AUDIO WATERMARKING USING SUBBAND MODULATION SPECTRA

Sascha Disch[*], *Jürgen Herre*[*], *Julius Kammerl*[**]

[*]dsh@iis.fraunhofer.de, hrr@iis.fraunhofer.de, Fraunhofer IIS
[**]kammerl@tum.de, TU München

ABSTRACT

Subband modulation spectra (SB-MS) have proven to be useful for many applications in audio signal analysis and audio signal processing. This paper presents a novel blind watermarking scheme for embedding and detecting a watermark in the subband modulation spectral domain representation of an audio signal. Some results of audio quality, data rate and robustness are presented. The scheme is shown to be robust to a number of attacks while providing excellent subjective audio quality.

Index Terms— *Audio Systems, Amplitude Modulation, Frequency Modulation, Robustness*

1. INTRODUCTION

Subband modulation spectra (SB-MS) have recently gained a growing interest in the audio processing community. SB-MS have been proposed by Greenberg et al. [1] for audio signal analysis. There were earlier publications reporting on perceptual masking effects in the modulation domain [2] described by a model similar to the well-known frequency-domain masking models [3]. This led to various publications in which SB-MS was used for audio signal analysis and processing: Vinton and Atlas suggested a codec based on SB-MS for scalable audio compression [4], Avendano and Goodwin presented a method of signal quality enhancement [5]. Furthermore SB-MS have been proposed for fingerprinting audio content identification [6].

This paper presents first investigations into the use of SB-MS for audio watermarking. It is structured as follows. First, the underlying signal model is given and two different types of SB-MS analysis/synthesis approaches are reviewed. This will be followed by a brief description of some relevant background on audio watermarking. The main part of the paper focuses on a novel system for embedding/detecting a watermark in the subband modulation spectral domain and presents results of perceptual quality and robustness of the watermark at a given transmission data rate.

2. SUBBAND MODULATION SPECTRA

2.1. Analysis

A decomposition of a signal into subband modulation spectra consists of a first analysis transform, a demodulation operation on each subband signal (i.e. dismantling the signal into a carrier and a modulator) and a second analysis transform on the envelope of each subband signal yielding the desired subband modulation spectra [7]. The demodulation step is a blind de-multiplication problem in the time domain (or a blind de-convolution operation in the spectral domain) which in general has an infinite number of solutions. Thus, some assumptions on the carrier signal or the modulator signal have to be made.

In the following it is assumed that the discrete-time signal has been decomposed by a DFT into P complex subband signals $\hat{x}_p$. These are to be separated into a modulator m_p and a sinusoidal carrier c_p, each

$$\hat{x}_p(n) = x_p(n) + jH\{x_p(n)\} = m_p(n) \cdot c_p(n) \quad (1)$$

$$c_p(n) = A_p \cos(\omega_p n + \varphi_p) \quad (2)$$

with $H\{\ \}$ denoting the Hilbert transform.

A straightforward assumption is to always consider a DFT's bin center frequency to be the carrier frequency and demodulate the analytic signal obtained from the DFT's one-sided spectrum asynchronously by a simple magnitude operation [7]. This leads to the first type of demodulation: The *asynchronous demodulation* is accomplished by

$$m_p(n) = \frac{1}{A_p}\left|\hat{x}_p(n)\right| \quad (3)$$

$$\forall m_p \in \mathbb{R}, m_p \geq 0$$

Note that the modulator containing the Amplitude Modulation (AM) is restricted to be real and non-negative. The Frequency Modulation (FM) is contained in the phase of the analytic signal and is usually not subjected to any further processing.

An alternative possibility for demodulation is so-called *synchronous demodulation* [7]. Here, the carrier signal is estimated in each subband and a demodulation towards the bin center frequency is performed by a complex multiplication:

$$m_p(n) = \hat{x}_p(n) \cdot \frac{1}{A_p} \exp(-j(\omega_p n + \varphi_p)) \quad (4)$$

$$m_p^{AM}(n) = \Re\{m_p(n)\}; m_p^{FM}(n) = \Im\{m_p(n)\}. \quad (5)$$

In this scenario, two signals are obtained to be further processed, i.e. the AM part and the FM part of the signal. Note that the analysis transform used to obtain the subband signals is not ideal ('brick wall' characteristic) in its frequency separation capability. If, for example, an AM modulated sinusoidal carrier signal of a given frequency is split into P frequency bands with fixed bandwidth and frequency borders, the AM sideband pairs that are located around the estimated carrier signal on a frequency axis might be damped asymmetrically by the corresponding subband filter or might even not be situated completely in the passband of the specific subband filter. In this case, undesired AM-FM and FM-AM conversion w.r.t. the subband signal takes place.

