OPEN ACCESS Freely available online
J. Lindfors, J. Liski, and V. Välimäki,
"Loudspeaker Equalization for a Moving Listener"
J. Audio Eng. Soc., vol. 70, no. 9, pp. 722–730, (2022 September).
DOI: https://doi.org/10.17743/jaes.2022.0020

Loudspeaker Equalization for a Moving Listener

JOEL LINDFORS,[1], JUHO LISKI,[1] AND
(lindforsjoel@gmail.com) (juho.liski@alumni.aalto.fi)

VESA VÄLIMÄKI,[1]* *AES Fellow*
(vesa.valimaki@aalto.fi)

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

When a person listens to loudspeakers, the perceived sound is affected not only by the loudspeaker properties but also by the acoustics of the surroundings. Loudspeaker equalization can be used to correct the loudspeaker-room response. However, when the listener moves in front of the loudspeakers, both the loudspeaker response and room effect change. In order for the best correction to be achieved at all times, adaptive equalization is proposed in this paper. A loudspeaker-correction system using the listener's current location to determine the correction parameters is proposed. The position of the listener's head is located using a depth-sensing camera, and suitable equalizer settings are then selected based on measurements and interpolation. After correcting for the loudspeaker's response at multiple locations and changing the equalization in real time based on the user's location, a loudspeaker response with reduced coloration is achieved compared to no calibration or conventional calibration methods, with the magnitude-response deviations decreasing from 10.0 to 5.6 dB within the passband of a high-quality loudspeaker. The proposed method can improve the audio monitoring in music studios and other occasions in which a single listener is moving in a restricted space.

0 INTRODUCTION

Reliable audio reproduction is essential for a multitude of applications, ranging from media consumption and music production to scientific purposes, such as listening tests. High-quality loudspeakers and accompanying equipment, including converters, amplifiers, and cables, all operate with a relatively flat frequency response. However, when the listener moves with respect to the loudspeaker, the response always varies. This work proposes a method to equalize the coloration caused by such slight changes in the listening position.

A major portion of the overall coloration of a sound reproduction system results from room modes and reflections [1, p. 550]. Other less-significant causes of coloration include amplifiers, cables, and the loudspeakers themselves. The effect that the room has on the sound varies depending on the room acoustics [1, pp. 511–545], including the size and shape of the room, the acoustic treatment installed, and any surfaces that are placed inside the room.

One approach for mitigating the room effect on sound reproduction is to use loudspeaker equalization, which appropriately prefilters the input audio signal [2, 3]. Loudspeaker equalization whitens the frequency response of the loudspeaker based on an on-site frequency response measurement. Automated loudspeaker calibration software that tailors the equalization with little human interaction are a commonplace solution in professional recording studios. Various commercial products utilize different methods for calibration. For example, one such product measures the room impulse response in the listening position to calibrate its loudspeakers [4], whereas another product measures multiple points around the listening position to create an average correction for optimizing the loudspeakers [5].

Although correcting for the coloration of the converter, amplifier, cables, loudspeaker, and the average or approximate acoustical imprint of the room, the listener's position significantly affects the frequency response of the system [3; 1, p. 515]. Peaks and nulls, caused by standing waves that form based on the room dimensions, behave in a seemingly sporadic manner as the listener moves within the room. It is known that the magnitude-response peaks especially are easily noticeable, leading to coloration [6]. Similarly, furniture and other inconsistent surfaces create reflections that bring about a multitude of filtering effects, further unbalancing the frequency response. Additionally, every loudspeaker shows significant angle-dependent coloration, depending on the direction of the listener with respect to the loudspeaker [1, pp. 463–511].

*To whom correspondence should be addressed, e-mail: vesa.valimaki@aalto.fi. Last updated: May 9, 2022

Room inverse filters, whose purpose is to flatten the magnitude response of a loudspeaker in a specific room, have been studied extensively in the past 40 years [7–10]. A least-mean-squares or direct inversion of the room impulse response can be performed to arrive at a non-parametric room equalization design, in which autoregressive-moving average modeling can be used for a parametric model [11]. Karjalainen et al. [11] also differentiate between single and multiple position designs, where the latter can be achieved in a few different ways [12, 8]. In this paper, a localized multiple position design is presented.

Some methods have been developed for localized audio reproduction. Galvéz et al. used a linear loudspeaker array for dynamic audio reproduction by tracking a moving head in front of 28 loudspeakers [13]. Galvéz and Fazi have also studied room reflection compensation for a similar loudspeaker array [14]. This system also uses the user's position with respect to the loudspeaker array to work. However, to the best knowledge of the authors, no studies have attempted to use spatial tracking of the user to calibrate a stereo loudspeaker setup.

