4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

STUDY OF MUTUAL INFORMATION IN PERCEPTUAL CODING WITH APPLICATION FOR LOW BIT-RATE COMPRESSION

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

In this paper we analyze this aspect of redundancy reduction as it appears in MPEG1-Layer 1 codec. Specifically, we consider the mutual information that exists between filter bank coefficients and show that the normalization operation indeed reduces the amount of dependency between the various channels. Next, the effect of masking normalization is considered in terms of its compression performance and is compared to linear, reduced rank short term ICA analysis. Specifically we show that a local linear ICA representation outperforms the traditional compression at low bit rates. A comparison of MPEG1 Layer 1 and a new ICA based audio compression method are reported in the paper. Certain aspects of ICA quantization and its applicability for low-bitrate compression are discussed.

1. INTRODUCTION

Perceptual coding algorithms belong to the class of lossy compression algorithms. The performance of a lossy compression for a given bitrate is often measured by the reconstruction error, which should be minimal so that the reconstructed data is as similar to the source as possible. In the case of perceptual coders the similarity measure is defined by the human ear, and accordingly the coder must exploit psychoacoustic knowledge about human hearing to make the reconstruction error inaudible.

Perceptual audio 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 investigate the redundancy reduction idea as it occurs in traditional psychoacoustic coder. Specifically, we measure the amount of *Mutual Information* (MI) between the different filter bank subbands on a natural sound. We show that the psychoacoustic model indeed helps remove some redundancies between the coefficients. Next, we use the idea of minimization of MI in order to design a new architecture for a low bit-rate audio coder. One of the open research questions regarding ICA is whether it can be used for data compression. It is well known that PCA is optimal for data reduction in terms of reconstruction error. Further more if the source is gaussian then optimal bit allocation can be achieved by using PCA. However, for non-gaussian variables where PCA is not optimal, ICA might give better resultsSpecifically, we consider the question of low bit-rate quantization using a combined PCA and low rank ICA representation.

In our method of low bit-rate quantization we use a combined PCA and low rank ICA representation. This achieves several advantages for the compression task: 1). Coarse first step quantization by rank reduction gives a lower overall error compared to low bit-rate quantization of the complete set of coefficients. 2) The reduced rank representation is very sparse and allows an adaptive transmission of the transform coefficients without increasing the overall bitrate. 3). The bit allocation is performed on approximately independent channels, a situation which is required by rate-distortion theory. It should be noted that no psychoacoustic model is employed since the ICA vectors do not correspond to the masking properties of the human ear. The superior performance of our method suggests 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].

2. PERCEPTUAL CODING

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.

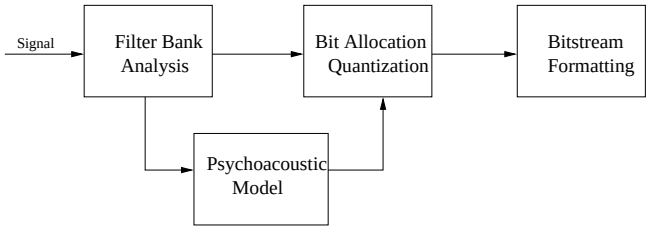

Fig. 1. Basic perceptual audio coder architecture.

MPEG-1 Layer-1 audio coder is an example for a simple perceptual coder which implements such scheme. The signal samples are first transformed to the time-frequency domain using polyphase filter bank. The filter bank is a set of 32 equally spaced band pass filters and the output of the filter bank are 32 subbands $\{S_i\}_{i=1}^{32}$ where S_i corresponds to the original signal filtered with band pass filter H_i.

Psychoacoustic model then measures the masking threshold $\{M_i\}_{i=1}^{32}$ introduced in each subband. The masking threshold is calculated by considering tonal and non tonal components which appear both in the particular subband and in adjacent subbands. The output of the psychoacoustic model is a measure of the signal to mask ratio (SMR) measured in dB in each subband. The SMR will be high in subbands where the masking threshold is small relative to the signal power in that subband, and small for subbands with relative high masking threshold.

The bit allocation algorithm uses the SMR to allocate the available bits to the subbands such that the quantization noise introduced by the quantizer will be inaudible. This is done by minimizing the noise to mask ratio (NMR) which is defined as SMR - SNR where SNR is the signal to noise ratio. If a subband has high SMR then we would like to assign more bits to the subband in order to get high SNR and to make the noise introduced by the quantization below the masking threshold in the band and thus inaudible.

