Lindfors, Joel; Liski, Juho; Välimäki, Vesa

User Position-Based Loudspeaker Correction

Published in:
Proceedings of the 2nd NordicSMC conference

Published: 01/01/2021

Document Version
Publisher's PDF, also known as Version of record

USER POSITION-BASED LOUDSPEAKER CORRECTION

Joel LINDFORS (joel.lindfors@aalto.fi) [1], **Juho LISKI** [1], and **Vesa VÄLIMÄKI** [1]

[1] *Acoustics Lab, Dept. of Signal Processing and Acoustics*, **Aalto University**, Espoo, Finland

ABSTRACT

In our work, we present and study a novel loudspeaker correction system. This correction system uses the location of the user to determine the calibration parameters. By correcting for the loudspeaker's response at multiple locations and changing the calibration in real-time based on the user's location, we expect a less colored frequency response compared to no applied calibration and static calibration methods. The developed method, User Position-Based Loudspeaker Correction (UC), produces a flatter frequency response than that of no applied calibration: for example, in one of the measurement conditions the averages of the ranges of the frequency response went from 10.3 dB in the non-corrected setting to 4.7 dB in the UC setting. Further, it is shown to outperform a static method of correcting for the frequency response of each point in space by using calibration derived from measurements from a predetermined listening position. Finally, by interpolating the EQ gains for the calibration from a set of measurements for the suggested correction method, the system's resolution could be increased with the resulting calibration still outperforming the static correction method.

1. INTRODUCTION

Truthful audio reproduction is essential for a multitude of media applications, ranging from media consumption and media production to scientific purposes (e.g., listening tests). However, listening with loudspeakers has its downsides. The overall coloration of the system results primarily from loudspeaker setups being inside acoustically imperfect rooms [1].

One approach for further mitigating the effect of the room on sound reproduction is to use loudspeaker calibration [2]. Speaker calibration uses a calibration filter based on an onsite measurement of the loudspeaker setup. Correction can be done based on a single-point measurement [3] or on an average of multiple measurements around the room [4]. Previously loudspeaker calibration was done by using analog graphic EQs to counteract any peaks and troughs in the frequency response of the whole listening setup [5], but recently automated loudspeaker calibration software that tailors the equalization with little human interaction has been the standard (for example [6]).

While used by many in domestic and professional domains, current state-of-the-art loudspeaker correction has a significant defect. While correcting for the coloration of the converter, amplifier, cables, loudspeaker and average or approximate acoustical imprint of the room, the listener's position significantly affects the perceived sound [1, 2]. This effect is caused by the room dimensions and construction, irregular surfaces inside the room (e.g., furniture) and the directivity pattern of the loudspeaker. These properties behave in a seemingly sporadic manner as the listener moves within the room [1].

Some methods have been developed for localized audio reproduction. A linear loudspeaker array was used for dynamic audio reproduction in [7] by tracking a moving head in front of 28 loudspeakers. The authors used a Microsoft Kinect to track the head in space and created dynamic filters to create listener-adaptive audio reproduction. In [8], room reflection compensation for a similar loudspeaker array was studied. This system also uses the user's position with respect to the loudspeaker array to work. However, to the best knowledge of the author, no studies have attempted to use spatial tracking of the user to calibrate a stereo loudspeaker setup.

In order to mitigate the issues present in static loudspeaker correction systems, this work aims to develop software for tracking the user in a 3D space by implementing spatial tracking for the measurement and calibration system. This tracking system uses a depth-sensing camera to track the measurement process and the listener in three dimensions in real-time. The loudspeaker is calibrated based on the measurements for each spatial position using a standard whitening process with a state-of-the-art multi-band graphic EQ [9]. The measurements are verified by testing the equalization in the measurement positions to establish the accuracy of the method. A program is implemented to test the correction in real-time, though significant subjective tests will not be performed due to time constraints.

This paper is organized as follows. Section 2 presents the principle behind the suggested correction system. Section 3 defines the measurement methodology. Section 4 shows the measurement results. Finally, the conclusions are presented in Section 5.

2. USER POSITION-BASED LOUDSPEAKER CORRECTION

2.1 Variable Loudspeaker Calibration

The developed method, User Position-Based Loudspeaker Correction (UC), allows changing the corrective equalization of the loudspeakers to correspond with the position of the receiver.