This paper proposes how to mitigate the irregularities of the loudspeaker response that changes with the user's location. It is suggested to first measure the loudspeaker response at many points in the listening area, design an appropriate equalizing filter for each point, and use computer vision for tracking the user's movements. This allows selection of the best equalizer at every moment. A depth-sensing camera is used to track the measurement process and the listener in three dimensions in real-time.

In this work, 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 equalizer (EQ) [15, 16]. A cascade structure was selected for this work, but a parallel one is also an option [17]. The accuracy of the method is verified by testing the equalization at the measurement points and between them. A program was also implemented to test the head tracking and equalization in real-time using a regular PC. Some results from this work have previously been presented in a Master's thesis [18] and student conference [19].

The rest of this paper is organized as follows. SEC. 1 presents the principles of the developed correction system. SEC. 2 describes the measurement methodology. SEC. 3 shows the measurement results from the proposed method and non-corrected case. SEC. 4 outlines the results between the proposed method and other correction methods. Finally, conclusions are given in SEC. 5.

1 USER POSITION–BASED LOUDSPEAKER CORRECTION

The perceived frequency response of a sound reproduction system is highly dependent on the position of the user with respect to the loudspeakers. Room modes, room reflections, and loudspeaker directivity all govern ways by which the response of a speaker changes, when the listener moves to a different location. A novel method, User Position–Based Correction (UC), alleviates these perturbations by taking the user's position into account when applying the loudspeaker correction. The intended application for this system is a modern audio production studio, where the user's head movements are slight, and thus, the user's head rotations are not accounted for.

1.1 Variable Loudspeaker Correction

For individual points in space, the equalization is realized much like conventional loudspeaker calibration. This study uses the swept-sine method to measure the frequency responses in all experiments [20]. In short, a 4-s sine sweep, increasing exponentially in frequency, is fed into the system and recorded with a measurement microphone. The measured signal is then deconvolved with the sine sweep to extract the impulse response of the system. The frequency response is obtained by applying the discrete Fourier transform to the impulse response. Finally, the frequency response is smoothed with one-third octave smoothing for display.

After the analysis, a whitening EQ that counteracts perturbations in the frequency balance is inserted into the signal chain. A recent digital third-order graphic EQ design is used to compute infinite impulse response (IIR) filter coefficients with low error [16]. The EQ comprises minimum-phase filters, and therefore, they cause only minimal time delay. However, the phase responses of these filters are not linear, and some phase distortion may occur. In this work, the effects of the phase distortion are considered to be negligible. If any phase distortion is seen as problematic, a recent linear-phase graphic EQ is an option [21].

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 [22–24] 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. For the lowest frequency bands up to 50 Hz in the anechoic chamber and 30 Hz in the recording studio room measurements, the command gains were fixed to 0 dB. These frequencies were selected based on the cutoff frequency of the speakers that were used in both situations.

The movement of the receiver is taken into account with object tracking. ArUco markers developed by Muñoz and Garrido are used to define the position of the measurement microphone in the measurement and testing phases [25]. A printed black-and-white ArUco marker allows estimating its position and rotation from a camera image. The Intel D415 depth-sensing camera [26] was used for determining the microphone distance from the camera itself, allowing for object tracking in three dimensions. Open source libraries (OpenCV and dlib) were used to extract the location of the receiver from a video feed [27, 28]. These libraries include functions that allow for resource-efficient tracking

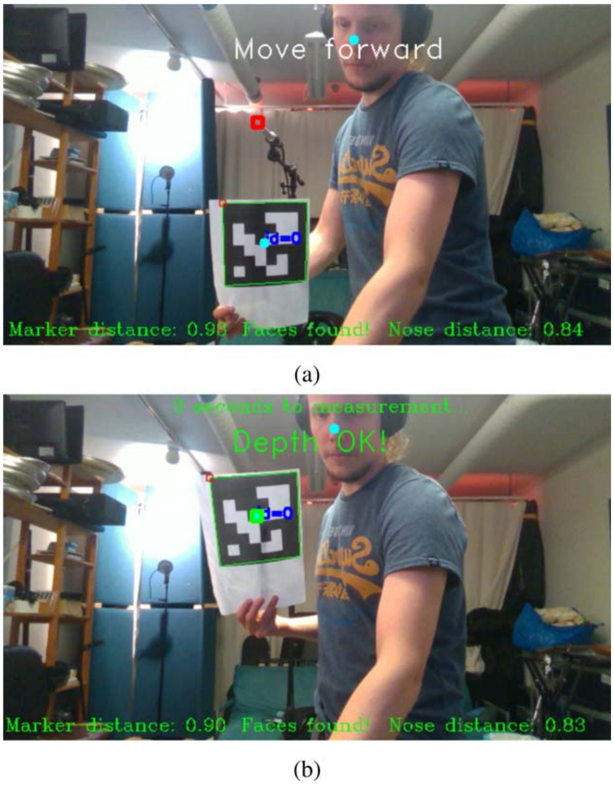

Fig. 1. Screenshot of the program (a) when the marker is at a wrong place and (b) when the marker is positioned properly.

of both the ArUco markers and facial landmarks for video sources.

