Packet Loss Concealment for Multichannel Audio Using the Multiband Source/Filter Model

Kiki Karadimou, Athanasios Mouchtaris, and Panagiotis Tsakalides *
Institute of Computer Science
Foundation for Research and Technology-Hellas
Vassilika Vouton, P.O. Box 1385
GR-711 10, Heraklion, Crete, Greece
{karadsun, mouchtar, tsakalid}@ics.forth.gr

Abstract

We recently proposed a multichannel audio coding method using a multiband source/filter model, which results in a compact representation of the original recording. Our method can reproduce the original recording using only one audio channel and side information for the remaining channels in the order of 5 KBps/channel. Here, we examine packet loss concealment strategies for use within our model, so that we can derive a complete system for low-bitrate multichannel audio streaming through the Internet or wireless channels.

1 Introduction

Multichannel audio recordings offer improved sound realism for audio reproduction compared to stereo recordings. They are created using a large number of microphones (usually more than the final number of channels), resulting in a number of recordings that are then mixed in order to create the final channels of the multichannel recording. Multichannel recordings before or after the mixing procedure, contain a much higher degree of information to store or transmit compared to stereophonic recordings. On the other hand, the increase in the number of channels offers opportunities to explore inter-channel similarities, in addition to the intra-channel redundancy that is explored in today's perceptual audio coders. Inter-channel similarities are partly explored in current multichannel audio standards such as Mid/Side Coding [1] which is implemented as part of MPEG-2 AAC (Advanced Audio Coding), and Intensity Stereo Coding [2] which is implemented as part of Dolby AC-3. Currently, these implementations require 64 KBps per channel for high-quality coding. Another approach, Binaural Cue Coding (BCC) [3], belongs to a wide area of research known as Spatial Audio Coding which attempts to drastically exploit inter-channel redundancy by encoding one audio channel as reference, and only a small degree of side information for the remaining channels. In this sense, BCC achieves an approximate bitrate of 6 KBps per channel as side information (*i.e.* excluding the reference channel which can be encoded using *e.g.* a monophonic perceptual audio coder). Recently, we proposed a novel approach for multichannel audio coding [4] which follows the philosophy of spatial audio coding, achieving a bitrate of 5 KBps per channel for the side information. In contrast to BCC, our method allows resynthesis of the individual microphone signals (before mixing) at the receiving end, which can be expected to offer a more realistic reproduction of the multichannel recording. Additionally, in contrast to most multichannel audio coding methods, our approach is suitable for applications such as distributed musicians collaboration and remote mixing.

Our source/filter model allows for low bitrate coding of multichannel audio recordings, before or after mixing. It is useful both for stored recordings (an analog of MP3 encoding for multichannel recordings), or for streaming applications. One application we consider is multichannel audio streaming through the Internet or wireless channels. In such cases, it is often possible that some packets of the transmitted information might be lost or delayed due to channel conditions. Especially for audio applications, not only lost packets but delayed packets as well are a serious problem. In any case, the missing information will result in an audible degradation of quality in the audio signal. This problem has been examined mostly in the Voice Over IP framework, where several directions have been proposed. One approach is to introduce redundant information to the transmitted packets, which can be extra bits for forward error correction (FEC), side information, or multiple packets corresponding to the same information. An issue with these systems is that they introduce redundancy thus increase in the bitrate, and also delays to the encoding/decoding process. Even under the presence of these approaches, it is possible that some packets of information might be considered as lost. Packet loss concealment (PLC) attempts to reconstruct the lost information exclusively at the receiver side, without redundancy, without any overhead to the transmitter, and with small overhead and delays for the total encoding/decoding process.

