Multichannel Audio Modeling and Coding Using a Multiband Source/Filter Model

Kiki Karadimou, Athanasios Mouchtaris, and Panagiotis Tsakalides *
Institute of Computer Science (FORTH-ICS)
Foundation for Research and Technology-Hellas
GR-711 10, Heraklion, Crete, Greece
{karadsun, mouchtar, tsakalid}@ics.forth.gr

Abstract

In this paper we propose a source/filter model for achieving low bitrate transmission of multichannel audio signals, in which the filter part corresponds to the specifics of each microphone information while the source part contains mostly the interchannel similarities. Using the appropriate filter for each channel and the source part of only one of the microphone signals, we can resynthesize a high quality approximation of each channel; thus, the filter part of each channel need only be encoded. Low datarates can be achieved in the order of few KBits/sec/channel focusing on applications such as remote mixing or distributed musicians collaboration.

1 Introduction

During the last decade, stereophonic audio is progressively being commercially substituted by multichannel audio. The increase of the audio channels, both at the capturing and rendering sides, led to the creation of more realistic audio recordings that immerse the listener into the acoustic scene, but increased the requirements for transmission datarates as well. For this reason, many compression techniques have been proposed in order to give efficient solutions in several storage and transmission constraints. Consequently, multichannel audio compression algorithms have been developed which not only reduce the intra-channel redundancies, but the inter-channel redundancies as well (a discussion regarding the reduction of inter-channel redundancy for various multichannel audio coding systems can be found in [1]). These algorithms, while very effective, remain highly demanding for many practical applications. Another relevant issue is the fact that in order to create the multiple channels for multichannel audio rendering, a large number of microphones in a venue is used. These microphone signals are then mixed into a smaller number of channels that constitute the final multichannel audio recording. For various applications, methods that allow for remote mixing are of great interest in the music industry. Within our goals is to design an algorithm that can be used towards addressing such issues. Another area within our interests is remote collaboration of geographically distributed musicians, a field of great significance with extensions to music education and research. Current experiments have shown that high datarates are needed so that musicians can perform and interact with minimal delay [2].

In this paper, we propose a source/filter model of multichannel audio which takes advantage of the redundancy among the channels in order to achieve low datarate requirements. The coding section of our model is based on the speech coding scheme of [3]. Our method is tailored towards the transmission of the various microphone signals of a multichannel recording *before* they are mixed and thus is of special interest in applications such as remote mixing and distributed musicians collaborations. This innovative approach relaxes the current bandwidth constraints of these demanding applications, enabling their widespread usage and more clearly revealing their value.

2 Multichannel Audio Modeling

Initially, a brief description is given of how the multiple microphone signals for multichannel rendering are recorded. In this paper, we mainly focus on live concert hall performances. A number of microphones is used to capture several characteristics of the venue, resulting in an equal number of *microphone signals*. Spot microphones are microphones that are placed close to the sound source; the recordings of these microphones depend largely on the instruments that are near the microphone and not so much on the acoustics of the hall. Resynthesizing the signals captured by these microphones, therefore, involves enhancing certain instruments and diminishing others, which in most cases overlap both in the time and frequency domains. Reverberant microphones are the microphones placed far from the sound source, that mainly capture the reverberation information of the venue. Under our approach, ideally only one microphone signal could be used to resynthesize all the remaining signals, using some additional microphone-specific information of small datarate demands. It is important to note that in this paper we are interested only in the spot microphones resynthesis. The reverberant microphones problem has been treated in our previous work [4], where it has been shown that these signals can be resynthesized from a reference recording using specially designed LTI filters.

In our methodology, we propose a tradeoff between the *accuracy* and *objectives* of the final multichannel recording. We propose that it is possible to achieve low datarates by substituting some microphone signals with others, which, although they are different acoustically, they however retain the "objectives" of the initial recording. By the term "objectives"

*This work has been supported by the Greek General Secretariat of Research and Technology under Program ΕΠΑΝ, and by a Marie Curie International Reintegration Grant within the 6th European Community Framework Program.