For individual points in space, the calibration is done much like loudspeaker calibration is done conventionally. In this study, we use the Swept-Sine-Method (SSM) outlined by Farina to measure the frequency responses in all experiments [10]. In short, a sine sweep, increasing exponentially in frequency, is fed into the system and recorded with a measurement microphone. The measured output is then deconvolved to extract the frequency balance of the system. Finally, the frequency response is smoothed with one-third octave smoothing.

After the analysis, a whitening EQ counteracting perturbances in the frequency balance is inserted into the signal chain. A digital graphic EQ designed by Liski *et al.* [9] is used to extract Infinite Impulse Response (IIR) coefficients with low error in the inter-band crossover error. In an ideal scenario, this method should whiten the frequency balance of the loudspeaker system. This EQ takes the desired EQ gains as function parameters and outputs a set of IIR coefficients in 31 one-third octave bands that will accurately generate those gains in the predetermined center frequencies. The method uses an interaction matrix that contains information about how much each band leaks to other bands. With this information, the application can produce a set of IIR coefficients with vastly reduced errors for each band's gains.

The movement of the user (i.e. the microphone) is taken into account with object tracking. Figures 1 and 2 show the simulated tracking of the user where the ArUco marker's center position replaces the center of the head of the user in these measurements. In these figures, one can also witness the object tracking properly locating the user's nose - although for the measurements this capability is not used. Open source libraries (OpenCV and dlib) were used to extract the location of the receiver from a video feed. ArUco markers developed by Rafael Muñoz and Sergio Garrido were used to mark the position of the measurement microphone in the measurement and testing phases. ArUco markers allow estimating the position and the rotation of printed black-and-white markers (various such markers are depicted in Fig. 3). An Intel D415 depth-sensing camera [11] was used for the distance of the microphone from the camera itself, allowing for object tracking in three dimensions.

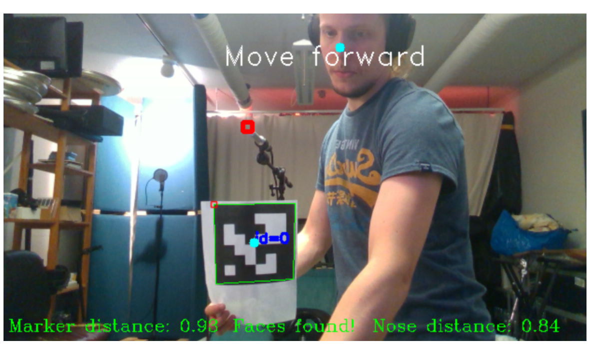

Figure 1: Screenshot of the program when the marker is not positioned properly.

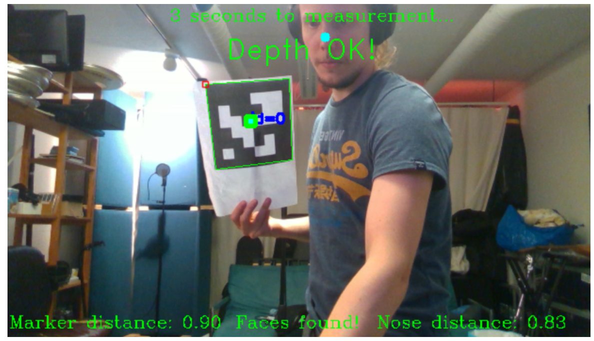

Figure 2: Screenshot when the marker is positioned properly.

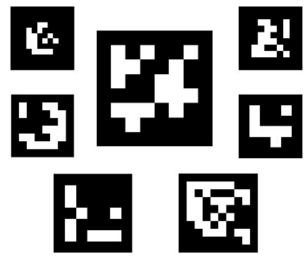

Figure 3: Examples of ArUco markers. (Adopted from [12].)

2.2 Interpolation of Equalizer Gains

In this paper equalizer gain interpolation was also implemented and tested. Although hypothetically inferior in performance to the UC, interpolation methods have a clear advantage in practice. Measuring the frequency responses for each point can be very labour intensive and error-prone in itself. Increasing mesh size to improve the resolution of the UC grows to impractical proportions when the mesh width, height and depth (as in, how many points are measured in each dimension) are increased beyond the measurement resolution. Moreover, interpolation could provide a continuous experience for the real-time application of the UC. The step-wise nature of the correction changing in real-time is audible for the user and could be viewed as distracting from the benefits of the correction method.

