Audio Compression: Discussion of an Alternative Approach

João Manuel Rodrigues, Ana Maria Tomé

***Abstract* – In this paper we present and dicuss a new algorithm for the compression of audio signals in which we are currently working. Backward-adaptive quantization, a fundamental innovation that differentiates this from other perceptual audio coders, is justified for its multiple advantages. We conclude with a discussion of some issues related to the design of the various components of the coder we are developing.**

***Resumo* – Neste artigo expomos e discutimos um novo algoritmo de compressão de sinais áudio em que estamos a trabalhar. A utilização de quantização retro-adaptada, uma inovação fundamental que diferencia este de outros codificadores perceptuais de áudio, é justificada pelas suas múltiplas vantagens. Concluímos com uma discussão de algumas questões relacionadas com o projecto dos vários componentes do codificador que estamos a desenvolver.**

I. Introduction

Since the introduction of the Compact Disc, digital coding of audio signals has become a common and popular technology. The simple 16-bit linear PCM format used in CDs, however, is now regarded as a very inefficient representation of audio content with its bit rate of 706 kbit/s per channel. As a consequence, new coding algorithms have been developed that can "compress" the audio information into a fraction of the bit rate with little or no degradation in perceived quality. This new generation of coding systems owes its great efficiency to the use of *perceptual coding* principles, i.e.: coding the signal in such a way that the injected noise is rendered inaudible by exploiting the limitations of the auditory system [1]. Since relevant psychoacoustic phenomena are highly dependent on the spectral content of the signal, it is not surprising that most high-quality digital audio coders such as the standard MPEG-Audio Layers I, II, III [2], and AAC [3], are based on sub-band or transform coding techniques. These coders share a common generic structure: a multirate filter bank or a lapped transform splits the input signal into subsampled frequency bands; a psychoacoustic model dynamically estimates the amount of noise that can be added to each band while still being masked by the signal itself; this *masking threshold* as well as bit rate constraints are then used to compute new step sizes and bit allocation for the sub-band quantizers; finally, entropy coding and multiplexing of the sub-band samples, step sizes and allocation information generates the output bit stream. At the other end of the communication channel the bit stream is parsed and demultiplexed, and the quantized sub-band samples are recovered. The sub-band sequences are then combined by an inverse transform or filter bank to produce the output signal.

In this paper, we present a perceptual audio coder with an alternative structure, shown in figure 1. It is based on the same frequency-domain coding principle, but differs from others in essentially one respect: the adaptation of the quantizers is derived, under perceptual considerations, not from the original signal but from previously quantized samples. That is: the system is backward-adaptive. Since this is not a common approach, we devote the next section to address the advantages and summarize some results that support the use of backward adaptation. We then discuss proposals for the implementation of the key components of the system: the filter bank, the perceptual adaptation algorithm, the quantization and entropy coding.

II. Backward Adaptation in Perceptual Coding

All common perceptual audio coders use a forward-adaptive quantization scheme. Access to the uncorrupted input signal can, in theory, lead to a more accurate adaptation. However, the need of embedding adaptation parameters (either bit allocation and/or quantizer scale factors) in the transmitted information will, in practice, compromise this advantage. The problem is that in order to reduce the amount of this side information, it must be quantized and decimated, thereby reducing its original accuracy.

In the proposed system, on the contrary, the quantizer adaptation parameters ($\mathbf{\Delta}$ in figure 1) are derived through perceptual and bit rate considerations from the previously quantized signal. The obvious advantage of this backward adaptation scheme is that no side information must be transmitted since the decoder replicates the encoder procedure to reproduce the adaptation parameters. Freed from the constraints imposed by limited channel capacity, there is no need to reduce the time, frequency, or magnitude resolution of the adaptation information, so it can evolve smoothly sample-by-sample. Algorithm design and implementation are much simplified because there is no side information to quantize, encode and multiplex. This should be a significant advantage since the optimal selection of adaptation parameters in an advanced forward-adaptive coder is not a trivial matter [4].

There are, of course, disadvantages in pure backward-adaptive systems. First of all, computational requirements of the decoder are increased by the inclusion of the adaptation algorithm. A related problem is that any future improvements in psychoacoustic modeling cannot simply be integrated into an encoder but must be included in every decoder too. In multimedia applications, programmable devices are the norm and downloading program updates is quite usual. Even in cheap, portable solid-state music players, there is a current trend towards programmable, multifunction devices [5], so this does not seem to be a major difficulty. Still, these problems can be mitigated, and some

