Chronicle

The XV Symposium New Trends in Audio and Video Technology Wrocław, Poland, September 25 – 27, 2015

In 2014 the XV Symposium New Trends in Audio and Video organized by the Polish Section of Audio Engineering Society under the honor patronage of the National Broadcasting Councils (KRRiT) and Acoustical Committee of Polish Academy of Science was held in Wrocław in September. The main topics of the conference covered mostly all domains of audio and video engineering, i.e. the sound visualization, measurement technique, perception and quality assessment, data reduction in audio and video signals, sound reinforcement, mixing and mastering technology, and many others. The extra attention has been paid for the problems of loudness of audio programs in radio and TV broadcasting. Over 50 people from different branches of audio and video technology from Poland as well as from abroad participated this Symposium and shared their knowledge and experiences during the paper sessions, technical tours, workshops and special presentations performed by Dolby Poland on the field of Object Audio in wide spectrum. Below there is a selection of abstracts of the papers presented at this event.

Abstracts

Object-based data compression for massive multichannel audio

Bartkowiak Maciej, mbartkow@multimedia.edu.pl
Wereszczyński Adam

Poznań University of Technology
Poznań, Poland

A new data compression scheme is considered for multichannel audio signal that typically represents a dense sound field. The technique exploits similarities between channels by jointly encoding "objects" extracted from them. For this purpose, dominant energy sources are identified in the spatio-temporal domain, separated from channel data, and encoded jointly, with taking care of appropriate temporal and spectral alignment between channels. The remaining residua signal is encoded in each channel. Experimental results show a significant efficiency increase w.r.t. encoding individual channels without extracting objects. Applications include 3D audio and offline Wave Field Synthesis systems.

⋆ ⋆ ⋆

Remote measurement of the parameters of a professional mixing audio console

Białczyk Andrzej, andrzej.bialczyk@tvp.pl

TVP S.A. (Polish Public Television)
Warszawa, Poland

One of the goal of the technical control section in TVP in Warszawa are video and audio measurements of the TVP equipments and transmission lines. Most of the measurement is performed on site in Warsaw. Last year studios and control room in regional station in Kraków were modernized. New digital mixing console CRESCENDO with NEXUS audio routing system was installed. In studios in Warszawa another digital mixing consoles AURUS with NEXUS routing system is implemented. Functionally this consoles are different but the same type of the modular electronic cards are used. All regional TVP station are using A/V contribution/distribution system built on the occasion switching analog to digital terrestrial televisions. This system is capable to distribute compressed video signal with embedded compressed and uncompressed PCM audio. The basic idea was to used this transmission lines for remote measurement of the high quality mixing console with state of the art A/D and D/A converters and digital I/O interfaces. Comparison of the measurement in studio in Warsaw and remotely made from Warsaw in studio in Cracow is described.

⋆ ⋆ ⋆

Examining influence of video framerate and audio/video synchronization on audio-visual speech recognition accuracy

Bratoszewski Piotr, bratoszewski@sound.eti.pg.gda.pl
Łopatka Kuba,
Czyżewski Andrzej

Gdańsk University of Technology
Gdańsk, Poland

The problem of video framerate and audio/video synchronization in audio-visual speech recognition is consid-

ered. The visual features are added to the acoustic parameters in order to improve the accuracy of speech recognition in noisy conditions. The Mel-Frequency Cepstral Coefficients are used on the acoustic side whereas Active Appearance Model features are extracted from the image. The feature fusion approach is employed. The initial video framerate equals 100 frames per second. The test signals were recorded with a specialized hardware for synchronous registration of audio and video data. In a practical implementation, however, it is difficult to achieve a high rate of images per second and maintain the precise audio/video synchronization. Therefore, in this work it is assessed, how the lowered framerate and lack of synchronization between audio and video data impairs the performance of the recognition engine. The lowered video framerate is enforced by downsampling the visual data. The lack of synchronization is simulated programmatically in the feature fusion process. The experiments are conducted employing the HTK engine (Hidden Markov Toolkit). Word Error Rate, correctness and accuracy measures are considered and a small dictionary of 11 words (numerals) is employed.

★ ★ ★

Selected aspects of level monitoring in public address systems

Dziechciński Paweł, pawel.dziechcinski@pwr.edu.pl

Wrocław University of Technology
Wrocław, Poland

In this paper selected aspects of the level monitoring of acoustic and audio signals are presented. These issues are discussed mainly in relation to public address systems. Information presented in this study may also be useful for other applications connected with control of the signal level. This paper is focused mainly on the measurement of signal peaks. Loudness and volume unit meters are not discussed. A great part of this paper is devoted to the crest factor of audio signals. As shown in the work the very important think in designing public address systems is a knowledge of designers about characteristics of acoustic signals. Lack of such knowledge may lead to realizations of systems that either do not meet requirements, or the costs of their implementation are greatly overestimated.

