyright information in these digital content without altering its quality. In other words, digital contents are “marked” with copyright information transparently using data hiding methods.

Although initial applications of data hiding was mainly in the ownership protection, there are now applications where other information is hidden in the digital content. One example is the transmission of metadata with the audio signal. The metadata may carry for example artist name, composer, genre, or even the URL of the artist site. These kind of applications are actively being pursued nowadays due to the ease of handling of the content, and possible new scenarios to utilize this metadata.

Most of the recent audio data hiding methods take advantage of the human auditory system (HAS) in order to hide information into the host speech or audio signal without causing significant perceptual disturbances. Two properties of HAS which is utilized most in information hiding are temporal masking and frequency masking properties. These are also used in the MPEG audio coding standards [1].

Echo data hiding has been regarded as an effective method to embed data with minimum impact on the perceived quality [2]. The echo data hiding method embeds data by introducing a small amount of echo in the audio signal. Data is hidden by varying the initial amplitude, decay rate, and the offset of the echo. If the delay is small enough, the human ear cannot distinguish the original signal from the echo, i.e., “fusion” of the original signal and the echo occurs. Although this “fusion” seems to be signal-dependent, it is generally said that in most cases, fusion occurs at around a delay of a few msec. Note that the reason for HAS being relatively insensitive to small amounts of echos can be related to both temporal and frequency masking.

Previously, we proposed a stereo audio data hiding method which uses the property of the HAS mentioned above to code embedded data in the relative polarity of the added echo between two stereo channels [3]. We added this echo to control the coherence between channels. This inter-channel coherence is known to control the “width” of the stereo image of the auditory object [4].

In this paper, we evaluate both the robustness and the embedded audio quality of the proposed data hiding method. We evaluated the robustness against additive noise, MP3 coding, sample rate conversion, and random sample cropping. We also evaluated the subjective quality of the embedded audio. These results were compared with popular data hiding methods, including data hiding with direct spread spectrum [5], and with echos, as mentioned above [2].

In the next section, the proposed data hiding algorithm is briefly described. In section 3, a description of the audio signals used for evaluation tests as well as the tested hiding algorithms are described. In section 4, the robustness evaluation procedures as well as its results are described.
5, the subjective quality of the data-embedded audio is evaluated. Finally, in section 6, the conclusion is given.

2. THE DATA HIDING METHOD

Fig. 1 shows the proposed data embedding scheme. The input is first split into two bands with a band-split filter with a cut-off of 1.5 kHz. The low-band signal is used as is. The Inter-Channel Coherence (ICC), \( C_{LR} \), is calculated between the high-band left signal, \( x_{LH} \) and the high-band right channel signal, \( x_{RH} \). The average of these signals gives the mid-channel \( x_m \). Echos are added to these signals to adjust the ICC to match the original values.

Fig. 2 shows the Lauridsen decorrelator used to adjust the ICC [4]. The decorrelator delays the mid-channel, \( x_m \), by \( D \), scales by factor \( G \), and both adds and subtracts this to/from \( x_m \), giving the left and right channel signals \( x'_{LH} \) and \( x'_{RH} \).

\[
x'_{LH}(k) = x_m(k) \pm Gx_m(k-D)
\]
\[
x'_{RH}(k) = x_m(k) \mp Gx_m(k-D)
\]

\( G \) is adjusted so that the ICC of the generated signal equals \( C_{LR} \).

\[
G = \sqrt{\frac{1 - C_{LR}}{1 + C_{LR}}}
\]

The echo added to each channel only needs to have reverse polarity, and the absolute polarity is arbitrary. Thus, this polarity is controlled to embed data. In all following experiments, we defined the polarity of the echoes as follows. When \( d_i = 0 \),

\[
x'_{RH}(k) = x_m(k) + Gx_m(k-D)
\]
\[
x'_{LH}(k) = x_m(k) - Gx_m(k-D)
\]

Otherwise when \( d_i = 1 \), the polarity of the added echo is reversed.

When the embedded data, \( d_i \), changes, the polarity of the echo reverses, which in some cases may be audible. We chose to smooth this transition using a linear window with a length of \( \frac{5}{16} \) of the frame length.