Some PLC methods are based on particular models for the signals, which in the case of speech signals can be either source/filter models (*e.g.* methods in [5, 6, 7]) or sinusoidal models (*e.g.* [8]). For audio and speech signals, without knowledge of the underlying coding method, PLC is usually achieved by repeating the last received samples or by interpolation of previous packets; for speech signals this can be improved by additional knowledge of the pitch evolution. For a survey of several PLC methods for speech and audio signals the reader is referred to [9]. In streaming audio, model-based approaches that depend on the coding procedure also exist similarly to the speech case (*e.g.* [10]). For multichannel au-

*This work has been supported by the European Social Fund and Greek National Resources, and by a Marie Curie International Reintegration Grant within the 6$^{\text{th}}$ European Community Framework Program.

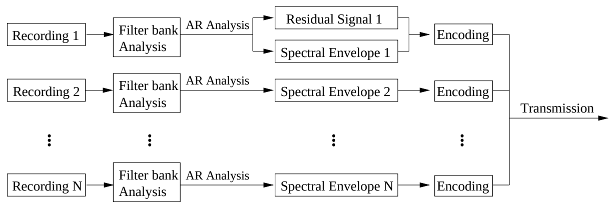

Figure 1: Diagram of the proposed encoding approach.

dio, a PLC method on the raw PCM-coded signal space has been developed in [11], where audio inter-channel similarities have been explored. Our objective in this paper is to design a model-based PLC scheme for multichannel audio, building on our previously proposed coding method, so that we can derive a complete system that can be used for low-bitrate encoding/decoding of multichannel audio, for transmission through the Internet or wireless channels. One of the PLC methods we propose is a novel approach based on exploring cross-channel correlations regarding the LSF vectors, in addition to intra-channel correlations of adjacent (in time) LSF vectors.

2 Multichannel Audio Model

In this section we give a brief description of the system we recently proposed [4] for multichannel audio coding. Initially, it is important to explain how multichannel audio recordings are produced and at what stage of this process our system is introduced. Our focus is on live concert hall performances, without loss of generality for our methods.

Multichannel audio recordings are created by initially using a large number of microphones in the venue. As an example, in our experiments we use a multi-microphone recording from a US concert hall which has been recorded using a total of 16 microphones. These microphones can be divided in two categories, spot microphones which record specific parts of the orchestra (e.g. the violins, the chorus, etc.), and reverberant microphones which are placed far from the orchestra and mainly capture the acoustic properties of the venue. The latter category can be sufficiently modeled using high-order linear time-invariant filters [12], thus it represents no significant challenges regarding coding under bandwidth constraints. The former category has been the main focus of our recent work, and we have shown that all spot microphone signals can be resynthesized successfully from only one of these microphone signals (termed as the *reference* signal), using a small amount of information (in the order of 5 KBps/sec) for each of the remaining microphone signals [4]. In general, the final multichannel recording is produced by a mixing process, where the recorded signals are mixed using aesthetic criteria and each of the resulting channels is led to the corresponding loudspeaker. Consequently, our coding process precedes the mixing process, in contrast to all other multichannel audio coding systems that encode the final channels that are the result of mixing. Thus, our system allows for a variety of applications that other systems cannot accommodate for, such as remote mixing and distributed musicians collaboration. Since our system encodes the initial microphone signals rather than simply binaural cues as in BCC, it can be expected to result in a more realistic reproduction of the multichannel audio recording. In both our system and BCC, the rate needed for each of the additional audio channels (excluding the reference channel) is in the order of 5 KBps. Our system involves a higher number of channels (before mixing); on the other hand, for our system only spot microphone signals require this side information, since for the reverberant microphone signals we only need to transmit a LTI filter with practically zero needs in bitrate.

