IEEE WASPAA2007 Demo Session Program

Demo Session MD1

Monday, October 22   20:00-22:00, West Dining Room

[MD1-1] (Related Paper: ML2-2)

Multi Target Acoustic Source Tracking using Track Before Detect

Maurice Fallon (Department of Engineering, Cambridge University)

To effectively track the location of a moving acoustic source (a person speaking, generally) or sources using an array of microphones, it is necessary to update all prior information on-line using information garnered from the current audio stream to give a complete (posterior) picture of the overall situation at that time frame. Such an update can be carried out using Bayes' Theorem. Particle Filtering or Sequential Monte Carlo (SMC) is a point-wise numerical estimation technique which can be used to approximate such an update.

In this demonstration, we simulate target dynamics and then use the Steered Beamformer Function (SBF) of the current audio frame as a localising function for the particle filter. We will demonstration the on-line tracking performance of an Acoustic Source Tracking algorithm (using recorded data) for one and two sources and discuss the algorithm's limitations.

Demonstration Videos and more detail will also be available here:
http://www-sigproc.eng.cam.ac.uk/˜mff25/research.html

[MD1-2]

Optimal Sensor Placement for Passive Source Localization

Jan Neering (Ecole des Mines de Paris)

In order to optimize the performance of passive source localization a variety of closed-form and iterative procedures have already been proposed within the last four decades. Most of them are based on two step localization processes: in a first step the time delay of arrivals (TDOAs) between the sensors are measured. In a second step those TDOAs are then used to estimate the position of the source. While a variety of powerful methods for these two steps already exist and keep on appearing, very little attention has been given to the sensor configuration, yet. Therefore, this demonstration intends to point out the utility of a third, initial step: the optimal sensor configuration.

The main focus of this demonstration lies on optimally placing the sensors with respect to the condition number of a linear-least-squares estimator. While other objective (cost) function, such as the Cramer Rao Lower Bound, could also be used, this approach has some very useful properties. It optimizes the performance of a closed-form estimator, which is computationally very efficient. Further, no assumptions about the sensor noise needs to be made. The found sensor configuration will be robust with respect to sensor displacement.

[MD1-3] (Related Paper: ML2-3)

Blind Sparse Channel Identification for Time Delay Estimation and Speech Dereverberation

Yuanqing Lin (GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania), Jingdong Chen (Bell Laboratories, Alcatel-Lucent), Youngmoo Kim, (Electrical and Computer Engineering Department, Drexel University), and Daniel D. Lee (GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania)

The demonstration is to illustrate the applications of our newly proposal blind sparse channel identification algorithms in real acoustic environments. The applications include time delay estimation

and speech dereverberation. The related papers can be found at
http://www.seas.upenn.edu/~linyuanq/WASPAA2007.pdf
http://www.seas.upenn.edu/~linyuanq/NIPS2007.pdf

[MD1-4] (Related Paper: TP1-01)

Blind Source Separation for Underdetermined Convolutive Mixtures in a Real Environment

Shoko Araki (NTT Communication Science Laboratories, NTT Corporation)

We present a blind source separation demo which is related to the paper ”A Two-Stage Frequency-Domain Blind Source Separation Method for Underdetermined Convolutive Mixtures” by H. Sawada et al.. In the demo room, we record the utterances by three speakers with a consumer IC recorder which has two (stereo) microphones. Then we separate each speaker utterance from these two mixtures.

[MD1-5]

Pocket-Size Real-Time Blind Source Separation Module

Keiichi Osako, Hiroshi Saruwatari, and Shigeki Miyabe (Nara Institute of Science and Technology)

To realize hands-free system, we have developed real-time BSS system using pocket-size DSP module. This module can separate multiple sound sources in real-time, and can maintain high quality. We combine our SIMO-ICA and flexible post-processing including direction-of-arrival (DOA) estimation of each source. This allow us to apply a new omni-directional time-frequency masking to the sources with arbitrary DOAs. We are willing to show the real-time separation demo in the session.

[MD1-6] (Related Paper: ML1-1)

Spatial Audio Scene Coding for Improved Headphone Virtualization

Michael M. Goodwin and Jean-Marc Jot (Creative Advanced Technology Center)

Recently, we proposed a spatial analysis-synthesis framework which we refer to as Spatial Audio Scene Coding (SASC). In SASC, directional cues that describe an auditory scene are derived on a time-frequency basis where the scene could be a natural scene comprising various sound sources and reverberation, or could correspond to a multichannel audio signal. In the latter case, in contrast to standard spatial audio coding schemes, SASC parameterizes the spatial characteristics without reference to the channel format. This format-independent representation enables efficient compression of arbitrarily complex scenes, robust multichannel format conversion, and flexible high-quality modifications such as source manipulation and scene rotation.

