Adapting sound reproduction to listener position with dynamic loudspeaker equalization

Matthieu Kuntz, Bernhard U. Seeber

Audio Information Processing, Technical University of Munich
E-Mail: {matthieu.kuntz; seeber}@tum.de

Introduction

Sound field synthesis techniques, particularly higher-order Ambisonics (HOA), are increasingly common in hearing research, as they enable an accurate reproduction of a desired sound field. This allows designing experiments with multiple sound sources, potentially in movement and reverberated. Coupling sound field synthesis techniques with loudspeaker-based reproduction allows the participants' movements to directly influence their perception without needing dynamic binaural synthesis. More so, factors like head-above-torso orientation do not have to be modeled, as they happen physically.

A well-known restriction of HOA is the size of the so-called "sweet-spot", the area where the reconstruction is accurate. It is typically located around the center of the loudspeaker array, with a frequency-dependent radius. Additional deviations from ideal loudspeaker arrays, such as loudspeaker directivity and residual reflections on hardware, further restrict the sweet-spot size. *In situ* measurements of the SOFE loudspeaker array of the Audio Information Processing group at the Technical University of Munich [1] show that the measured sweet-spot size radius is about half as large as the theoretically estimated one [2].

To allow participants to move freely in the reproduced sound field without affecting the reproduction accuracy, the employed sound field synthesis technique should be able to accurately reproduce sound fields over a sufficiently large area or a smaller area that can be steered to follow the listener's movements. Analytical approaches extend the wave field synthesis (WFS) technique and allow the translation of the area of correct reconstruction inside the loudspeaker array. Local WFS by spatial band-limitation [3] recomputes the loudspeaker driving functions based on a circular harmonics expansion around a different point inside the loudspeaker array. Local WFS by virtual secondary sources [4] creates a virtual loudspeaker array, ideally placed around the desired position for accurate reproduction. The virtual loudspeakers are then reproduced as focused sources using WFS. Numerical approaches, such as [5, 6], invert previously measured transfer functions to optimize the reproduction accuracy in an arbitrary region.

The approach investigated here consists of equalizing the loudspeakers in level, time and phase at a desired point in the loudspeaker array, coinciding with the listener's position. Coupled with a motion tracking system, the listener's position can be used to compute the corresponding equalization (EQ) filters for every loudspeaker. This approach has been shown to work well via simulations, with a sweet-spot radius at 2 kHz of at least 22 cm in an area of 1 m x 1 m around the center of the loudspeaker array [7].

Methods

Experimental setup

The experiment took place in the SOFE loudspeaker array [1], using the 36 horizontally distributed loudspeakers (Dynaudio BM6A mkII, Dynaudio, Skanderborg, Denmark). A microphone (NTI Audio) was tracked with twelve optical motion-tracking cameras (OptiTrack Prime 17W, NaturalPoint Inc. Corvallis, Oregon, USA), running at 44100/123 = 358.6 Hz, synchronized to the sound card via a word-clock signal. The microphone was connected to an RME Micstasy (Audio AG, Haimhausen) for power supply, pre-amplification, AD-conversion and recorded with an HDSPe soundcard (Audio AG, Haimhausen). The microphone was placed on a boom. The experimenter stood outside the loudspeaker array and maneuvered the microphone around the center of the array, covering an area of about 1 m by 2 m.

Filter computation

The loudspeaker EQ filters were computed from impulse responses measured in 1° azimuth steps from a sample loudspeaker placed in the anechoic chamber, but outside the SOFE loudspeaker array. The impulse responses were inverted, and the filter gain was limited to 10 dB. This resulted in a set of 360 1024-tap FIR filters.

During the playback, the microphone position was called from the motion capture system and the relative loudspeaker positions were computed. The relative azimuth angle was used to pick the corresponding EQ filter from the filter set, resulting in a matrix of 36 filters, one for each loudspeaker. The relative distance was used to compensate for the time of arrival differences and 1/R decay. The delay was applied by shifting the filter in the time domain by an integer sample number. The remaining time delay was applied by linear phase shifting in the frequency domain. In the manuscript, this approach is referred to as dynamic EQ.

In addition to the dynamic EQ, we used a set of static EQ filters, a set of 36 1024-tap FIR filters, measured in the center of the SOFE loudspeaker array, that were not changed during the playback.

Stimuli

A 60s white noise from 100 Hz to 15 kHz was played back at a level of 65 dB SPL from either a single loudspeaker at 0° or panned to 13° using 17th-order Ambisonics with the *basic* decoder (for a more detailed description of the HOA playback, please refer to [2]).