A total of 4 different interpolation methods were tested: trilinear, cubic, modified Akima piecewise cubic Hermite, and spline interpolation. These four interpolation methods shed light on the overall performance of interpolating the gains for UC. For the sake of keeping things neat and tidy, only the results for the best performing interpolation method (trilinear) are shown.

2.3 Real-time implementation

Although these results only consider offline measurements, a real-time version of the program was developed to a point where the principle could be tested. However, no statistically significant subjective evaluation of the program was made due to time constraints. In the real-time version of the program no particular artifacts of the filters changing could be detected by the author. Adding to this, the program was able to run both the object tracking algorithms and the 31-band IIR EQ filter in real-time with no apparent issues in computation.

3. MEASUREMENT METHODOLOGY

The primary aim of this study is to quantitatively assess the effectiveness of UC. In the measurements, the performance of the UC against a non-corrected condition (NC) and other related methods was evaluated. These related methods were a static method of correction (SC), UC, and methods with equalizer gain interpolation). Static correction refers to the conventional method of correction, where the loudspeakers are corrected based on a single measurement at the listening position. The full testing routine can be characterized by four steps:

1. Measurement. The measurement program is used to get the location-specific swept sine responses.
2. Equalizer filter design. A Matlab script is ran to compute the one-third octave band filter coefficients either on the location of the measured sweep or an interpolated version from multiple locations.
3. Equalizer testing. A measurement program is used again utilizing the designed filters.
4. Analysis. The test results are analyzed.

We performed three quantitative inquiries. The tests used were the following:

1. On-location performance. For this test, filters were generated for each measured location and then tested for performance. This test measures the absolute performance of UC.
2. Conventional loudspeaker correction performance (SC). Here the equalizers generated from the listening position were also tested in other points in the mesh.
3. Interpolation performance. For this test, multiple sets of measurements were used to compute interpolated filter coefficients. These coefficients were tested in the intended locations for performance.

The UC measures the points in a 3D mesh depicted in Fig. 4. The performance of the UC is evaluated in each point of the mesh. The interpolation performance is evaluated in a 2x2x2-mesh inside the depicted 5x5x5-mesh. The points in the mesh were 5 cm from each other for this particular measurement location.

4. MEASUREMENT RESULTS

The test results were measured in three different rooms. By using multiple rooms, different aspects of the performance of the correction could be isolated. For this paper, results only from one of the rooms are presented. The room is a recording studio workshop with a reasonable amount of acoustic treatment.

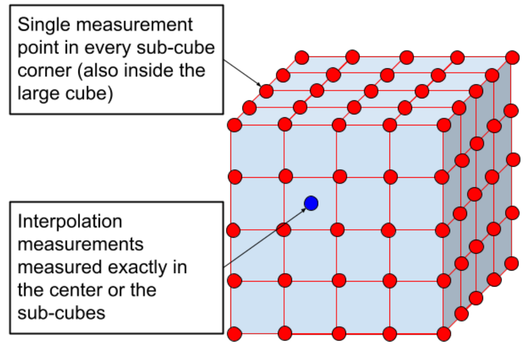

Figure 4: The cuboid structure of the measurements. The taxonomy of the measurement points can be seen in selected points.

Figure 5: A picture of the recording studio tests.

4.1 Single-point and Average Performance

The UC performed better than NC. An example of this is depicted in Fig. 6, which shows the uncorrected and corrected frequency responses in a single point that the UC corrects for. UC adequately finds the perturbances in the frequency response and applies counteracting equalizers for each frequency band resulting in a overall flatter frequency response and a smaller range (the difference between the lowest and the highest decibel value). Ranges for the NC and UC in this point can be seen in Table 1.

The average of the ranges is obtained by computing the range of each single-point response and then averaging over the number of measurement points. In one of the measurement locations (a recording studio room), the average of the decibel ranges in each point for the left channel went from 10.32 dB (NC) to 4.91 dB (UC), marking a 5.41 dB reduction between the lowest and the highest decibel value in a 35Hz–20000 Hz range. This reduction in the average of the ranges grants insight into the overall performance of the UC when compared to NC. These results are depicted in Fig. 7.