Figs. 1(a) and 1(b) show the simulated tracking of the user, in which 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 this capability was only used for the real-time implementation of the program and not the measurements.

1.2 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 labor-intensive and error-prone in itself. Increasing the mesh size to improve the resolution of the UC method grows to impractical proportions if the distance between measurement points is the same as the measurement resolution. The mesh is a three-dimensional cuboid with a width, height, and depth. A rendition of the mesh can be seen in Fig. 2. Moreover, interpolation provides a continuous experience for the real-time application of the UC. The step-wise changing of the correction in real-time is audible for the user and reduces the benefits of the correction method.

Four different interpolation methods were tested: trilinear, cubic, modified Akima piecewise cubic Hermite, and spline interpolation [29–32]. The differences between the performance of the four interpolation methods were small.

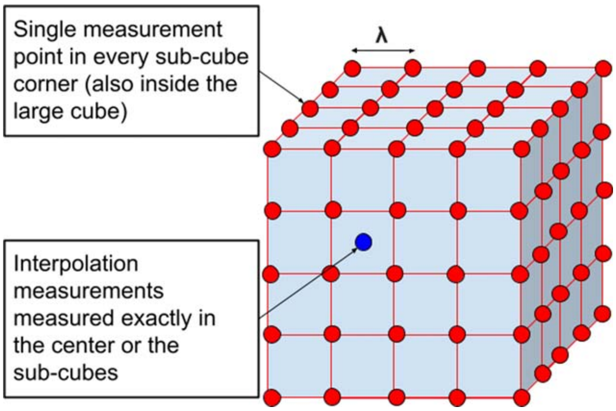

Fig. 2. Cuboid structure of the measurement mesh. The equalizer coefficients for static correction (SC) were measured from the center of this cube. The distance λ between two measurement points (dots) was 5 cm in all directions for the recording studio measurements and 35, 15, and 15 cm in width, height, and depth dimensions, respectively, for the anechoic chamber measurements.

For the sake of simplicity, only the results for the best-performing interpolation method (trilinear) are shown in this paper, see SEC. 4. The results for the rest of the interpolation methods are found in [18].

1.3 Real-Time Implementation

A real-time version of the UC program was developed to a point in which the principle could be tested in practice. The PC used for testing has a Windows 10 operating system with the Intel Core i5-10310U CPU system with 16 GB of RAM and integrated Intel UHD graphics. With the required image and audio processing running, the program is able to perform in real time at the audio sampling rate of 44.1 kHz. This included both the head tracking algorithms and two 31-band IIR EQ filters, one for each stereo channel. In the real-time version of the program, no artifacts caused by filter parameter changes could be detected. This was tested by using different crossfade times and ensuring that the selected value did not cause any audible artefacts in informal listening sessions.

The real-time implementation uses the open-source framework JUCE [33] for audio processing, a machine learning toolkit dlib [28], and a computer vision toolkit OpenCV [27] for image processing. The real-time implementation applies a previously computed whitening EQ filter to the loudspeakers' two audio streams, based on the current position of the user. When the user moves enough for the filter to change, a short linear crossfade is applied to avoid discontinuities. The crossfade length of 100 ms was sufficient to suppress audible artifacts. The rest of this paper reports results of offline measurements, carried out with a separate measurement program, which is otherwise identical to the real-time version.

For the sake of demonstrating the subjective experience of the UC method, the authors have made videos that allow one to hear what the correction system sounds like in-situ. To contrast this, the videos also show the non-corrected condition. These videos can be found online [34]. The use of headphones is advised for the best experience.

2 MEASUREMENT METHODOLOGY

The focus of this study is to assess the steady-state objective performance of the proposed UC method. This section explains the test procedure in which the UC method is compared with three other equalization methods. The tested rooms and spatial resolution of the listener tracking are also described.

Differences in the measurement setups are tailored to underline specific research questions. For example, the anechoic chamber tests had large distances between the measurement points to assess a wider range of travel around the loudspeaker and thus measure the suppressing effect of the system on the loudspeaker's directivity pattern.

A third room, a listening space in Aalto Acoustics Lab, was also measured. The results were in line with the results from the recording studio and thus are not presented in this paper. These results can be observed in [18].

2.1 Testing Procedure

The effectiveness of the UC method against a non-corrected condition (NC), static correction (SC), and a method using linear interpolation of the EQ gain (LIC) was measured. The SC method corresponds to a conventional single-point correction, in which the loudspeakers are equalized based on a measurement at a single listening position. LIC refers to a UC-like calibration method, in which the EQ gains for a point between four measured UC points were linearly interpolated from them instead of using the data point itself, i.e., LIC equalizer gains are approximations of the UC equalizer gains and will likely perform worse than UC at that point.

