Aalto University
School of Science
School of Electrical Engineering

THE UNIVERSITY of York

EPSRC
Engineering and Physical Sciences Research Council

The York-Aalto Auralization and Spatial Audio Workshop
21-22 May, Department of Media Technology, Aalto University

Monday 21st May (Odeion)

09.15 **Coffee**

09.30 **Introduction**

09.45 **Ville Pulkki: Technologies and psychophysics of spatial sound**

An overview of Directional Audio Coding (DIRAC), its use in teleconferencing and as a generic audio format for sound recording. It also has applications in virtual worlds and game audio. Microphone techniques are discussed, and higher order microphone inputs for parametric audio. HRTF measurement and calibration based on pressure and velocity measurements using very small microphones. Finally laser based impulse response modeling was introduced.

10.00 **Mikko-Ville Laitinen: Fine-tuning Directional Audio Coding**

Reproduction of surrounding applause-type signals is known to produce audible artifacts with parametric spatial audio coders. This arises from inadequate temporal resolution. Also decorrelation can add reverb, especially with dry speech. Better quality results if the temporal resolution is high enough to analyze each handclap separately, which can be obtained by using different temporal windows for different frequencies. Formal listening tests confirm the improved subjective quality. Spaced mics can help, no decorrelation needed at high frequencies.

10:15 **Juha Vilkamo: Covariance-matrix-based techniques for spatial audio**

In perceptual processing of spatial audio, a typical assumption is that the spatial aspect of a loudspeaker-reproduced sound is determined especially by the energies and the time-aligned dependencies of the audio channels, which are expressed by the real part of the signal covariance matrix, in perceptual frequency bands. A generally applicable method to robustly process spatial sound in this domain is presented. This includes an adaptive mixing solution to reach the given target covariance matrix by best usage of the independent components in the input channels, and means to inject the necessary amount of decorrelated sound energy when the target is not achieved otherwise. The method was implemented as a part of a perceptually motivated spatial microphone technique, and numerous further use cases, including upmixing, have been identied.

10:30 **Olli Rummukainen: Audiovisual reproduction in surrounding display: effect of spatial width of audio and video**

Multimodal perception strives to integrate information from multiple sensorial channels into a unified experience, that contains more information than just the sum of the separate unimodal percepts. As a result, traditional quality metrics for unimodal services cannot reflect the perceived quality in multimodal situations, and new quality estimation methods are needed. In this work, audiovisual perception was studied with an immersive audiovisual display. The audiovisual display consisted of a video screen with field of view of 226 and 3D sound reproduction with 20 loudspeakers. The aim of the study was to observe the crossmodal interaction of auditory and visual modalities, when the spatial widths of audio and video reproduction were limited. A subjective study was organized, where the overall perceived degradation of the stimuli was evaluated with Degradation Category Rating in four different types of audiovisual content. In addition, free descriptions of the most prominent degrading factors were collected. The participants' individual tendencies to experience immersion were screened prior to the experiment with a questionnaire. The results show that video width is the dominant element in defining the degradation of a stimulus. Also audio width had an impact when the video width was at maximum. Individual tendency to experience immersion was not found to have significant impact on perceived degradation in this study. Slight content effects were observed. Constrained correspondence analysis of the free description data suggests the reasons for highest perceived degradation to be caused by wrong audio direction, reduced video width and missing essential content.

10:45 **Olli Santala: Effect of source signal peakedness on localization in spatially complex sound scenarios**

Research into localisation in spatially complex sound scenarios. Directional perception of distributed sound sources with results showing that transients help localisation when in the presence of background noise.

11:00 **Ville Pulkki: Binaural auditory models based on anatomical, neurophysiological and psychoacoustical knowledge**

A discussion on the human hearing system and how various aspects can be modeled in different ways to give a better understanding of how it works – with obvious implications for much of the work we are all engaged in.

11:15 **Damian Murphy: Virtual Acoustics at York, Modelling, Auralization, Interaction**

An overview of various projects ongoing at the AudioLab York, to be expanded upon in the presentations following. Hybrid acoustic modeling is at the centre of this work with a collaboration with York Computer Science to implement this on the cloud (YouShare) with Blender used as the design front end/interactive rendering wrapper. Applications in Digital Heritage were introduced, the importance of accuracy and context, and this leads on to the OpenAIR project – an online database of impulse responses from a variety of spaces. Finally we are beginning to explore outdoor acoustics (rather than traditional architectural acoustic studies) through soundscape studies and auralization of sonic crystal structures.