In this demonstration, we use SASC to improve headphone virtualization. In standard headphone virtualization of stereo or multichannel recordings, channel-dependent interaural relationships based on head-related transfer functions (HRTFs) are imposed on each input channel in the binaural mix. SASC provides a new paradigm: the input content is analyzed for channel-independent time-frequency positional information, and the binaural signal is generated by applying appropriate HRTF cues to each time-frequency component. This enables a higher spatial resolution than in channel-centric virtualization methods: sources panned between channels are rendered with the correct HRTFs instead of a combination of the channel HRTFs.

[MD1-7]

Binary Area Matching: A Robust Search Method for Audio and Video Signals

Takayuki Kurozumi, Ryo Mukai, and Kunio Kashino (NTT Communication Science Laboratories, NTT Corporation)

Searching audio contents by an audio query is becoming an important function for information retrieval. This demonstration introduces our media search method called binary area matching (BAM). To show its robustness with respect to noise and distortion, we capture various music fragments played by a speaker and use them as queries. The system identifies the music title even if the query sounds are severely distorted. A video-matching version of the system is also presented, where you can identify the video title by a video fragment captured by a USB camera.

Demo Session TD2

Tuesday, October 23 20:00-22:00, West Dining Room

[TD1-1] (Related Paper: MP1-02)

Ad Hoc Microphone Array

Zicheng Liu (Microsoft Research)

Many people carry laptops and other devices to meetings. Most of these devices have built-in microphones and Wi-Fi cards. Therefore these microphones can easily form an array which we call ad hoc microphone array. The goal of this project is to leverage such ad hoc microphone arrays to improve meeting experience. Compared with conventional microphone arrays, the ad hoc microphone array has the advantage that the microphones are usually distributed and they are closer to the human speakers. But ad hoc microphone arrays present many new technical challenges including: the microphone gains are unknown, the array geometries are unknown, and the channels may not be synchronized. We have developed technologies to automatically estimate microphone gains and 2D locations of the microphones based on human speech signals. During the meeting, we estimate the 2D positions of the speakers. Such information, for example, can be sent to the remote participants to improve their meeting experience.

[TD1-2] (Related Paper: MP1-08)

Electronic Pop Protection for Microphones

Steven Backer (Sennheiser Research), Gary W. Elko, and Jens Meyer (mh acoustics)

An algorithm for pop-noise suppression in microphones will be demonstrated in real-time. A microphone typically used in conferencing and public address applications is connected to a custom DSP platform, which outputs two signals. One signal is unprocessed (simply amplified), and the other is processed using the pop-noise suppression algorithm. Demo attendees can listen to and compare both outputs over a loudspeaker (or headphones) for subjective evaluation of the algorithm. Technical details and an objective performance evaluation may also be presented.

[TD1-3]

Audio Object Processing using Independent Component Analysis

Shigeki Miyabe, Yuki Haraguchi, Hiroshi Saruwatari, Kiyohiro Shikano (Nara Institute of Science and Technology), and Toshiyuki Nomura (NEC Corporation)

We will demonstrate audio object processing using independent component analysis (ICA) for compressive coding and remixing of multichannel audio signals. It is difficult to apply ICA to broad-band signal separation (blind source separation; BSS) when the sound source components of the audio signal is more than the reference signals, referred to as under-determined source separation. However, in each of the narrow bands, the number of the source components are smaller than that of the sound sources because of the signal characteristics so-called sparseness. Thus in each narrow band, frequency-domain ICA can separate audio components composed of large number of sources into each of source components. Such feature of ICA is applicable to many kinds of audio processings.

First, we will demonstrate multichannel audio coding based on ICA. ICA can separate source components in each narrow band even in the under-determined cases. This separation can clarify sparseness among sources, which is an important feature of audio coding. It will be shown that the process of ICA improves quality of our proposed audio coding.

Next, we will demonstrate the processing of source localization in stereo audio. Localization of each source is controlled as if the user can manipulate the audio console.

