Time Base Modulation: A New Approach to Watermarking Audio and Images

Jonathan Foote and John Adcock
FX Palo Alto Laboratory, Inc.
3400 Hillview Avenue
Palo Alto, CA 94304
{foote,adcock}@fxpal.com

ABSTRACT

A novel method is presented for hiding information in audio and image files by subtly and inaudibly compressing or expanding time regions of an audio file. By comparing the altered file with a reference copy, compressed and expanded regions can be detected. The location and duration of compression and expansion encodes information in the modified file. This approach is novel and has many advantages over other methods. Principal advantages are that, unlike most watermarking techniques, it is theoretically noiseless, introduces no spectral distortion, and is robust to all compression and transmission schemes. The approach can be used to watermark images as well, and we discuss how the approach can be made "self-clocking" so as not to depend on a reference file.

Keywords: Watermarking, data hiding**,** steganography, audio, time scale modification

1. INTRODUCTION

Many techniques exist for the time-scale modification of audio. This refers to changing the time duration of an audio sample without changing the pitch or other spectral characteristics. Because the pitch is not changed, small amounts of time scale modification are typically not noticeable, which is the basis for the watermarking approach described here. By modulating the time base of an audio file, information can be undetectably encoded in it. As shown in Figure 1, short time regions of the file are either compressed or expanded by an imperceptible amount (exaggerated in the Figure for illustration). We call this method "time base modulation" as the underlying time basis is modulated by the watermark function. By matching the amount of compression and expansion, the overall length of the file is not altered. The location and degree of compression or expansion are the variables used to encode information. The watermark is detected by comparing the watermarked copy with the reference (unmarked) audio. Time-alignment of the watermarked and reference audio produces a "tempo map" that indicates how the time base of the watermarked audio has been altered. In regions of compression or expansion, the tempo map will deviate from a straight line and the embedded watermark data may be recovered from these deviations. A similar technique can be used to watermark images or scanned text by invisibly compressing or expanding image regions.

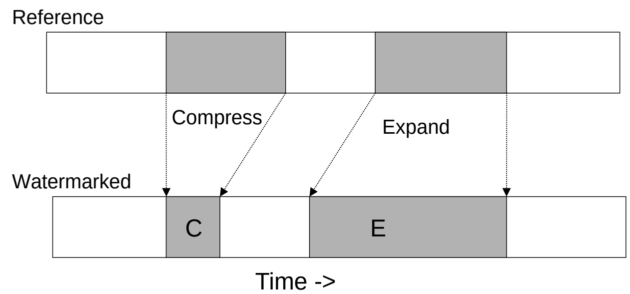


Figure 1. Watermarking a signal by time base modulation

The watermark can encode copyright information, a cryptographic signature, or information that specifically identifies a particular copy of the source audio or image. This is highly useful, for example, to hide encryption keys or for digital rights management. If each legitimate user of a copyrighted work is given a file with a unique watermark, the watermark found in illicitly distributed copies then can identify the source. Another application is to encode a cryptographic "hash" of the signal to verify its authenticity. Any alteration in the signal (such as inserting or deleting material) will generate a different hash value, which can be compared with the encoded value so that tampering can be detected [1]. Watermarks are especially suited to high-security applications: if each copy of a sensitive document is watermarked, the source of illicit copies can be more easily identified[1]. In a later section, it is shown how the time base watermarking procedure can be implemented directly in a photocopier or printer, allowing every individual copy to be uniquely watermarked.

Many different patterns of compression and expansion are available to encode information. It may be preferable to have the same amount of compression and expansion so that it is the watermarked version is the same size as the reference file, though this is not strictly necessary. Experiments have shown satisfactory results with compression/expansion ratios on the order of one to two percent, though this could be increased at the risk of introducing detectable artifacts. Overall, an encoding rate on the order of 8 bits per second is feasible in audio, and is limited only by how objectionable time-scale modification becomes in the extreme.

[1]N.B.: the authors do not believe that watermarking is a practical solution to current digital rights management problems, and are firmly opposed to the use of this or other technologies that infringe the Fair Use rights of the public.

For many applications, such as speech, compression/expansion ratios of up to five or ten percent may be usable, leading to a corresponding increase in the encoding rate.