1-4244-0132-1/05/$20.00 ©2005 IEEE

we refer to the aesthetic reason for a particular microphone placement (*e.g.* a microphone might be placed close to the chorus of an orchestra for placing emphasis on this part of the orchestra). With our algorithm, we attempt to resynthesize the recording of this "chorus microphone signal" using another microphone signal of the same performance (*e.g.* from a microphone placed close to the violins). The resynthesized signal can be of very good quality and might sound as if it was recorded with a microphone placed close to the chorus albeit different when compared to the actual "chorus microphone signal". This is the case when the new signal retains the "objectives" of the recording, with a loss of "accuracy" (*i.e.* the new recording does not sound the same as the original recording). We claim that with our model it is possible to achieve low datarates, good audio quality, and retain the sense of realism, without significant sacrifices regarding the accuracy of the multichannel recording.

Our proposed methodology is based on a multiband source / filter representation of the multiple microphone signals. Each microphone signal is segmented into a series of short-time overlapping frames using a sliding window. Under the source/filter model, for each frame the signal $s(n)$ at time n is related with the p previous signal samples by the following autoregressive (AR) equation

$$s(n) = \sum_{i=1}^{p} a(i)s(n-i) + e(n) \tag{1}$$

where $e(n)$ is the modeling error, and p is the AR filter order. The $p+1^{th}$-dimensional vector $\boldsymbol{a}^T = [1,\ -a_1,\ -a_2,\ \cdots,\ -a_p]^T$ is the low dimensional representation of the signal spectral properties. In the general case, the AR all-pole filter, which can be found using linear predictive (LP) analysis [5], gives only an approximation of the signal spectrum, and more specifically the spectral envelope. The error signal can be obtained from the original signal by inverse filtering with the estimated AR filter. This error signal is also referred to as the residual.

Consider now two microphone signals of the same music performance, which have been placed close to two different groups of instruments of the orchestra. Each of these microphones mainly captures that particular group of instruments, but also captures all the other instruments of the orchestra. Assume we model each of the two audio frames with the source/filter model, resulting in two different AR filters and residual signals. It is apparent that if the AR vector could capture the exact envelope (shape) of the spectrum of the particular audio segment, then the two different residual signals would contain the same flat harmonic frequency components; thus they would be equal (differences in total gain are of no interest for this application). Therefore, we would be able to resynthesize each of the two audio frames using only the AR filter that corresponds to that audio frame, and the residual signal of only one of the two microphones. If we used similarly the source/filter model for all the spot microphone signals of a single performance, we would be able to completely resynthesize these signals using their AR vector sequences (one vector for each audio frame) and the residual of only one microphone signal. This would result in a great reduction of the datarate of the multiple microphone signals.

In practice, the AR filter is not an exact representation of the spectral envelope of the audio frame, and the residual signals for the two microphone signals will not be equal. We can improve the modeling performance of the AR filter by using filterbanks. We divide the spectrum of the audio signals and apply LP analysis in each band separately (subband signals are downsampled). A small AR filter order for each band can result in much better estimation of the spectral envelope than a high-order filter for the full frequency band. The multiband source/filter model achieves a flatter frequency response for the residual signals. Then we can use one of them for resynthesizing the other microphone signals, in the manner explained in the previous paragraph. However, the error signals cannot be made exactly equal, thus the resynthesized signals will not sound exactly the same as the originally recorded signals. This corresponds to the loss of "accuracy" for the multichannel recording that was discussed; on the other hand, the proposed model results in audio signals of high-quality which retain the "objective" of the initial recordings, in the sense that was explained.

Our claims can be verified experimentally, not only for audio signals, but for other cases of harmonic signals as well, such as speech signals. We note that some specific types of microphone signals, such as percussive instruments and signals from microphones far from the source, present different challenges that we addressed in previous work [4]. The method proposed in this paper focuses on the large class of audio signals that can be modeled using a short-time analysis approach with emphasis on their spectral envelope.

