AES 34th INTERNATIONAL CONFERENCE

New Trends in Audio for Mobile and Handheld Devices

Jeju Island, Republic of Korea

August 28–30, 2008

Technical Sessions*

**This preliminary program is accurate as of press time.*

Thursday, August 28 9:30 am

KEYNOTE SPEECH

Service Innovation by KTF—*Byungki Oh*, KTF Co., Seoul, Korea

KTF, one of the biggest mobile carriers in Korea, provides the SHOW third generation (3G) services that intend to deliver not only voice calls but also visual communication features such as video calling and data services. Details of the current SHOW 3G services and expectations of future services are discussed.

Thursday, August 28 10:00 am

INVITED PAPERS

DisplayPort: Digital Multimedia Display Interface Standard for PCs and Mobile Devices—*Kewei Yang*, Analogix Semiconductor

The newly-developed DisplayPort standard is intended to provide an interface suited to a very wide range of applications, including both external ("box to box") and internal (e.g., notebook PC panel interface) connections. With the widespread industry support, DisplayPort will replace just about every major digital multimedia display interface for PCs that are in existence today. Previous display interfaces, such as DVI, LVDS, and VGA will no longer be included on displays in the future, as the PC industry vision is to have ultra-sleek displays with just a DisplayPort connector on the back of the PC. DisplayPort allows high-definition digital audio to be available to the display device over the same cable as the digital video signal. It delivers true plug-and-play with robust interoperability. When the optional content protection capability is active, DisplayPort will support viewing high definition television, video, and audio types of protected content. This paper provides an overview of the proposed standard and its basic technical details.

ICA: Independent Component Analysis and its Appliction to Mobile Audio— *Te-Won Lee*, Qualcomm

Thursday, August 28 11:00 am – 12:00 noon

PAPER SESSION 1: SIGNAL PROCESSING, PART 1

1-1 Discrimination of Music Signals for Mobile Broadcasting Receivers—*Myungssuk Song, Hong-goo Kang*, Yonsei University, Seoul, Korea

This paper proposes a Gaussian mixture model (GMM)-based music discrimination system for mobile broadcasting receivers. The objective of the system is to automatically archive music signals from audio broadcasting programs that are normally mixed with human voices, acoustic noises, commercial advertisements, and so on. To enhance the robustness of the system performance and to sharply cut the starting/ending-point of the recording, we also introduce a post-processing module whose features consist of signal duration, energy dynamics, and local variation of feature statistics. Experimental results to various input signals verify the superiority of the proposed system.

1-2 Non-Linear Signal Processing for Low Frequency Enhancement—*Pauli Minnaar*, AM3D, Aalborg SV, Denmark

A lot of attention has been directed at designing various sounds that are treated as noise, such as automobile acceleration sounds and vacuum cleaner sounds. The reason is that the idea of sound being a normal part of product operation has permeated society. We focused on sound design and evaluated it for 11 kinds of button sounds. First, an impression was extracted by the semantic differential (SD) method, and the relevance with a time frequency analysis was investigated. Next, we confirmed whether or not the impression changed when a sound that generated a bad impression was processed using adaptive control into a sound that generated a good impression.

Thursday, August 28 1:30 pm – 3:00 pm

PAPER SESSION 2: CODING FOR AUDIO AND SPEECH, PART 1

2-1 Framework for Unified Speech and Audio Coding —*Eunmi Oh, Miyoung Kim*, Samsung Advanced Institute of Technology (SAIT), Suwon, Korea

The purpose of this study is to propose a framework of unified speech and audio coding that can compress speech and music equally well, and then to verify the feasibility of a highly efficient low-rate coding scheme. In this paper a coding scheme is introduced by utilizing flexible time and frequency representation of a filter bank called Frequency Varying Modulated Lapped Transform (FV-MLT). The time/frequency resolution of FV-MLT is determined by psychoacoustic model. The output of the filterbank is quantized by considering rate/distortion optimization. The high temporal resolution coding tool can be used depending on the characteristics of input signal.