2. PRIOR WORK

This method has several significant advantages over previous approaches. First of all, for most audio it will be virtually undetectable, because of the human auditory system's insensitivity to extremely low frequency modulation. At the same time it is exceptionally robust to transmission and compression, because current digital audio technology has a time precision on the order of several microseconds in an hour. Typical audio sources such as speech or music have enough natural variation that the artificial tempo changes induced by the watermarking will be neither audible nor easily detectable. The exception might be strictly rhythmic music produced by a computer sequencer or other mechanical device. In this case, fine analysis of the beat-to-beat spacing might reveal the tempo modifications, though they might not be perceptible to a listener. Tempo changes such as those inherent in analog recording and reproduction equipment will not interfere with the watermark. A straight tempo change, caused, for example, by an inaccurate playback speed will not affect the watermark. Analog recording imperfections commonly called "wow" and "flutter" occur on a timescale significantly shorter than the watermark tempo changes, and will tend to average out, leaving the watermark unaffected [6]. A possible limitation is that the watermark can be partially obscured or degraded by intentionally changing the time scale of audio regions, or superimposing another tempo-based watermark, but this will not actually remove the watermark unless the exact inverse of the compression and expansion is known and used (which requires the reference audio to determine). If the watermark is being used as a digital signature, this alteration will invalidate both the watermark as well as the signature, and will be easily detectable. Note that few if any other watermarking schemes are robust under the application of multiple watermarks.

2.1 Existing audio watermarking techniques

Prior-art audio watermarking techniques can be classified into data-domain and frequency-domain methods. Data-domain methods work by modifying the actual audio data, such as dithering the least significant bit in a PCM representation or hiding data in compressed-domain representations. Frequency-domain methods work by modifying the spectral content of a signal, for example by removing a particular frequency component or adding information disguised as low-amplitude noise. Time base modulation is significantly novel enough that it fits into neither category. Also significant is that time base modulation has distinct advantages over both prior-art approaches: unlike data-domain methods, it is robust to transmission, encoding, and compression, and unlike frequency-domain methods it does not alter the frequency spectrum nor add noise. In fact, time base modulation is completely "orthogonal" to prior-art methods: time base watermarking can be done independently of, and will not interfere with, other watermarking techniques, and could be combined with any of the prior-art methods to increase the information capacity and robustness of the watermark. For audio, prior-art watermarking techniques can be classified into data-domain and frequency-domain methods.

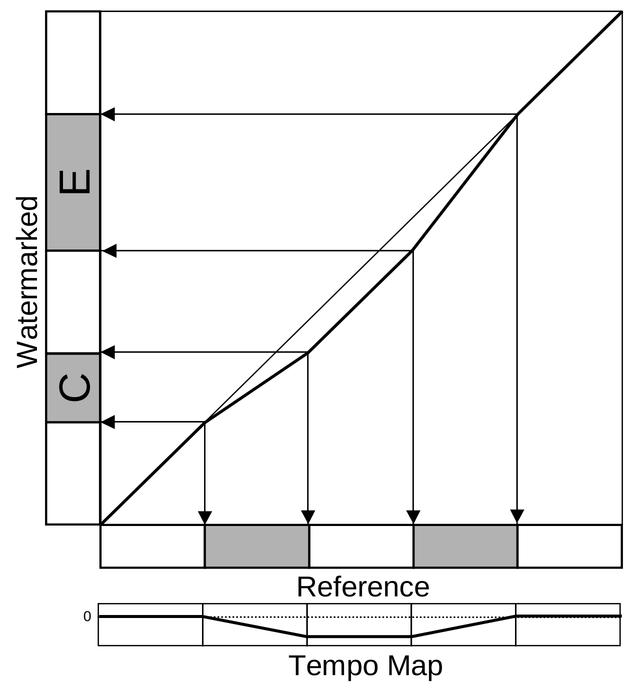


Figure 2. Recovering the "tempo map" from the dynamic alignment path of watermarked and reference signals.

Data-domain methods include compressed-domain methods; bit dithering; frequency-domain methods, including spread-spectrum and phase modulation; amplitude modulation; and short-time echo and autocorrelation modification. There may be other methods that have not have been disclosed, and are possibly trade secret. None of these resemble the methods described here; in fact the tempo-modulation approach can be combined with any of the prior-art methods to increase the information capacity and robustness of the watermark signal. Data-domain watermarking techniques include:

•**Compressed-domain watermarking**. In this technique, only the compressed representation of the data is watermarked, and is thus not persistent [10]. When the data is uncompressed, the watermark is not available, unlike the present invention, where the watermark persists across all analog and digital reproduction and compression techniques.

•**Bit dithering**. In this method, information is encoded by modulating the least significant bits of the time-domain or compressed representation. While this potentially has a large data rate, it is not robust to compression or analog transmission and reproduction, and introduces noise into the signal.

•**Amplitude modulation**. In this method, signal peaks are modified to fall within predetermined amplitude bands [7]. This technique introduces modulation distortion, and is not robust to amplitude compression, which is widely used in analog and digital telephony, broadcasting, sound reinforcement, and noise reduction.