The full testing routine for all tests 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 run 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 swept-sine signals processed with the designed filters.
4. Analysis: The test results are analyzed.

Three quantitative inquiries were performed. The tests used were the following:

1. On-location performance: 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. This equalizer is the same for all measurements for both measurement locations.
3. Interpolation performance: Multiple sets of measurements were used to compute interpolated filter coefficients. These coefficients were tested in the intended locations for performance.

Fig. 3. Measurement setup in the large anechoic chamber of the Aalto Acoustics Lab.

Fig. 4. Recording studio tests.

2.2 Test Rooms

The test results were measured in two different rooms. By using multiple rooms, different aspects of the performance of the UC method could be isolated. The rooms measured were the anechoic chamber in the Aalto University Acoustics Lab, Espoo, Finland (see Fig. 3), and the first author's music production workshop studio in Espoo, Finland (see Fig. 4).

Measuring the system in these two rooms allowed for assessing the UC's performance against different types of difficulties. The anechoic chamber tests were conducted to isolate the UC's performance in correcting for the loudspeaker's directivity pattern whereas the recording studio room tests evaluated the effectiveness of the system in a real-world scenario.

2.3 Spatial Resolution

The UC method measures the points in a 3D mesh depicted in Fig. 2. The mesh points were 5 cm apart for the studio room tests. For the anechoic chamber tests, the points were 35 cm apart from each other in the x dimension (horizontal/side-to-side) and 15 cm apart in the y dimension (vertical/up-down) and z dimension (depth/front-to-back). The performance of the system is evaluated at each point of the mesh. The interpolation performance and static correction filters are evaluated in a $2 \times 2 \times 2$ mesh inside the depicted $5 \times 5 \times 5$ mesh, neglecting every second point.

The UC system measures and calibrates each loudspeaker separately. For the anechoic chamber measurements, a single loudspeaker was used. In the recording studio tests, both loudspeakers in a stereo setup were separately measured in the same spatial locations.

The Intel D415 depth-sensing camera was used to track the user in three dimensions. With a proper framework for fetching the data from the camera (the authors used Intel RealSense SDK 2.0 [26]), there is no apparent latency in the video or depth feeds. The camera runs at 30 frames per second, and its field of view is $65^\circ \times 40^\circ$ (horizontal $\times$ vertical). The minimum operating distance for the camera is 45 cm. The depth sensor accuracy is stated to be 2% at 2 m, meaning that the error at the usual operating range is about 2 cm.

3 MEASUREMENT RESULTS

In this section, the UC method is tested for the correction of the different isolated factors contributing to the perturbed magnitude responses. These factors include the attenuation caused by the loudspeaker directivity and changing room modes, when the listener moves with respect to the loudspeaker system.

3.1 Compensating the Loudspeaker Directivity

The loudspeaker directivity pattern is one of the reasons for the frequency response changes with the listener's position. Measurements were carried out in an anechoic chamber to discern whether the system accurately corrects for the directionality of the loudspeaker. Fig. 3 shows the measurement setup from the anechoic chamber measurements. In these measurements, a Genelec 8331 speaker was measured and equalized using the UC method. Magnitude response curves included in the loudspeaker product manual show characteristic step-wise attenuation at high frequencies when the listening angle increases [35].

In Fig. 5, the loudspeaker was measured at a 30° angle with and without UC. From this experiment, most notably it can be seen that the slight high-end attenuation caused by the relatively wide listening angle is well compensated by the UC method. On top of this, some notable local peaks and nulls at around 600, 1,500, 5,000, and 13,000 Hz are corrected.

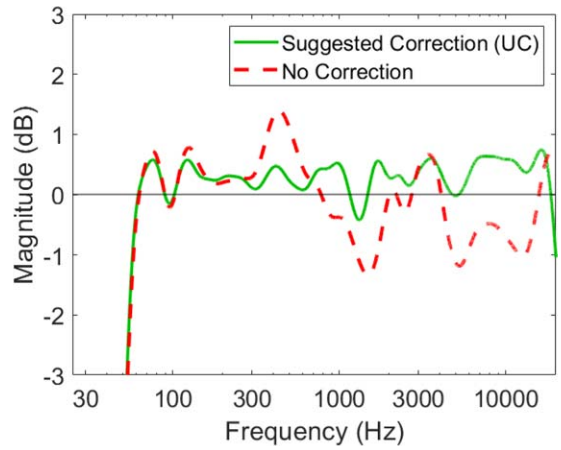

Fig. 5. Magnitude response with and without User Position–Based Correction (UC) at a single point in the anechoic chamber at a 30° angle.

