Tutorial for ISIMP-2001

Recent Developments in Advanced Audio Processing

C.-C. Jay Kuo

Integrated Media Systems Center and Department of Electrical Enigneering-Systems
University of Southern California, Los Angeles, CA 90089-2564

Tutorial 1: High Fidelity Multichannel Audio Coding

When DVD and home theater systems become more popular these days, high fidelity multichannel (5.1 channel or 10.2 channel) audio systems are well received in the market. Compared with the traditional mono or stereo audio, multichannel audio requires a much more efficient coding scheme for its storage and transmission. This talk will present two new multichannel audio coding techniques: (i) the use of the Karhunen-Loeve Transform (KLT) to decorrelate signals of multiple channels; and (ii) the use of bit-layer coding to achieve an fully embedded bitstream. We exploit the inter-channel redundancy inherent in most multichannel audio sources, and prioritized the transformed channel transmission policy. Experimental results show that, compared with MPEG AAC (Advanced Audio Coding) algorithm, the proposed MAACKL (Modified Advanced Audio Coding with KLT) algorithm not only reconstruct better quality of the multichannel audio material at regular low bit rate of 64 kbits/sec/ch, but also achieves quality scalability in the single multichannel audio bit stream. The embedded multichannel audio coding system inherit the efficient inter-channel de-correlation block in MAACKL algorithm, and add an progressive quantization coding block and a context-based QM noiseless coding block. The final bit stream generated by this embedded multichannel audio coding system has a fully progressive property, which can terminate the encoding or decoding at any arbitrary point. Experimental results show that, compared with MPEG AAC algorithm, the reconstructed multichannel audio using the proposed algorithm achieves better performance not only with objective MNR (Mask-to-Noise Ratio) measurement, but also with subjective listening test at various bit rates.

Tutorial 2: Internet Audio Streaming with Adaptive Time-Scale Modification

Recent research efforts on Internet audio streaming have been focused on error control and delay concealment in the presence of delay jitter and packet loss. Given a fixed receiver buffer and a tight end-to-end delay bound, some packets sent to the receiver may still be discarded since the receiver buffer is adjusted to accommodate the average end-to-end delay. A delay spike happens when several consecutive packets are delayed and arrive at the receiver almost simultaneously. It happens when audio packets piled behind a large Internet load. In this work, we extend the silence interval-based adaptive playout algorithm by exploiting the time-scale modification scheme, focusing on fast adaptation to delay jitter and minimization of packet droppings at the receiver while maintaining a low buffering delay. Time-scale modification (i.e. contraction and/or expansion) modifies the time duration of an acoustic signal without changing of its acoustic attributes, such as pitch, timber, and so on. By applying a varying degree of stretching for each packet (although it is important to maintain the stretching factor within a talk-spurt), every packet could contribute in adapting to the network delay jitter/spike as well as packet loss.

For time-scale modification, the synchronized overlap-and-add (SOLA) scheme is adopted in our approach and modified into a fast and packet-oriented version. We manipulate time-scale modification based on the estimation of packet delay under the framework of adaptive playout. The time-scale modification factor will be estimated for each packet depending on the delay constraint, delay statistics, and the number of late-arrival packets. This estimated stretching is then bounded by certain upper/lower bounds that is calculated based on the speech contents. It is shown that the proposed adaptive playout mechanism is adaptive to the observed delay jitter and/or packet loss up to a certain degree.

Tutorial 3: Encryption, Authentication and IP Protection of Audio/Speech Data

The goal of audiovisual data protection should aim at protecting the content of the data, not the binary bitstream itself. Under this principle, faster algorithms can be developed by selectively encrypting only some part of the bitstream. We propose a selective encryption scheme for ITU G.723.1 speech coding. For systems that are not suitable for selective encryption, another type of fast encryption algorithm is needed. A novel audiovisual data encryption scheme base on multiple entropy coders will be presented. Modern digital media editing and processing technology allows high quality forgery to be created at a relatively low cost. Judging the authenticity of audiovisual data by human perception alone is not enough anymore. Computer-aided methods are increasingly needed for fail-proof message authentication. A speech data integrity algorithm is proposed to protect the integrity of speech content instead of the bitstream itself. Speech features relevant to the semantic meaning are extracted, encrypted and attached as the header information. The receiver decrypts feature values and compares them to features extracted from the received data. A digital audio watermarking scheme of low complexity is proposed as an effective way to deter users from misusing or illegally distributing audio data. We advocate the importance of the synchronization attack caused by casual audio editing or malicious random cropping, which is a low-cost yet effective attack to watermarking algorithms developed before. The proposed scheme is based on audio content analysis with the wavelet filterbank while the watermark is embedded in the Fourier transform domain. A blind watermark detection technique is developed to identify the embedded watermark under various types of attacks.

Tutorial 4: Technology for 3D Immersive Audio Synthesis

Recently, there has been a proliferation of 3D (or immersive) audio technologies intended for desktop computers. Many sound cards, multimedia speakers, video games, audio software, and CD-ROMs are marketed as having some sort of the 3D capability. In addition, a new technology called acoustic environment modeling has emerged. It combines the basic 3D technology with reverberation and other effects in order to simulate natural acoustic scenes. A 3D audio system has the ability to position sounds all around a listener. The sounds are actually created by loudspeakers (or headphones), but listener's perception is that sounds come from arbitrary points in space. This tutorial will address how 3D audio systems work. In particular, techniques such as HRTF measurements, binaural synthesis, crosstalk cancellation, reverberation algorithms and head-tracking for 3D audio by using loudspeakers will be described in detail.