2-2 Using Salient Envelope Features for Audio Coding—*Joachim Thiemann, Peter Kabal*, McGill University, Montreal, Quebec, Canada

In this paper we present a perceptual audio coding method that encodes the audio using perceptually salient envelope features. These features are found by passing the audio through a set of gammatone filters and then computing the Hilbert envelopes of the responses. Relevant points of these envelopes are isolated and transmitted to the decoder. The decoder reconstructs the audio in an iterative manner from these relevant envelope points. Initial experiments suggest that even without sophisticated entropy coding a moderate bit rate reduction is possible while retaining good quality.

2-3 **Personalized Music Service Based on Parametric Object-Oriented Spatial Audio Coding**—*Yangwon Jung, Hyun-o Oh*, LG Electronics, Seoul, Korea

From the development of spatial audio coding such as MPEG Surround, the concepts of object-oriented spatial audio coding emerged, and now, efforts are made for the standardization of MPEG SAOC (Spatial Audio Object Coding). The key target applications of MPEG SAOC were suggested as backward compatible/interactive re-mix, and teleconferencing/telecommunications, and gaming. Among those, we pay attention to the interactive re-mix applications and propose personalized music service systems.

Thursday, August 28 3:00 pm – 4:00 pm

POSTERS

P-1 A Bit Reduction Algorithm for Spectral Band Replication Using the Masking Effect—
Sang Bae Chon,[1] *Mingu Lee*,[1] *Hee-Suk Pang*,[2] *Koeng-Mo Sung*[1]
[1]Seoul National University, Seoul, Korea
[2]Sejong University, Seoul, Korea

Spectral Band Replication (SBR) is a state-of-the-art technology to enhance audio or speech codecs especially at low bit rates based on harmonic redundancy in the frequency domain. With SBR, it is possible to generate high frequency components of a full-band audio signal with a bit rate of a few kbps. In this paper a bit reduction algorithm for SBR is proposed using the Masking Effect and threshold in quiet. The proposed algorithm reduces the SBR bit rate by modifying the envelope data of SBR so that the modification cannot be perceived in the subjective sense. Experiments show that the proposed algorithm achieves about 10-to-12 percent bit reduction for the envelope data of SBR based on the 3GPP Enhanced aacPlus codec at the bit rate of 24 kbps with no perceptive sound quality degradation.

P-2 Bit-Rate Reduction Using Efficient Difference Coding of Sinusoid Amplitude—*Nam Suk Lee,* Samsung Electronics Co., Ltd., Republic of Korea

In this paper we develop new method for sinusoid amplitude coding. In MPEG-4 SSC (SinuSoidal Coding, parametric coding for high quality audio), audio signal is analyzed by transients, sinusoids, and noise. The frequency, amplitude, and phase of each sinusoid are extracted. Then, each sinusoid of current frame is connected with each sinusoid of the previous frame using similarity of the frequency, amplitude, and phase. Sinusoid connected to sinusoid of the previous frame is continued and the sinusoid that is not connected to the sinusoid of previous frame is born. Continuations and Births are quantized and coded by Huffman entropy coding method. In our paper we suggest a new method of sinusoid amplitude in birth. Bit-rate is reduced using this method. In SSC parametric codec implemented by Samsung Electronics, bit-rate is reduced by 15.89 percent of sinusoid amplitude in birth.

P-3 Implementation of 3-D Sound Using Grouped HRTF—*Bo-Kug Seo, Il-Hyun Ryu, Hyung-Tai Cha*, SoongSil University, Seoul, Korea

Head Related Transfer Function (HRTF) databases, including the information of the sounds arriving at the ears, are generally used to make the 3-D sound. The convolution of the HRTF with the original sound is a ➥

general method for sound image localization. However, to use the non-individual HRTF can cause confusion in perception of the directions of the source and a degrading the moving sound effects due to each listener's unique characteristics. In this paper we propose a new HRTF method that will generate a 3-D-sound by grouping and averaging the HRTFs existing in the vicinity of the direction of localization. The MOS (Mean Opinion Score) test results show that the proposed method is much better than the conventional methods in both sound localization characteristics and moving sound effects.