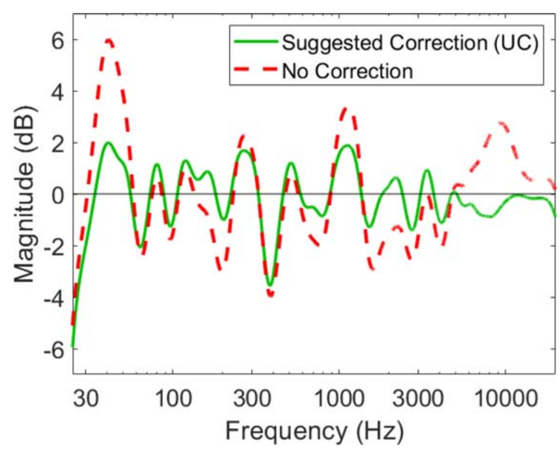

Fig. 6. Magnitude response of the left loudspeaker with and without correction at a single point in the recording studio. UC, User Position–Based Correction.

3.2 Room Modes and Reflections

When the listener moves inside the room, they move through a changing field of excited room modes. In addition to this, reflective surfaces induce comb filtering and partial absorption of frequencies that both depend on the receiver's relative location in the room.

In Fig. 6, the uncorrected and corrected frequency responses can be observed in a single point in the studio room. The UC method adequately finds the perturbances in the frequency response at this spatial position and applies counteracting equalizers for each frequency band, resulting in an overall flatter magnitude response and smaller range, i.e., the difference between the lowest and highest decibel value. Ranges for the NC and UC methods at this location are presented in Table 1. It should be noted that since nulls are caused by phase cancellation at specific frequencies, it is often considered implausible to completely correct for them with an equalizer. Two of the most common reasons for these nulls are standing waves and desktop reflections. In the recording studio measurements, the setup included

Table 1. Ranges for the User Position–Based Correction (UC) and non-corrected condition (NC) for the left-most, top-most, and hind-most point in the interpolation mesh in the recording studio. The best result in each frequency range is highlighted in bold.

Correction Method	35 Hz–20 kHz	100 Hz–20 kHz
NC	10.0 dB	7.4 dB
Suggested method (UC)	**5.6 dB**	**5.4 dB**

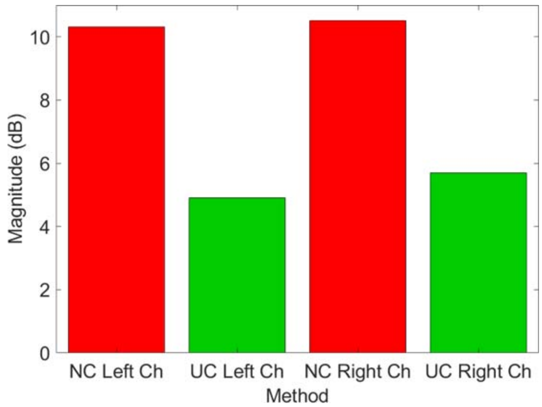

Fig. 7. Average magnitude response range for the uncorrected condition and after applying the proposed User Position–Based Correction (UC) method measured in the recording studio in the frequency range of 35–20,000 Hz. Smaller value is better. NC, non-corrected condition.

a desktop between the speakers and microphone, which, in conjunction with the nulls caused by the standing waves, will inherently result in undulation in the magnitude response of the calibrated condition.

3.3 Average Magnitude Response Range

The average range of a loudspeaker magnitude response is obtained by computing the range of each single-point response and then averaging 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 was reduced from 10.3 dB (NC) to 4.9 dB (UC), marking a 5.4-dB reduction between the lowest and highest decibel value in the 35 Hz–20 kHz range. This reduction in the average range grants insight into the overall performance of the UC method when compared with NC. These results are collected in Fig. 7.

4 COMPARISON WITH OTHER METHODS

This section presents a comparison of the proposed UC method with two other techniques (SC and LIC) and with the noncorrected loudspeakers. In addition to specific test cases, an average range of magnitude response deviation is presented for each method.

4.1 Static Correction

The responses for the UC and SC for the left loudspeaker in one point of the measurement mesh in the recording studio can be viewed in Fig. 8. Although both correction methods are working relatively well, the UC performs better at evening out the frequency response. The biggest difference between the two can be seen at 1 kHz, in which the frequency response of the SC shows a considerable peak. The magnitude response ranges for both methods are presented in Table 2.

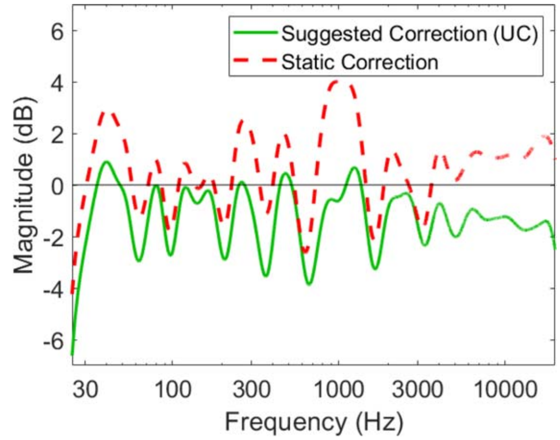

Fig. 8. Magnitude response of the left speaker of the recording studio room with proposed and static correction. This measurement is located in the left-most, top-most, and hind-most point in the measurement mesh and is roughly 22 cm from the center of the mesh. UC, User Position–Based Correction.