2.2. Synthesis

To allow signal modifications in the SB-MS domain, the analysis transform needs to be invertible, preferably assuring the perfect reconstruction (PR) property. For the asynchronously demodulated signals the (modified) magnitude is recombined with the phase prior to a transformation back into the time domain. In case of the synchronous demodulation, the AM and FM parts are re-joined to form a (modified) complex signal.

3. AUDIO WATERMARKING

A watermark in signal processing is defined as non perceptible, robustly embedded additional information in a carrier signal. Many approaches to audio watermarking have been published so far, including 'echo hiding' [8] where the binary watermark information is embedded in the carrier signal as additive echoes having a certain delay [9], the 'spread spectrum' methods where the binary data is contained in a pseudo noise sequence which is shaped according to psychoacoustic criteria and added to the carrier signal [10]. Another class of methods depends on the manipulation of least significant bits in a binary representation of the carrier signal [10][11]. In this paper we present a novel method of embedding a watermark in the SB-MS domain.

4. WATERMARKING IN SB-MS DOMAIN

4.1. System overview

In Figure 1 an overview of a SB-MS watermarking chain is depicted. It consists of a watermark embedder, a transmission channel and a watermark detector. In the embedder the input audio signal x(n) is transformed into the modulation domain by the analysis stage in a block-wise

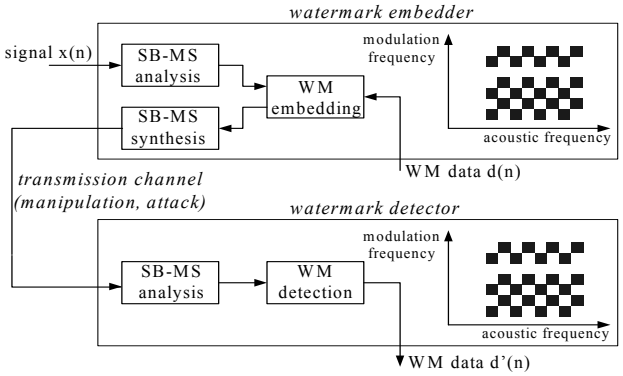

Figure 1. Overview of SB-AMS watermarking system.

fashion. The analysis stage itself is composed of the first analysis transform T1, an asynchronous demodulation stage and the second transform T2 (Figure 2). Subsequently the watermark is embedded into the modulation spectrum by application of suitable weights to certain tiles of the frequency/modulation frequency plane according to the binary watermark data sequence d(n). More precisely, the key idea of the SB-MS watermark for embedding the hidden information at the embedder side is to perform a time variant modulation frequency notch filtering. The modified (watermarked) time domain signal is obtained after passing the synthesis stage.

At the detector side, an analysis stage provides the modulation spectrum in an analogous way like described for the embedder. Usually, most real world audio signals are characterized by a broad and continuous energy distribution in the modulation domain. Thus, the detection of the watermark can be accomplished by comparing the mean energy measured at embedding (notch) positions in the modulation spectrum with an energy estimate derived from the surroundings of the respective notch area. A decision stage finally yields the recovered watermark binary sequence d'(n).

For improved detection stability, a set of notches along acoustic and modulation frequency is embedded, constituting a checkerboard pattern. In this case, the embedding/detection is favorably carried out in a differential fashion: Two nested checkerboard patterns are used, each covering similar frequency/modulation frequency tiles and having the same surface area in the SB-MS. Thus, the mean energy in the covered areas of both patterns tends to be similar for non-watermarked signals. Therefore, the measured modulation energy difference converges towards zero with increasing size and resolution of the checkerboard patterns. For embedding the binary symbols of the watermark, the fields of one checkerboard pattern are engrained by reducing energy in the selected areas. For recovering the embedded information, the algebraic sign of

the energy difference of the two nested checkerboard patterns is evaluated.

4.2. System enhancement

The effective temporal scope of an SB-MS coefficient is determined by the temporal support of both analysis stages T1 and T2. For common choices of parameters for T1 and T2, this amounts to time resolutions in the order of several hundred milliseconds and thus inevitably leads to pre- and post-echo artifacts [12] in watermarked signals containing dominant transient components. To overcome this problem, a multi-band gain control has been employed to shape the processed signal in order to match the signal's original envelope within the temporal granularity of the first analysis transform. The location of the gain control within the system is shown in Figure 2. The gain factors α_i are derived from the spectral power distribution of the input signal of the embedder and detector, respectively.

