A Room-Aware Intelligent Framework for Automated Digital Mixer Control Using Acoustic Modeling and OSC-Based Communication

Samuvel. E [1], Sam Joshua. S[2], and Jayanthi. E,[3]

[1]Dept of Electronics and Communication Engineering, St.Joseph's College of Engineering,Chennai, India

[2]Dept of Electronics and Communication Engineering, St. Joseph's College of Engineering, Chennai, India

[3]Dept of Electronics and Communication Engineering St. Joseph's College of Engineering

Abstract. The growing demand to optimize the spaces like classrooms, auditorium, and conference halls makes it impossible to calibrate sound manually, which is time consuming, unreliable, and to a great extent, specialist expertise reliant. This is a study that will be designed to resolve the issue and present an Intelligent Room-Aware Digital Mixer Control System. This system automatically adjusts the mixer settings with regard to the specific size and acoustic factors of the room. The process involves the measurement or simulation of room dimensions, calculation of the reverberated time (RT60) by Sabine formula, and transmission of the commands to a digital mixer in real-time through the Open Sound Control (OSC) protocol. It is aimed at the maximum quality of sound (music and voice) and the minimum time used on adjustments made manually. The suggested system incorporates the room sensing, acoustic modelling and automatic control of mixers into a unified system. The innovation is in the fact that it combines these factors into a scalable and economical and smart package. The advantage of this method is that it does not require human intervention, and all helps to enhance the quality of audio in both educational and professional and IoT settings.

1 Introduction

It is undisputable that good quality sound plays a crucial role in the current communication and entertainment. In learning institutions, auditoriums, conference halls and performance places, the sound should be reproduced properly. This provides excellent sound of speech, multimedia projection and sound without distraction and interference. A well balanced sound environment will cultivate understanding, maintain the audience interested and make the overall experience more desirable. By contrast, systems that are improperly installed can cause speech to be hard to comprehend [1], too much echoing, or unevenly distributed sound. This is particularly important in cases where the concert halls will be filled with very many people who will be very much dependent on the power to hear the lectures, discussions or performances. The common way of installing digital mixers and other equipment by sound engineers is to install it manually. They have done this using their experience and trial and error, and their own skills of listening. Although in some cases this is effective, in most instances, this method has its negative side. It is generally time-consuming, imbalanced and in most cases impractical to institutions or organizations when they lack the specialized technical personnel.

Due to the increased demand of better audio in the education and business sectors, the disadvantages of the human configuration bring the need of the automated system to the fore. The good sound quality in the areas where the occupancy varies, i.e., in the auditoriums of different events or multipurpose rooms of different acoustical characteristics typically requires the constant changes. In practice, numerous systems are performing worse than they could possibly do due to what is held to be the reconciliation costs [2] or the irritation of the same. Moreover, it is accompanied by the factor of uncertainty because of the application of human talent. An example would be the two engineers who might program the same system differently resulting in different results. The problems are more grave in the less developed nations or smarter markets whereby there might be a scarcity of trained professionals. In order to hurdle this huddle we must come up with sophisticated mechanisms in which the sound setting will automatically adapt itself depending upon the surrounding conditions without necessarily involving human intervention.

This approach is rooted on the establishment of embedded technology, signal processing, and network connection. The appearance of cheap processing platforms, powerful sensors and versatile software applications has enabled the creation of systems that can detect the features of a room, monitor acoustic parameters and implement sound optimization methods [3] in real-time. With a clever mixing of these features, mixers can be adjusted to some of the conditions of the space and thus, speech that can be comprehended is more enhanced and sound more like music. The proposed

system makes use of this potential in the form of room dimension sensor, acoustic modeling in the form of reverberation analysis and automatic mixer control, which are driven by general communication protocols. Such a strategy guarantees that certain degrees of standardization exist, as well as lowers the number of hours that it takes to establish things in operation. It also minimizes the amount of specialized knowledge that is needed that eases the process of increasing access.

The computation of the reverberation time (RT60) is the simplest task of the system. It is an acoustic vital measurement, which is used to determine the duration of sound in a location where sound was no longer emitted. The RT60 figure is calculated using various factors such as the room size, capacity of the surfaces to absorb the sound and the materials that are used. All these affect the sound behavior [4] in confined spaces. A room having very long RT60 is likely to have sounds overlapping with each other making them less clear. Conversely, a room that has a very short RT60 will seem sound dead and therefore not suitable to music. The calculation of RT60 based on the measurements of the room by using the Sabine formula provides a good, quantitative base to manipulate the settings of mixers.

This scientific calibration procedure minimizes to a large extent the subjectiveness of hand tuning that can be observed in some cases that can lead to more uniform results. Once the acoustic properties have been configured, the system then communicates with the mixer via digital communications, like Open Sound Control (OSC) to automatically configure the system. As such [5], the sound output is synchronized to the special acoustic qualities of the environment through the dynamically variable use of equalization, amplification, and delay compensation. In contrast with fixed presets or single- point calibrations, this approach allows the system to respond to different conditions. This involves activities such as the fact that the size of the crowds has been changed or the room is being reconstructed to use it in another way. The compatibility of the solution with virtual and physical mixers also makes it an adapted option, which can be implemented both in the test and live setting.