Table 2. Ranges for User Position–Based Correction (UC) and static correction (SC) in the left-most, top-most, and hind-most part of the interpolation grid in the recording studio room.

Correction	35–20,000 Hz	100–20,000 Hz
SC	6.6 dB	6.6 dB
Proposed (UC)	**4.8 dB**	**4.5 dB**

Table 3. Ranges for User Position–Based Correction (UC) and linear interpolation of the equalizer gain (LIC) in the left-most, under-most, and hind-most part of the interpolation grid.

Correction	35–20,000 Hz	100–20,000 Hz
LIC	5.1 dB	5.0 dB
Proposed (UC)	**5.0 dB**	**4.3 dB**

4.2 Correction With Interpolated Equalizer Gains

The responses for the UC and LIC for the left loudspeaker are shown in Fig. 9. These results were obtained from the recording studio measurements. Although both the UC and LIC perform well, the UC still produces a flatter frequency response. The trilinear interpolation that was used for the LIC measurements interpolates the EQ gain coefficients from four surrounding UC measurement points, which are all at a 0.43-cm distance from the LIC measurement point. The magnitude response ranges for both methods are presented in Table 3.

4.3 Average Range Comparison

In Fig. 10, the average decibel ranges from the recording studio measurements for each method are shown. This

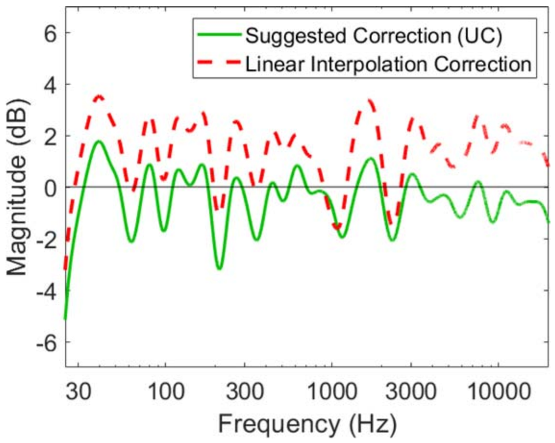

Fig. 9. Frequency response of the studio room with User Position–Based Correction (UC) and static correction for the left loudspeaker. This point is in the left-most, top-most, and fore-most point in the interpolation mesh. This result was obtained from the recording studio room.

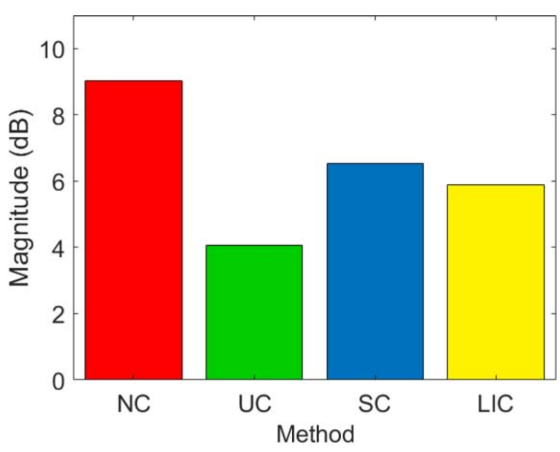

Fig. 10. Averages of each method in the frequency range of 50–20,000 Hz. Smaller values are better. LIC, linear interpolation of the equalizer gain; NC, non-corrected condition; SC, static correction; UC, User Position–Based Correction.

confirms the hypotheses that the UC method outperforms all other methods and that NC performs worse than any other method. The SC and LIC methods also provide some improvement over no correction, but it is only about 2 and 3 dB in average for the SC and LIC methods, respectively. However, the proposed UC method reduces the magnitude response variations by 5 dB in average.

5 CONCLUSION

A loudspeaker calibration system equalizing the magnitude response based on the user's location was introduced and studied. The calibration system uses a depth-sensing camera, object tracking, and a state-of-the-art graphic EQ design. The calibration mesh defines the listening zone and spatial resolution of the calibration. Within this mesh, individual points are defined in a cuboid, where the corners of each sub-cuboid are calibration points translating to physical locations in the listening zone. The calibration system requires a measurement phase to estimate the magnitude response of the speakers in the room for all of these points.

The proposed system was compared with 1) no correction, 2) static single-point correction, and 3) interpolated EQ gains. The proposed system was shown to measurably improve the average magnitude response of the loudspeakers with respect to all other systems. Although the reduction from NC in the average magnitude response ripple in the listening zone was several decibels for SC (a reduction of 3.6 dB) and LIC (4.3 dB), the proposed UC method (5.4 dB) clearly outperforms all other methods in this regard.