The first stage of our system is the modeling process, which corresponds to passing all N microphone signals (see Fig. 1) through an M-band subsampled filterbank, resulting in M*N signals. Using critical subsampling, this introduces no additional overhead regarding the total number of samples. Each of the subband signals is processed independently of the others, using the linear predictive (LP) source/filter model. This process results in a more accurate LP model rather than applying the LP model in the full-band (0-20 kHz). In practice, we have found that octave-spaced filterbanks produce best subjective results (using listening tests), which can be attributed to the fact that this type of filterbank is more accurate for the lower-frequencies which are more acoustically significant. The novelty of our model has been the fact that for each of the subband signals, its source (or error) signal can be essentially substituted with the source signal of one of the N initial microphone signals (reference signal) of the corresponding subband. This is due to the subband LP analysis which succeeds to capture most of the important information using only few coefficients (the LP filters), leaving the error signals common for all microphone signals (for each corresponding subband). Then, we need to encode one full audio signal (the reference signal), while for the remaining N-1 signals we only need to encode the LP filters for all subbands, which translates to a large degree of information reduction without degradation of the audio quality.

The second part of our system consists of coding the model parameters using as low number of bits as possible. We have shown that for coding the LP filters for each microphone signal (side information) we can use the approach of [13]. Briefly, the process includes converting the LPC's to LSF's (Line Spectral Frequencies) which are more robust for coding, and then using a trained GMM (Gaussian Mixture Model) for clustering each LSF vector into a Gaussian class. For each class, the Gaussian parameters (mean and covariance matrix) can be used for KLT (Karhunen-Loeve) decorrelation. This is followed by scalar non-uniform quantization of the decorrelated vector parameters, which produces the final bitstream to be transmitted. The procedure is inverted for each vector in the transmitter side for each Gaussian class, in order to classify the vector to the class which results in the smallest quantization distortion. In the receiver, the inverted process is used to reconstruct the transmitted vector. We have shown that use of this procedure in our system results in bitrates in the order of 5 KBps for encoding all the subband LP parameters for each microphone signal with high-quality. The reference signal, on the other hand, can be encoded as a monophonic audio source, using e.g. a perceptual audio coder such as MP3, without any perceptually noticeable artifacts in the final resynthesized recording.

3 LSF Estimation Methods

As mentioned, the information that is transmitted to the receiver for reconstructing the N microphone signals includes the LSF vectors and one full audio (reference) channel. Thus, we assume that each of the transmitted packets contains either an LSF vector or a segment of PCM samples of the reference channel. The reference channel might be encoded in practice by a perceptual audio coder, but this would not change the PLC methods we propose here. In the remaining of this section we describe four different approaches for estimating the lost LSF vectors. Regarding the error signal (extracted from the reference channel), we assume that it has been correctly received.

Regarding the case when a part of the error signal might be lost as well, this issue can be addressed by existing methods for monophonic audio loss concealment, but is also a subject of our ongoing research. For all methods described in this paper, it is assumed that the packet corresponding to time t is lost, and is estimated using only previous packets (for minimal delay).

3.1 Single-Channel Estimation

Let $\boldsymbol{x}_t$ be the lost packet (LSF vector) that we wish to estimate and $\boldsymbol{x}_{t-1}$ be the previous in time LSF vector. In order to estimate $\boldsymbol{x}_t$, we use the GMM-based joint density estimation method of [14], and will be referred to in this paper as "1 ch est" (since the lost packet is estimated using one previous packet from the same-channel recording). We only use the previous in time LSF vector under this approach. We apply a Gaussian Mixture Model (GMM) for modeling the probability density function of the joint vector-space $\boldsymbol{z} = [\boldsymbol{x}_{t-1}^T \ \boldsymbol{x}_t^T]^T$,

$$\mathrm{g}(\boldsymbol{z}) = \sum_{i=1}^{M} p(\omega_i)\mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_i^z, \boldsymbol{\Sigma}_i^{zz}), \tag{1}$$

where $p(\omega_i)$ is the prior probability of class ω_i and $\mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The parameters of the GMM (mean vectors, covariance matrices and prior probabilities of each Gaussian class), can be estimated from the observed data (a training dataset) using the expectation maximization (EM) algorithm. Then, the lost packet $\boldsymbol{x}_t$ is estimated in minimum mean-squared (MMSE) sense from the relation:

$$\hat{\boldsymbol{x}}_t = \sum_{i=1}^{M} p(\omega_i|\boldsymbol{x}_{t-1}) \left[\boldsymbol{\Sigma}_i^{x_t x_{t-1}} \left(\boldsymbol{\Sigma}_i^{x_{t-1} x_{t-1}}\right)^{-1} \boldsymbol{x}_{t-1}^o + \boldsymbol{\mu}_i^{x_t}\right] \tag{2}$$

where

$$\boldsymbol{x}_{t-1}^o = \boldsymbol{x}_{t-1} - \boldsymbol{\mu}_i^{x_{t-1}} \tag{3}$$

and

$$p(\omega_i|\boldsymbol{x}_{t-1}) = \frac{p(\omega_i)\mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_i^{x_{t-1}}, \boldsymbol{\Sigma}_i^{x_{t-1}x_{t-1}})}{\sum_{j=1}^{M} p(\omega_j)\mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_j^{x_{t-1}}, \boldsymbol{\Sigma}_j^{x_{t-1}x_{t-1}})}. \tag{4}$$

3.2 Repetition

Another very simple receiver-based PLC strategy is to replace the lost packet $\boldsymbol{x}_t$ with a copy of the packet that contains the LSF vector of the previous time instant $\boldsymbol{x}_{t-1}$:

$$\hat{\boldsymbol{x}}_t = \boldsymbol{x}_{t-1}. \tag{5}$$

This technique will be referred here as "copy".

3.3 Linear Regression

Under a linear regression approach, the lost packet is defined as a linear combination of some previous in time LSF vectors. In our experiments we used three past packets:

$$\hat{\boldsymbol{x}}_t = a \ \boldsymbol{x}_{t-1} + b \ \boldsymbol{x}_{t-2} + c \ \boldsymbol{x}_{t-3} \tag{6}$$

where a, b, c are scalars and are estimated using least-squares estimation from a training dataset. For compactness we refer to this method as "regr".

3.4 Cross-Channel Estimation

Here we propose an extension of the single-channel estimation approach into cross-channel estimation. In the problem we examine, there are correlations not only between adjacent LSF vectors of the same microphone signal, but also across the LSF vectors of the various microphone signals. Thus, it may be possible to enhance the single-channel estimation method we described, using the LSF vectors of other microphone signals. In order to show this, we examine a 2-channel estimation approach (denoted by "2 ch est" for compact notation). Thus, we train a GMM using triplets of LSF vectors, *i.e.* using a supervector $\boldsymbol{z} = [\boldsymbol{x}_{t-1}^T \ \boldsymbol{y}_t^T \ \boldsymbol{x}_t^T]^T$, where $\boldsymbol{x}_t$ is the vector of one channel that we need to estimate (target vector), $\boldsymbol{x}_{t-1}$ is the previous vector of the same channel assumed as known, and $\boldsymbol{y}_t$ is the LSF vector corresponding to the current frame of another channel (of the same multichannel recording), also assumed as known.

If we denote as $\boldsymbol{w}_{t-1} = [\boldsymbol{x}_{t-1}^T \ \boldsymbol{y}_t^T]^T$, then we can obtain $\hat{\boldsymbol{x}}_t$ using again (2) and (4), where $\boldsymbol{x}_{t-1}$ in those equations is now substituted by $\boldsymbol{w}_{t-1}$. It should be noted that in this case, $\boldsymbol{w}_{t-1}$ has double dimension than $\boldsymbol{x}_{t-1}$, since it is a supervector containing both the LSF vector of the previous packet (same channel) and the LSF vector of the current packet of another channel. However, it is a simple procedure to derive the various correlation matrices and mean vectors using the GMM parameters obtained for $\boldsymbol{z}$.

