60ᵗʰ International Conference

Dereverberation and Reverberation of Audio, Music, and Speech
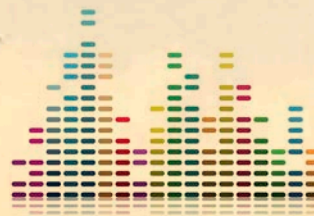
Leuven, Belgium
February 3–5, 2016

CONFERENCE REPORT

Delegates from around the world gathered in the beautiful city of Leuven to network, hear about the latest research into reverberation and dereverberation, listen to demonstrations, and discuss future prospects for the field. Some delegates were also excited to visit a local pub with a selection of over 2000 beers—a world record.

Toon van Waterschoot, conference chair, welcomed the delegates to the conference. The conference topic was motivated by recent work on reverberation and dereverberation, including a number of review articles, the 2014 REVERB (REverberant Voice Enhancement and Recognition Benchmark) challenge, and the DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech) research project. The scope of reverberation and dereverberation itself covers a large range of topics and experts in many areas including signal processing, psychoacoustics, room acoustics, and machine learning gave presentations during the conference. Toon reminded delegates

of the call for papers for a special issue of the *AES Journal* (see p. 169), thanked the committee members for their hard work to prepare for the conference, and thanked the European Commission for its financial support.

The conference had an ambitious technical program. Each morning, a keynote lecture was presented to the delegates, arranged ably by Timo Gerkmann. The keynote speakers gave excellent overviews of their respective topic areas, ensuring that everyone who attended was up to speed with the basic principles and history of reverberation synthesis and room modeling, perception, and dereverberation techniques. The technical content of the conference was mainly delivered in nine paper sessions with a total of 30 invited and contributed papers, coordinated by papers chairs Stefan Goetze and Ann Spriet, supported by a demonstration session. The topics covered by the technical program aligned with the keynote lecture topics, and are described here by theme.

DREAMS
Dereverberation and
Reverberation of
Audio, Music, and Speech

CARL VON OSSIETZKY universität OLDENBURG
Imperial College London
KU LEUVEN

B&O BANG & OLUFSEN
oticon PEOPLE FIRST
Fraunhofer
NXP
NUANCE

The conference was held at the Leuven Institute for Ireland in Europe, dating from 1607.

The conference organizing committee, (L–R) Stefan Goetze, Timo Gerkmann, Ann Spriet, Jan Østergaard, Aldona Niemiro-Sznajder, Toon van Waterschoot, Enzo De Sena, Simon Doclo

REVERBERATION SYNTHESIS

The first keynote was given by Vesa Välimäki, entitled "More Than 50 Years of Artificial Reverberation." The talk traced the development of synthetic reverberation over the last 55 years or so, but also mentioned approaches from the last few years. He played a number of audio examples throughout the talk, which was sometimes challenging as the lecture hall itself was very reverberant. Reverberation has been synthesized in a number of ways, beginning with classical Schroeder reverbs, delay networks, and convolution algorithms. Schroeder worked on reverb in his spare time at Bell labs, and modern Feedback Delay Networks (FDNs) are effectively a generalization of this kind of reverberator. Recent work in 2015 has considered how to choose the feedback matrix and how to connect the FDN parameters to the real world. Convolution reverbs can alternatively be used, where a dry sound is convolved with a room impulse response (RIR). Convolution is computationally very expensive, especially as the channel count increases. Fast convolution techniques can be used at the cost of latency, and partitioned fast convolution can help reduce the latency by first partitioning the RIR into frames.