It can be shown [5] that the optimal bit allocation which minimize the noise to mask ratio is given by:

$$R_k = R + \frac{1}{2}log_2(\frac{S_k}{M_k})^2 - \frac{1}{2}log_2\left(\left(\prod_{k=1}^{N}\left(\frac{S_k}{M_k}\right)^2\right)^{\frac{1}{N}}\right) \quad (1)$$

Where R is the total number of bits available for all subbands, N is the number of subbands and M_k is the masking threshold in each subband. From this equation we can observe that the masking threshold in each subband is a normalization factor for the subband coefficients. If we denote $\hat{S}_i = \frac{S_i}{M_i}$ then the bit allocation becomes:

$$R_k = R + \frac{1}{2}log_2\frac{\hat{S_k}^2}{(\prod_{k=1}^{N}\hat{S_k}^2)^{\frac{1}{N}}} \quad (2)$$

Which is the optimal bit allocation without masking.

Normalizing the filter bank coefficients with the masking thresholds reduced the quantization noise to be inaudible. However, in addition this process also reduces information redundancy within the coefficients. It is quite clear that the different subbands are not statistically independent. Almost every sound will exhibit some correlation between the output of the band pass filters. For example for human voiced speech, the subband which contains the pitch would have high energy response and subbands which contains the pitch partials would also have response. These subbands would clearly have statistical dependency between their outputs. To measure the amount of redundancy within coefficients we measured the *Mutual Information* between the different filter bank subbands on a natural sound (sound of a cat). The mutual information was calculated twice. First, the mutual information between the filter bank coefficients S_i was calculated and then we calculated the mutual information between the normalized coefficients $\hat{S}_i$. Figure 2 shows both results. The top image shows the mutual information between the filter bank subbands coefficients. It can be seen that there exists some redundancy between the low subbands (white boxes in the image). The bottom image shows the mutual information of the coefficients normalized with the masking threshold. The information redundancy between the low subbands was almost completely removed. We expect that ICA might be a better tool in removing this information redundancy between the different subbands.

3. INDEPENDENT COMPONENT ANALYSIS

In this work Independent Component Analysis (ICA) is applied for extracting an efficient signal representation in terms of statistically independent components [6]. 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:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (3)$$

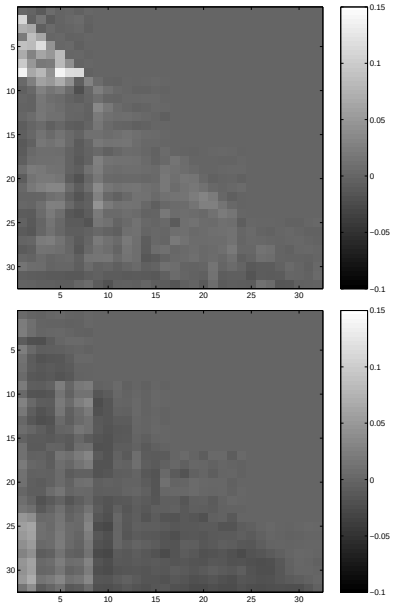

Fig. 2. Mutual Information between different subbands. The top image shows the mutual information between the filter bank different subbands while the bottom image shows the mutual information of the subbands after the masking normalization process. In both images if the subbands i and j have no mutual information so $I(i,j) = 0$ and it is showed as black box.

where $\mathbf{s} = (s_1, s_2 \ldots s_n)$ 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

$$< \log(p(x)) >_q \quad \propto \quad - KL(q \parallel p) \tag{4}$$

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

$$KL(q \parallel p) = KL(q \parallel \prod q_i) + KL(\prod q_i \parallel p) \tag{5}$$

$\prod q_i$ 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.

In order to consider the use of ICA for data compression, we considered the following example (Figure 3). The non-Gaussian distribution is quantized at low bit-rate using ICA and PCA analysis. One can see that for the ICA case, all the quantization points fall on the axes, thus giving a smaller reconstruction error. In the PCA case, the wrong orientation of the principal axes results in quantization points that are far from the true data points. It is interesting to note that in the high bit-rate case, the PCA quantization might outperform ICA in the Mean Square Error (MSE) sense even for non-Gaussain distribution. When many quantization points are available (the PCA grid is very dense), PCA tends to locate the quantization points more optimally in terms of MSE compared to ICA.