4 Results

For our experiments, we use two microphone signals from a US orchestra, *before* those are mixed into the final multichannel recording. In particular, the first signal was recorded near the male voices of the orchestra's chorus, while the second one was recorded near the female voices. These signals are very easy to distinguish acoustically. Our objective is to test the four LSF estimation methods described in the previous paragraph, assuming that the error signal frame is correctly received, examining various channel conditions. In this section, we give objective results of the reconstructed signals using the Log-Spectral Distortion (LSD) measure. LSD is the distance between LP-based spectral envelope of the original LSF vector and the reconstructed LSF vector, defined as

$$LSD = \left(\frac{1}{F_s}\int_{0}^{F_s}\left[10\log_{10}\left(\frac{S(f)}{\hat{S}(f)}\right)\right]^2 df\right)^{\frac{1}{2}}, \tag{7}$$

where F_s is the sampling rate, $S(f)$, $\hat{S}(f)$ are respectively the LP power spectra corresponding to the original vector $\boldsymbol{x}_t$ and the reconstructed vector $\hat{\boldsymbol{x}}_t$. The results given here are based on LSF vectors before the coding/quantization process. The effect of LSF quantization to the resynthesized audio signals has been examined in our previous work [4].

The sampling rate for the audio signals is 44.1 kHz; we divided the frequency spectrum into 8 octave subbands using 40^{th} order Daubechies wavelet filters. In each subband we applied a Hamming window of 256 samples, with 75% overlapping, while the LP filter order was 8. These parameters were shown in our previous work to be a good selection given

# Band	1 ch est	copy	regr	2 ch est
1	3.0171	2.9522	2.9217	2.3350
2	2.9630	2.9633	2.9397	2.3354
3	2.5260	2.5416	2.5080	2.2251
4	2.0886	2.1003	1.9626	1.8935
5	1.6999	1.7190	1.6188	1.5732
6	1.4971	1.5467	1.4512	1.4393
7	1.3582	1.4455	1.3305	1.3203
8	1.2832	1.3836	1.2756	1.2496
Average	2.0542	2.0815	2.0010	1.7964

Table 1: Average LSD (dB) values for each subband, in the scenario that the current LSF vector is lost given that the previous LSF vector was received.

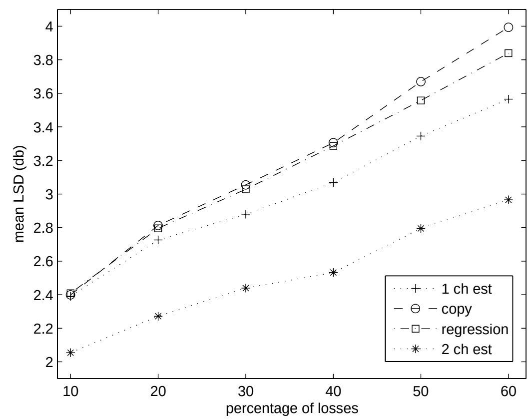

Figure 2: Average LSD (dB) values for various choices for the percentage of channel losses, for the four PLC methods.

the trade-off regarding audio quality *vs.* total information reduction. When applying the GMM estimation methods ("1 ch est" and "2 ch est"), a GMM of 16 components was used. The parameters of the GMM were estimated using a training audio dataset of about 92 000 pairs of LSF vectors (approximately 2 min of audio) from all the subbands. In the "1 ch est" case, each pair consisted of two consecutive in time LSF vectors ($t-1$ and t); in "2 ch est", triplets of vectors were considered instead of pairs, corresponding to the previous LSF of the same channel, the current LSF of another channel, and the target LSF vector, based on what was described in the previous section. For training, we used recordings of the same performance as the data we estimated (a different part of the recording than the one used for testing). In the regression method, we used the same training dataset for least-squares estimation of the regression parameters. The testing audio data used was about of 1 min duration.