Newer synthesis methods including velvet noise-based reverberation algorithms, scattering delay networks, and modal reverberation were then introduced. Velvet noise consists of a sparse nonuniform tertiary sequence, a structure that is efficient to implement compared to convolution, and gives quite a flat frequency response. Recent work has tried to remove flutter echoes and other artifacts when using velvet noise reverberators. Scattering delay networks approximate geometric ray tracing using a digital waveguide network. Scattering junctions are placed at wall reflections and a delay network is set up between these junctions. The first-order reflections are very accurate, and error increases for higher-order reflections. Alternatively, RIRs can be modeled in terms of a modal structure, describing the amplitude, frequency, and damping of many resonant filters in parallel. Three methods can be used to find the parameters: fitting mode parameters to measurements; deriving mode parameters from a physical room; and selecting parameters according to the desired response.
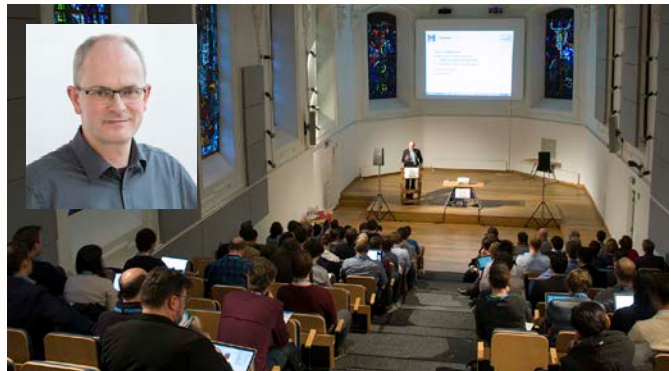
Finally Vesa shared his perspective on the state and future of synthetic reverbs. FDNs and convolution reverbs are still popular, but scattering delay networks, velvet noise-based methods, and modal reverb are exciting new possibilities. There is growing interest in modeling outdoor environments and multichannel reverberation. In summary, artificial reverberation research is still going strong and

The stained glass windows of the Irish College in Leuven frame Toon van Waterschoot's introductory presentation.

Thomas Brand gets ready to explain perception of reverberation to a busy auditorium full of delegates.

Vesa Välimäki explains how to add artificial reverberation to sound.

there are more inventions yet to come.

A few papers considered the topic of artificial reverberation. Talks were given covering the opportunities and challenges of parametric reverberation for object-based audio, creating a sense of a larger room inside a smaller one using loudspeakers, and approaches to set the reverberation levels for automatic mixes. In addition, two papers were presented on the related topic of numerical methods for room acoustics, including considerations of the boundary conditions for different absorption and irregular geometric conditions.

PERCEPTION OF ROOM ACOUSTICS

The keynote on the second day explored the question "How do humans benefit from binaural listening when recognizing speech in noisy and reverberant conditions?" and was given by Thomas Brand. This talk provided an audiologist's perspective on the perception of reverberation. He began by summarizing the measurement of speech intelligibility (SI) and relating it to human spatial processing. SI modeling began with telephony in 1922, this making more than 90 years of intelligibility models compared to only 50 years of artificial reverberation. SI is often linked to the signal-to-noise ratio (SNR). The rate at which 50% of listeners can understand the speech is called the speech reception threshold (SRT), and varies by language.

Binaural auditory models use combinations of filter banks, compression, and modulation with noise also added in the internal processing. This internal representation must then be interpreted by the model. The filter bank can be used to explain three kinds of masking: energetic masking, modulation masking, and informational masking. There are also binaural effects including unmasking and better-ear-listening that affect SI. SI models can be based on SNR, modulation, correlation, or automatic speech recognition models.

Thomas then discussed the effect of reverberation on SI. For near-field speech, reverberation affects the SRT, and the strongest effects are caused by strong early reflections. Spatial separation of target and interferer generally benefits intelligibility, but early reflections can destroy this benefit. In the far field, early reflections (before 50–100 ms) may however be considered useful for speech intelligibility. Detrimental reflections can be canceled by the binaural system when they come from a different direction than that of the target speech. Another interesting result is that for nonnative speakers the overall effects of reverberation are the same, however native speakers have lower absolute SRT.