11:30 **Gavin Kearney: Distance Perception in Higher Order Ambisonics**

An investigation into the perception of source distance in interactive virtual auditory environments in the context of First (FOA) and Higher Order Ambisonic (HOA) reproduction. In particular, the accuracy of sound field reproduction over virtual loudspeakers (headphone reproduction) with increasing Ambisonic order. Performance of 1st, 2nd and 3rd order Ambisonics in representing distance cues is assessed in subjective audio perception tests. Results demonstrate that 1st order sound fields can be sufficient in representing distance cues for Ambisonic-to-binaural decodes.

11:45 **Jude Brereton: The Virtual Singing Studio**

The Virtual Singing Studio (VSS) is a a loudspeaker-based room acoustics simulation for real-time musical performance under development at the AudioLab, University of York. An objective evaluation of the VSS is made and relative differences between T30 and EDT values of the input (*real*) and output (*virtual*) Room Impulse Responses are reported. Although differences in T30 values fall within the double tolerance subjective limen (10%), EDT values of the simulation are less well matched to the input RIR. The process of designing and implementing such a 'vocally interactive' virtual acoustic environment is also considered and compared to 'off-line' auralization techniques.

12:00 **Lunch**

13:00 **Seb Jouan (Regional Director, Acoustics, AECOM)**

An industry perspective of various projects worked on previously at Arup, and now at AECOM. This included the acoustic design of a number of public buildings, auralization used as part of contemporary art and urban design.

13:15 **Aglaia Foteinou: Acoustic Reconstruction of Heritage Sites**

This presentation considers various aspects of the problem of recreating and optimizing acoustic simulations where the original structure is not available for comparision – although key in the design of new buildings, this is also important for the accurate recreation of buildings from the past. To help with this work the medieval St. Margaret's Church in York, UK, has been studied, which after renovations, is now used for music performances and conferences. The particular advantage being that the church's physical acoustical characteristics can be easily changed through variable acoustical panels and drapes arranged throughout the space, depending on the acoustic requirements of the activity within venue. A 3D model of the space based on geometric algorithms was created and has been optimised according to the acoustical configurations used during the actual measurements. The measurement and modelling work is presented, together with first case subjective testing. EDT has also been explored as a particular parameter that might be used as being key for fine-tuning a model. This is explored in a the limited, but highly controllable, example of a simple shoebox model where the effects of small changes on this parameter can be evaluated.

13:30 **Jelle van Mourik: Geometric Acoustics using Blender**

Blender has been explored as a tool that might be enhanced to facilitate acoustic design through the simple extension of the basic source code through its plugin enhancements, coupled with simple standalone executables. A number of examples are presented based on variations of the basic ray-tracing method.

13:45 **Sorrel Hoare: Auralizing Sonic Crystal Sound Barriers**

Sonic crystals have been investigated in recent years both as a potential form of noise barrier, and as a form of sonic art aimed at enhancing perception of the surrounding acoustic environment. The broader aim of this research is concerned with the auralization of these structures, which has, as yet, rarely been attempted. Prediction of the acoustic wave propagation through 2-D arrays of solid, cylindrical scatterers embedded in air has been performed in 2-D Finite Difference Time Domain (FDTD) simulations. The model has been extended into the third dimension and the results are compared with those obtained in the previous experiment. In both the 2-D and 3-D simulations the location of the fundamental band gap corresponds with the predicted location - predictions being based on simple theoretical considerations relating the frequency of the transmission gaps to the array configuration. A broader consideration is made in terms of how such structures might be used in the environment and a number of supporting soundscape studies have also taken place with sound examples presented in this presentation.

14:00 **Iain Laird: Virtual Performance Spaces**

A Virtual Performance System (VPS) is a real-time 3D auralisation system which allows a musician to play in simulated acoustic environments. Such systems have been used to investigate the effect of stage acoustics on the performance technique of musicians. This presentation explores the process of calibrating a VPS using energy-based quantities and goes on to verify this technique by comparing known acoustic quantities measured in a test space with a virtual version of the same space. This work has demonstrated that calibrating a VPS using metrics based on Support will result in an accurate simulation of a test space according to known acoustic metrics such as T30. A comparison of quantities referring to earlier parts of the response, such as Early Decay Time (EDT), show some errors which are thought to be caused by the non-anechoic nature of the reproduction space.