3 Multichannel Audio Coding

The next step in our algorithm is to quantize the spectral envelopes for each of the microphone signals. This is done *separately* for each of the frequency bands in which we divided the microphone signal. We follow the quantization scheme of [3], developed for vector quantization of speech line spectral frequencies (LSF's). We transform the AR coefficients of each microphone signal to LSF's, since LSF's are more resistant to quantization errors. Next, we model the sequence of LSF's that we obtain from each microphone signal with the use of a Gaussian Mixture Model (GMM)

$$g(\boldsymbol{x}) = \sum_{i=1}^{m} p_i N(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{2}$$

where $N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the normal multivariate distribution with mean vector μ and covariance matrix $\boldsymbol{\Sigma}$, m is the number of clusters and p_i is the prior probability that the observation $\boldsymbol{x}$ has been generated by cluster i. The Karhunen Loeve Transform (KLT) is adopted for the LSF's decorrelation. KLT is especially fit for GMM-modeled parameters since it is the optimal transform for Gaussian signals in a minimum-distortion sense. Using GMM's, each LSF vector is assigned to one of the Gaussian classes using some classification measure, thus can be considered as approximately Gaussian and can be best decorrelated using the KLT.

Using the GMM modeling of the spectral parameters, it holds for the covariance matrix of each class that it can be diagonalized using the eigenvalue decomposition as

$$\boldsymbol{\Sigma}_i = \mathbf{Q}_i \boldsymbol{\Lambda}_i \mathbf{Q}_i^T \tag{3}$$

where $i = 1, \ldots, m$ and $\boldsymbol{\Lambda}_i = diag(\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,p})$. In other words, $\boldsymbol{\Lambda}_i$ is the diagonal matrix containing the eigenvalues, and $\mathbf{Q}_i$ is the matrix containing the corresponding set of orthogonal eigenvectors of $\boldsymbol{\Sigma}_i$, for the i^{th} Gaussian class of the model. Then, KLT substitutes each spectral vector for time segment k, $\boldsymbol{z}_k$, with another vector of decorrelated components $\boldsymbol{w}_k$

$$\boldsymbol{w}_k = \mathbf{Q}_i^T(\boldsymbol{z}_k - \boldsymbol{\mu}_i). \tag{4}$$

Similarly, the inverse KLT procedure (IKLT) reconstructs $\boldsymbol{z}_k$

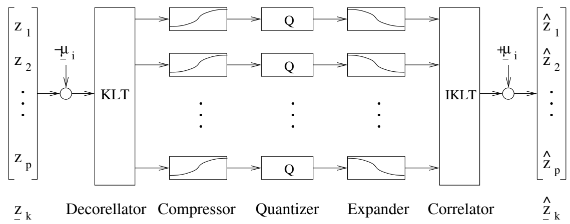

Figure 1: Quantization among clusters.

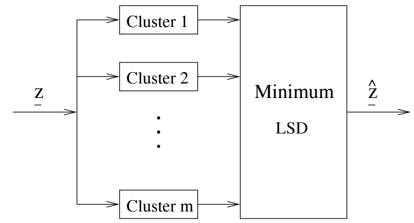

Figure 2: Overall Quantization.

from $\boldsymbol{w}_k$ using the inverse relation

$$\boldsymbol{z}_k = \mathbf{Q}_i \boldsymbol{w}_k + \boldsymbol{\mu}_i. \tag{5}$$