★ ★ ★

Developing a system for melody line tracking and vowels recognition in monophonic a capella singing

Gołębiewski Radosław
Łukasik Ewa, ewa.lukasik@cs.put.poznan.pl

Poznań University of Technology
Poznań, Poland

The problem of monophonic a capella singing analysis writes into the domain of sung speech recognition. The audio files usually analyzed are popular music, mono- and polyphonic music or singing interleaved with instrumental music. Therefore the tasks are rather complicated and usually adapt HMM based methods of speech recognition to

the singing voice. The related research usually concentrates on: – phonemes and words recognition in music sound files, – automatic alignment between lyrics and a melody. The problem undertaken in this paper is somehow simplified as it concerns Gregorian chant, which is sung in Latin and the text is only necessary to detect moments of pitch changes (especially in melismas) or phoneme changes when several syllables are sung on one note. Although the recognition of the full text would be beneficial, for the prototype application we decided to recognize only vowels in short-length consecutive windows. The results are aligned with the pitch track. A prototype system was tested on the Gregorian chant exemplary interpretations and will be further developed.

★ ★ ★

Subjective perception of music genres in the field of music information retrieval systems

Hoffmann Piotr, phoff@sound.eti.pg.gda.pl
Kostek Bożena, bokostek@audioacoustics.org

Gdańsk University of Technology
Gdańsk, Poland

The aim of this paper is to evaluate the relationship between perception of music genres and subjective features of music that can be assigned to them. For this purpose a group of subjective features such as loudness, melody, rhythm, volume, instrumentation was chosen to describe music genres. A group of 30 listeners with normal hearing, ranging from 20 to 40, was created. Each subject participating in listening tests was asked to choose perceptual features of music which best correspond to the given music genre. Music genres were limited to six most typical ones for this group of listeners, i.e.: Classical, Electronic, Jazz, Pop, Rap, Rock. On the basis of the listeners' answers the statistical V-Cramer test was conducted to find the correlation between evaluated features and music genres. Identification of potential features that may be important for music recognizing can result in a more efficient automatic classification of music genres. Results of the carried out subjective tests and statistical analysis were included.

★ ★ ★

Visualization of sound field produced by mobile device using acoustic vector sensors

Kotus Józef, joseph@sound.eti.pg.gda.pl

Gdańsk University of Technology
Gdańsk, Poland

The results of the acoustic energy distribution measurements around the mobile devices were presented in the paper. The acoustic energy distribution was obtained in free field using sound intensity measurement technique in the frequency domain. The 3D Acoustic Vector Sensor was applied for this purpose. Each particle velocity sensor is sensitive in only one direction, so three orthogonally placed particle velocity sensors had to be used. In combination with a pressure microphone, the sound field in a single point is fully characterized and also the acoustic intensity vector,

which is the product of pressure and particle velocity, can be determined. This intensity vector indicates the acoustic energy flow. Measurement sensor was placed in fixed points using a Cartesian robot. Sound field was produced employing loudspeakers built in mobile device. Obtained intensity measurement results were used to prepare visualizations of acoustic energy distribution around the mobile device.

⋆ ⋆ ⋆

Modeling of environmental noise propagation and its influence on hearing system using pl grid infrastructure

Kotus Józef, joseph@sound.eti.pg.gda.pl
Szczodrak Maciej, Czyżewski Andrzej,
Kostek Bożena

Gdańsk University of Technology
Gdańsk, Poland

Specific computational environments, so called domain grids, are being developed within the PL Grid Plus project. The services developed for the domain grid "Acoustics" were presented in the paper. The prepared services concern two interrelated subjects. The first one, called "Noise Map", can be used for calculating the noise map of large city areas. The second one, called "Hearing" service, enables simulations of noise impact on the human hearing system. Using the PL Grid infrastructure in is possible to calculate environmental noise maps for linear noise sources (road noise) or point sources (industrial noise, outdoor acoustics events). Based on the calculated noise levels, it is possible to conduct further simulations showing the effects of noise-induced temporary threshold shift. The developed tools provide flexible opportunities for defining variety of computing scenarios which have essential cognitive and educational virtues, especially for students and researchers.

