IMPROVED LOW BIT-RATE AUDIO COMPRESSION USING REDUCED RANK ICA INSTEAD OF PSYCHOACOUSTIC MODELING

Adiel Ben-Shalom, Michael Werman

School of Computer Science
Hebrew University
Jerusalem, Israel.
{chopin,werman}@cs.huji.ac.il

Shlomo Dubnov

Department of Communication Engineering
Ben-Gurion University
Be'er-Sheva, Israel
dubnov@bgumail.bgu.ac.il

ABSTRACT

Traditional audio coding is based on a perceptual compression paradigm that exploits psychoacoustic information to efficiently encode audio signals. Recently, extensive research has been conducted in order to understand how the brain encodes natural signals. These results suggest that the encoding process is very efficient in terms of redundancy reduction of the signal information. It could be that the psychoacoustic effects (such as the masking effect) are only a special case of a more general redundancy reduction mechanism that exists in the auditory pathway. Motivated by this work we propose a new audio coding scheme that is based on improved sound representation found by Independent Component Analysis. Using a local linear, low rank, non-orthogonal transform, we remove additional redundancies in the signal. At low bitrates this coding scheme gives results superior to a legacy perceptual encoding scheme for different kinds of audio signals.

1. INTRODUCTION

Perceptual audio coders exploit psychoacoustical knowledge about the human auditory system to efficiently encode audio signals. These coders exploit a phenomenon known as the 'masking effect', which was discovered in psychoacoustics experiments. Extensive research has been conducted over the last years which aims to understand how the auditory sensors encode the information in our brain. Recent results show that the signals are efficiently encoded by the auditory sensors in terms of redundancy reduction along the auditory pathway. Several models have been proposed to describe the behavior of this efficient coding process [1, 2].

In this work we use the redundancy reduction idea in order to design a new architecture for a low bit-rate audio coder. We simulate the audio encoding process in a manner which we assume is done along the auditory pathway. Redundancy reduction is done using ICA, which is a recently developed statistical tool for data analysis. We optimally decompose the signal into a reduced rank representation whose basis vectors are "as independent as possible". This achieves several advantages for the compression task: 1). The reduced rank representation is very sparse and allows an adaptive transmission of the transform coefficients without increasing the overall bitrate. 2). The bit allocation is performed on approximately independent channels, a situation which is required by rate-distortion theory 3). No psychoacoustic model is employed since the ICA vectors do not correspond to the masking properties of the human ear. Nevertheless, the superior performance of our method suggests an interesting idea that the the classical psychoacoustical masking of pure tones could be a special case of a more general redundancy reduction mechanism of the auditory pathway [3].

The other components of our encoder are a standard filter bank used for time to frequency mapping, bit allocation and a uniform quantizer. We compared our encoder to a basic perceptual encoder and it shows superior results in objective listening tests.

2. PRELIMINARIES

2.1. Perceptual Coding

Perceptual coding algorithms belong to the class of lossy compression algorithms. The performance of a lossy algorithm is often measured by the reconstruction error. We would like the reconstruction error to be minimal so that the reconstructed data is as similar to the source as possible. This situation is not true for perceptual coders. As in other lossy coders, the goal of perceptual coders is that the reconstructed data will be similar to the source. However, the similarity measure is defined by the human ear, thus, the coder must exploit psychoacoustic knowledge about human hearing to make the reconstruction error inaudible.

An important aspect of the human hearing is the masking effect. The masking effect [4] states that the threshold of hearing of the different frequencies arises in the presence of a masking tone or noise. Masking curves depicts the threshold of hearing neighboring frequencies in the presence of the tone or noise masker. The masking effect is used by perceptual audio coders to make the reconstruction error inaudible.

Figure 1 depicts the structure of a basic perceptual coder. The signal samples are first processed using a time to frequency mapping. The output of the filters are called subband samples or subband coefficients. The subband coefficients are then used to calculate the masking thresholds for each band. The bit allocation algorithm assign bits to the different bands so that the noise, which is introduced by the quantization process will be below the masking threshold, thus inaudible by the listener.

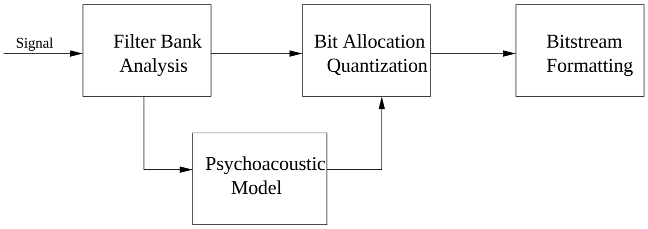

Fig. 1. Basic perceptual audio coder architecture.