A nonuniform quantizer is achieved by a combination of a compressor, a uniform quantizer and an expander. The decorrelated vectors are processed using a logarithmic compression function, quantized by a uniform quantizer and expanded using the inverse of the compression function. The companding method of [6] was used, since this function resulted in robust quantization in our experiments. A bit allocation scheme for the uniform quantizer is needed in order to allocate the total available bits (denoted by b_{tot} and specified by the user) for quantizing the source, among the various clusters of the GMM. Let b_i be the bits for quantizing cluster i, and q_i the quantity

$$q_i = \left[\prod_{j=1}^{p} \lambda_{i,j}\right]^{\frac{1}{p}}, \quad i = 1, \ldots, m, \tag{6}$$

where p is the dimensionality of the LSF vector. In the *fixed rate* bit allocation scheme the length of the codewords is fixed and can be easily found to satisfy the constraint $2^{b_{tot}} = \sum_{i=1}^{m} 2^{b_i}$. Subject to this constraint, the optimal bit allocation which minimizes the total average mean square distortion is given by

$$b_i = b_{tot} - \log_2 \left[\sum_{j=1}^{m} (p_j q_j)^{\frac{p}{p+2}}\right] + \frac{p}{p+2} \log_2(p_i q_i), \; i = 1, \ldots, m. \tag{7}$$

After the evaluation of the cluster allocated bits, we calculate the bit allocation among the cluster dimensions as

$$b_{i,j} = \frac{b_i}{p} + \frac{1}{2} \log_2 \left[\frac{\lambda_{i,j}}{q_i}\right], \quad i = 1, \ldots, m \quad j = 1, \ldots, p \tag{8}$$

where $b_{i,j}$ is the allocated bits to the j^{th} component of the i^{th} cluster and $\lambda_{i,j}$ is the j^{th} eigenvalue of cluster i. In our implementation we rounded $b_{i,j}$ in the nearest integer number for more accurate bit allocation. The procedure we follow for coding the LSF vectors of each frequency band is:

A. Cluster Quantization. The quantization of an LSF vector with the parameters of i-th cluster (Fig. 1) consists of the following stages: (i) We subtract from the LSF vector $\boldsymbol{z}_i$ the mean $\boldsymbol{\mu}_i$ of the cluster i; (ii) we decorrelate the resultant vector using the matrix $\mathbf{Q}_i^T$; (iii) we pass the vector's components through a nonuniform quantizer (compressor, uniform quantizer, expander); (iv) we reconstruct the correlated version of the quantized vector using the matrix $\mathbf{Q}_i$; (v) we finally add the cluster mean $\boldsymbol{\mu}_i$ to obtain the quantized value of $\boldsymbol{z}_k$ by the i^{th} cluster, $\hat{\boldsymbol{z}}_k$.

B. Overall Quantization. A specific LSF vector $\boldsymbol{z}_k$ is quantized with the use of every cluster of the GMM as described. In order to choose the GMM cluster that best models a particular LSF vector, we evaluate the relative distortion value for the vector and we choose the one with the minimum distortion (Fig. 2). Here, we employ the Log Spectral Distortion (LSD) as a measure of distance as in [3]

$$LSD(i) = \left(\frac{1}{F_s} \int_0^{F_s} \left[10 \log_{10} \left(\frac{S(f)}{\hat{S}^{(i)}(f)}\right)\right]^2 df\right)^{\frac{1}{2}} \tag{9}$$

where F_s is the sampling rate, $S(f)$, $\hat{S}^{(i)}(f)$ are respectively the LPC power spectra corresponding to the original vector $\boldsymbol{z}_k$ and the quantized vector $\hat{\boldsymbol{z}}_k^{(i)}$, for each cluster $i = 1, \ldots, m$. We choose $\hat{\boldsymbol{z}}_k$ corresponding to the cluster of minimum LSD and transmit its bitwise representation. At the receiver, we convert the quantized LSF vector into its corresponding LPC value and use the LPC vector to resynthesize the audio signal of segment k.

4 Results

For our experiments, we have obtained microphone signals from a US orchestra hall by placing 16 microphones at various locations throughout the hall. Our objective is to indicate that the model and the coding method we propose result in a high quality recording with low datarate requirements. For this purpose, we use two of these microphone signals, where one of the microphones mainly captures the male voices of the chorus of the orchestra, while the other one mainly captures the female voices. These recordings are very easy to distinguish acoustically. The efficiency of the proposed algorithm has been tested via objective and subjective tests, that indicate that our method results in a very good quality microphone signal that retains the objective of the initial recording with a very small loss of accuracy.

