CONFERENCE REPORT

AES 41st International Conference

audio For games

2-4 February 2011, London, UK

The 41st AES International Conference, *Audio for Games*, was held at the prestigious headquarters of the British Academy of Film and Television Arts (BAFTA) in the heart of London's glamorous West End. Just yards away from the bright lights of Piccadilly Circus, the brightest lights of the game audio industry gathered to hear about the latest innovations and discuss current issues.

Conference chair Michael Kelly and the conference committee (Steve Martz, Damian Murphy, Becky Stewart, Russell Mason, Pete Harrison, and Kazutaka Someya) put together a compelling array of workshops, papers, posters, and events.

WORKSHOPS

Following the successful format of the previous Audio for Games conference, the first day was devoted to a series of workshops, expertly organized by Steve Martz. The first of these was an introduction to sound and music for games, presented by Jonatan Crafoord. He described an attitude of some that music and sound don't contribute much to a game, but he countered that they are essential to increasing the realism and enhancing the emotion of the gaming experience. He set out a number of challenges to game audio. One of these is the difficulty in developing an audio pipeline within a game. Depending on the complexity of the game, this can require as many as 25 programmers working for up to 2 years. Cost savings can be made by developing a simpler audio pipeline, but this often makes the sound design and audio implementation more difficult and hence more expensive. Jonatan argued that a potential solution is to licence middleware as the audio pipeline for a game. A second challenge that Jonatan discussed was that games are expected to be replayed many times, so a large amount of material is required in order that repetition of sound does not become annoying. He outlined many methods for reducing the repetition, including randomly selecting from a number of samples or sample segments and real-time variations to the sound such as modulation and pitch shifting. He also discussed the amount of music that may be required for a large game project, giving an example of 20 hours of music available for the game *World of Warcraft*.

Education in game audio was discussed by Jeanine Cowen and Michael Sweet. They described the background of their institution, and explained that while technology rapidly changes, the basic knowledge and understanding remains relatively constant. From a background of education in music and audio, Jeanine and Michael explained the aspects of game audio that are common with other programs in audio, music technology, composition, performance, and the unique aspects of game audio such as composition of music that can be applied in a dynamic and interactive manner, and the difficulties of synchronizing music and audio with game content that is under the control of the player as opposed to the

From left: conference chair Michael Kelly, papers chair Damian Murphy, workshops and tutorials chair Steve Martz, and AES president Jim Kaiser, who gave a short introductory speech.

fixed structure and timing of films. One particularly exciting aspect of the game audio program was the opportunity for collaboration with students at other institutions studying game design and programming; through this students could gain experience of the whole development environment, learn about collaborative work, and gain knowledge from students with differing backgrounds.

Nicolas Fournel gave an overview of the dramatic increase in complexity in game audio, from the simple beeps of Pong to current games that require the capture and organization of significant audio resources. Predicting that future games are likely to be even more complex, he demonstrated the difficulties now faced by game audio sound designers and programmers, and gave examples of tools that can be used to assist the game audio designer. These include feature-extraction algorithms that can represent the audio using more intuitive attributes such as pitch, allowing a sample with required characteristics to be found more rapidly, and a tool for modeling and resynthesizing sounds that allows the sound designer to modulate a wide range of parameters.

A panel of presenters, Bike Suzuki, Masataka Nakahari, Tomoya Kishi, Kenji Kojima, and Steve Martz, gave an introduction to reverberation generation. They started by summarizing the main available techniques, such as modeling virtual environments, capturing impulse responses of real rooms, and artificial reverberation developed for musical applications. In a modification of a relatively simple statistical approach to room modeling, they outlined a technique similar to the classic Sabine formula for calculating the decay time, but includes a statistical method to estimate the direction of the reflections that can be used to create the spatial characteristics of the reverberation. Compared to the conventional method of creating multichannel reverberation by generating a number of uncorrelated decays, this technique has the advantage of relatively low processing complexity in conjunction with the incorporation of additional room detail.

The advantages of emotion-driven audio were presented by Maciej Zurawski. He described a number of ways that music evokes emotions, including innate brainstem reflexes (such as the shock from a sudden loud sound), evaluation conditioning or association based on experience (such as a melody that evokes a memory of an emotion), and visual imagination (such as the imagination of movement based on the flow of a musical line). In contrast to traditional composition techniques, Maciej proposed an alternative where segments or lines of music are composed based on their emotion, and these are triggered by variables with the game through which the instantaneous intended emotion can be tracked.

For a wide range of games genres, the speech is critical, and poor quality or unintelligible speech can seriously detract from the gaming experience. Gordon Durity discussed this in detail, giving examples of aspects that are important to the perceived quality of the speech in a game, including limiting repetition (frequently repeated phrases rapidly become annoying), and keeping the speech relevant to the action, but structuring this in a way that maintains the conversational thread or narrative of the speech. He gave examples of football commentary, where the commentary was arranged to switch between general banter and specific comments on the action, depending on the instantaneous parameters of the game. Natural speech patterns such as this can be difficult to program, as it requires the system to store a wide range of contextual information so that the dialogue can return to previous topics when there is little action on the football pitch. Dieter Piltz described the steps involved in the capture, editing, and organization of the speech. A game that includes an interactive commentary can contain well in excess of 30,000 speech extracts that need to be logged and accurately tracked.