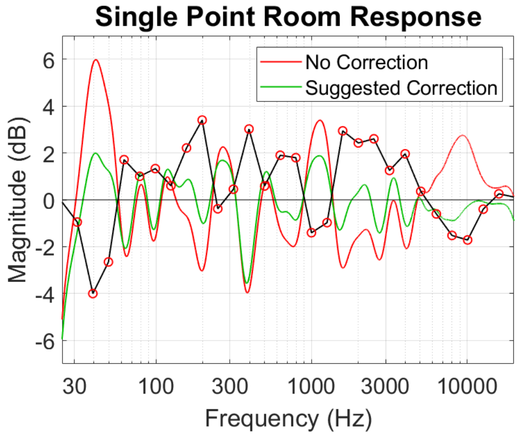

Figure 6: Frequency response of the left loudspeaker with and without correction in a single point in the mesh. The point here is measured in the leftmost, topmost and backmost point in the mesh from the listener's perspective. The EQ gains and the expected EQ cascade response are denoted with red circles and the black line.

Bandwidth	35–20000 Hz	100–20 kHz
NC	9.96 dB	7.37 dB
UC	**5.56 dB**	**5.44 dB**

Table 1: Ranges of NC and UC for the left loudspeaker. The point here is measured in the leftmost, topmost and backmost point in the mesh from the listener's perspective.

4.2 Evaluation and Comparison With Other Methods

The responses for the UC and SC for the left loudspeaker can be viewed in Fig. 8. Although both correction methods are working relatively well, the UC seems to be doing a better at evening out the frequency response. For the left loudspeaker, the range for frequencies between 35–20000 Hz is 6.6 dB for SC and 4.8 dB for UC.

The responses for the UC and Linear Interpolation Correction (LIC) for the left loudspeaker can be viewed in Fig. 9. While both the UC and the LIC are performing well, the UC still produces a flatter frequency response as hypothesized.

In Fig. 10, the average decibel ranges for each method are shown. Confirming the hypothesis, the order of performance is (from best to worst), UC, LIC, SC and NC.

5. CONCLUSIONS

In this study, a loudspeaker calibration system tracks the user's location and aims to calibrate the loudspeakers based on that was implemented and studied. The tracking system uses a depth-sensing camera, object tracking, and a state-of-the-art graphic EQ design. A 3D mesh is defined in space where the program is operational. Within this mesh, individual points are defined in a cuboid structure, where the corners of each sub-cuboid are calibration points that translate to physical locations inside the room. The program requires a measurement phase to estimate the initial conditions in the room for all of these points. These measurements can be used as is, or they can be used to generate interpolated points between them to increase the resolution of the 3D mesh.

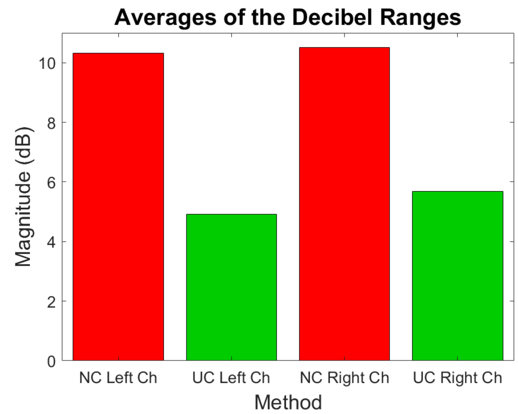

Figure 7: Averages of the decibel ranges in the frequency response for the uncorrected condition and UC in the frequency range of 35–20000 Hz.

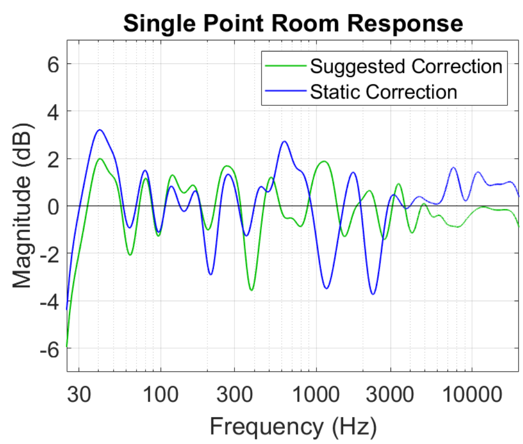

Figure 8: Frequency response of the studio room with UC and SC for the left loudspeaker. This point is in the leftmost, topmost and backmost point in the interpolation mesh.