4.1 Modeling Performance

In this section, we show that the use of the proposed method results in a modeled signal that is objectively and subjectively very close to the original recording. For this purpose, we use the two microphone recordings of the male and female voices of the chorus, as mentioned. The objective is to resynthesize one of these recordings using its corresponding low-dimensional model coefficients along with the residual of the other recording.

From initial listening tests it has been clear that using a number of bands around 8 for our model produced high quality resynthesis without loss of the objective of the initial recording. For example, we have been able to resynthesize the male voices recording based on the residual from the female voices. Without the use of a filterbank, the resulting quality of the resynthesized signal greatly deteriorated with a complete loss of the recording objective. In order to show this objectively, we measured the distance between the residual signals of the two recordings, using the normalized mutual information as a distance measure. The intuitive claim is that decreasing the distance of the two residuals will in-

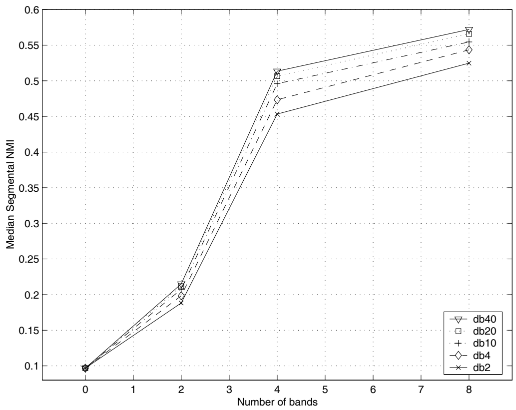

Figure 3: Normalized Mutual Information between the residual signals from the reference and target recordings as a function of the number of bands of the filterbank, for various Daubechies (db) filters.

crease the quality of the resynthesized recording. Our listening tests indicated that increasing the number of subbands in our model, and consequently improving the model accuracy, resulted in much better quality of the resynthesized signals. While several measures were tested, the normalized mutual information proved to be very consistent in this sense.

The use of mutual information $I(X;Y)$ as a distance measure between random variables X and Y is very common in pattern comparison. Since our interest is in comparing two vectors X and Y (Y being the desired response), it is useful to use a modified definition for the mutual information, the Normalized Mutual Information (NMI) $I_N(X;Y)$ which is the mutual information normalized by the entropy of Y, so that $0 \leq I_N \leq 1$. The NMI obtains its minimum value when X and Y are statistically independent and its maximum value when $X = Y$. The NMI does not constitute a metric since it lacks symmetry, however it is invariant to amplitude differences which is very important when comparing audio waveforms.

In Fig. 3 we plot the NMI between the power spectra of the two residual signals with reference to the number of different subbands used, for different orders of the Daubechies wavelet filters, which were used for our tree-structured filterbank [7]. As a result, our filterbank has the perfect reconstruction property, which is essential for an analysis/synthesis system, and also octave frequency-band division, which is important since the LP algorithm is especially error-prone in lower frequency bands. For our implementation, we used 32^{nd} order LP filter for a 1024 sample frame (corresponding to about 23 msec. for 44.1 kHz sampling rate) for the full band analysis. For the subband analysis, we used an 8^{th} order filter for each band, with a constant frame rate of 256 samples for each band (thus varying frame in msec.). The amount of overlapping for best quality was found to be 75% for all cases. These parameters were chosen so that the total number of transmitted coefficients for the resynthesized recording remains the same for both the full band and the subband cases. For the particular number of parameters used, the total number of coefficients used for the resynthesis is eight times less than the total number of audio samples. The coefficients that we intend to code for each microphone signal are the line spectral frequencies (LSF's) given their favorable quantization properties.