Using the aforementioned parameter values and in order to evaluate the estimation accuracy of the proposed PLC methods ("1 ch est", "copy", "regr", "2 ch est"), we carried out experiments based on two scenarios. In the first scenario, the current LSF vector is lost given that the previous in time vector was received. Thus, we applied this scenario to *all* the LSF vectors. The obtained average LSD values between the true and the estimated LP spectra for each subband are shown in Table 1. The results in the table show that the cross-channel estimation method performs best (least distortion), which was expected since we use an extra channel of audio for our estimation method, and given that there is correlation between the various channels of the same recording. However, from the table we can also see that the various methods perform almost equally well, which gives obvious preference to using simpler methods (such as repetition) rather than estimation methods which need a training procedure.

The second scenario we examined can be considered as a more realistic simulation regarding the channel conditions. In the second scenario we assumed that a percentage of the transmitted packets is lost. The packet losses were simulated using a Gilbert loss model [15], which is usually applied to such problems. This is a more realistic scenario since it is possible that two or more consecutive packets might be lost. In consecutive losses, we obtain the current estimate using the previous packet estimate as being the actual packet, and so forth. The results obtained are the LSD values averaged over all the *lost* packets. In Fig. 2 we plot the average LSD values with reference to the percentage of losses, for the four PLC methods described in Section 3. These results show that as the probability of packet losses increases (*i.e.* worse channel conditions), the cross-channel GMM-based estimation becomes more advantageous than the other three methods, introducing clearly less distortion in the LSF vectors. These results give an indication that the additional overhead of the cross-channel estimation method is well-justified, since audio quality is our main concern in this application. On the other hand, the single-channel estimation method is shown to perform slightly better than the simpler methods only in very bad channel conditions. The listening tests that follow give us a better indication regarding the performance of each of the four PLC methods.

4.1 Listening Test

We conducted a listening test in order to evaluate the quality of the audio signals that were resynthesized using the four PLC methods that were objectively tested in the previous paragraphs (*i.e.* single channel estimation, cross-channel estimation, repetition, and regression). In our test 17 listeners volunteered to participate, and listened to 1 audio segment (about 10 sec) that was resynthesized using the 4 PLC methods for various channel conditions (more specifically 10%, 40%, and 60% probability of packet loss). The channel conditions for these packet loss probabilities were simulated using a Gilbert loss model [15], as in the previous section. Thus, the total number of the audio files that were tested was 12 (1 audio file$\times$4 PLC methods$\times$3 percentages for packet loss). The testing method was a Degradation Category Rating (DCR) listening test which is often employed in speech coding [16]. In this type of subjective testing, each listener is presented (we used high-quality headphones) with two audio segments at a time, one of which corresponds to the actual audio recording while the other one is the signal that was resynthesized with one of the PLC methods). Then, the listener is asked to grade the resynthesized audio signal compared to the original signal, using a 5-grade scale with the following description for the grades: 5 corresponds to "No quality degradation perceived" (compared to the original recording), 4 to "Quality degradation perceived but not annoying", 3 to "Quality degradation perceived and is slightly annoying", 2 to "Quality degradation perceived and is annoying", and 1 to "Quality degradation perceived and is very annoying". Considering the various probabilities of packet loss, for each listener 12 DCR tests were conducted.

The average results (and the 95% confidence intervals) of the DCR test are shown in Fig. 3 for the three choices of

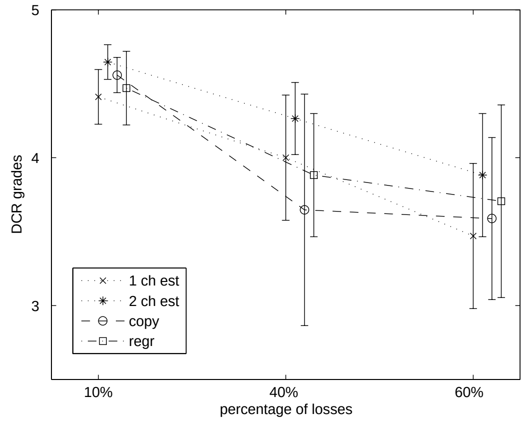

Figure 3: Results from the DCR listening test, for the four PLC methods.