•**Echo hiding**. In this method, discrete copies of the original signal are mixed in with the original signal. The echo time is short enough and the copy amplitude is low enough to be inaudible, yet the echo can be detected via autocorrelation [8]. This method introduces spectral distortion because of phase cancellation at frequencies whose periods are multiples of the echo delay. Also, this technique may not be robust under compression, as imperceptible echoes are likely to be discarded by perceptual coding.

Frequency-domain watermarking techniques include:

•**Phase coding**. This technique relies on the human auditory system's relative insensitivity to phase. The signal is windowed, as in a spectrogram, and the magnitude and phase of each window is computed. An artificial absolute phase signal, which encodes the watermark, is introduced into the first window. The phase information for subsequent frames is iteratively computed from the phase differences from each frame and the absolute phase. The resulting phases are combined with the original magnitudes to construct the watermarked signal [8,8]. This method introduces phase dispersion into the signal, and is probably not robust under compression.

•**Frequency band modification**. In this method, information is encoded by removing or enhancing particular spectral bands, removing a narrow spectral band using a notch filter, or encoded into frequency band differences [11]. This method introduces spectral distortion, may not be robust to perceptual encoding, and does not work unless the altered frequency components are well-represented in the source audio.

•**Spread spectrum**. In this method, a signal carrying the watermark information is modulated into wideband noise by multiplication with a pseudorandom sequence. Because the modulation function is known (or can be regenerated), the watermark signal can be demodulated [9]. This technique adds noise to the watermarked signal, and the low amplitude of the spread spectrum signal means it may be likely to be discarded under perceptual coding. In addition, the sampling frequency is commonly used as the modulation carrier frequency to avoid having to synchronize the receiver. In this case, resampling or analog transmission is likely to destroy the synchronization, and hence the watermark.

Compressed-domain watermarking schemes are not relevant because they do not work, as time base modulation does, on the uncompressed audio signal. Many other schemes, particularly dithering and frequency domain approaches, are not robust to audio compression. This is especially problematic, as to be a good watermark the frequency modifications must be perceptually inaudible, yet this is precisely the information that is lost or altered when perceptual compression schemes such as MP3 are used. In contrast, time base modulation has shown to be robust over both perceptual encoding and analog reproductions, and unlike prior art, introduces no noise, echo, spectral or short-time phase distortion to the watermarked signal, nor does it remove or otherwise alter frequency components or signal amplitudes. To reemphasize, the time base modulation approach presented here can be done theoretically noiselessly and with no alteration of the signal spectrum or phase.

2.2 Methods for watermarking images

There has been considerable work in watermarking images. Most approaches are quite similar to those described above, for example spread spectrum techniques can be used for images as well as audio [9]. One relevant prior-art approach for watermarking text modulates white space between words and sentences [8]. This method needs to detect word boundaries, and is not applicable to common images other than scanned text (such as Figure 8). The Glyph technology developed at Xerox PARC encodes information into digital hardcopy using tiny marks that can be dithered to produce gray shades [11]. The "Patchwork" watermarking system alters the intensity of random pairs of points in the image [8]. A method called texture block coding encodes information by copying areas of random texture. These areas can be found by autocorrelation [8]. As with audio, these methods are all orthogonal to time base modulation (for images, a better name might be spatial modulation), and could be used in concert.

3. WATERMARKING VIA TIME BASE MODULATION

Techniques for scaling the pitch of an audio signal are well-known and in common use. These techniques can be used equally well for changing the length of an audio recording without the objectionable ("chipmunk" or "Darth Vader") pitch modification introduced by a simple rate change. A frequent application is to play back audio at a higher rate, so that it may be auditioned in less time. A common time-scaling technique is based on the Short-Time Fourier Transform, but other methods such as the Phase Vocoder, Time Domain Harmonic Scaling, and Pitch-Synchronous Overlap Add (PSOLA) are also widely used [2,3,4,]. Any time-scaling method can be used for the watermarking procedure; preferred methods can compress or expand by ratios very near one with few audible artifacts.

Consider the signal $x(t)$ expressed as the concatenation of K non-overlapping blocks, $x_{1\ldots K}$, with concatenation denoted by **C**.

$$x(t) \equiv \mathop{\mathbf{C}}_{k=1}^{K} x_k$$

The watermarked signal, $x_w(t)$, is generated by performing a compression or expansion of each block, x_k, by an amount denoted by E_k, and concatenating the modified blocks.

$$x_w(t) = \mathop{\mathbf{C}}_{k=1}^{K} TSM(x_k, E_k)$$

Where $TSM(x_k, E_k)$ indicates the time scale modification of block x_k by the amount E_k. In practice, care may be required to avoid introducing audible discontinuities at the block boundaries. This may be achieved by using a time scale modification algorithm that leaves data at or near the block boundaries unchanged, or by overlapping segments slightly and averaging data within the overlapping region during the construction of the watermarked signal.