P-4 An Improved Weighting Curve Based on Equal-Loudness Contour—*Inseok Heo, Koeng-Mo Sung*, Seoul National University, Seoul, Korea

If the signals the loudness of which is to be determined are pure tones or narrow band noise, it is sufficient to measure their frequencies and sound pressure levels; from this data the loudness level can be determined by using the equal-loudness contour. An equal-loudness contour is a measure of sound pressure, over the frequency spectrum, for which a listener perceives a constant loudness when presented with pure steady tones. However, more convenient are sound level meters containing a filter that imitates more or less the shape of these curves. There are some curves defined in the International standard IEC61672:2003 for measuring such a filter. Among those curves, A-weighted curve and ITU-R 468 weighting are commonly used these days. Both curves are based on the inverse shape of Equal-loudness contour—whose standard curve is revised as defined in ISO 226:2003. In this paper an improved equation of weighting curve is discussed. Compared to pre-existing weighting curves, more similarities are presented between proposed method and the inverse shape of standard curve (ISO 226).

P-5 Low Carrier Frequency Noise-Shaper for Digital Amplifier—*Park Kyoungsoo*,[1] *Koeng-Mo Sung*[2]
[1]NeoFidelity, Inc., Seoul, Korea
[2]Seoul National University, Seoul, Korea

Noise-shapers have not been a severe concern in digital amplifiers that convert PCM data to PWM digitally because other factors—clock jitter, power supply regulation, and power stage linearity—are more dominant to get an audiophile quality SNR. A simple noise-shaper with order of a few is good enough for most audio applications if the over sampling ratio is 8 and the sampling frequency is 48 kHz, which is the most typical value in real chip manufacturing. For more power sensitive applications, the PWM carrier frequency or over sampling ratio can be reduced to enhance power efficiency by lowering the switching loss of power stage, which is one of the most important sources of power loss. Lowering switching loss also makes it possible to increase maximum attainable power output with fixed power supply voltage, which is another concern in mobile applications. This can make, however, high frequency noise smear more to audio band from the low over sampling ratio and, therefore, a more crafty design of noise-shaper is required to overcome the artifacts. A low over sampling ratio noise shaper can be designed by spreading the zeros of noise transfer function optimally with parameters of sampling frequency, over sampling ratio, requantization resolution, and noise-shaper order. This technique can use up the noise within the band budget given that a minimum SNR requirement for an application is defined. Optimal or marginal noise shaper design also makes it necessary to consider nonlinearity at a small signal input. Proper dithering is investigated to avoid that nonlinearity at small signal input keeping just-fitting SNR with limited carrier frequency.

Thursday, August 28 4:00 pm – 5:00 pm

PAPER SESSION 3: IMPLEMENTATION

3-1 Smooth PCM Clipping of Audio—*Mohammed Chalil, Pushparaj Dominic, Deepak Raghtham*, Analog Devices, Bangalore, India

The PCM clipping problem with respect to audio is discussed in this paper. Two solutions to avoid this problem is discussed and results are compared.

3-2 Advanced Terrestrial DMB System Structure for Multichannel Audio Services—*Hyun Wook Kim, Kyungsun Cho, Han Gil Moon*, Samsung Electronics, Suwon, Korea

An advanced Terrestrial Digital Multimedia Broadcasting (T-DMB) system structure is proposed for multichannel audio services. Multichannel audio signals are divided into stereo audio signals and additional multichannel side information. Each audio signal and multichannel side information is transmitted through each elementary stream, respectively. The advanced T-DMB system mainly proposes a separated OD Stream and protected ES_DescriptorUpdate command for conveying multichannel side information. The first OD Stream for stereo audio signals can convey a command for a new object descriptor and a second OD Stream uses a protected command for additional multichannel side information.