Lively discussions took place between delegates during the breaks.

Also on the subject of speech production, Roberto Pomoni and Francesco Zambon discussed the procedures and difficulties of recording speech in different languages to create international versions of predominantly English-language

games. As noted above, sometimes a very large number of speech extracts are required in a game, and with up to ten different languages this could result in over 150 hours of audio and 100,000 audio files. With work of such complexity, care needs to be taken over the management of the recorded resources, and the different languages need to be recorded in a way that allows easy transition: closely recorded audio with little captured ambience or reverberation, of a high quality without distortions or problems, and a close match to the characteristics and the file structure of the original audio.

A panel of industry experts discussed the difficult topic of level and loudness in games. Panel chair Steve Martz set out the problem that users complain about the variable loudness of games, posed the question about why this has arisen, and asked the panellists to suggest potential solutions. Steve Root discussed the issue from the game studios' point of view, and explained that different types of games may need to have different levels; a car racing game needs to be loud and exciting, whereas a puzzle game should often be quieter. Mark Pascoe compared the problem in the gaming industry to that in the broadcast industry. He explained that it is an ongoing problem in broadcast, and that a number of approaches to standardizing the recorded level of programs and adverts had been tried; he also commented that with the increasing convergence of services within a single device (e.g. a games console that also plays DVDs and streams online broadcast services), a common solution for all of these industries needs to be found. However, the problems faced by the panelists were the difficulty of defining a standard loudness level, the confusion about where in the signal chain the level or loudness should be specified, and the practical issue of regulating such a standard.

KEYNOTE SPEECH

The keynote speech was given by John Broomhall, a composer and audio producer who specializes in music and audio for games. This was one of the most musical keynote speeches in AES history, as he accompanied his talk with excerpts on piano, and included a cameo role by computer-game pioneer Ian Livingstone on harmonica.

John gave a summary of his career from the early days when resources were comparatively limited: in these situations composers had to try to do a lot with very little. As storage capacity and processing power increased, audio and music designers could achieve much more, however the increase in complexity meant that sometimes the creativity and finesse suffered in the rush to get it all finished in time for release. In addition, the ability to do a lot more meant that sometimes the end result was over-complicated. John quoted a former colleague: "Now we can have all the sound that we want, now we have to figure out what we should leave out."

He went on to discuss the use of sound in movies, and encouraged the games industry to take more cues from this. Games often strive to achieve realistic sound, but he showed a number of exam-

Top row, from left: Adam Philp of Lionhead Studios talks to Michael Kelly and Steve Martz; Jaroslaw Beksa and Pawel Barszcz with their poster on multiplatform audio games; Gavin Kearney, Marcin Gorzel, and Claire Masterson of Trinity College Dublin; Jordi Janer and Oscar Mayor in front of their poster on audio transformation technologies. **Bottom row**, from left: Bike Suzuki, Steve Martz, Masataka Nakahara, Tomoya Kishi, and Kenji Kojima, who gave a paper on reverberation processing for games; Loudness workshop panelists Jonatan Crafoord, Steve Root, Mark Pascoe, Steve Martz, Francesco Zambon, and Garry Taylor.

Michael Kelly, right, thanks keynote speaker John Broomhall for his enlightening contribution to the conference.

ples in which movie sound was deliberately distorted or altered to increase the tension or emotion of a scene. For instance, sound objects can be removed from a scene in order to focus on one aspect or to enhance the impact of a sudden loud sound.

John finished by stating that the key to good music and sound for games is the relationship between the creative input of the composer/sound designer and the technical input of the audio programmer, and he praised the best audio programmers as the unsung heroes of games.

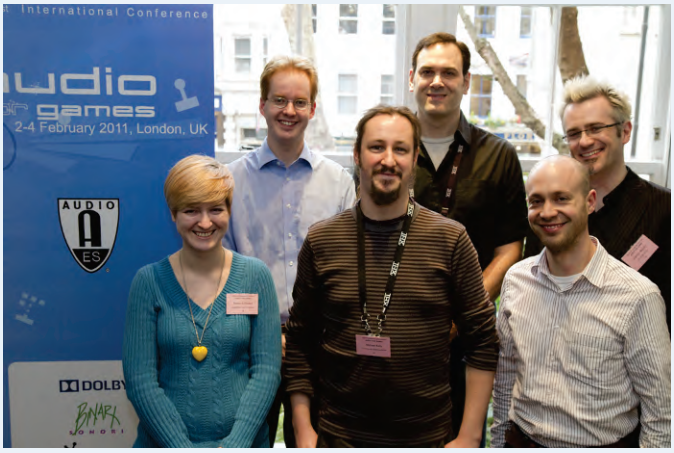
PAPERS SESSIONS

Papers chair Damian Murphy put together a densely packed program of lecture and posters presentations, over six varied sessions covering aspects of education, perception, reverberation generation, pitch detection, and spatial audio among others.