Measurement procedure

After initializing the playback loop, the microphone position and a time stamp were prompted from the motion capture system. After the updated EQ filters were computed, they were sent to the convolver engine of the rtSOFE room simulation and auralization software [8] via TCP. After the TCP message was sent, the loop was restarted, leading to a microphone position update. The loop took, on average, 16 ms, with minimal variance during and across measurements.

Running in parallel, the convolver read the white noise signal from a file, convolved it with the current EQ filter matrix, and streamed the output to the soundcard. Newly received filters were linearly cross-faded over 256 samples.

The microphone signal was recorded synchronously and bandpassed between 100 Hz and 15 kHz. The measured sound pressure level was analyzed in 125 ms frames, centered around the position update time from the motion tracking system.

Results

Sound pressure level

Figure 1 shows the measured sound pressure level (SPL) error inside the loudspeaker array, following the path of the tracked microphone.

For the static EQ, we observe SPL errors from -5 dB to 3 dB, the absolute SPL decreasing with increasing distance to the loudspeaker, following the 1/R law. When applying the dynamic EQ, these errors are reduced to -2 dB to 1 dB, with slightly larger errors towards the sides of the loudspeaker array.

Figure 2 shows the measured SPL error for the HOA playback. We observe errors ranging from -6 dB to 4 dB for the static EQ. They also follow the 1/R decay closely but show more local variations. Especially the 13° axis shows a slightly higher SPL than the surroundings. The dynamic EQ reduces the error range from -3 dB to 2 dB. The distribution of SPL errors also changes, with positive errors appearing on the right-hand side and negative errors on the left-hand side of the 0° axis, showing higher SPLs closer to the main loudspeakers.

Figure 3 shows the SPL error distribution as a function of distance from the center of the loudspeaker array. For the static EQ, both with single loudspeaker and HOA playback, the error range spreads out as distance increases, which is due to the directional nature of the 1/R decay. At the same distance from the array center, points closer to the main loudspeaker show a higher SPL than points further away. The correlation between SPL error and distance from the center nearly disappears when applying the dynamic EQ. There is a tendency towards negative errors, i.e., lower SPLs over the area, which is likely due to destructive interference.

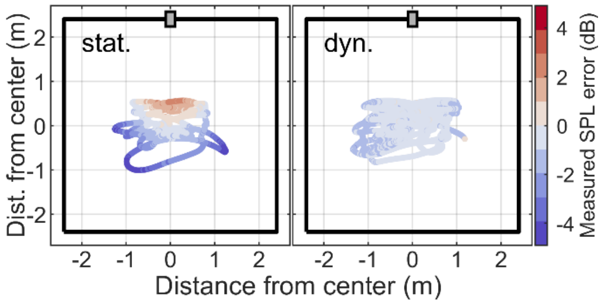

Figure 1: Dynamically measured broadband SPL errors inside the loudspeaker array with single loudspeaker playback. Each dot represents a microphone position, the color indicating SPL error. The left panel shows the static EQ and the right panel shows the dynamic EQ.

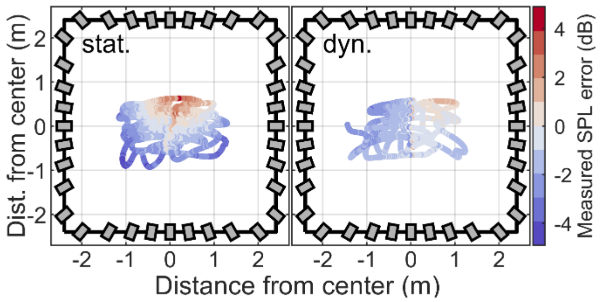

Figure 2: Dynamically measured broadband SPL errors inside the loudspeaker array with HOA playback. Each dot represents a microphone position, the color indicating SPL error. The left panel shows the static EQ and the right panel shows the dynamic EQ.

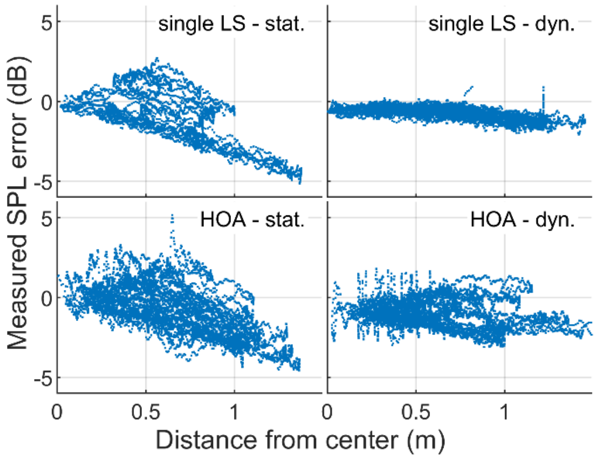

Figure 3: SPL errors for different EQ and playback methods as a function of distance from the center of the loudspeaker array.