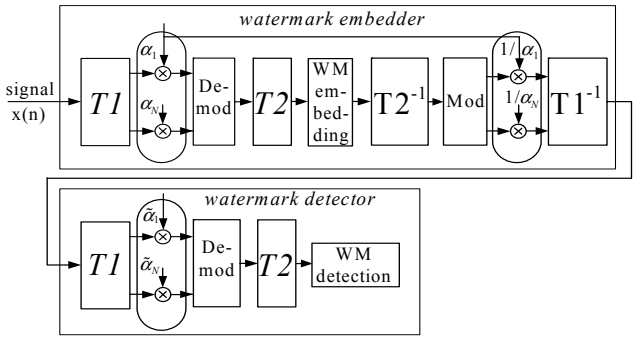

Figure 2. Gain control placement in watermark system.

Taken into account the interaction between gain control and second transform stage, the gain control is also calculated in the detector in the same way.

4.3. System parameters

For both transforms T1 and T2, an oddly stacked DFT/IDFT with a 256 tap window length and 50% overlap was chosen, yielding modulation spectra with 128 acoustic frequency and 128 modulation frequency bands resolving modulation frequencies up to ~172 Hz (signal sampled at 44.1 kHz).

The watermark data was embedded into the area of 2 - 22 kHz (acoustic frequency) and 20 - 100 Hz modulation frequency. To obtain continuous synchronization of the embedded data sequence in the detector's analysis stage, an accompanying clock-signal is additionally embedded in the 110 - 150 Hz modulation frequency band. Hence, robustness for attacks like "time shift" and "time scaling" can be achieved.

4.4. System operation point

The operation point (OP) of any audio watermark system has to be positioned somewhere inside the trade-off triangle of audio quality, data rate and robustness [13]. This is illustrated in Figure 3.

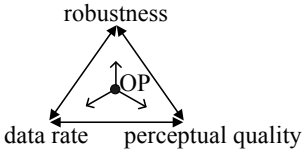

Figure 3. Watermark system trade-off triangle [13].

For the measurements presented in the following section we chose to embed one bit per short-time modulation spectrum. Given the selected processing block size for the two transformation stages, a data rate of 4.5 bits/s is achieved, while providing high transparency and strong robustness against several attacks, as our tests have proven. The data rate could be increased by e.g. subdividing the aforementioned checkerboard patterns into several zones dedicated to accommodate one bit each. However, at a given perceptual quality, this will result in a decrease in robustness as depicted in the trade-off triangle.

5. RESULTS

A watermarking system has been built according to the previously explained principle structure and tested for its subjective audio quality and robustness.

5.1. Audio Quality

The audio quality of the watermarked signals was tested using an ITU-T BS.1116 standardized listening test [14]. The test set, which is known from MPEG Audio coding standardization, includes mostly critical music material:

- t1: Pop music
- t2: Bag pipe
- t3: Glockenspiel
- t4: Pitch pipe
- t5: Castanets
- t6: Classical music
- t7: Male speech
- t8: A-capella (Suzanne Vega)

The signal was presented via headphones in a dedicated, acoustically isolated listening lab. Eight experienced listeners participated in the test. The results are depicted in Figure 4 as mean values and 95% confidence intervals of subjective difference grades (i.e. the difference of signal-under-test score minus reference-signal score). The grades are according to the ITU-R BS.562 five grade impairment scale, ranging from 5.0 ('imperceptible') to 1.0 ('very annoying'). Thus, a difference grade of 0.0 corresponds to an imperceptible impairment.

The results show that the watermark was not found to be perceptible for most test items except for the a-capella item

't8' where a small but statistically significant degradation was perceived (the confidence interval does not include the "0"point).

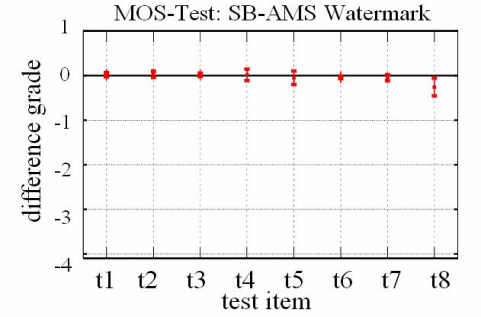

Figure 4. Subjective listening test results.