Mark Sarisky presented a paper that considered the place of education in game audio. His opinion is that game-audio education should be contained within generic audio programs, rather than being structured as a separate entity. He examined the recommendations of the education arm of the Interactive Audio Special Interest Group, and drew up a proposal in which the majority of the relevant concepts can be covered in two modules to be added to a generic audio program.

Questions from the audience helped to ensure stimulating discussions.

The loudspeaker placement for an air hockey game was investigated by Bill Kapralos. This style of game requires a very different user interface arrangement in contrast to conventional computer games; rather than the users sitting facing a screen in front of them, the screen is laid horizontally and the users can be positioned all around the screen. Two four-channel loud-



Conference committee members: from left, Rebecca Stewart, Russell Mason, Michael Kelly, Steve Martz, Pete Harrison, and Damian Murphy.

speaker arrangements were tested, and for one of the configurations examined there was no visual puck but the users were guided by the position of the sound. Of the two loudspeaker arrangements tested, the diamond arrangement was preferred and Bill concluded that this demonstrated that games can be used for psychoacoustic testing.

There were a number of papers on reverberation systems for computer games. Enzo de Sena explained the need for a compromise between accuracy and low computational complexity for real-time applications such as game audio. He described a technique that used a scattering delay network that models reflection and scattering using a small number of nodes. He explained that this can be used to simulate frequency-dependent wall absorption and sources and receivers with a range of directivities. The success of the system was evaluated by comparison with theoretical reverberation-time calculations and estimated modal characteristics.

Simon Shelley gave an overview of an open acoustic impulse-response library. This allows users to upload impulse responses together with the associated information about the capture method and any associated information and photographs. The server then measures a number of acoustic properties of the impulse responses and convolves it with an example anechoic recording so that the users can hear it before download. The impulse responses are then available for download and may be useful for game audio.

Diffraction is an area that is often overlooked in virtual environments such as those used in computer games. More often than not, occlusion (such as when a small wall is between the sound source and the receiver) blocks the sound completely, whereas in reality the sound can often be heard because some of the sound will diffract around the obstruction. Bill Kapralos described a method for modeling the diffraction so that sources can still be heard through available openings in the environment. As the method is computationally complex, a graphics processing unit was employed due to its efficient and highly parallel structure. The resulting system is capable of rapid calculation of the diffraction effects, and Bill demonstrated the resulting sound from this technique in comparison to conventional methods.

One of the major issues in game audio is the limited storage capacity, especially on more limited formats such as smartphones. A potential solution to this problem is to use recorded sound samples for more than one purpose. Mario Cáceres presented a paper that examined using environmental sounds in a game as sources for music composition. A number of sounds, such as explosions, gunshots, and tonal sound effects were filtered and modulated for use as the musical sounds within a game. Examples included a grenade-launcher sound effect used as a bass drum and a gunshot used as a snare drum. Mario demonstrated that this approach can be used to successfully render a music track, drastically reducing the storage requirements for the music samples.

Another approach to limited capacity on the game device was discussed by Sascha Grollmisch and Estefanía Cano. They described a system that used signal processing on a remote server to reduce the processing load on the local machine. The server undertakes real-time pitch detection for use in Internet-based games: this attempts to detect the variation in pitch over time, and there are a number of versions that have been customized for specific instrument or voice types. Using this system, the pitch information can be provided to the gaming device with a latency of around 60 ms, with a substantially reduced load on the local machine's processor.

Among the poster presentations, there was a description of the development of location-aware interactive games. These use GPS-enabled smartphones to overlay information and sound on a user's natural environment. Examples were given of using this system to integrate interactive audio in a tourist attraction for a medieval castle. Other examples of audio-based games were given in the posters: Song2See is a game to assist music students with practice, as they can view their pitch and tuning in real time; and a game where users investigate a virtual world to follow a story, which as it is predominantly audio-based, can be used by both sighted and visually impaired users.

Overall, the AES *Audio for Games* conference was a fantastic way to keep abreast of the latest developments in game audio, and to discuss problems and solutions with colleagues working in this area. The newest social-networking option for AES members—the ability to comment on conference and convention papers, in addition to *Journal* papers (https://secure.aes.org/forum/pubs/)—was introduced at the 41st, allowing the discussion among authors, attendees, and AES members unable to attend the conference to continue in the interval until the next AES *Audio for Games* conference.

Rebecca Stewart with the student volunteers.

Editor's note: Purchase the AES 41st papers on CD-ROM at www.aes.org/publications/conferences. Get individual papers in the AES E-Library at www.aes.org/e-lib.