packet loss. From these results we can conclude that indeed the cross-channel estimation method performs best, especially in higher percentages of packet loss, which is consistent with the results of Fig. 2. The figure is also consistent with the objective results of Fig. 2 in the sense that there is no clear advantage for one of the other three PLC methods (even for 60% losses, the three methods performed very close objectively). In general, we can conclude that the clear advantage of cross-channel estimation in the objective results was confirmed subjectively, whereas the other three methods performed almost equally well objectively and subjectively.

5 Conclusions

We presented a new approach for packet loss concealment for multichannel audio streaming, using our previously proposed multiband source/filter model for multichannel audio coding. The objective and subjective results showed that simple methods, such as repeating the last received frame when the current frame is lost, perform worse than more complex methods of missing frame estimation, mostly when realistic channel conditions were considered in our simulations and for high percentage of packet losses. The cross-channel estimation method which was introduced in this paper, was shown to result in the best PLC performance both objectively and subjectively. We are currently investigating the case when, in addition to lost LSF packets, some packets containing the reference signal might be lost as well. In this case, strategies that have been developed for loss concealment of monophonic audio can be applied to the reference audio channel directly, or to its LPC residual signal which is needed for resynthesis of the multiple channels using our model.

References

[1] J. D. Johnston and A. J. Ferreira, "Sum-difference stereo transform coding," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, (San Francisco, CA), pp. 569–572, 1992.

[2] J. Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding," in *Proc. 96th Convention of the Audio Engineering Society (AES), preprint No. 3799*, (Amsterdam), February 1994.

[3] F. Baumgarte and C. Faller, "Binaural cue coding - Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 509–519, November 2003.

[4] K. Karadimou, A. Mouchtaris, and P. Tsakalides, "Multichannel audio modeling and coding using a multiband source/filter model," in *Conf. Record of the Thirty-Ninth Asilomar Conf. Signals, Systems and Computers*, (Pacific Grove, CA), pp. 907–911, November 2005.

[5] J. Lindblom, J. Samuelsson, and P. Hedelin, "Model based spectrum prediction," in *Proc. IEEE Workshop on Speech Coding*, (Delavan, WI), pp. 117–119, September 2000.

[6] C. A. Rodbro, M. N. Murthi, S. V. Andersen, and S. H. Jensen, "Hidden Markov model-based packet loss concealment for voice over IP," *IEEE Trans. Speech and Audio Processing*, Accepted for future publication 2006.

[7] M. A. Kohler and R. K. Yarlagadda, "Markov chain prediction for missing speech frame compensation," in *Proc. IEEE Workshop on Speech Coding*, (Delavan, WI), pp. 75–77, September 2000.

[8] J. Lindblom, "A sinusoidal voice over packet coder tailored for the frame-erasure channel," *IEEE Trans. Speech and Audio Processing*, vol. 13, pp. 787–798, September 2005.

[9] C. Perkins and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, pp. 40–48, September/October 1998.

[10] S. Wabnik, G. Schuller, J. Hirschfeld, and U. Kraemer, "Packet loss concealment in predictive audio coding," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz, NY), pp. 227–230, October 2005.

[11] R. Sinha, C. Papadopoulos, and C. Kyriakakis, "Loss concealment for multichannel streaming audio," in *ACM Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, (Monterey, CA), pp. 1–10, June 2003.

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

[13] 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.

[14] A. Kain and M. W. Macon, "Spectral voice conversion for text-to-speech synthesis," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, (Seattle, WA), pp. 285–289, May 1998.

[15] M.Yajnik, S. Moon, J. Kurose and D. Towsley, "Measurement and modeling of the temporal dependence in packet loss," in *Proc. IEEE Conf. Computer Communications*, vol. 1, pp. 345–352, 1999.

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