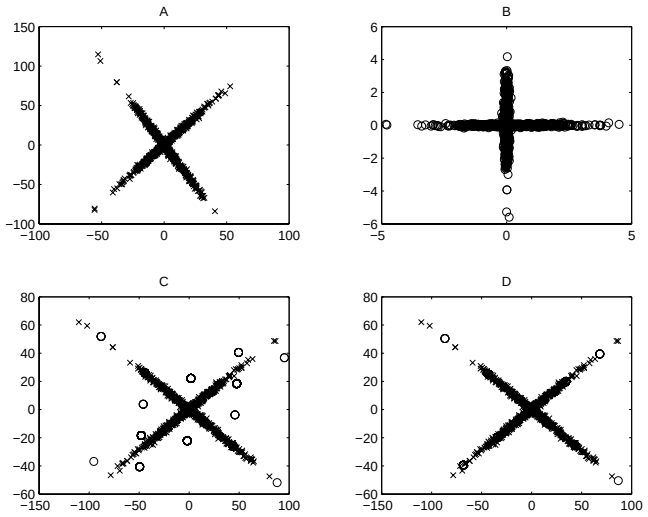

Fig. 3. Quantization of non-gaussian distribution with PCA and ICA. A. Recovered sources as was found by PCA. B. Recovered sources found by ICA which is the true original distribution. C. Quantizing PCA coefficients results in points far from the true data points. D. ICA reconstruction of the data results in correct reconstruction of the data thus giving small reconstruction error.

In order to see the utility of ICA for compression, one must observe that rate-distortion theory requires to minimize MI between the signal and a quantized version of the signal for a given distortion constraint. This is different from ICA that tries to minimize MI between linearly transformed signal components without considering quantization. Given an ICA decomposition $x = As$ and a quantizer $\hat{s} = Q(s)$, one can show that the log-likelihood of the data x given $\hat{s}$ is given by $\mathcal{L}(p(x|\hat{s})) = \sum I(s_i, \hat{s}) - I(s)$, where s_i are the components of the vector s. This is in contrast to the original maximum likelihood ICA formulation, where in addition to minimization of the MI of s, one must maximize MI between the individual components and the quantized variable.

Using this formulation, we approximately solve the ICA compression problem by sub-optimal three step approach. We first do a large reduction step by implicitly assuming a Gaussian distribution of the original filter-bank coefficients. Maximizing the capacity between the complete and a re-

duced rank set of coefficient is done by PCA. Next we minimize $I(s)$ of the reduced rank coefficient set using traditional ICA methods, and finally we maximize the quantization-related MI by applying a standard bit-allocation procedure.

4. LOW BIT-RATE ENCODING ALGORITHM USING ICA

Our audio compression algorithm is comprised of several building blocks (Figure 5). 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.

4.1. 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]:

$$s_i[t] = \sum_{n=0}^{511} x[t-n]H_i[n] \tag{6}$$

where

$$H_i[n] = h[n]cos\left[\frac{(2i+1)(n-16)\pi}{64}\right] \tag{7}$$

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

4.2. 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 :

$$\mathbf{X^T} = \mathbf{USV^T} \tag{8}$$

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 $\mathbf{V}$

$$\mathbf{Y^T} = \mathbf{X^T V_r} \tag{9}$$

where $\mathbf{V_r}$ is a matrix which contains the first r column vectors from $\mathbf{V}$.

The reduced dimension r is chosen adaptively in each block. This is done by inspecting the singular values $\lambda_1 \ldots \lambda_n$ (diagonal of S) and choosing the first r basis vectors in V such that the corresponding r eigenvalues $\lambda_1 \ldots \lambda_r$ satisfies:

$$\frac{\sum_{i=r+1}^{n} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \leq E \tag{10}$$

where E is the error introduced by the dimension reduction procedure. Figure 4 shows the singular values calculated for 1 second of Pop music. It can be seen that the singular values decay rapidly to zero. The value of E can be chosen during the encoding process to adjust the reconstruction error. In our experiments we chose $E = 0.2$. We emphasis that the dimension reduction process is not inaudible. However, at low bit-rates the error introduced by the dimension reduction is perceptually better perceived than an error introduced by a quantization procedure.

$\mathbf{Y}$ 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 $\mathbf{Y}$:

$$\hat{\mathbf{Y}} = \mathbf{W V_r^T X} \tag{11}$$

$\mathbf{W}$ is the unmixing matrix obtained by ICA. $\hat{\mathbf{Y}}$ is the reduced rank independent component representation of the

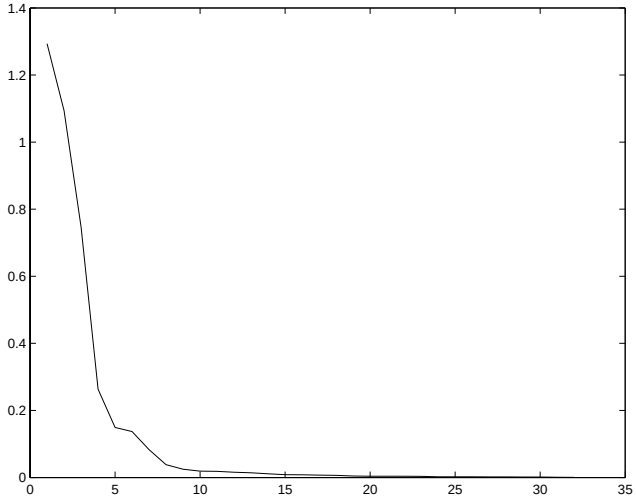

Fig. 4. The singular values decay rapidly. For each ICA block we consider the first r basis vectors which correspond to the first r singular values