Room inverse filtering is widely used in a modern recording studio for precise audio monitoring. The proposed system improves on the existing technology, offering greater accuracy in audio reproduction for mixing engineers and hi-fi enthusiasts alike. Even small movements of the engineer over a mixing desk can counteract the room response measurements, hindering their confidence in their craft. State-of-the-art face tracking that utilizes pose estimation can be considered for the future iterations of the system, by integrating the rotation of the user's head into the correction [36]. With the EQ system outlined in this paper, the perturbations caused by the listener's movement can be significantly reduced.

6 ACKNOWLEDGMENT

This research has been funded by the Nordic Sound and Music Computing Network—NordicSMC (NordForsk project number 86892).

7 REFERENCES

[1] L. Beranek and T. Mellow, *Acoustics: Sound Fields, Transducers and Vibration* (Academic Press, Cambridge, MA, 2019).

[2] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, "Comparison of Loudspeaker Equalization Methods Based on DSP Techniques," *J. Audio Eng. Soc.*, vol. 47, no. 1/2, pp. 14–31 (1999 Jan./Feb.).

[3] F. Toole, "The Measurement and Calibration of Sound Reproducing Systems," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 512–541 (2015 Jul./Aug.). https://doi.org/10.17743/jaes.2015.0064.

[4] Genelec, *GLM 4 System Operating Manual* (Genelec, Iisalmi, Finland, 2021).

[5] Sonarworks, *Sonarworks Reference 4 User Manual* (Sonarworks, Riga, Latvia, 2017).

[6] R. Bücklein, "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, vol. 29, no. 3, pp. 126–131 (1981 Mar.).

[7] J. Mourjopoulos, "On the Variation and Invertibility of Room Impulse Response Functions," *J. Sound Vibr.*, vol. 102, no. 2, pp. 217–228 (1985 Sep.). https://doi.org/10.1016/S0022-460X(85)80054-7.

[8] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907 (1989 Nov.).

[9] S. Cecchi, A. Carini, and S. Spors, "Room Response Equalization—A Review," *Appl. Sci.*, vol. 8, no. 1, paper 16 (2018 Jan.). https://doi.org/10.3390/app8010016.

[10] S. Cecchi, A. Terenzi, V. Bruschi, A. Carini, and S. Orcioni, "A Subband Implementation of a Multichannel and Multiple Position Adaptive Room Response Equalizer," *Appl. Acoust.*, vol. 173, paper 107702 (2021 Feb.). https://doi.org/10.1016/j.apacoust.2020.107702.

[11] M. Karjalainen, T. Paatero, J. N. Mourjopoulos, and P. D. Hatziantoniou, "About Room Response Equalization and Dereverberation," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 183–186 (New Paltz, NY) (2005 Oct.). https://doi.org/10.1109/ASPAA.2005.1540200.

[12] S. Bharitkar and C. Kyriakakis, "A Cluster Centroid Method for Room Response Equalization at Multiple Locations," in *Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 55–58 (New Paltz, NY) (2001 Oct.). https://doi.org/10.1109/ASPAA.2001.969541.

[13] 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 Apr.). https://doi.org/10.17743/jaes.2019.0007.

[14] M. F. S. Gálvez and F. M. Fazi, "Room Compensation for Binaural Reproduction With Loudspeaker Arrays," in *Proceedings of EuroRegio* (Porto, Portugal) (2016 Jun.).

[15] J. Liski, J. Rämö, and V. Välimäki, "Graphic Equalizer Design With Symmetric Biquad Filters," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 55–59 (New Paltz, NY) (2019 Oct.). https://doi.org/10.1109/WASPAA.2019.8937168.

[16] J. Rämö, J. Liski, and V. Välimäki, "Third-Octave and Bark Graphic-Equalizer Design With Symmetric Band Filters," *Appl. Sci.*, vol. 10, no. 4, paper 1222 (2020 Feb.). https://doi.org/10.3390/app10041222.

[17] J. Liski, B. Bank, J. O. Smith, and V. Välimäki, "Converting Series Biquad Filters Into Delayed Parallel Form: Application to Graphic Equalizers," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3785–3795 (2019 Jul.). https://doi.org/10.1109/TSP.2019.2919419.

[18] J. Lindfors, *User Position-Based Loudspeaker Correction*, Master's thesis, Aalto University, Espoo, Finland (2021 Dec.).

[19] J. Lindfors, J. Liski, and V. Välimäki, "User Position-Based Loudspeaker Correction," in *Proceedings of the 2nd Nordic SMC Conference*, paper 17 (Online) (2021 Nov.).

[20] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), paper 5093.