2.2. Independent Component Analysis

Independent Component Analysis (ICA) is a recently developed statistical tool for extracting statistically independent components from a random vector [5]. ICA can be used to solve the classical 'cocktail party' problem in which n sensors record a mixture of n people speaking simultaneously. ICA is used to recover the original speaker signals from n mixtures. Other applications of ICA are audio analysis, natural images analysis, financial data, medical data and other inverse mixing problems [6].

In addition to unmixing problems ICA has been shown to be a useful tool for feature extraction and data representation. Formally, let $\mathbf{x} = (x_1, x_2 \dots x_n)$ be the observed data vector. ICA's goal is to find the matrix A such that:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{s} = (s_1, s_2 \dots s_n)$ are statistically independent components. The columns of the matrix $\mathbf{A}$ can be thought of as basis vectors and the vector $\mathbf{s}$ is the representation of $\mathbf{x}$ in this basis. ICA analysis for feature extraction and data representation was studied in [2, 1]. For natural audio signals it was shown that ICA analysis results in a local vector basis which resembles short waveforms in the original signal [2].

The ICA problem can be formalized as maximum likelihood estimation problem. We wish to find a matrix A and set of sources s which best explains the empirical variables x. From Information theory we know that

$$< \log(p(x)) >_q \;\; \propto \;\; -KL(q \parallel p) \tag{2}$$

where q is the empirical distribution of the sources, p is the hypothesized distribution of the sources and KL stands for the Kullback-Leibler divergence. One can show that

$$KL(q \parallel p) = KL(q \parallel \prod q_i) + KL(\prod q_i \parallel p) \tag{3}$$

$\prod q_i$ are the marginal product of the empirical distribution. The second term is minimized when we choose $p = \prod q_i$. This reduces the problem to minimize $KL(q \parallel \prod q_i)$. The KL distance between a distribution vector and its marginal probabilities is called the *Mutual Information*. Eventually, we wish to find a matrix which will make the empirical sources as independent as possible.

Several algorithms have been proposed to solve the ICA problem. A comprehensive overview of the algorithms can be found in [6]. In our experiments we used the Jade [7] algorithm.

3. ENCODING ALGORITHM

Our audio compression algorithm is comprised of several building blocks (Figure 2). We use subband decomposition to perform an initial time to frequency mapping. The subband coefficients are then grouped to blocks and ICA analysis is computed on each block. The output of the ICA analysis is both reduced rank ICA coefficients and ICA mixing/demixing matrix. The ICA coefficients are then quantized and packed in frames. The ICA transform matrix is quantized and sent as side-information for each block.

3.1. Subband Decomposition

For subband decomposition we adopt the polyphase filter bank used in the MPEG coding standard [8, 9]. This filter bank is a pseudo-QMF, cosine modulated filter bank which splits the PCM input audio samples into 32 equally spaced bands. The filter bank gives good time resolution and reasonable frequency resolution [9].

We denote by $x[n]$ the input sample at time n and by $s_i[t]$ the output of the i'th filter bank band at time t. The filter bank is critically sampled, which means that for every 32 input samples the filter bank outputs 32 samples. Since the output of each band is sub-sampled by a factor of 32 then t is a multiple of 32 audio samples. The output of each filter can be written [10]:

$$s_i[t] = \sum_{n=0}^{511} x[t-n] H_i[n] \tag{4}$$

where

$$H_i[n] = h[n]cos\left[\frac{(2i+1)(n-16)\pi}{64}\right] \tag{5}$$

and $h[n]$ corresponds to analysis window coefficients.

3.2. Reduced Rank ICA Coding

The filter bank output coefficients are grouped into blocks for ICA processing. When choosing the block length we have to consider two factors. On one hand, we want a true realization of the redundancy reduction process in the auditory pathway, which constrains us to short blocks. On the other hand, the ICA matrix must be sent along with each block of data as side information so using short blocks gives us more overhead. We found that using blocks of approximately 1 second is a sufficient trade-off.

ICA analysis is comprised of two steps. The first step includes dimension reduction of the data, and the second step consists of ICA analysis on the reduced rank coefficients. We denote the filter bank coefficients block by $\mathbf{X}$. $\mathbf{X}$ is a $32 \times L$ matrix where $\frac{32 \times L}{SR} = 1\ second$. If we consider $\mathbf{X^T}$ we can view the columns as variables and the rows as time instants of these variables. Each row is a vector of dimension 32 which is a time instance of the filter bank output. These variables are highly correlated and we would like to represent them in a basis on which there will be no correlation between the variables.