Thursday, August 28 5:00 pm – 6:00 pm

INDUSTRIAL SOLUTION INTRODUCTIONS

Friday, August 29 9:00 am

INVITED PAPER

Nokia's View on Mobile Audio Development—*Juha Backman*, Nokia, Finland

Audio has become an increasingly important component of wireless handsets. From the viewpoint of a leading mobile phone company, we will explore future directions of some aspects of hardware including electronics, loudspeakers, microphones, signal processing, connectivity, services, and contents.

Friday, August 29 9:30 am – 10:30 am

PAPER SESSION 4: EVALUATION AND TESTING

4-1 Impression Evaluation Model for Button Sounds Using a Neural Network—*Gen Onishi*,[1] *Shunsuke Ishimitsu*,[2] *Koji Sakamoto,*[3] *Takayuki Arai*,[3] *Toshikazu Yoshimi,*[4] *Yuichi Fujimoto,*[4] *Kenichi Kawasaki*[4]
[1]Hiroshima International University, Hiroshima, Japan
[2]Hiroshima City University, Hiroshima, Japan
[3]University of Hyogo, Hyogo, Japan
[4]Pioneer Corporation, Saitama, Japan

This paper presents an impression evaluation model for button sounds generated when users press the buttons on car audio equipment using a neural network. The dynamic characteristics of 11 kinds of button sounds obtained by their wavelet transform frequencies and sound pressure values are fed into the network model inputs. The model then responds with three factor scores, “esthetic,” “force,” and “metallic,” and an evaluation value of “offensive–pleasant” as the outputs. By analyzing the inside functions of the neural network after training, we confirmed the model acquired a mechanism that extracted four impression evaluation values from the sound characteristics, thus showing the model could attain automation of button sound design.

4-2 **A Study of Evaluating the Button Sounds Using Wavelets**—*Shunsuke Ishimitsu,*[1] *Koji Sakamoto*,[1] *Toshikazu Yoshimi,*[2] *Yuichi Fujimoto,*[2] *Kenichi Kawasaki*[2]
[1]Hiroshima City University, Hiroshima, Japan
[2]Pioneer Corporation, Saitama, Japan

A lot of attention has been directed at designing various sounds that are treated as noise, such as automobile acceleration sounds and cleaner sounds. The reason is that the idea of sound being a normal part of product operation has permeated society. We focused on sound design and evaluated it for 11 kinds of button sounds. First, an impression was extracted by the semantic differential (SD) method, and the relevance with a time frequency analysis was investigated. Next, we confirmed whether or not the impression changed when a sound that generated a bad impression was processed using adaptive control into a sound that generated a good impression.

Friday, August 29 11:00 am – 12:30 pm

PAPER SESSION 5: CODING FOR AUDIO AND SPEECH, PART 2

5-1 **Introduction to the OpenCORE Audio Components Used in the Android Platform**—*Javier Tapia, Jim Kosmach, Dusan Veselinovic, Greg Sherwood, Ralph Neff*, PacketVideo Corporation, San Diego, CA, USA

Audio and speech codecs such as MP3, AAC, and AMR are used extensively on mobile devices throughout the world. In the ideal case, such codecs rely on hardware acceleration. However, it is also very common to see software audio codecs running on the main application processor, which is often an ARM core processor. Such codecs must be memory efficient, processing cycle efficient, portable to multiple operating systems, robust to data loss, and must also have a modular interface. In this paper we introduce the OpenCORE multimedia framework and associated optimized audio codecs which are a part of the Android platform. We show how these components meet the challenging requirements for use in mobile devices. The OpenCORE audio components are currently available from the Open Handset Alliance as part of the Android SDK, and the source code for these components is scheduled for release in late 2008. The components are thus freely available for use in mobile device projects, and for non-mobile projects as well.