The sequence of expansion/compression values, E_k, encodes the watermark, and can be recovered by comparing the watermarked version with the reference (unaltered) audio. This is done by finding the time-warping function that best matches the reference file to the watermarked file. Subtracting the linear component of the alignment path yields the "tempo map" from which the water-

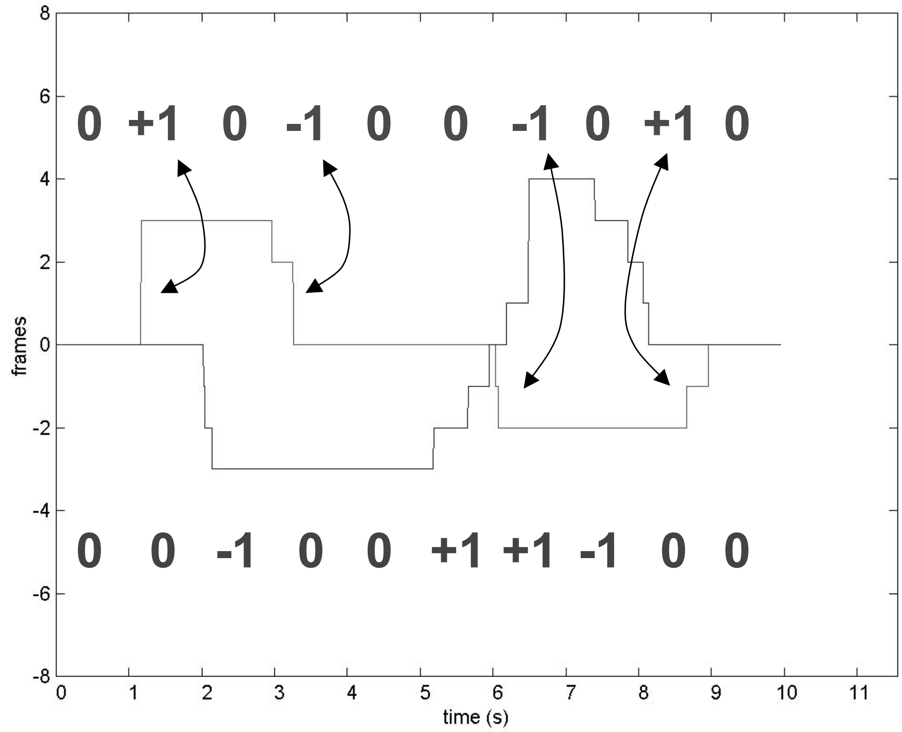


Figure 3. Two different watermarks recovered by dynamic programming. Compressed regions are indicated by +1, while expanded regions are indicated by -1.

mark information can be determined. After removal of the linear component, the path has positive slope in the compressed regions, negative slope in the expanded regions, and slope zero in unaltered regions, but may be offset from zero by preceding compressions or expansions. This is illustrated in Figure 2. The slope of the recovered tempo map corresponds to the values of E_k used during encoding. This fact is used to recover watermarks later in this Section: a tempo map can be estimated from the derivative of E_k and compared with the recovered watermark.

In the current implementation, reference and watermarked signals are matched using short-time spectral magnitude features. The analysis frames are 128 samples wide. For audio sampled at 22.05 kHz, this results in a 5.8 ms frame width and a frame rate of 172 frames per second. Each analysis frame is windowed with a Hamming window and converted to the frequency domain via a fast Fourier transform (FFT). The logarithm of the magnitude of the result is used as an estimate of the power spectrum of the windowed frame. The resulting vector of spectral components characterizes the spectral content of a window. The sequence of spectral vectors, represents the frequency content of the signal over time. Some frequency components may be optionally discarded if they are not as useful for the similarity calculation, for example extreme low or high frequency bands which may not have substantial power.