Delegates were left to ponder a few final remarks. The binaural auditory system should be taken into account in modeling, and it

can be applied to all kinds of situations, not just for speech. There are different measures and models to quantify SI, based on SNR, modulation, correlation, or automatic speech recognition models. Of these, the latter approach seems to be a good one. Eventually, SI has to be measured with human listeners, and interesting new questions always arise from understanding the differences between the measured and the modeled scores.

A number of papers explored various aspects of the perception of reverberation in rooms and in music production. Some considered producer or listener preference for the level of reverberation and compression added to a source, or perceived loudness of reverberant content. Another talk described a localization experiment that attracted 893 listeners who were attendees at the UK's Royal Society Summer Science Exhibition in 2015. Another contribution suggested a method for perceptual assessment of room acoustics, allowing direct switching between rooms. Papers on SI explored the prediction of speech clarity and the effects of reverberation on SI in public spaces.

Yosra Mzah talks about the effect of public spaces on speech intelligibility.

DEREVERBERATION

The final keynote of the conference was delivered by Emmanuël Habets, entitled "Fifty Years of Reverberation Reduction: From Analog Signal Processing to Machine Learning." He visited three main topics: reverberation reduction, reverberation suppression, and direct estimation of the dry signal. Reverberation reduction and suppression techniques first try to estimate the channel and then apply equalization. In both cases, small errors in the channel estimation can severely degrade the performance.

Several approaches exist for reverberation reduction, including FIR and IIR filter models and spatial filtering. Challenges in the FIR model are mainly around the underlying assumptions, signal coloration, and high numbers of coefficients required. Better performance can be achieved by processing in the time–frequency domain. IIR models consider the room response as an autoregressive process, which holds for multiple sources. Challenges including reducing the effect of early reflections and maintaining control of the desired signal. Spatial filtering approaches can be useful, especially when the early reflection directions are known, but these techniques are hard to adapt to moving sources.

Reverberation suppression algorithms assume that reverberation is additive and that the reverberant signal is uncorrelated with direct signal. This signal model leads to a family of approaches including data-independent beamforming, single-channel spectral enhancement, multichannel spectral enhancement, and data-dependent beamforming. Challenges here include achieving even more reverberation reduction and handling unknown reverberant sound fields.

The last approach explained was direct estimation where dereverberation is performed based only on a source model. Here we can apply techniques like neural networks, although there are

Emmanuël Habets describes signal flows for dereverberation.

challenges of generalizability and the resources required for training.

There are some practical challenges for dereverberation. One is that the processing tends to decrease the perceived loudness, especially for multiple sources. Also, coloration due to early reflections is common, and there can be a perceived unbalance between the early reflections and the removed late part. Furthermore, interference (such as echoes, noise, and interfering sources) can be introduced or appear more prominent after the signal has had reverberation reduced. There may also be application-specific requirements for real-time processing and low latency. In conclusion, we were shown that there have been significant advancements in the last 15 years, but it is still a challenging problem, especially for machine hearing. Future directions are likely to focus on reducing early reflections, processing signals with lower SNRs and greater levels of reverberation, and application-specific challenges such as binaural processing for hearing aids. There is also a need for unified and perceptually meaningful evaluation metrics for researchers.

Many of the papers presented focused on techniques and applications of dereverberation. Some researchers have been working on estimating the room impulse response or parameters to describe it, including the direct-to-reverberant ratio and RT60. Techniques proposed included kernel regression, recurrent neural networks, polynomial roots, an orthonormal basis function model, and iterative blind estimation. Parameter estimation for late echo suppression was also considered.

Papers directly describing dereverberation included those advancing blind techniques and those incorporating additional prior knowledge. The relationship between data-dependent beamforming and multichannel linear prediction was explored. Features that might be known and can be incorporated into dereverberation algorithms include spectral estimation, noise statistics, a model of spatial coherence, room geometry, and surface absorption. Prior knowledge has to be applied with care—for instance in hearing aids where wrongly estimating a talker position can nullify advances in beam steering techniques. Blind techniques included adaptive kurtosis maximization and blind separation of direct sound from the early reflections. Sparse methods were also considered. Finally, one presenter took a higher-level approach to dereverberation by canceling all sound in a quiet zone, using an underlying numerical model to simulate the acoustic propagation paths in the room.