The NMI values in Fig. 3 are median values of the segmental NMI between the two residual signals using an analysis window of 6 msec. The residual signals are obtained using an overlap-add procedure so that they can be compared using the same analysis window. Our claim, that using a subband analysis with a small LP order for each band will produce much better modeling results than using a high LP order for the full frequency band, is greatly justified by the results shown. For the full band analysis we obtain a NMI value of 0.0956 while for a 8-band filterbank the median NMI is 0.5720 (40^{th} order wavelet filters). In Fig. 3 we plot the median NMI for different orders of the Daubechies filters. We can see that increasing the filter order results in slightly better results. Intuitively this was expected; an increase in the filter order results in better separation of the different bands, which is important since we model each subband signal independently of the others. In a similar experiment, we compared the residual signals in the time-domain and found that the median NMI doubles when using the 8-band system when compared to the full-band case. The results for both the frequency and time domains are similar regardless of the analysis window length for obtaining the NMI segmental values. When increasing the window size the NMI drops, which is expected since more data are compared. The decrease is similar for the various numbers of bands we tested.

	ABX-1	ABX-2	ABX-3
Results correct	86%	63%	10%

Table 1: Results from the ABX listening tests.

In order to test the performance of our method, we also employed subjective (listening) tests, in which a total of 17 listeners participated (individually, using good quality headphones). We used the two concert hall recordings from the same performance as mentioned earlier (one capturing the male voices and one capturing the female voices of the chorus). We chose three parts of the performance (about 10 sec. each, referred to as Signals 1-3 here) where both parts of the chorus are active so that the two different microphone signals can be easily distinguished. For each signal we designed an ABX test, where A and B correspond to the male and female chorus recording (in random order), while each listener was asked to classify X as being closer to A or B regarding as to whether the male or female voices prevail in the recording. We tested 3 different types of wavelet-based filterbanks, namely 8-band with filters db40 (test ABX-1) and db4 (ABX-2), and 2-band with db40 (ABX-3). For each of these 3 tests, we used all three of the chosen signals, thus a total of 9 ABX tests was conducted per listener. The results are given in Table 1. We can conclude that the objective results, as well as the various claims made in the previous sections regarding the model, are verified by the listening tests. It is clear that the 8-level wavelet-based filterbank produces excellent results when aliasing is limited (*i.e.* db40 case), although there is certainly room for improvement and further enhancement to our model is currently underway. On the other hand, when aliasing is high or when the number of bands (and thus the modeling accuracy) drops, the performance of the proposed method greatly deteriorates, not only in the sense of enhancing the male voices, but also regarding final quality (which most listeners noticed during the experiments). We note that we obtained very similar results (using the NMI as well as informal listening tests) with a Laplacian pyramid filterbank, which is a different type of octave-spaced filterbank [8]. The choice of filterbank and whether octave-spaced filterbanks are indeed better for our model is a subject of our ongoing research. In the following section we show that coding with bitrates in the order of 10 KBits/sec per channel

LSD (dB)	KBits/sec
1.1598	10
0.4641	20
0.1529	30
0.1020	40

Table 2: LSD for various choices of total bitrate.

Band Nr.	Bits/frame	LSD (dB)
1	28	2.0025
2	28	0.8583
3	28	1.2951
4	28	0.9006
5	33	0.4759
6	28	0.6990
7	10	1.7439
8	10	1.3029

Table 3: Number of bits chosen and corresponding LSD for each frequency band (10 KBps case).

are possible regarding the side information for good quality audio. However, these are only preliminary results, and we are confident that the bitrates can be further reduced.