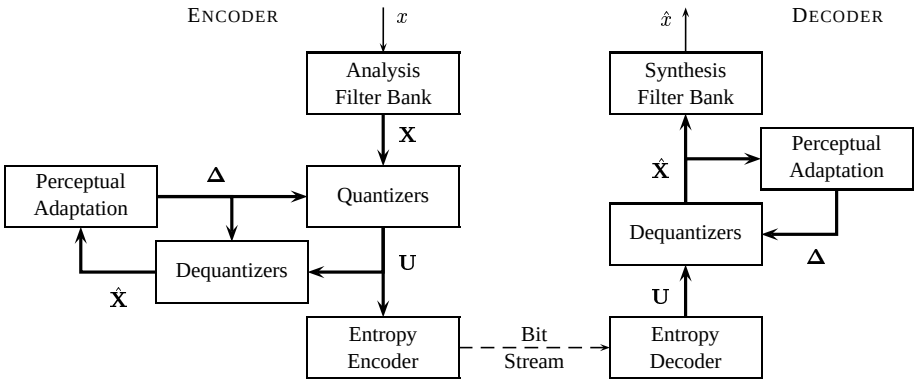

Figure 1 - A frequency-domain perceptual coder with backward-adaptive quantization.

versatility regained by transmitting small amounts of side information, in a hybrid backward- and forward-adaptive fashion. This has also been recognized and implemented by the designers of AC-3 [6], which can be considered a backward-adaptive system, although not quite to the same extent as ours.[1]

Another potential problem of a different, more fundamental nature, is the possibility that the noise introduced by the quantization may disturb the process leading to grossly inaccurate adaptation and poor performance. This could be a serious drawback because, at high compression ratios, large amounts of noise are introduced, and furthermore, this could be aggravated by the nonlinear psychoacoustic computations included in the adaptation loop. In fact, results from previous work with a backward-adaptive coder show that the effect of quantization noise on the backward estimation of psychoacoustic parameters is very small even when very coarse quantization is employed [7]. Also, the performance of that coder was favorably compared to an idealized forward-adaptive version [8].

III. Design Issues

In this section, we discuss some coder design issues, taking into account that the option for a backward-adaptive scheme conditions every component of the system.

A. The Filter Bank

The filter bank is used to decompose the input into several subsampled band-limited signals. There are fundamentally two reasons for doing this. Firstly, it is well known that audio signals are highly correlated and that sub-band coding is an effective means to exploit that redundancy [9]. The other reason is that it provides a direct way to control the spectral shaping of the introduced noise, which is convenient to take advantage of the masking properties of the ear.

The filter bank is crucial to the overall performance of the coder, and it should satisfy several requirements such as critical subsampling, perfect reconstruction, and temporal and spectral resolution compatible with auditory masking constraints [10]. A modulated lapped transform (MLT) [11], which is a particular type of multirate filter bank, is a particularly attractive option because it meets the first two requirements with relatively low delay, and allows fast implementations. It is not surprising that the MLT, also known as the modified discrete cosine transform (MDCT) or time-domain aliasing cancellation (TDAC) filter bank, is one of the most popular transforms in audio coding. The MLT, however, decomposes the signal into equally subsampled, equal bandwidth channels, while the time and frequency resolution of the ear is known to vary widely: at low frequencies, the critical bands (a measure of the frequency selectivity of the ear) are narrow and temporal masking lasts longer; at high frequencies the opposite occurs. Nonuniform decompositions, generally based on tree structures, may be used to approach these multiresolution properties, but can give rise to higher complexity and delay, as well as difficulties in getting proper frequency responses. Another, more common, approach is window-switching, that is: commuting between high and low dimension transforms to trade between temporal and spectral resolution when needed. This is simpler but still it requires a transient detection algorithm and additional constraints on the design of the transform windows.

We believe that in our samplewise adaptive coder the resolution constraints are not so stringent, and a simple fixed-dimension MLT may very well be effective. Assuming 44100 samples/s signals, the frequency resolution of a 256-band MLT, for instance, will be 86 Hz, which is narrower than the lower critical bands, and, at 5.8 ms, the time resolution is certainly good enough to avoid violating post-masking thresholds even in high frequency bands. This contrasts with forward-adaptive systems such as MPEG Layer I where, despite much coarser frequency resolution with only 32 bands, quantizer adaptation only occurs once every 8.7 ms.

B. Quantization and Coding

In backward-adaptive coders, it is very important that quantizers support a large dynamic range because there is no advance information about sudden attacks in the signal. On the other hand, it is psychoacoustically acceptable to introduce larger absolute errors in larger amplitude signals. Therefore, nonuniform quantization with a near-logarithmic companding rule seems quite appropriate. It is also essential to use mid-tread quantizers because their signal-to-noise ratio is never below 0 dB, but also because

[1] In fact, the bit allocation algorithm in AC-3 uses only a fraction of the transmitted information, resulting in coarser adaptation.

a mid-rise quantizer could eventually make the adaptation loop turn unstable. Two degrees of freedom can easily be controlled in a logarithmic quantizer, which determine the maximum absolute error introduced in small signals and the maximum relative error in large signals. The adaptation algorithm can manipulate either or both of these parameters for each quantizer.