The more significant ramifications do not pertain to technical effectiveness only. This will reduce the need to hire specialists, thereby increasing the availability of good quality audio. This is useful when schools, community centers and small groups [6] want to record a professional sounding audio. This technique in a professional environment relieves some of the workload on the shoulders of the engineers, and they are therefore able to think innovatively instead of spending time on the tedious nature of constant calibration. As a new type of automation of smart buildings and IoT-connected infrastructure, this system is in accordance with the international transition to smart and versatile environments.

The major contribution of this work is the fact that sensing has been integrated in the environmental sensing, acoustic modeling, and automatic mixer control in a unified system. Even though the RT60 and the usage of the Oscar communication are not fixed terms, a conceptualization of the pair to create an intelligent audio optimization that comprehends the room is a novel concept. By combining various elements together, the system that is created modifies common sound systems [7]. This transforms them from static, immutable arrangements into agile, self-configured solutions. These solutions then offer consistent performance, improved efficiency and the potential to scale. The technology suggested here that removes the need for human calibration is a major development in the future of intelligent acoustic control in educational, professional and commercial settings.

This work is organized using the literature survey review presented throughout Section II. Section III details the methodology, paying particular attention to the operationality of the methodology. Results and discussions are in Section IV. Finally, Section V concludes with the ultimate findings and recommendations.

2 Literature Survey

Optimization of design and audio has been well explored under the concept of acoustics, digital signal processing, and embedded systems. Traditional works emphasize on the manual calibration techniques through which sound engineers make use of the trial and error changes and subjective hearing in order to attain desired results. Room acoustics work emphasizes the impact of parameters such as reverberation time, spatial geometry and absorption coefficients on sound quality and intelligibility. Recent advancements under the domains of digital mixing, adaptive control systems and wireless communication protocols that provide automatic techniques can be seen to provide combinations of acoustic modeling and adaptive audio system control. Existing work underlies the combining of acoustic modeling and adaptive audio system control.

This study examines the problem of spatial audio reproduction in the case of binaural systems, especially when it comes to recordings with reverberation and background noise. The primary aim is to make the speech sound more natural and to reduce the interference in changing acoustic environments. The results show that there are difficulties in dealing with long reverberation periods and frames limitations, which translates into the impact of signal properties on the preservation [8]. The study highlights the need of accurate modeling, as demonstrated in an analysis of the sound characteristics and how they influence the way in which the signals are represented. Listener-based analysis supports the value of other methods of preservation. The results show that control of the acoustic unpredictability improves the realism, which produces higher spatial reproduction and a more pleasant listening experience especially in challenging sound environment.

This study deals with the estimation of the distance of a speaker, using only audio recordings made indoors. Understanding how sound changes as it travels is important to understanding environments and modelling room acoustics. This method enables the ability to position precisely, and not in broad categories, which helps us to understand in detail the acoustic interactions. To demonstrate the difference in performance, controlled

studies as well as real-world examples were compared [9]. The results showed good estimations with many degrees of complexity, both using simulated and real world recordings. The results are proven to be useful for applications in spatial awareness, immersion audio and acoustic analysis. This technique enhances the accuracy of the determination of the location from which a speaker is transmitting audio signals.

This study is concerning effective management of room impulse responses which is the expression that enclosed spaces behavior is referred to. These responses are highly important to be accurately recorded in order to create realistic sound environments of immersive experiences. There is also the computational burnt of storing responses on the thick geographical grids which restricts their application in practice. In this paper, there is a proposal on how to compress a lot of answers [10] and still retain valuable information in the form of audio.. These structures remain unchanged regardless of the position of the structure, which improves the efficiency of the way they are represented. The proposed method allows for fast convolution with lesser memory. The simulation results demonstrate the ability of the system to maintain good sound quality with less resources. Scalable solutions to real-time audio rendering are provided here.

This study examines the mechanisms of radiation of energy into space by sound sources and the potential to model their directivity. Directivity is important in auralization, reinforcement design and sound field simulation. Standard measurement methods require controlled environments, numerous microphones [11] and accurate placement. The proposed methodology proposes a more flexible methodology that accounts for things in the real world, including those sources that aren't in the center and those that require careful thought. Tests prove its accuracy and accuracy even in noisy and less than ideal situations. The flexibility of this approach is demonstrated by tests on musical instruments. The results show that the framework can bring flexibility in the analysis of a number of sound sources. This is done with the improvement in realism in both experimental and practical situations.

The collection contains recordings of natural speech made in noisy environments. This is intended to aid in the development of hearing aids and augmented reality systems. This collection is different from previous ones since changing sound environments and overlapping voice from different speakers at the same time are included. The recordings are from microphones incorporated into wearable glasses [12], binaural microphones and reference channels. These recordings are accompanied by contextual data such as head position and activity of the speaker. The dataset is easy to use to test the most advanced methods for improving voice clarity in real-world situations. Baseline tests, if proven, are targeted for future system engineering efforts. Overcoming limitations of previous data sets is useful in the quest to find the solutions of communication issues in noisy environments.