5.2. Robustness

A set of common types of attacks on the watermarked signal have been simulated to characterize the robustness of the proposed method. These methods of attack were

- Time shift (constant offset between embedding clock and detection clock)
- Perceptual coding MPEG-1/2 Layer 3
- DA-AD conversion (by Soundblaster AWE64 Sound Card)
- Filtering (lowpass)
- Echo addition
- Dynamic range compression (commonly used by radio broadcast stations)
- Changed playback speed (without/with pitch correction)

All signal processing to compute filtering, echo, dynamic compression and speed change was conducted by using the "Audacity" software package [15]. The results in terms of bit error rate (BER) are listed in Table 1.

Type of signal manipulation	BER
Time shift (< 1/4 block length)	< 1%
MPEG-1/2 Layer 3 Codec @ 64kbit/sec/channel	12%
DA-AD conversion	6%
Echo addition (feedback 0.3; t=5..100ms)	< 7%
Additive white noise (-60dB)	7%
Dynamic range compression	< 2%
LP filter (cutoff frequency 3kHz)	< 2%
Changed playback speed (< 3%)	< 30%
Changed playback speed (< 3%) + pitch correction	< 8%

Table 1. Robustness test results.

From the results it can be concluded that the watermark scheme under test is robust with respect to most common types of manipulations/attacks. Due to the presence of a synchronization sequence even attacks like "time shift" and "time scaling" (changed playback speed plus pitch correction) are handled well as expected. Nevertheless, there is a noticeable sensitivity of the watermark if the playback speed is changed without applying any pitch correction. This can be explained easily by the frequency shift of the fixed grid checkerboard embedding pattern due to the speed change. In future versions of this technology, this could possibly be overcome by employing some signal adaptive frequency shift compensation on the detector side.

6. CONCLUSION

Audio watermarking in the subband modulation spectral domain is a promising new approach to robustly hide data in audio signals. It proved to be resistant to most common signal manipulations (attacks) usually involved in music transmission or storage. At a watermark data rate of 4.5 bits/sec, listening test results indicated excellent perceptual audio quality.

7. REFERENCES

[1] S. Greenberg, and B. E. Kingsbury, "The modulation spectrogram: in pursuit of an invariant representation of speech", *ICASSP*, 1997, pp. 1647--1650

[2] T. Houtgast, "Frequency selectivity in amplitude modulation detection", *Journal of the Acoustical Society of America 85*, vol. 4, 1989, pp. 1676--1680

[3] E. Zwicker, "Psychoakustik", *Springer*, Berlin/Heidelberg, 1982

[4] M. S. Vinton and L. E. Atlas, "A scalable and progressive audio codec", *ICASSP*, 2001, pp. 3277--3280

[5] C. Avendano and M. Goodwin, "Enhancement of audio signals based on modulation spectrum processing", *Proc. 117th AES Convention*, San Francisco, 2004

[6] S. Sukittanon and L. E. Atlas, "Modulation frequency features for audio fingerprinting", *ICASSP*, vol. 2, 2002, pp. 1173--1176

[7] L. E. Atlas and C. Janssen, "Coherent modulation spectral filtering for single-channel music source separation", *ICASSP*, 2005, pp. 461--464

[8] D. Gruhl, L. Anthony and W. Bender, "Echo hiding", *Information Hiding: First International Workshop*, Cambridge, UK1174, 1996, pp. 293--315

[9] L. Boney et al., "Digital watermarks for audio signals", *International Conference on Multimedia Computing and Systems*, 1996, pp. 473--480

[10] W. Bender et al., "Techniques for data hiding", *IBM Systems Journal 35*, 1996, no. 3,4

[11] P. Prandoni and M. Vetterli, "Perceptually hidden data transmission over audio signals", *ICASSP*, vol. 6, 1998, pp. 3665--3668

[12] AES CD-ROM, "Perceptual Audio Coders: What to Listen For", *Demonstration CD-ROM on Audio Coding Artifacts*, AES Publications, 2002

[13] C. Neubauer, "Übertragung digitaler Wasserzeichen in unkodierten und kodierten Audiosignalen", *PhD Thesis*, Technical Fac. of the University Erlangen-Nuremberg, 2001

[14] ITU-R, "Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems", *Recommendation BS.1116*, 1993

[15] http://sourceforge.net/projects/audacity/