To find the best warping path, spectrograms for both the reference and test signals are computed. The spectrograms are aligned by warping the reference to match the test signal. This is done using dynamic programming (DP). Because DP is well-documented in the literature [4,5], the algorithmic details will not be reproduced here, but it can be shown to find the optimal alignment path in quadratic time. The DP algorithm is especially well suited to recovering the watermark, and easily adapts to special conditions of this application. For example, DP gracefully handles the case when test and reference files do not start and end at exactly the same time. For example, if the watermarked signal were extracted from a continuous broadcast, it is not necessary to specify the start and end points exactly. Another feature is that the frame spectra need not match exactly, as long as they are more similar than neighbors, thus the system is robust to reasonable spectral distortion. Also, in this application the expected displacement is very small; that is, the best path does not significantly deviate from the diagonal. In this case, the DP algorithm can be made effectively linear time by "cutting corners," that is by computing only paths very near the diagonal [4]. Similarly, an overall time modification (caused, e.g., by sampling rate conversion or incorrect analog reproduction speed) is gracefully handled by the DP algorithm. In this case, the watermark is recovered by subtracting the diagonal of the rectangle formed by the cross-product of the two signals, rather than the square.

The overall data rate of the watermark is a trade-off with the detectability of the watermark and the degradation of the signal. For the purposes of explanation, we call the minimum length of a compression/expansion interval a "block." For simplicity, let us assume that all blocks are the same length (though uncompressed blocks might usefully be smaller than altered blocks). Each block can be compressed or expanded by a factor $1 \pm \varepsilon$. If ε is suffi-

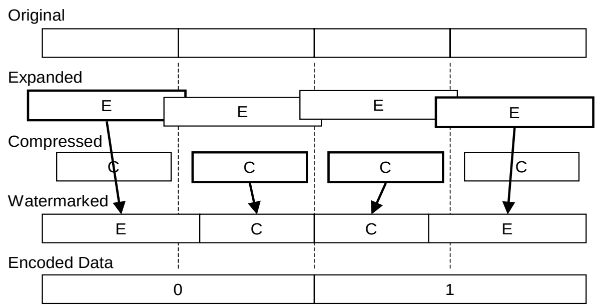


Figure 3. Construction of a 2-bit watermarked signal from compressed and expanded segments.

ciently small, then compression/expansion can be discretized to an integral factor of ε, i.e. $1 \pm n\varepsilon$, whereby n is a small integer, possibly negative. Blocks can be left uncompressed, *i.e.:* $n = 0$. To reduce audible artifacts, it is advisable (though not strictly necessary) that the magnitude of n be limited to less than some small value, say N. For the same reason, the change in n should be small between adjacent blocks. To preserve the time length of the file, it is preferable, although again not strictly necessary, that n sum to zero across all blocks in the signal, so the amount of compression exactly equals the amount of expansion. Thus every block k has an associated multiplier, n_k, such that $-N \le n_k \le N$ and a corresponding compression/expansion value $E = 1 + n_k\varepsilon$. A particular watermark thus corresponds to the sequence $E_0, E_1, \ldots, E_K$. This sequence can subsequently be recovered by quantizing the derivative of the tempo map. Experiments have shown that reasonable values are a block length of 0.5 seconds, $\varepsilon \approx 0.01$, and $N = 2$. Thus each second of audio can encode roughly $2\log_2(2N + 1)$ bits, or slightly more than 8 bits. While this is not a huge data rate, note that a typical popular song is 180 seconds long. At a data rate of 8 bps, this could encode 180 bytes, or more than enough data for song title, artist, publisher, and an ID. When used as a watermark, 180 bytes yields more than 10^{400} individual identifiers, which would be more than enough for any conceivable combination of source identifiers, device identifiers, and timestamps.

Figure 4 shows the results of an initial watermarking demonstration. Two copies of a 10-second Beatles song excerpt were watermarked. Compression and expansion ratios were two percent, over 1-second regions, thus a total displacement of 20 ms, or 3.44 frames. The first copy was expanded at 1 and 8 seconds and compressed at 3 and 6 seconds. The other copy was compressed at 2 and 7 seconds and expanded at 5 and 6 seconds. Tempo maps from the dynamic programming algorithm are shown in Figure 4, which plots deviation from linear time in spectrogram frames. The compression and expansion regions are easily detectable, as are the plateaus of time offset (but normal tempo) caused by the compression. The time difference was recovered to within ± 1 frame, indicating that an additional level of compression/expansion could be used to effectively double the information capacity. This approach has been tried on other audio domains, including soundtrack speech and orchestral music with similarly good