The quantized outputs are coded using an arithmetic code [12]. Arithmetic coding is the most efficient form of entropy coding known, not being restricted to encode each symbol with an integral number of bits like Huffman codes. Furthermore, it promotes a clear separation between coding and statistical source modeling, which allows an easy integration of reasonably complex, highly dynamic, context-adaptive source models. Some preliminary measurements confirm that the statistical distributions effectively vary both in time and from band to band, so adaptive models are advisable.

C. Perceptual Adaptation Algorithm

Auditory models found in the perceptual audio coding literature are traditionally based on the concept of masking threshold, i.e., they dynamically estimate the amount of noise that may be added to a signal without causing audible distortion. In spite of its appeal, it is recognized that this concept, or at least its simplistic interpretation, suffers from several problems [13]. Another approach to auditory modeling is to quantify the ability of the ear to perceive differences between two signals—e.g. the input and output of a coding system—by comparing some internal representation of these signals. This strategy has been fruitfully applied in audio quality measurement but not in audio coding. The reason for this is certainly the higher complexity that it involves. However, for any given coder, there is some a priori knowledge of the kind of distortion that will be introduced. Therefore, an internal representation model for audio coding may not need the full generality of those used in quality measures. We are currently exploring the applicability of this approach to audio coding. The studied model computes the excitation pattern from the spectral representation obtained by the coder transform, and it is relatively simple but plausible since it is based on a recent audio quality measure [14]. An approximate formulation has been derived that permits the evaluation of the effect of quantization noise on the excitation patterns predicted by this model. This derivation and some preliminary validation results will be presented in [15].

IV. Conclusions

Several high-quality audio coding algorithms are in widespread use nowadays, but all follow a similar strategy. This paper proposes an alternative approach to audio compression in which backward adaptation plays a significant role. We discussed the implications of this option and laid out research directions for the development of the system. Previous results with a preliminary version show the viability of the approach, and we expect to achieve very good performance with a low complexity.

References

[1] Peter Noll, "Wideband speech and audio coding", *IEEE Communications Magazine*, pp. 34–44, Nov. 1993.

[2] Karlheinz Brandenburg, Gerhard Stoll, et al., "The ISO/MPEG-Audio codec: A generic standard for coding of high quality digital audio", in *92nd AES-Convention*, Vienna, Mar. 1992, Audio Engineering Society, Preprint 3336.

[3] Marina Bosi, Karlheinz Brandenburg, Schuyler Quackenbush, Louis Fielder, Kenzo Akagiri, Hendrik Fuchs, Martin Dietz, Jürgen Herre, Grant Davidson, and Yoshiaki Oikawa, "ISO/IEC MPEG-2 advanced audio coding", *Journal of the Audio Engineering Society*, vol. 45, no. 10, pp. 789–813, Oct. 1997.

[4] Ashish Aggarwal, Shankar L. Regunathan, and Kenneth Rose, "Near-optimal selection of encoding parameters for audio coding", in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, May 2001.

[5] Jason Kridner, Mark Nadeski, and Pedro Gelabert, "A DSP powered solid state audio system", in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Mar. 1999, pp. 2283–2286.

[6] Craig C. Todd, Grant A. Davidson, Mark F. Davis, Louis D. Fielder, Brian D. Link, and Steve Vernon, "AC-3: Flexible perceptual coding for audio transmission and storage", in *96th AES-Convention*. Audio Engineering Society, Feb. 1994, Preprint 3796, (also in http://www.dolby.com/tech/ac3flex.html).

[7] João Manuel Rodrigues and Ana Maria Tomé, "A backward-adaptive perceptual audio coder", in *EUSIPCO*, Trieste, Italy, Sept. 1996, vol. II, pp. 1007–1010.

[8] João Manuel Rodrigues and Ana Maria Tomé, "On the use of backward adaptation in a perceptual audio coder", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, pp. 488–490, July 2000.

[9] Nuggehally S. Jayant and Peter Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, N.J., 1984.

[10] James D. Johnston and Karlheinz Brandenburg, "Wideband coding—perceptual considerations for speech and music", in *Advances in Speech Signal Processing*, Sadaoki Furui and M. Mohan Sondhi, Eds., chapter 4. Marcel Dekker, Inc., New York, 1991.

[11] Henrique S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Norwood, MA, 1992.

[12] Ian H. Witten, Radford M. Neal, and John G. Cleary, "Arithmetic coding for data compression", *Communications of the Association for Computing Machinery*, vol. 30, no. 6, pp. 520–540, June 1987.

[13] Raymond N. J. Veldhuis, "Bit rates in audio source coding", *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 1, pp. 86–96, Jan. 1992.

[14] Thilo Thiede, William C. Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G. Beerends, Catherine Colomes, Michael Keyhl, Gerhard Stoll, Karlheinz Brandenburg, and Bernhard Feiten, "PEAQ—the ITU standard for objective measurement of perceived audio quality", *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, Jan. 2000.

[15] João Manuel Rodrigues, Ana Maria Tomé, and Tomás Oliveira e Silva, "Auditory models in audio coding", in *111th AES-Convention*, New York, Sept. 2001, Audio Engineering Society, (To be presented.).