DEMONSTRATIONS

In the afternoon of day two, there was a demonstration and poster session with 11 different demonstrations covering a range of the topics discussed above. Many demonstrations were directly related to papers that had been presenting in the morning or on the previous day. Delegates engaged enthusiastically with demonstrators, and the session provided an excellent platform for headphone listening and longer discussions about the work. The demonstration session was also an excellent opportunity for networking and striking up new collaboration. Such sessions, as organized here by Jan Østergaard and Enzo De Sena, are to be highly recommended to future AES conference organizers.

Researchers from Hanyang University, South Korea, engage in a lively discussion about their demonstration of a source localizer and tracker.

Demonstrations concerning reverberation included recordings in concert halls, reverberation using an object-based renderer, perception of hyper-compression by untrained listeners, room acoustics experiments using virtual reality, and perceptual evaluation of reverb in music production. Other demonstrators showed off work that can be applied to dereverberation, including separation of direct sound from early reflections, source localization and tracking, room acoustics characterization, speech dereverberation in hearing aids, and joint noise reduction and dereverberation for speech enhancement. There was also a demonstration of the system that had been used to gather data for large-scale auralized sound localization experiments.

On the final day, Brecht de Man from Queen Mary University of London, UK, gave a presentation to accompany his demonstration of perceptual evaluation of reverb in music production. He gave an overview of different kinds of reverb currently used by music producers, emphasizing that reverbs are not necessarily tied into the real world. The aim of the perceptual evaluation was to link perceptual descriptions used by the producers to physical features that engineers can estimate. A browser-based user interface was designed to allow producers to rate different reverbs alongside one another. The perceived amount of reverb can be expressed using loudness and reverb time. The equivalent impulse response, estimated from the signal, seems to correlate with the perceptual measurements.

SOCIAL EVENTS

The conference committee laid on social events on both evenings of the conference, with local arrangements in the capable hands of committee member Aldona Niemiro-Sznajder. On the first day, after intense technical talks and discussions during the day, delegates were treated to a reception in the conference venue. The reception took place in the ancient cloisters of the Irish College, although protected

João Felipe Santos shows off his demonstration.

Mathieu Baqué explains how to separate direct sound from early reflections.

Delegates relax during the conference's social events.

from the elements by a rather more modern glass wall. Luckily, it had been possible to snap a photograph of all delegates earlier in the day. The reception was conducive to more informal discussions about the day's topics. Later in the evening, after a break for dinner, an organ recital was held at St. Gertrude's Church in the Small Beguinage district of the city. The church was considered as a wonder of Leuven in the 17th century due its tower being built without using any nails. Organist Wouter Dekoninck performed a number of pieces before chatting with delegates about reverberation in organ music.

On the second evening of the conference everyone was treated to a banquet, also held at the conference venue. The menu was both Belgian and Irish-inspired, reflecting the dual nationality of our hosts. Fruit beer was served upon arrival, and delegates enjoyed a salmon starter and duck main course, both complemented with carefully matched wine. The chocolate dessert, which aroused much excitement when the menu was announced, did not disappoint.

SUMMARY

Toon van Waterschoot gave some closing remarks to draw the conference to a close. He thanked the speakers and demonstrators for their contributions and all delegates for attending. Overall, the fields of reverberation and dereverberation are active and there are many exciting new techniques being proposed and evaluated. A number of challenges remain to be solved in both domains, and any new ideas will require perceptual models and listening tests to determine whether or not they are effective. Interested readers are encouraged to read the conference proceedings and to keep an eye out for the special issue of the *Journal of the Audio Engineering Society* (planned for 2017 Jan/Feb).

Editor's note: the papers presented at this conference can be obtained from http://www.aes.org/publications/conferences/?confNum=60

Delegates took advantage of a break in the weather to gather outside for a group photo, in front of a building dating from 1617.