The first step is to reduce the dimension of the data. We do it by reducing the dimension of the row space of $\mathbf{X^T}$ by using the singular value decomposition (SVD) method. $\mathbf{X^T}$ can be decomposed to :

$$\mathbf{X^T} = \mathbf{USV^T} \tag{6}$$

where U is an $m * m$ matrix and V is an $n * n$ matrix and S is a diagonal matrix which contains the singular values of $\mathbf{X^T}$. In our scheme, $m = L$ and $n = 32$. To reduce the dimension of the row space of $\mathbf{X^T}$ to a lower dimension r, we project $\mathbf{X^T}$ on the first r column vectors of $\mathbf{V}$

$$\mathbf{Y^T} = \mathbf{X^T V_r} \tag{7}$$

where $\mathbf{V_r}$ is a matrix which contains the first r column vectors from $\mathbf{V}$. $\mathbf{Y}$ now is an $r * m$ matrix in which the rows contain the representation of the filter bank coefficients in the reduced rank basis. The rows of $\mathbf{Y}$ are not statistically independent. To achieve independence we apply ICA analysis on the rows of $\mathbf{Y}$:

$$\mathbf{\hat{Y}} = \mathbf{W V_r^T X} \tag{8}$$

$\mathbf{W}$ is the unmixing matrix obtained by ICA. $\mathbf{\hat{Y}}$ is the reduced rank independent component representation of the subband coefficients. The matrix $\mathbf{B} = (\mathbf{V_r^T W})^\sharp$ is encoded as side information for each block and used by the decoder to decode the samples by

$$\mathbf{X_{rec}^T} = \mathbf{\hat{Y}_q^T B} = \mathbf{\hat{Y}_q^T}(\mathbf{V_r^T W})^\sharp \tag{9}$$

The sign $\sharp$ stands for the pseudo-inverse matrix.

3.3. Bit Allocation and Quantization

Rate distortion theory shows that a signal can be compressed, for a given distortion D, in a rate that is lower-bounded by the minimal mutual information between the original signal and the quantized signal. In order to obtain an optimal quantizer Q, knowledge of the complete multi-variate probability distribution of the source vector is necessary. This requires exponentially large codebooks. Due to practical considerations, the quantization is performed componentwise, a situation which is optimal only if the variables are mutually independent. In case of Gaussian variables, statistical independence is achieved by PCA. In case of non-Gaussian signal statistics, this is approximately achieved using ICA.

The output of the ICA analysis step is a set of r statistically independent bands. Our hypothesis is that in our representation the different bands closely resemble the coding information sent by the auditory sensors to code audio signals. Thus, we do not introduce any other perceptual measure in the bit allocation process as was done in the legacy audio coder. The quantization of the different bands here should be optimal in term of minimum reconstruction error of the coefficients.

If we denote by R_{avg} the average number of bits used to encode samples in the block, R_k the average bit rate used to encode samples in the k'th band and by σ_k the variance of the coefficients on the k'th band. Then the optimal bit allocation for the different bands is given by [11]:

$$R_k = R_{avg} + \frac{1}{2}log_2 \frac{\sigma_k^2}{\prod_{k=1}^{r} (\sigma_k^2)^{\frac{1}{r}}} \tag{10}$$

The bit allocation according to equation 10 is optimal in terms of the reconstruction error. The problem is that R_k might be negative or not an integer number. To solve this problem we use an iterative algorithm for bit allocation with positive integer constraint similar to the one described in [11].

Using the bit allocation information we quantize the ICA coefficients with a uniform quantizer. We assign 8 bits to quantize the ICA mixing matrix samples. In our experiments the dimension r was chosen to 5. Thus, the ICA matrix size is 32×5 which results in overhead of 160 bits assigned for each block of data. We compensate this overhead with the dimension reduction of the filter bank coefficients. The scalefactors which are used by the decoder for re-quantization are quantized with 6 bits.

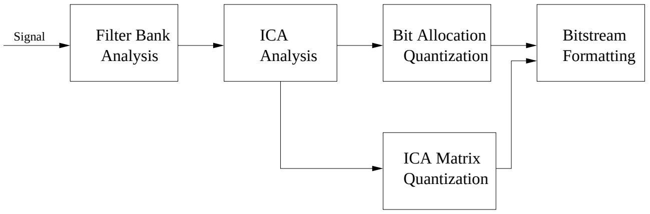

Fig. 2. Architecture of the proposed encoder.