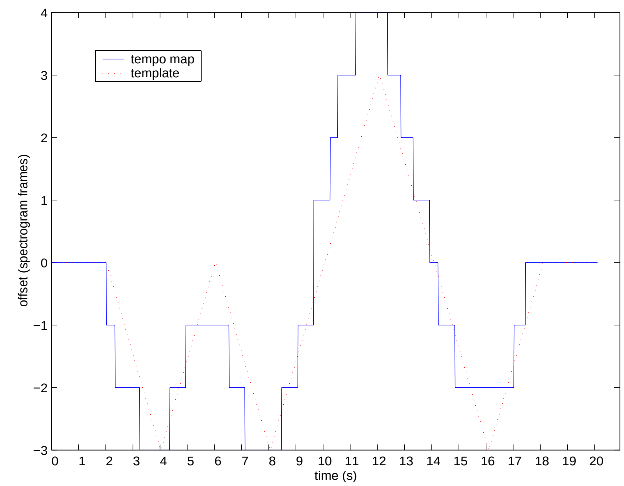


Figure 4. Recovered tempo map and expected template for binary 0010 watermark

results[1]. Informal listening tests from 13 volunteer listeners have confirmed the inaudibility of the time base modulation, though one "golden-eared" listener was able to detect artifacts from the time compression in direct A/B comparison with the reference audio. Superior time compression methods may have fewer artifacts. An additional experiment demonstrated that the watermark easily survived 64 kB MP3 encoding and decoding.

Another experiment tested the recoverability of watermarks. In this case, the reference signal was the first 20 seconds of the song "Magical Mystery Tour" by the Beatles, converted to a monophonic representation at 20,050 Hz sampling rate. (The method will, of course, work with any sampling rate and can be applied to stereo or multichannel audio in parallel across channels.)

A naive encoding scheme was used to encode a unique 4 bit watermark in 16 different copies of the audio. In this scheme, two two-second blocks were used to encode one bit of information. A compression followed by an expansion represented a binary "one," while the reverse order indicated a "zero". See Figure 3. Obviously there are more efficient coding schemes, particularly those that use a region of no time scale modification to encode an additional state.

Given this coding scheme, it is simple to generate a watermarked signal on the fly by concatenating compressed and expanded regions of the signal. In this case, a compressed and expanded versions of the signal were generated with a ratio of 2.5%; thus the expanded version had a duration of 20.5 seconds. Each version was evenly divided into 10 equal blocks, each two seconds long. A watermarked signal was created by the simple method of concatenating compressed and expanded blocks. Blocks at the beginning and end of the signal were not compressed, thus only the middle 16 seconds were altered. (In certain cases, this resulted in audible artifacts from mismatched block edges; in a practical

[1]Original and watermarked audio examples may be heard at http://www.fxpal.com/people/foote/musicr/watermark/index.htm

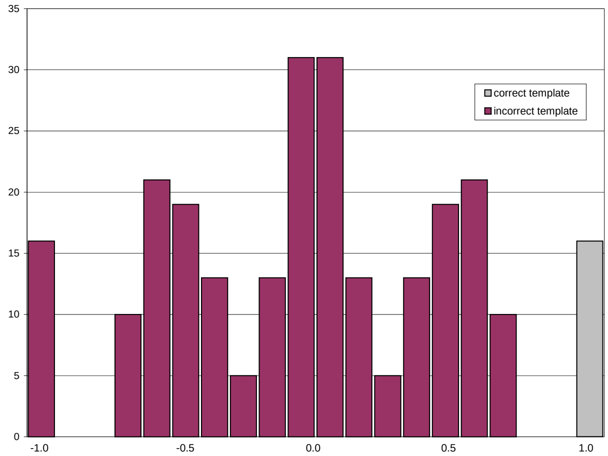


Figure 5. Score histogram for watermark recovery experiment

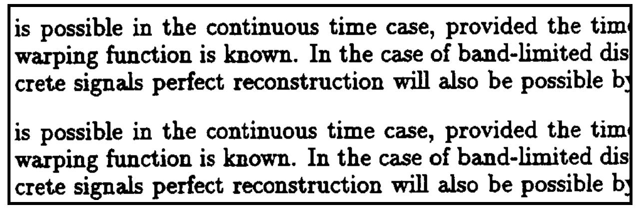
system these could be eliminated by cross-fading the blocks over a short period.)

Given a known sequence of compression and expansion, it is straightforward to estimate what the tempo map should be in each case. Given a region of compression followed by expansion, the tempo map will speed up then slow again to zero, indicating a binary “one.” Conversely, a binary zero will result in a tempo map feature that dips below zero in a “V” shape. A tempo map is the derivative of the compression/expansion rate. Because these are blockwise constant, the resulting tempo map can be predicted by blockwise-linear ramps. Accordingly, “templates” were constructed having linear ramps corresponding to the expected tempo changes. Figure 4 shows the template and recovered tempo map for the copy watermarked with binary “0010.” This corresponded to a compression/expansion sequence of CECEECCE, where “C” and “E” represent a compressed or expanded two-second block.