Effect of microphone speed

Figure 4 shows the distribution of errors as a function of microphone speed. The speed was computed for a given update frame by considering the distance traveled and time elapsed since the previous position update. The errors are independent of microphone speed, with correlation coefficients below 0.2.

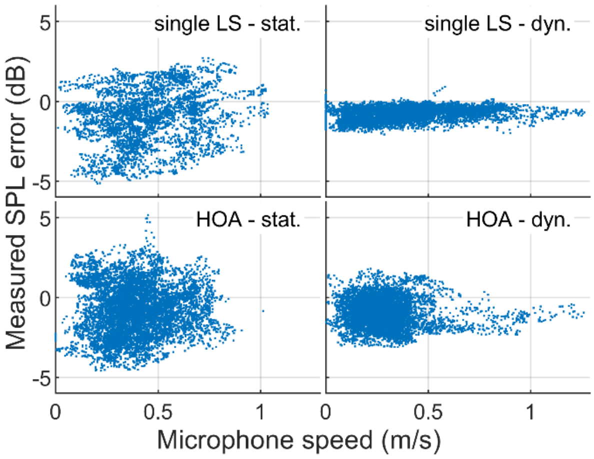

Figure 4: SPL errors for different EQ and playback methods as a function of microphone speed.

Discussion

Measurements of sound pressure level show that dynamically changing loudspeaker EQ filters in a loudspeaker array reduces the observed SPL errors over a wide area. Our system's update rate of 16 ms is fast enough to keep observed errors small.

We observe higher errors for the HOA playback than for the single loudspeaker. This is due to strong destructive interferences at high frequencies, where the sweet-spot is too small to be useful. This can also be observed in the SPL map (Figure 2, left panel), where points next to the 13° axis show a slightly lower SPL due to destructive interferences.

The dynamically equalized HOA playback, shown in Figure 2 and Figure 3, yields an unexpected error pattern. Points close to the 0° axis show slightly higher errors than their surroundings, with deviations up to 1.5 dB, which seem to also originate from high frequency interference. However, it is unclear why that pattern follows the 0° axis in particular.

While the current method of computing filters from a directivity measurement shows promising results, the errors could probably be reduced even further by adding a reference measurement in the center of the array. This additional measurement can equalize differences between the individual loudspeakers and ensure they are entirely phase-aligned in the center.

Conclusion

This work shows that errors in the reproduced sound pressure level in a loudspeaker array over single loudspeaker playback or HOA can be reduced by adapting the loudspeaker equalization filters to the microphone position. This yields promising results for the application in hearing experiments, where the EQ filters would be changed based on a listener's position. The FIR filter changes can be handled by the rtSOFE convolver, ensuring a high update rate and making an integration with the room simulation and auralization straightforward.

Acknowledgements

The SOFE system was financed by the BMBF Bernstein Center for Computational Neuroscience, 01GQ1004B. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 352015383 – SFB 1330 C 5.

Literature

[1] Seeber, B.U., Clapp, S.: Interactive simulation and free-field auralization of acoustic space with the rtSOFE. The Journal of the Acoustical Society of America 141(5) (2017), 3974.

[2] Kuntz, M., Bischof, N.F., Seeber, B.U.: Sound field synthesis for psychoacoustic research: In situ evaluation of auralized sound pressure level. The Journal of the Acoustical Society of America 154(3) (2023), 1882-1895.

[3] Hahn, N., Winter, F., Spors, S.: Local wave field synthesis by spatial band-limitation in the circular/spherical harmonics domain. Proc. of the 140th AES Convention (2016), paper 9596.

[4] Spors, S., Ahrens, J.: Local sound field synthesis by virtual secondary sources. Proc. of the 40th AES International Conference (2010), paper 6-3.

[5] Kirkeby, O., Nelson, P.A., Orduna-Bustamante, F., Hamada H.: Local sound field reproduction using digital signal processing. The Journal of the Acoustical Society of America 100(3) (1996), 1584-1593.

[6] Kolundžija, M., Faller, C., Vetterli, M.: Reproducing sound fields using MIMO acoustic channel inversion. Journal of the Audio Engineering Society 59(10) (2011), 721-734.

[7] Kuntz, M., Seeber, B.U.: Moving the Ambisonics sweet-spot by adapting loudspeaker equalization and Ambisonics decoding. Proc. of Forum Acusticum 2023 (2023), 1805-1808.

[8] Seeber, B.U., Wang, T.: real-time Simulated Open Field Environment (rtSOFE) software package. (2021), available on Zenodo: 10.5281/zenodo.5648305.