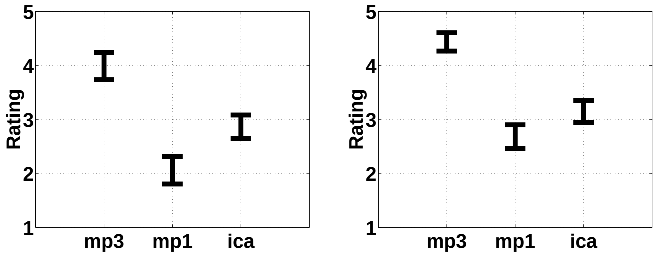

Fig. 3. Encoders mean ranking value with 95% confidence interval. The left figure corresponds to encoding in 32kbps and 44.1khz sampling rate. The right figure corresponds to encoding in 32kbps and 32khz sampling rate.

4. EXPERIMENTS RESULTS

We compared our algorithm with two perceptual audio coders. MPEG-1 layer 1 and MPEG-1 layer 3 (MP3) [8]. Layer 1 algorithm is simple yet uses perceptual measures such as the masking effect to encode audio signals efficiently. Layer 3 contains several enhancements such as improved hybrid filter bank, noise shaping procedure, and huffman coding. Since we compared our encoding algorithm to perceptual coding algorithms, the test was carried out using a psychophysical experiment.We performed two sets of tests. In both tests the encoder bit-rate was 32kbps. In the first test the sampling rate was 44.1Khz which results in 0.7256 bits per sample, and in the second test we used sampling rate of 32Khz which results in 1 bit per sample. The participants were asked to rate the encoder given a reference source with a 1 to 5 scale where 5 stands for imperceptible encoding and 1 stands for a very annoying encoding. We carefully selected the music test files to cover wide range of audio data. Figure 3 depicts the mean rating value for each of the encoders. Table 1 shows pairwise comparisons between the encoder ranking results. It can be seen that for both sampling rates the ICA coder was rated higher than Layer-1 and less than Layer-3. Moreover, as we go up with sampling ratio ICA encoder is significantly better than Layer-1. The test files, which were used in the experiment can be downloaded from http://www.cs.huji.ac.il/∼chopin/ica-encoder/index.html

	32kbps, 44.1khz SR			32kbps, 32khz SR		
	mp3	mp1	ica	mp3	mp1	ica
mp3	0	47	43	0	50	48
mp1	1	0	9	0	0	11
ica	5	35	0	2	37	0

Table 1. Pairwise comparison between encoders ratings. Entry (i, j) in the table correspond to the number of times that encoder i got better ranking than encoder j.

5. CONCLUSION

In this paper we have shown new architecture for a low bit-rate audio coder motivated by new results from auditory research. Our results show that representing audio data as independent components can reduce the audible noise in audio compression. The superior results of MP3 over our algorithm can be argued to be because of the advanced coding algorithms used in MP3. MP3 adds very efficient noise shaping algorithm, which together with huffman coding gives superior results. We have implemented the same coding blocks as in Layer-1. Thus, comparison with Layer 1 is more appropriate. The ICA encoder had superior results than Layer-1 for different music files. This leads us to the conclusion that using ICA might be equivalent or better than psychoacoustic modeling.

6. REFERENCES

[1] A. J. Bell and T. J. Sejnowsky, “The ’independent components’ of natural scentes are filters,” *Vision Research*, , no. 37, pp. 3327–3338, 1997.

[2] A. J. Bell and T. J. Sejnowski, “Learning the higher-order structure of a natural sound,” *Network: Computation in Neural Systems*, , no. 7, pp. 261–266, 1996.

[3] O. Schwartz and E. P. Simoncelli, “Natural sound statistics and divisive normalization in the auditory system,” in *NIPS*, 2000, pp. 166–172.

[4] T. Painter and A. Spanias, “A review of algorithms for perceptual coding of digital audio signals,” *DSP-97*, 1977.

[5] A. J. Bell and T. J. Sejnowsky, “An information maximisation approach to blind separation and blind deconvolution,” *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[6] A. Hyvarinen, “Survey on independent component analysis,” *Neural Computing Surveys*, , no. 2, pp. 94–128, 1999.

[7] J. F. Cardoso, “Jade for real-valued data,” http://sig.enst.fr/ cardoso/guidesepsou.html.

[8] “Information technology-coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbits/s-part-3:audio,” ISO/IEC Int’l standard IS 11172-3.

[9] K. Brandenburg, “”the iso/mpeg-audio codec: A generic standard for coding of high quality digital audio,”,” in *92nd AES Convention, preprint 3336,*, Audio Engineering Society, New York,, 1992.

[10] D. Pan, “A tutorial on MPEG/audio compression,” *IEEE MultiMedia*, vol. 2, no. 2, pp. 60–74, 1995.

[11] K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann Publishers, 1996.