⋆ ⋆ ⋆

Transmission of audio signals at concert venues

Kozłowski Piotr Z., piotr.kozlowski@pwr.edu.pl

Wrocław University of Technology
Wrocław, Poland

Nowadays systems designer and realized at concert venues are based on different topologies. Technology of audio signal's transmission influence deeply on ideas used to send the signals between different parts of system. Mainly topology serves functional needs, but signal transmission technology caused some tiny structural differences between systems built in period even a few years. Main topic of paper is topology of systems realized at concert venues (commercial and educational). Discussion presented at paper is based on audio systems design experience for concert venues constructed or repaired last years such as: Wrocław Opera House, Musical Academy in Katowice, Musical Academy in Wrocław, Teatr Rozrywki in Chorzów, Opera House in Poznań, new house for Polish National Radio Symphony Orchestra in Katowice.

⋆ ⋆ ⋆

Influence of sound on concentration and perception of video dynamic

Krawczyk Dawid, dawid.krawczyk@pwr.edu.pl
Brachmański Stefan

Wrocław University of Technology
Wrocław, Poland

The paper presents the experimental results which were carried out to examine the influence of sound which appears in a video sequence on the subjective feeling of a person. The matters regarding the selection of specific samples and the method of their creation were presented. The issues connected with the methodology of carried out process were discussed. The experiments were carried out in two stages. In the first stage the influence of sound on memorizing and concentration of a given person was examined. In the second stage the influence of sound on the subjective feelings of video dynamics was investigated. The experiments results didn't justify to unambiguously claim that eyesight plays decisive role in the information interpretation. It can be however noticed that researches revealed that both senses of hearing and sight complement each other.

⋆ ⋆ ⋆

The acoustic properties of different types of earplug used by sound engineers

Kruk Bartłomiej, bartlomiej.kruk@pwr.edu.pl
Łuczyński Michał

Wrocław University of Technology
Wrocław, Poland

The main aim of this paper is to test various types of earplugs used be sound engineers. At live events, when sound engineer needs to use earplug for healthy reason, it is very important to hear correct anyway. The linear frequency characteristic allows to avoid mistakes when work with sound. Earplugs were tested for suppression depending on the frequency. The authors tested earplugs in the different methods: subjectively using pure tone audiometry and objectively using the designed and created ear canal. Research allowed to choose the appropriate earplugs for sound engineering purpose.

⋆ ⋆ ⋆

Aperture effects in audio PCM converters

Kulka Zbigniew, z.kulka@ire.pw.edu.pl

Warsaw University of Technology
Warszawa, Poland

Sampling and reconstruction are fundamental operations in digital audio signal processing which are performed by conventional pulse-code modulation (PCM) analog-to-digital (A/D) and digital-to-analog (D/A) converters. Sampling is the first step towards digitizing continuoustime audio signal. The reconstruction on the other hand is the operation of converting a sampled signal back to a continuous-time signal. The aperture effects during these operations create the time and frequency response errors and finally the sound distortion. In the paper, the theoretical basis and practical aspects of sampling in A/D and reconstruction in

D/A audio PCM converters have been presented including analysis of the errors resulting from the aperture effects and discussion of correcting methods. Keywords: aperture effect, time and frequency aperture errors, aperture jitter, pulse-codemodulation (PCM) A/D and D/A audio converters.

⋆ ⋆ ⋆

Analysis of noise assessment of selected areas in Gdańsk

Marciniuk Karolina, karmarci@sound.eti.pg.gda.pl
Kotus Józef, Kostek Bożena

Gdańsk University of Technology
Gdańsk, Poland

The subject of this study focuses on the noise condition analysis of selected areas in the city of Gdańsk with the use of a numerical model implemented on the supercomputer infrastructure. The numerical model of traffic noise and its main sources are briefly characterized. Then the correctness of the model based on the acoustical conditions of the chosen areas obtained by the system of sensors located in the city of Gdańsk is analytically verified.

⋆ ⋆ ⋆

Automatic mastering procedure basing on the existing phonographic recordings

Mickiewicz Witold, witold.mickiewicz @zut.edu.pl

West Pomeranian University of Technology Szczecin
Szczecin, Poland

The paper presents the idea of automatic mastering procedure, which uses the database of existing phonographic recordings of classical music. Basing on the objective parameters calculated for every member of such database: mainly amplitude spectrums and maximum to average amplitude ratio – the tonal balance and the dynamical spread of the new recording of the same classical music piece can be automatically adjusted. It is especially useful for 2-microphones "life" recordings made for the archiving purposes and can help to obtain better subjective assessments of the recordings. In the paper the details of the computer mastering algorithm are presented. The results of subjective assessment of symphonic orchestra recordings made originally with one main microphones pair and corrected using proposed procedure exhibit the usability of presented method.