5-2 **An Efficient Forward Prediction Order Selection**—*Choong Sang Cho,*[1] *Je Woo Kim,*[1] *Seok Pil Lee,*[1] *Byeong Ho Choi*,[1] *Jin Ah Kang,*[2] *Hong Kook Kim*[2]
[1]Korea Electronic Technology Institute (KETI), Kyunggi, Korea
[2]Gwangju Institute of Science and Technology (GIST), Gwangju, Korea

Recently, the users of multimedia products required high quality audio services. Such services can be realized by lossless audio coding of multichannel audio. The MPEG-4 audio lossless coding (ALS) is a good candidate for doing this. However, a proper forward prediction order selection is crucial to the performance of MPEG-4 ALS since the compression ratio of MPEG-4 ALS highly depends on the prediction order. In this paper we propose a new method that can minimize the total bit rate with respect to the prediction order. That is, the proposed method estimates a prediction order where the bit rate increment due to the prediction order becomes higher than the bit rate reduction of the residual encoding, resulting in the increase of the total bit rate. For this end, we first compute the mean squared errors (MSEs) of residual corresponding to two successive prediction orders, and then estimate the prediction order by comparing the relative reduction in MSEs with a predefined threshold that corresponds to the bit rate for the quantization of a reflection coefficient. The performance of the proposed method is evaluated by measuring the compression ratio. As a result, it is shown from the experiments that the proposed method has better compression ratio than the normal method in MPEG-4 ALS and also it has a comparable compression ratio to the adaptive method in MPEG-4 ALS.

5-3 **Segmented Dimensionality Reduction Coding on Frequency Domain Signal**—*Minje Kim, Seungkwon Beack, Taejin Lee, Daeyoung Jang, Kyeongok Kang*, ETRI (Electronics and Telecommunication Research Institute), Daejeon, Korea

This paper proposes schemes of compressing frequency domain acoustic signals using dimensionality reduction methods. Dimensionality reduction methods that work on a two-dimensional matrix usually result in high compression ratio since they not only allow us to represent the input matrix with smaller amount of data, but exploit intrinsic information of the original data. Frequency domain signals can be seen as a (number of frequency bands) X (number of total frames) input matrix of dimensionality reduction methods. However, in this case, real-time encoding is not possible and encoder-side delay is inevitable, which amounts to the length of the whole input signal. To minimize the delay this paper proposes a coding scheme that conducts multiple dimensionality reduction on segments of input data frames serially.

Saturday, August 30 9:00 am – 10:30 am

PAPER SESSION 6: SIGNAL PROCESSING, PART 2

6-1 **A Perceptual Measure for Audio Source Eeparation**—*Mingu Lee,*[1] *Insuk Heo*,[1] *Nakjin Choi*,[2] *Keong-Mo Sung*[1]
[1]Seoul National University, Seoul, Korea
[2]Samsung Electronics Co., Ltd., Korea

In this paper an improved method on evaluating the performance of blind audio source separation is discussed. Based on Vincent’s measures, defined by SDR, SIR, SNR, and SAR, several well-known psychoacoustic characteristics, such as equal-loud- ➥

ness contours and masking effect, are considered to make these measures more relevant to human auditory system. Correlation between the results of the proposed methods and those of carefully designed listening tests is presented for verification.

6-2 Tiny DSP: DSP Core, Algorithm Development, and "Device Mastering"—*Nathan Bentall, Peter Eastty, Duncan Stott*, Oxford Digital Ltd., Stonesfield, Oxfordshire, UK

Market expectations of small-size, low-cost, and, in many cases, a requirement for very low power create difficult challenges in the electric and acoustic design of consumer devices that include loudspeakers such as mobile phones, laptop computers, and flat panel displays. Developments in the field of "sound improvement" algorithms can go a long way to improving the listening experience, often making an attempt at acoustic correction, but even on high ticket-price items, fierce competition results in very high sensitivity to component cost, which can rule out many DSP devices. Rapid development is mandatory due to development cost and large variety of new models. A combination of processor design and tool set are described, which simultaneously address these issues; an implementation of a commercially-available device is described; an example usage is outlined for improvement of a consumer-device; a real time parameter adjustment tool is presented which enables real-time tuning of developed algorithms to facilitate the ultimate aim: a combination of low cost DSP and algorithm that can be rapidly individually tailored to any given device—a process described here as "Device Mastering."