14:15 **Coffee break**

14:45 **Demos in Acoustics Lab**

Olli Rummukainen: Audiovisual Surrounding Display

Akis Politis: Parametric Spatial Audio Effects (to be presented at DAFx12)

Juha Vilkamo: Upmixing with covariance-based techniques

Mikko-Ville Laitinen: Head-tracked DirAC

16:00 **Day 1, Close**

Tuesday 21st May (Odeion)

09.15 **Coffee**

09.45 **Lauri Savioja: How to Achieve Optimal Acoustic Conditions**

A consideration and overview of room acoustic simulation strategies, context, goals and future work. The definition of optimal considers a simulation that should be efficient, full audio bandwidth and accurate. Currently this is based on FDTD and Image-Source for the Early Part of the Impulse Response, followed by the ART method for the late reverberation. Simulations should be good enough for the resulting application, but a question remains as to what is an optimal solution and how should this be evaluated. Could a fitness function be defined in subjective terms so that the simulation system might be optimized based on this? Would it be possible to solve the inverse problem such that given an audio recording could the space it was captured in be determined?

10:00 **Alex Southern: Developments in Low Frequency & Hybrid Acoustic Modelling**

A presentation of the WaveModeller acoustic modeling tool that takes a Google Sketchup 3-D room file and uses this as the basis for a FDTD simulation. This has been used in a number of related research projects and demos are available on YouTube - http://www.youtube.com/user/mrapsouthern/videos GPU acceleration is also enabled through the use of the NVIDIA CUDA library. Spatial encoding strategies for FDTD modeling was also covered, the perception of dispersion error, and hybrid modeling approaches and FDTD equivalence with the Image Source method. Latest work is based on forthcoming DAFx12 paper – spatial high frequency extrapolation method for room acoustic auralization, which bypasses the need for geometric acoustics in FDTD based auralization. Works well for large spaces where average absorption (boundaries and air) can be considered constant. Less good for smaller spaces where this cannot be assumed.

10:15 **Tapio Lokki: Perceptual, physical and virtual studies on concert hall acoustics**

An overview of work on the subjective and objective evaluation of concert hall acoustics including the significant study that has been ongoing over the last couple of years using the loudspeaker orchestra to study subjective preference in concert halls, based on an extended series of perceptual tests. Attempting to answer the question of which concert hall has the best acoustics and why this is considered to be the case. Room acoustics simulation also comes into this research as means to test some of these ideas. Also introduced was the work ongoing in audio augmented reality and auditory (eyes-free) user interfaces.

10:30 **Jukka Pätynen: Acoustic Measurements with a Loudspeaker Orchestra**

An overview of the work involved in the development of the loudspeaker orchestra as outlined above by Tapio Lokki. This involved making anechoic recordings for a full orchestra on an instrument-by-instrument basis, so that sections could be developed. Directivity characteristics were analysed and synthesized as part of this and the signals had to be processed in various ways – for instance in creating an orchestral section from a single solo recording. New methods of time-frequency

analysis have also been developed to help understand how the sound in a concert hall builds up, based on taking successively increasing time/frequency windows. This has been further developed into time/spatial analysis also. New publications forthcoming for these techniques.

Time-spatial analysis also.

10:45 **Sakari Tervo: Localization and Tracing of Early Reflections**

The newly proposed Spatial Decomposition Method for Room Impulse Responses (currently unpublished, to be submitted to IEEE TASLP) was presented and some excellent audio examples presented including comparisons to SIRR. General assumption is a wide band signal, small aperture size at the measurement probe microphone, high sample rate. Assume also that this result can be approximated using image-sources. The impulse response can then be represented and rendered by the image source method. Works for any microphone array and reproduction technique.

11:00 **Philip Robinson: Spatial Discrimination in Concert Halls**

11:15 **Demos in Odeion**

12:00 Lunch

13:00 **Planning future collaborations.**

Open facilitated discussion on possible areas of future research and collaboration. Areas highlighted included: (1) Optimisation strategies for perceptual/objective assessment – hybrid or new solutions are needed as anything related to perception is hard to test and implement. (2) Virtual common environment for all forms of collaboration - minimizing carbon footprint, sharing experiences, as seamless environment and user experience, related to telepresence research. Latency and accuracy of the virtual environment are potential problems that would have to be considered. Computational, transmission need to be optimized, prioritisation of data location. The presented environment needs to consider abstract vs. accurate/natural presentation. (3) Desirable augmented personal devices – to equalize or extend the human experience. 'Bionic' hearing would relate to issues in health, safety, longevity, connectivity. (4) Live (remote) listening devices – surveillance, uses in the natural world, an acoustic camera with tracking, source identification and separation. Raises questions of ethics and where are the research questions?

14:30 Coffee

15:00 Listening Test Opportunities

16:00 **Close!**