⋆ ⋆ ⋆

Measuring and analyzing audio levels in film, commercials and cinematic trailers using LKFS loudness model

Milarska Kamila, Zakrzewski Aleksander
Kostek Bożena, bozenka@sound.eti.pg.gda.pl

Gdańsk University of Technology
Gdańsk, Poland

The aim of this research study was to measure audio levels of commercials, cinematic trailers and during displaying of a motion picture. The background of this study is related to numerous complaints from customers, and even acquired hearing impairment after the screening at the cinema. The problem of too excessive level of sound in movie theaters is known for a long time, however recently interest in this topic is much higher. In the paper, first issues concerning sound level analysis in the movie theaters, i.e. recommendations and standards, which relate to the permissible sound levels, as well as the parameters and units that determine the emission of sound are presented. The next Section contains the description of the measurements performed, equipment used and the calibration process. Measurements are made with the division to cinema sound, YouTube service and audio contained in DVD and Blu-ray discs. Subsequently, the results of measurements and analysis related to the selected film screenings are shown. The last Section summarizes the results and presents general conclusions.

⋆ ⋆ ⋆

Research of loudness of perceptually-coded musical signals

Muras Łukasz, maurycy.kin@pwr.edu.pl
Kin Maurycy J.

Wrocław University of Technology
Wrocław, Poland

Paper presents the results of subjective and objective measurements of loudness impressions of perceptually coded musical signals. Research was done for six kinds of musical styles and five values of bit-rates. It turned out that loudness impression decreases when bit-rate value gets lower. For the same musical signals, the objective parameters: Program Loudness as well as True-Peak have been measured using a professional loudness meter. Comparison of these two results indicates that the correlation between subjective impressions and objective parameters exists although this relation is not univocal for various kinds of musical styles.

⋆ ⋆ ⋆

Control measurements of loudness of audio signal in broadcasting

Plaskota Przemysław, przemyslaw.plaskota@pwr.edu.pl
Dobrucki Andrzej B.

Wrocław University of Technology
Wrocław, Poland

In the paper, the analysis of the regulation of the National Broadcasting Council of 30 June 2011 concerning principles of advertising and teleshopping in radio and television services and the change of regulation of 18 December 2013 is discussed. The aim of analysis is a verification of a technical description of loudness measurement of the advertising. The technical requirements of measuring of loudness and the possible ambiguousness of results of measuring

executed on the basis of the indicated measuring methods are also presented.

⋆ ⋆ ⋆

Incremental shape from motion revisited

SKARBEK Władysław, w.skarbek@ire.pw.edu.pl

Warsaw University of Technology
Warszawa, Poland

The tutorial oriented on incremental Shape from Motion (SfM) focuses on mathematical background for integrated theoretical treatment of both batch and on-line scenarios. Key geometric properties, used for SfM algorithms design and analysis, are stated and proved. Moreover, the cases of rank two and three measurement matrix for non planar shapes are fully characterized and its role in shape ambiguity explored. An object-oriented description of SfM is presented to combine various algorithmic ingredients, like subspace tracking, and to integrate various scenarios for SfM. Experiments refer to head pose identification and 3D face animation proving high visual accuracy of incremental SfM.

⋆ ⋆ ⋆

Learning mixtures of mouth parts from unconstrained imagery

WĘSIERSKI Daniel, daniel.wesierski@pg.gda.pl

Gdańsk University of Technology
Gdańsk, Poland

Detecting human mouths in unconstrained imagery is an attractive task in computer vision. Human-machine interfaces, such as audio-video speech recognition, could take immediate advantage of a working markerless technology. Since human lips can deform heavily, we describe a method for learning diverse mixtures of mouth parts that rigidly encode visual deformations. Our method commences by splitting the training examples of mouths into parts, which contain a particular subset of keypoints annotated over lip contours. It then indexes the subsets that share corresponding keypoints and clusters them. Since the annotated keypoints might insufficiently reflect iconic appearance changes of the whole mouth, the system further clusters the representative lip shapes into mouth appearance clusters that visually represent particular expressions. Finally, these clusters serve to discriminatively train fine-grained appearance templates under SVM regime that perform well in detecting various mouth parts in unconstrained imagery.

⋆ ⋆ ⋆