Data can be detected by correlating the mid-channel signal delayed by \( D \) with the echo. The mid-channel can be estimated from the sum of both channels at the decoder, while the echo can be detected from the difference signal.

The embedded data can be detected from the polarity of the correlation between the sum and difference signal. If this correlation is positive, \( d_i \) can be estimated as 0, and if negative, \( d_i \) can be estimated as 1. The decoder configuration is shown in Fig. 3.

3. EXPERIMENTAL SETUP

We selected 14 stereo audio clips from the RWC Music Database: Music Genre [6], listed in Table 1. We embedded random bit patterns into these clips, and tested its robustness against additive noise, audio coding, sample rate conversion, and random sample cropping. All clips were approximately 2 minutes long, thus a total of about 28 minutes of audio data was used. We used a frame length of 2000 samples. This amounts to a total of 37,085 frames. Data was embedded in 35,120 frames, or approximately 94.7 % of all frames. The rest of the frames did not contain enough energy to embed data reliably. The average embedded bit rate is 22.08 bps.

We also included two conventional data hiding methods for comparison. The first method uses spread spectrum techniques to spread the data to all available frequency range, and adds this to the host signal after the power of the spread-spectrum data is adjusted to be ~20 dB relative to the host signal (to be noted dss) [5]. The data rate was set to about the same rate as the proposed method.

The second method adds echos with two different delay values to mark the host signal with hidden data (to be noted echo) [2]. The echo delays were set to 100 and 150 samples (approximately 2.3 and 3.4 msec at 44.1 kHz). The use of echos to embed data is similar to our proposed method. However, this method does not exploit the stereo redundancy, while the proposed method does. We were only able to embed data with this method at half the rate of the proposed method.

Note that for all methods tested, the error rates are for the raw embedded data. No error correction nor repetition was employed. If a more robust data hiding is required, such
Table 1. Evaluated sound sources.

<table>
<thead>
<tr>
<th>Genre</th>
<th>RWCDB no.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic</td>
<td>59</td>
<td>Mozart K. 485 Rondo D major</td>
</tr>
<tr>
<td>Classic</td>
<td>49</td>
<td>Handel Water Music Suite 2</td>
</tr>
<tr>
<td>Classic</td>
<td>53</td>
<td>Holst Jupiter from The Planets</td>
</tr>
<tr>
<td>Classic</td>
<td>59</td>
<td>Suza Stars and Stripes Forever</td>
</tr>
<tr>
<td>Jazz</td>
<td>30</td>
<td>Kitchen, HH Band</td>
</tr>
<tr>
<td>Jazz</td>
<td>33</td>
<td>The Ramp, H’s band</td>
</tr>
<tr>
<td>Jazz</td>
<td>35</td>
<td>Gypsy Eyes, 22 Project Band</td>
</tr>
<tr>
<td>Popular</td>
<td>1</td>
<td>Wasting Time, Shinya Iguchi</td>
</tr>
<tr>
<td>Popular</td>
<td>5</td>
<td>Taking Your Heart, Shinya Iguchi</td>
</tr>
<tr>
<td>Rock</td>
<td>7</td>
<td>Everyday Lovin’, Donna Burke</td>
</tr>
<tr>
<td>Rock</td>
<td>11</td>
<td>21st Century, Donna Burke</td>
</tr>
<tr>
<td>World</td>
<td>64</td>
<td>Blue Print, 22 Project Band</td>
</tr>
<tr>
<td>World</td>
<td>68</td>
<td>Grassy Dance, Shigekazu Kamaki</td>
</tr>
<tr>
<td>World</td>
<td>71</td>
<td>Desperate Little Man, Banjo Club</td>
</tr>
</tbody>
</table>

as with copyright watermarks, sophisticated error correction codes may be employed.

4. ROBUSTNESS EVALUATION

Fig. 4 shows the bit error rate (BER) vs. signal-to-added-noise-level ratio (SNR). As can be seen, “dss” is very robust to additive noise, while “proposed” and “echo” show modest number of errors. Although both of these use echos to embed data, “proposed” seems to be more robust, especially at SNR above 20 dB. Note that added noise larger than that giving 20 dB is quite audible, and renders the audio clip worthless.