4.2 Coding Results

Regarding the coding scheme proposed, our initial listening tests indicated that the final quantized version is acoustically of similar quality compared to the recorded signal, for bitrates as low as 10 KBits/sec. First, we give some objective results using the LSD measure. The sampling rate for the audio data is 44.1 kHz; we divided the frequency range into 8 octave subbands using 40^{th} order Daubechies wavelet filters which gave the best quality in the modeling results (the model parameters are the same as in the previous section). The LSF's of each band are modeled using a GMM of 16 components. The parameters of the GMM were estimated using a training audio dataset of about 220 000 vectors per subband. For obtaining this dataset, we employed a different window rate for each subband (for maximizing the number of training vectors). This database consists of recordings of the same performance as the data we encoded (but a different part of the recording than the one used for testing). For these parameters, and with varying choice of bitrate, we obtain the values of Table 2 (bitrate *vs.* LSD). An example of the total bits that were allocated in each band for the coding procedure, is given in Table 3 corresponding to the 10 KBits/sec case. This was found to be the minimal bitrate for high quality coding (similar quality to the original). For comparison, current compression algorithms for multichannel audio have minimal bitrate requirements in the order of 64 KBits/sec/channel, for achieving high quality coding.

We conducted DCR-based (Degradation Category Rating) [9] listening tests for evaluating the quality of the coded signals using a 5-grade scale in reference to the original recording (5 corresponding to being of same quality, and 1 to the lowest quality, when compared with the original male chorus recording). Subjects listened to three sound clips (originally recorded *vs.* coded Signals 1-3 from Section 4.1), where the coded signals were obtained using the best modeling parameters (8-level db40 wavelet-based). The results for 10 and 20 KBps are depicted in Fig. 4, where the 95% confidence interval are shown (x's mark the mean value and the two horizontal lines indicate the confidence limits). These results show that the coded signals in the 10 and 20 KBps cases are of high quality, similar to the quality of the original signals.

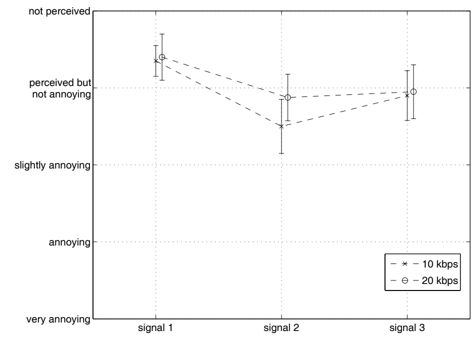

Figure 4: Results from the quality rating listening tests.

5 Conclusions

We presented a new approach for modeling and coding multichannel audio for low bitrate transmission. Our algorithm can encode a multichannel audio recording so that one audio channel with side information (in the order of few KBits/sec/channel) can be used to decode the multiple channels at the receiving end. Our algorithm was based on the "accuracy-objectives" tradeoff introduced, and thus mainly focuses on encoding the microphone signals *before* those are mixed into the channels of the final multichannel recording. The authors wish to thank the listening tests volunteers, and Prof. Kyriakakis of USC for his suggestions.

References

[1] D. Yang, H. Ai, C. Kyriakakis, and C.-C. J. Kuo, "High-fidelity multichannel audio coding with Karhunen-Loeve transform," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 365–380, July 2003.

[2] A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis, "From remote media immersion to distributed immersive performance," in *Proc. ACM SIGMM Workshop on Experiential Telepresence*, 2003.

[3] A. D. Subramaniam and B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 365–380, March 2003.

[4] A. Mouchtaris, S. S. Narayanan, and C. Kyriakakis, "Virtual microphones for multichannel audio resynthesis," *EURASIP Jrnl. on Applied Signal Process., Special Issue on Digital Audio for Multimedia Comm.*, vol. 2003:10, pp. 968–979, 2003.

[5] S. Haykin, *Adaptive Filter Theory.* Prentice Hall, 1996.

[6] J. A. Bucklew and N. C. Gallagher, "A note on the computation of optimal mean-squared error quantizers," *IEEE Trans. Comm.*, vol. 30, no. 1, pp. 298–301, 1982.

[7] G. Strang and T. Nguyen, *Wavelets and Filter Banks.* Wellesley-Cambridge, 1996.

[8] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, pp. 532–540, 1983.

[9] W. B. Kleijn and K. K. Paliwal, eds., *Speech Coding and Synthesis.* Elsevier Science, 1995.