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

$$\mathbf{X_{rec}^T} = \hat{\mathbf{Y}}_\mathbf{q}^\mathbf{T}\mathbf{B} = \hat{\mathbf{Y}}_\mathbf{q}^\mathbf{T}(\mathbf{V_r^T W})^\sharp \quad (12)$$

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

4.3. 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]:

$$R_k = R_{avg} + \frac{1}{2} log_2 \frac{\sigma_k^2}{\prod_{k=1}^{r} (\sigma_k^2)^{\frac{1}{r}}} \quad (13)$$

The bit allocation according to equation 13 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. We compensate the overhead of the ICA matrix transmission 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.

4.4. Coding delay

Computing the ICA matrix for each block is a time consuming task which adds coding delay to the scheme. The coding delay depends on the implementation of ICA. With fast ICA implementation the coding delay can reduce to the coding delay introduced by coders using psychoacoustic models.

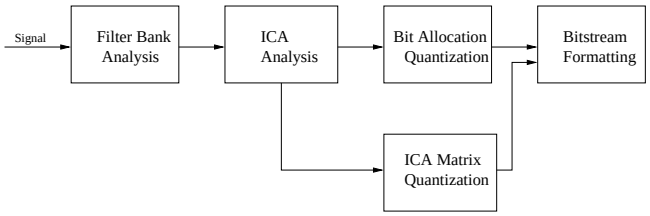

Fig. 5. Architecture of the proposed encoder.

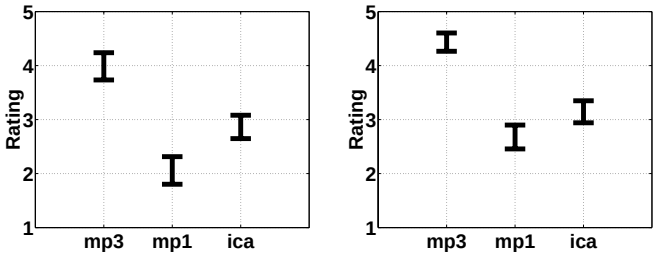

Fig. 6. Encoders mean ranking value with 95% confidence interval. The left figure corresponds to encoding in 32kbps and 44.1khz sampling rate. The right figure corresponds to encoding in 32kbps and 32khz sampling rate.

5. EXPERIMENTS RESULTS

We compared our algorithm with two perceptual audio coders. MPEG-1 layer 1 and MPEG-1 layer 3 (MP3) [8]. Layer 1

	Layer-1	ICA
Country	0.332263	0.217462
Pop	0.333203	0.322581
New Age	0.322254	0.190823
Classical	0.304873	0.166340
Percussion	0.382145	0.172121

Table 1. Empirical entropy for the quantized filter bank coefficients in the audio files used in the experiment

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. The number of participants in the experiment was 10. Figure 6 depicts the mean rating value for each of the encoders. 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.

Another measure that we tested was the empirical entropy of the quantized filter bank coefficients. This measure gives indication whether the coefficients can be further compressed using lossless compression to achieve the entropy lower bound. Table 5 shows the empirical entropies. It can be seen that for all sounds the entropy measured for the ICA quantized coefficients is lower than the one measured in Layer-1 which is based on the psychoacoustic model. This implies that the coefficients can be further compressed using lossless coding scheme, thus we can further reduce the bitrate for the ICA encoder.

The test files, which were used in the experiment can be downloaded from http://www.cs.huji.ac.il/∼chopin/ica-encoder/index.html

6. CONCLUSION

We showed that the psychoacoustic model which is used in perceptual audio coding can be interpreted in terms of reducing information redundancy in the signal by reducing the mutual information between the filter bank subbands. We argued that ICA is more appropriate for removing this redundancy. We showed that for low-bitrates an audio compression algorithm based on ICA is superior than legacy perceptual coding algorithm. 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 sound files. This leads us to the conclusion that using ICA in low bit-rates might be equivalent or better than psychoacoustic modeling.

7. 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] M. Bosi, "Filter banks in perceptual audio coding," in *Preceeding of the 17'th AES International Conference*, 1999, pp. 125–135.

[6] 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.

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

[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.