A tempo map was extracted for each watermarked version, and compared with the sixteen expected templates. Given a tempo map $\vec{m}$ and a template $\vec{t}$, a useful metric is the cosine of the angle between them.

$$D_C(\vec{m}, \vec{t}) \equiv \frac{\vec{m} \bullet \vec{t}}{\|\vec{m}\| \|\vec{t}\|}$$

.This has the property that it can yield a good similarity score regardless of the actual vector magnitudes. The experiment showed that the expected watermarks show a much higher cosine similarity with the expected template than with any other template. Of the 256 tempo map-template comparisons, the maximum cosine distance for the incorrect template was 0.618, while the minimum score for the correct template was 0.9094. Figure 5 shows the cosine distances for all 256 watermark-template comparisons. All correct templates had scores greater than 0.9, while all incorrect templates had scores less than 0.62. The score differences are proportional to the Hamming distance between the watermark and the template. To increase the score distance, and thus robustness, a subset of codes with larger Hamming distances can be used. For example, if the eight four-bit codes with odd parity are used (guaranteeing a Hamming distance of two or more)


is possible in the continuous time case, provided the tim
warping function is known. In the case of band-limited dis
crete signals perfect reconstruction will also be possible b

is possible in the continuous time case, provided the tim
warping function is known. In the case of band-limited dis
crete signals perfect reconstruction will also be possible b

Figure 6. Original (top) and watermarked scanned text.

the maximum incorrect template score is reduced to 0.238. Note that it is not necessary to use templates; the watermark can be extracted by straightforward thresholding methods. For example, inspecting Figure 4 shows the binary digits could be extracted if thresholds were set at +/1 frames

5. USING NON-LINEAR SCALE MODIFICATION FOR WATERMARKING IMAGES AND SCANNED TEXT

Besides audio and video, this technique can be used to watermark still images. Like audio, the watermarking process has the advantage that it is robust under lossy compression and analog reproduction. In addition, the watermark encoding can easily be implemented optically as well as digitally, even directly in the mechanism of a printer or photocopier. This might be especially valuable for high-security applications: a photocopier or printer could encode the time, date, and device or user ID invisibly into every copy made or printed. If an illicit copy is found, the watermark could help locate the source.

In this method, areas of the image are compressed or expanded by an imperceptible amount. Well-known digital resampling techniques can stretch or compress image regions by a small amount. Alternatively, mechanical or optical methods can be used, such as varying the speed of a drum or platen scanner, varying the print head speed in a printer, or by varying the speed of the cylindrical objective lens in a photocopier with respect to the drum. At least two axes of warping (vertical and horizontal) are available; it is also possible to differentially warp “stripes” across the image, although this may lead to more noticeable artifacts as straight lines not parallel to the warp axis will no longer be perfectly straight. For a small class of images this will lead to visible artifacts, particularly on images that have regular lines or grids that run diagonally. As in the audio case, the image can be analyzed to find the regions or mode of watermarking that will result in the least perceptible alterations. For example, a Fourier analysis of the image can find the angular direction with the lowest energy. Using this direction as the warping axis will minimize perceptual artifacts. For example, given an image of parallel lines, Fourier analysis can easily find the direction of the lines. Warping the image parallel to that direction will result in less perceptible artifacts. In general, given the small degree of warping, the watermarking will not be perceptible on the vast majority of images. In particular, scanned text is especially immune as the natural variation due to kerning and line-filling masks the warped regions particularly well. Figure 6 shows the watermarked and original scans of a paper document.

Figure 8. Original (left) and watermarked image (center), with image difference (right)

To recover the watermark, the reference image is compared with the watermark image and the warping recovered using dynamic programming. Unlike the audio case, the pixels can be compared directly. Columns of pixels perpendicular to the warp axis can be compared by Euclidean or other distance, just as spectral vectors are compared in the audio case.

6. FUTURE WORK

This method could be used to watermark video or motion pictures by watermarking the audio. It is straightforward to remove or duplicate frames to maintain synchronization with the time-altered audio. Only a very few frames would need to be altered as the time base modification is very small. We also show how a similar technique can be used to imperceptibly encode a small amount of information in still images or scanned text. This method could be used to watermark particular video or motion picture frames in addition to watermarking the audio

A particular advantage of this method is that it is computationally reasonable enough to encode the watermark on the client side, perhaps as part of a streaming media player. In this case, the server could send the watermark data as well as audio (possibly encrypted) to the client. The client would then uncompress/decrypt the streaming media and apply the watermark to the audio before buffering, playing or locally storing the signal, thus reducing the computational load on the server. In fact, the watermark signal could encode information such as the time the file was played or the IP address or other information that could identify the particular playback device