The system was compared to a non-corrected system, a system using static correction and systems using the suggested correction system with interpolated EQ gains. The system was shown to measurably improve the average frequency response of the loudspeakers with respect to the non-corrected system. Data supporting the system used with interpolated EQ gains outperforming a SC system was presented. Tests evaluating subjective improvements in sound remain to be done.

In the future, research should focus on the subjective measurement of the system. This includes further developing a version of the program that is capable of real-time play and blind listening tests.

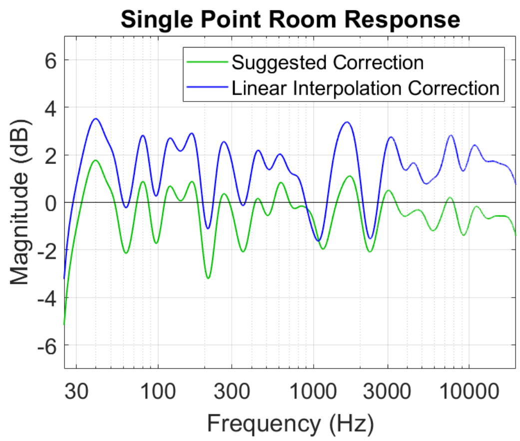

Figure 9: Frequency response of the studio room with UC and SC for the left loudspeaker. This point is in the leftmost, topmost and backmost point in the interpolation mesh.

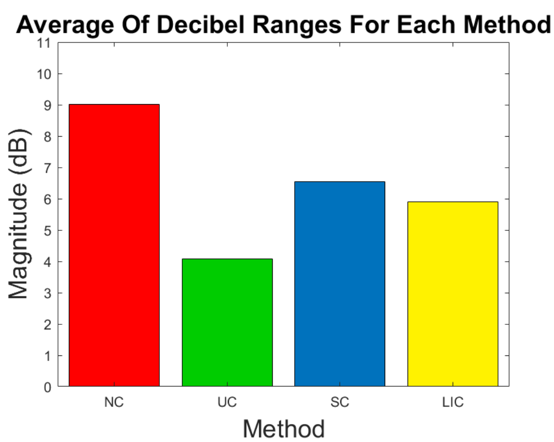

Figure 10: Averages of each method in the frequency range of 50–20000 Hz. Smaller value is better.

Acknowledgments

This work was done in the Aalto University Acoustics Lab between January 1st and July 31st of 2021. This work has been funded by the "Nordic Sound and Music Computing Network—NordicSMC", NordForsk project number 86892.

6. REFERENCES

[1] L. L. Beranek and T. Mellow, *Acoustics: Sound Fields and Transducers*. Academic Press, 2019.

[2] F. Toole, "The measurement and calibration of sound reproducing systems," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 512–541, 2015.

[3] P. M. Clarkson, J. Mourjopoulos, and J. K. Hammond, "Spectral, phase, and transient equalization for audio systems," *J. Audio Eng. Soc.*, vol. 33, no. 3, pp. 127–132, Mar. 1985.

[4] R. Wilson, "Equalization of loudspeaker drive units considering both on- and off-axis responses," *J. Audio Eng. Soc.*, vol. 39, no. 3, pp. 127–139, Mar. 1991.

[5] D. Ruch, "How to EQ a PA System," Mar. 31 2021. [Online]. Available: https://daveruch.com/advice/how-to-eq-a-pa-system/

[6] Genelec, *GLM 4 System Operating Manual*. Genelec, Nov. 2021.

[7] M. F. S. Gálvez, D. Menzies, and F. M. Fazi, "Dynamic audio reproduction with linear loudspeaker arrays," *J. Audio Eng. Soc.*, vol. 67, no. 4, pp. 190–200, 2019.

[8] M. S. Galvez and F. M. Fazi, "Room compensation for binaural reproduction with loudspeaker arrays," in *EuroRegio*, 2016.

[9] J. Liski, J. Rämö, and V. Välimäki, "Graphic equalizer design with symmetric biquad filters," in *Proceedings of the 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019, pp. 55–59.

[10] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[11] Intel, "Intel® RealSense™ Depth Camera D415," Jul. 12, 2021. [Online]. Available: https://www.intelrealsense.com/depth-camera-d415/

[12] OpenCV, "Detection of ArUco Markers." [Online]. Available: https://docs.opencv.org/4.5.2/d5/dae/tutorial_aruco_detection.html