6-3 Simple High-Band Extension Method Using Wavelet for Mobile Devices—*Sang-keun Oh*, LG Electronics, Seoul, Korea

This paper suggests the simple higher-frequency band extension (HBE) method based on digital wavelet transform (DWT). The method estimates the band missed through the perceptual coding process by using DWT filters. As the input signal for processing, we used the band passed signal wave (f) including the missed band that is used as the approximation coefficients. Then the method synthesizes the band passed signal wave (f) using DWT and reconstructs the output signal. In order to estimate the missed band from the approximation coefficients, we designed the FIR-type inverse filter of DWT analysis filter in integer numbers. We expect the method can behave in real time for variable types of mobile devises such as mobile handsets and MP3 players using ARM9 core as their processing units.

Saturday, August 30 11:00 am – 12:30 pm

PAPER SESSION 7: 3-D AUDIO AND SYNTHETIC AUDIO

7-1 Robust Crosstalk Cancellation Based on Energy-Based Control—*Young-Cheol Park, Junho Lee, Dae-Hee Youn,* Yonsei University, Seoul, Korea

This paper presents a robust crosstalk cancellation algorithm based on controlling acoustic energy density. Since it is known that energy density in a space is more uniform than squared pressure, the algorithm is useful for providing wider zone of crosstalk cancellation than is obtained by controlling the squared pressure. In the approach in this paper, both the acoustic velocity and the pressure are minimized. This yields a robust crosstalk cancellation in the vicinity of error sensor location. Simulation results confirm that the use of proposed algorithm improves zone of crosstalk cancellation, which contributes to the robustness in relation to movement of listener.

7-2 Designing Low-Dimensional Interaction for Mobile Navigation in 3-D Audio Spaces—*Till Schäfers, Michael Rohs, Sascha Spors, Alexander Raake, Jens Ahrens*, Technical University of Berlin, Berlin, Germany

In this paper we explore spatial audio as a new design space for applications such as teleconferencing and audio stream management on mobile devices. Especially in conjunction with input techniques using motion-tracking, the interaction has to be thoroughly designed in order to allow low-dimensional input devices, like gyroscopic sensors, to be used for controlling the rather complex spatial setting of the virtual audio space. We propose a new interaction scheme that allows the mapping of low-dimensional input data to navigation of a listener within the spatial setting.

7-3 A Consonance-Maximization Tuning Algorithm in Equal-Temperament Synthesized Tones—*Sang Bae Chon,*[1] *Sang Ha Park*,[1] *Ah Jin Yim*,[2] *Koeng-Mo Sung*[1]
[1]Seoul National University, Seoul, Korea
[2]Ewha Woman's University, Seoul, Korea

There are many different temperament tuning theories such as "Pythagoras tuning," "just intonation," and "equal temperament." Among them, the equal temperament is the most popular one for its key compatibility in spite of the general dissonance from its harmonic structure. That dissonance is the reason that no other chords except octaves sound perfectly harmonic in equal temperament tuning. We propose a tuning system for sound synthesis using equal temperament that maximizes the consonance of the chords. By removing dissonance (such as in beating and roughness) and rearranging the harmonic structure, it is possible to achieve more consonant sound, which is only possible in computer-based sound synthesis. To verify the performance of the proposed algorithm, listening tests were carried out with the most popular chords, i.e., perfect 5th, major 3rd, and major 6th, on synthesized piano, violin, and flute sounds. They confirmed that the modified major 3rd and major 6th chords using the proposed algorithm sounded more harmonic than the original ones. We believe that the harmonicity and consonance can be improved in commercial applications such as ring tone synthesis.

Saturday, August 30 1:30 pm – 3:00 pm

WORKSHOP

Audio in the IT Industry