[TD1-4]

Spatial Sensitivity for Listening Room Compensation

Stefan Goetze (Dept. of Communications Engineering, University of Bremen), Markus Kallinger (Fraunhofer Institute for Integrated Circuits), Alfred Mertins (University of Luebeck), and Karl-Dirk Kammeyer (Dept. of Communications Engineering, University of Bremen)

In many applications for room acoustics the distortions introduced by reverberation due to numerous signal reflections at the room boundaries have to be compensated or reduced. Filters for listening room compensation (LRC) are capable of reducing reverberation effects but are often designed for exactly known source- and microphone locations. Especially for a hands-free teleconferencing setup this assumption may be violated. This contribution demonstrates the influence of spatial errors on listening room compensators. Different LRC designs (including least squares equalizer and room impulse response shortener) are evaluated for different room positions, room reverberation times and filter orders by means of objective measures (D50, variance of equalized system, etc.). Thus, numerous conditions can be analyzed and the equalization results are presented visually and acoustically.

[TD1-5]

The Flexible Audio Extension for the IVS_VDT Virtual Reality System and Its Different Uses

Carlo Belardinelli (Virtual Development and Training Center, Fraunhofer IFF)

Aim of this demonstration is the presentation of VR platform audio core function as implemented by the Fraunhofer IFF as well as its capabilities. In particular, since our implementation correctly spatializes sound we will show how the scenario visual representation is enriched by 3D audio effects. This results in an improvement in the overall immersivness and efficiency of the VR rendering. All the different features of digital audio signal processing will be introduced and practically showed both on the authoring and on the user side.

The different scenarios provide examples about possible interactions with sound in VR, giving the perceiver important information about the surrounding environment as well as about the events taking place.

Finally, an overview on the most recent developments of our platform in the field of Active Noise Control and Active Noise Reduction in industrial environment will be offered. It will be showed how VR can be a valuable resource to test and validate DSP and manufacturing engineering techniques by means of the visual display of the sound field within 3D environments.

[TD1-6] (Related Paper: TL2-4)

MPEG-4 Enhanced Low Delay Audio Coding

Jürgen Herre and Markus Schmidt (Fraunhofer IIS Erlangen)

MPEG-4 Enhanced Low Delay AAC (AAC-ELD) is currently under standardization at ISO/MPEG. It has been designed with the aim of using low delay perceptual audio coding in the context of high quality communication. It also has been the subject of WASPAA paper #0023 ("Low Delay Filterbanks for Enhanced Low Delay Audio Coding"). The presented technology combines the characteristics of MPEG-4 AAC Low Delay (AAC-LD) and the spectral band replication (SBR) technique and, at the same time, optimizes the delay properties of both these techniques. This demo presents the audio quality of this codec.

By means of a diverse set of example audio files the strengths of AAC-ELD are demonstrated and can be experienced by the listener. The following established coding techniques are used as reference: AAC-LD, ITU-T G.722.2 (AMR-WB) and ITU-T G.722.1-C. The audio content is presented via headphones to allow the best listening experience. With the click of a button the listener is able to switch between the different coding algorithms and compare them directly.

[TD1-7]

A Musical Audio Search System Based on Self-Similarity – SAME: Self-Similarity Audio Structure Matching and Exploration –

Tomonori Izumitani, Ryo Mukai, and Kunio Kashino (NTT Communication Science Laboratories, NTT Corporation)

We demonstrate a musical audio search system based on audio signal matching. The system is able to search a database with queries that are played in different keys from the corresponding titles in the database. This kind of search is very inefficient with conventional musical audio search methods using audio fingerprinting due to distortion of spectral characteristics by key variations. The proposed system utilizes the self-similarity structure of musical audio signals. The self-similarity at a time point is defined by spectral similarities to different lags. This structure is applicable for audio matching because the relationship between two time points tends to be retained even when the music is played in a different key. In our experimentation, for example, the proposed method yields a precision and recall rates of around 0.75 even when the pitches in queries and stored signals differ from each other by seven semitones.

[TD1-8]

Simulation of Recruiting Hearing Impairment Using a Tree-Structured Allpass-Complementary ERB-Band Filter Bank

Ryan J. Cassidy and Julius O. Smith III (CCRMA, Stanford University)

The simulation of hearing impairment provides a useful tool for clinicians and researchers working to ameliorate such a condition. Prior work in this vein, which has focused primarily on speech listening, has examined a variety of phenomena associated with hearing impairment, including increased threshold, reduced dynamic range, and reduced frequency selectivity, to name a few. In this work, we focus primarily on the task of music listening. In particular, we present an auditory filter bank design based on a tree-structured, allpass-complementary approach. The overall analysis-synthesis system employed, in addition to having negligible magnitude distortion in the absence of modifications, matches ERB analysis bandwidths obtained from a popular Gammatone filter bank design with high precision. The bandwidth match is compared favorably to that of a prior hearing loss simulation technique. In addition to providing an accurate simulation of hearing impairment, the system can be implemented with reasonable latency.