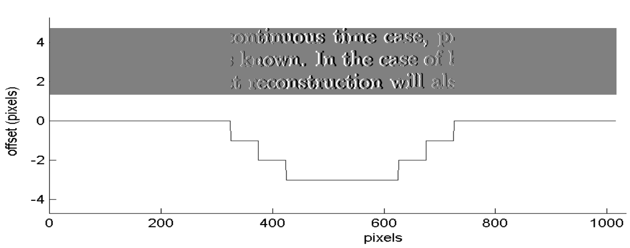


Figure 7. "Tempo" map recovered from scanned text

A drawback of this and many prior-art watermarking methods is the need for the reference signal to decode the watermark. While this is perfectly feasible for many applications (for example digital rights management where the copyright owner would have access to the unmarked signal), there are applications where it might be desirable to extract the watermark without reference to the unaltered signal. Time base modulation watermarking can be used in the absence of the reference signal if the actual tempo can be inferred or predicted. Many methods exist to analyze the tempo or speaking rate of audio [12]. From this, it is a simple matter to construct a rate predictor that will predict the signal at some short time in the future. The simplest predictor might be a first-order constant rate that assumes a constant tempo. A rate predictor can be used to encode and decode a time base watermark in the absence of a reference signal. The original signal is analyzed, and its tempo is altered to match that of the predictor using time scaling methods previously described. In the case of the first-order predictor, the rate-adjusted signal will have a constant tempo. For higher-order predictors, the rate-adjusted signal will have exactly the tempo prescribed by the predictor. This rate adjusted signal can now be watermarked using the methods described. To recover the signal, note that the only rate differences between the rate predictor and the watermarked signal are due to the watermark. Thus the difference between the watermarked signal rate and the predicted rate reveals the watermark. Note that the predictor need not be accurate as long as it is consistent, though the more accurate the rate predictor, the better the adjusted-rate signal will match the original.

7. REFERENCES

[1] J. Zhao, E. Koch and C. Luo, "Digital Watermarking In Business Today and Tomorrow," in *Communications of ACM.* Vol. 41, No. 7, July 1998.

[2] S. Sprenger, "Time and Pitch Scaling of Audio Signals," http://www.dspdimension.com/html/timepitch.html

[3] Roucos, S., et al, "High Quality Time-Scale Modification for Speech", Proceedings of the 1985 IEEE Conference on

Acoustics, Speech and Signal Processing, 1985, pp. 493-496.

[4] J. Kruskal and D. Sankoff, “An Anthology of Algorithms and Concepts for Sequence Comparison,” in *Time Warps, String Edits, and Macromolecules: the Theory and Practice of String Comparison*, eds. D. Sankoff and J. Kruskal, CSLI Publications, 1999

[5] J. Lacy, S. Quackenbush, A. Reibman, and J. Snyders, “Intellectual Property Protection Systems and Digital Watermarking,” in *Proc. Second Workshop On Information Hiding, Lecture Notes in Computer Science*, Vol. 1525, Springer-Verlag. 1998 pp. 144-158.

[6] S. J. Godsill, “Recursive Restoration of Pitch Variation Defects in Musical Recordings,” in *Proc. ICASSP 94*, Volume 2, pp 233-236, 1994

[7] Wolosewicz, et. al., “Apparatus and method for encoding and decoding information in analog signals,” United States Patent 5,828,325, Oct. 27, 1998,

[8] W. Bender D. Gruhl, N. Morimoto, A. Lu “Techniques for Data Hiding,” in *IBM Systems Journal*, Vol 35, Nos 3&4, 1996, pp 313 -336

[9] I. Cox, J. Kilian, T. Leighton and T. Shamoon, “Secure Spread Spectrum Watermarking for Images, Audio and Video,” in *Proc. of 1996 Int'l. Conf. on Image Processing (ICIP'96)*, vol. III, pp. 243-246, (1996).

[10] C. Xu & J. Wu, “Content-Based Watermarking for Compressed Audio,” in *Proc. Recherche d'Informations Assistee par Ordinateur* [RIAO], April 2000.

[11] D. Hecht, “Embedded Data Glyph Technology for Digital Hardcopy Documents,” in *Proc. SPIE*, vol. 2171, 1995

[12] J. Foote and S. Uchihashi, S., “The Beat Spectrum: A New Approach to Rhythm Analysis,” in *Proc. IEEE International Conference on Multimedia and Expo* (ICME) 2001 http://www.fxpal.com/people/foote/papers/ICME2001.htm