We also tested the robustness to MPEG audio coding. We encoded each data-embedded audio source by MPEG 1 layer 3 (MP3) codecs, and decoded them back. Lame ver. 3.98.2 with joint stereo coding at rates of 160, 128, and 96 kHz were used. Fig. 5 shows the BER with MP3 coding. Again, “dss” is very robust, showing almost no errors at all regardless of the bit rate. The proposed method also shows modest robustness to MP3, with BER less than 1%.

Fig. 6 shows the robustness to sample rate conversion. All samples were either downsampled or upsampled to the specified sample rate, 22.05, 32 or 48 kHz, and converted back to the original 44.1 kHz. We used the built-in sample rate conversion of the CoolEdit software (now Adobe Soundbooth). All filtering including the anti-alias filters were applied within this tool. This time, the dss method showed a much higher error rate than the proposed or the echo method. The errors in the dss method are probably caused by the anti-alias filters. Meanwhile, the proposed method shows less than 1 % error rate regardless of the sample rate.

Fig. 7 shows the robustness to random sample cropping. Random samples were cropped at a rate of 1, 0.5 and 0.1 %. Again, “dss” showed a much higher error rate than the proposed or the echo method. The errors in “dss” are probably caused by misalignment of pseudo-random bit sequence at spreading. Meanwhile, the proposed method shows less than 10 % error rate regardless of the sample rate.

5. EMBEDDED AUDIO QUALITY EVALUATION

We conducted the MUSHRA subjective listening tests [7] to quantify degradations caused by data hiding. We selected five short audio clips from mostly the EBU SQAM CD, listed in Table 2, and embedded random bit patterns with the proposed, dss, and the echo hiding methods, respectively. The clips ranged from 6 to 15 seconds in length. All were CD quality, sampled at 44.1 kHz, and in stereo.

We included the original audio clip as the reference, (ref), and a 3.5 kHz low-pass-filtered audio clip as the an-
Table 2. Sound sources for quality evaluation.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>abba</td>
<td>Electric instr. (Rock)</td>
</tr>
<tr>
<td>beethoven</td>
<td>Ode to Joy (Classics), orch. chorus</td>
</tr>
<tr>
<td>mozart</td>
<td>K. 270 (Classics), wind instruments</td>
</tr>
<tr>
<td>piano</td>
<td>Schubert (Classics), piano solo</td>
</tr>
<tr>
<td>trumpet</td>
<td>Haydn (Classics), Orch. &amp; trump.</td>
</tr>
</tbody>
</table>

We also included a 7 kHz low-pass-filtered clip (lpf7), and MP3 transcoded audio clips at 96 kbps (mp3_96), 64 kbps (mp3_64), and 48 kbps (mp3_48) for comparison. Ten subjects, all in their twenties with normal hearing, participated in the tests. All subjects were required to rate each processed audio, including the hidden reference on a 100-point scale. 100 to 80 points on this scale was categorized as “excellent,” 80 to 60 as “good,” 60 to 40 as “fair,” 40 to 20 as “bad,” and 20 to 0 as “very bad.” The subjects rated the processed audio for only one source in a single session. Thus, each subject conducted five sessions, one session per source. The rated scores were averaged for all subjects [7]. Fig. 8 shows the average score of all clips.

The proposed method shows audio quality of about 80 points, which is almost in the “excellent” category. The dss method was rated in the “bad” category due to its noisiness. The echo method was rated as “good,” and its quality was fairly close to the MP3 coding at 64 kbps. There were significant differences in the quality by sound source, but the order of the quality by method did not change.

6. CONCLUSION

We conducted a detailed evaluation of a data hiding algorithm for stereo audio signals which embeds data using the polarity of the reverberations added to the high-frequency channels. The performance was also compared to conventional data hiding using Direct Spread Spectrum (dss), and also a method using echos with different delays (echo). Robustness against additive noise, MP3 audio coders, sample rate conversions, and random bit cropping were evaluated and compared. Although dss proved to be very robust to additive noise and MP3 coding, significant errors were noted for other degradations. The proposed algorithm showed fairly constant robustness for all degradations tested. The embedded audio quality test using the MUSHRA standard method resulted in little noticeable quality degradation with the proposed algorithm, while the dss method showed a very poor quality, and the echo method showed modest degradation.

7. REFERENCES