[21] V. Bruschi, V. Välimäki, J. Liski, and S. Cecchi, "Linear-Phase Octave Graphic Equalizer," *J. Audio Eng. Soc.*, vol. 70, no. 6, pp. 435-445 (2022 Jun.).

[22] J. S. Abel and D. P. Berners, "Filter Design Using Second-Order Peaking and Shelving Sections," in *Proceedings of the International Computer Music Conference*, volume 2004 (Miami, FL) (2004 Nov.).

[23] R. J. Oliver and J.-M. Jot, "Efficient Multi-Band Digital Audio Graphic Equalizer With Accurate Frequency Response Control," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), paper 9406.

[24] V. Välimäki and J. Liski, "Accurate Cascade Graphic Equalizer," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 176–180 (2017 Feb.). https://doi.org/10.1109/LSP.2016.2645280.

[25] OpenCV, "Detection of ArUco Markers," https://docs.opencv.org/4.5.2/d5/dae/tutorial_aruco_detection.html (accessed Jul. 13, 2021).

[26] Intel, "Depth Camera D415," https://www.intelrealsense.com/depth-camera-d415/ (accessed May 11, 2022).

[27] G. Bradski, "The OpenCV Library," *Dr. Dobb's J. Software Tools*, vol. 25, no. 11, pp. 120, 122–125 (2000 Nov.).

[28] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758 (2009 Jul.).

[29] Yoder Dennis, "Trilinear Interpolation," https://www.grc.nasa.gov/WWW/winddocs/utilities/b4wind_guide/trilinear.html (accessed May 11, 2022).

[30] F. Lekien and J. Marsden, "Tricubic Interpolation in Three Dimensions," *Int. J. Numer. Methods Eng.*, vol. 63, no. 3, pp. 455–471 (2005 Mar.). https://doi.org/10.1002/nme.1296.

[31] C. Moler, "Makima Piecewise Cubic Interpolation," *MathWorks* (2019 Apr.). https://blogs.mathworks.com/cleve/2019/04/29/makima-piecewise-cubic-interpolation/.

[32] Encyclopedia of Mathematics, "Spline Interpolation," *Encyclopedia of Mathematics* (2020 Jun.). http://encyclopediaofmath.org/index.php?title=Spline_interpolation&oldid=48784.

[33] JUCE, "Jules' Utility Class Extensions," https://juce.com (accessed May 11, 2022).

[34] J. Lindfors, J. Liski, and V. Välimäki, "Companion Page to Loudspeaker Equalization for a Moving Listener," http://research.spa.aalto.fi/publications/papers/jaes-moving-listener/ (accessed May 5, 2022).

[35] Genelec, "8331A SAM Studio Monitor," https://www.genelec.com/8331a (accessed Aug. 8, 2021).

[36] G. Fanelli, T. Weise, J. Gall, and L. V. Gool, "Real Time Head Pose Estimation From Consumer Depth Cameras," in R. Mester and M. Felsberg (Eds.), *Pattern Recognition*, Lecture Notes in Computer Science, vol. 6835, pp. 101–110 (Springer, Berlin, Germany, 2011).

THE AUTHORS

Joel Lindfors

Juho Liski

Vesa Välimäki

Joel Lindfors was born in Espoo, Finland, in 1993. He received his Bachelor of Science degree in electrical engineering from Aalto University, Espoo, Finland, in 2019 and M.Sc. in electrical engineering in 2022. His Master's Thesis *User Position-Based Loudspeaker Correction* studies the same concept as this paper. Joel runs the Lindfors Productions studio in his hometown of Espoo, where he does a range of audio production work from composition and arrangement to mixing and mastering.

•

Juho Liski was born in Helsinki, Finland, in 1989. He received his M.Sc. and Doctor of Science in Technology degrees in electrical engineering from the Aalto University School of Electrical Engineering, Espoo, Finland, in 2016 and 2020, respectively. He has been working at the Acoustics Lab of the Aalto University School of Electrical Engineering since 2015 and is currently a post-doctoral researcher, focusing on audio signal processing and equalization. Dr. Liski is a Member of the Acoustical Society of Finland.

•

Vesa Välimäki is a Professor of Audio Signal Processing and Vice Dean for Research at Aalto University, Espoo, Finland. He received his M.Sc. and D.Sc. degrees from the Helsinki University of Technology in 1992 and 1995, respectively. In 1996, he was a Postdoctoral Research Fellow at the University of Westminster, London, UK. In 2001–2002, he was a Professor of Signal Processing at the Pori unit of the Tampere University of Technology. In 2008–2009, he was a visiting scholar at the Stanford University Center for Computer Research in Music and Acoustics. His research interests are in signal processing and machine learning applied to audio and music technology. Prof. Välimäki is a Fellow of the AES and IEEE. He was the General Chair of the 14th International Sound and Music Computing Conference in 2017. Since 2020, Prof. Välimäki has been the Editor-in-Chief of the *Journal of the Audio Engineering Society*.