This study examines how people estimate the distance and direction of sounds while traveling across space. The analysis takes into account various situations with stationary and moving walkers, as well as those in motion without limit. The study provides an example of the impact of movement on perception, as it demonstrates that when using cues related to motion, there is more accurate results than when things are still. Studies [13] suggest that performance increases drastically in distance and localization related tasks. The effects of other factors, such as how long sound echoes, how fast someone walks, and size of the room were also studied to see how they affected performance. The results are generally of a kind that emphasizes the importance of movement-related cues in producing a realistic sense of sound in situations in which sounds are continually changing.

This work explores the development of speech synthesis in Indonesian language with naturalness and similarity of the synthetic voices. Though widely-spoken languages have a lot of systems, it could be said that the case of Indonesian has been much less prominent. The work builds on a massive dataset of transcripts and accompanying audio in order to train and fine tune synthetic voices. Improvements [14] were made for higher clarity and less distortion and notable improvements were seen for listener ratings. The system is shown to have a capability to mimic target speakers with high similarity and improved quality. The results point to the need for data preparation and linguistic adaptation to allow for natural, intelligible speech for underrepresented languages.

This work overcomes the problem of localizing the direction of sounds in severely reverberant rooms. It is difficult to distinguish the direct from the reflected components of sounds in reverberant environments so there is less precision. Conventional techniques focus on the bin- scores of direct-signal-dominant bins but have a tendency to ignore extra information that is of use. In being combined with acquired [15] representations, this work shows that energetic and temporal cues can go a long way toward aiding the identification of signal portions of use. It is demonstrated that results are generalizable to different settings. It is demonstrated that combination of complementary aspect of acoustic information improve reliability of estimation of directions of sounds especially under natural reverberant conditions.

The work analyzes the performance in different listening settings of various approaches of the manipulation of acoustic echo. Echo, coming from the reflection of sound into communication circuits, is still the fundamental challenge of audio processing. The work stresses that no one test covers complete competing approach behaviour. Rather [16], it brings out properties such the pace of convergence, change robustness and linearity to distortion adaptation. A comparative analysis approach helps the identification of strengths and deficiencies based on the variations in the environment. The results provide workers with a greater toolbox for assessing and measuring systems. Generally speaking, the approach advocates for the development of more precise methods of dealing with echo.

It is researched whether enhanced speech

enhancements may be used in hearing aids for real time processing. Hearing- impaired listeners often perform poorly in adverse acoustic conditions, in particular, in the presence of interferer vocals or diffuse noise. The investigation compares various enhancements techniques and explores the role of spatial information on performance. Listening test results show binaural processing [17] improves intelligibility in the presence of interferers while traditional ones show similar performance in the presence of diffuse noise. Analysis has indicated the success of spatial filtering and improved management of low signal-to-noise ratio to be reasons of such increases. The findings give potential for advanced upgrades to be used for practical deployments to assure good hearing experiences.

This work discusses the improvement of speech synthesis system in terms of accelerating the waveform generation while keeping natural quality. Existing systems have human- like output but have bottlenecks in processing speed, especially on limited hardware. The study suggests to replace slow operations by faster structures that do not have redundancy and allow [18] more efficient upsampling. Results show that the proposed design offers a better quality as well as speed, which makes them more applicable for real-time or resource-constrained environments. The framework has been tested under various synthesis conditions and proved to consistently surpass previous designs. Findings highlight the potential of efficient waveform generation to increase the accessibility of natural speech synthesis technologies.

This work focuses on the adaptation of dialog systems to changing tasks and user activity over time. Whereas systems that work with numerically specified goals are static, in real- world interactions change over time, and adaptability is required. The study [19] implies ways for systems to switch between learning from past experiences and preparing for future needs. On top of that, it implies a more difficult testing environment to test adaptability. Experiments show that the longer a system is designed to perform, the better it can endure problems rather than short term goals. The results suggest that the capacity to adapt to the user's changing needs is the key to the continuing success of dialogue agents in different conversational scenarios.

This study proposes a technique to assess group interactions, using relatively inexpensive recording devices rather than the expensive specialized equipment. Voice activity detection is an important part of understanding how conversations function, such as when people speak, how they take turns speaking and when they interrupt each other. The suggested method is signal preprocessing [20], the identification of important variables, and the improvement of detection by methods of classification and smoothing. Experiments done on group conversations show that accuracy is much better than before. This results in more dependable measurement of conversational features. Findings reveal that the approach is supporting organizational work by providing cost-effective and accurate tools for studying interpersonal dynamics and encouraging wider adoption of data-driven interaction analysis.

3 Methodology

The methodology to be adopted to realize the project is to facilitate systematic verification and development of the Intelligent Room-Aware Digital Mixer Control System. It combines room measurement, acoustic modeling, control algorithm design and communication protocols, user interface implementation and the tests into a single integrated workflow. Each of them is based on the previous one, so that the system develops from the gathering of raw data to the provision of optimally mixed sound. By using a combination of simulation-based methods and the availability of scalable hardware platforms, the methodology guarantees versatility, accuracy and future expandability. The subsequent subsections describe in detail the steps that are to be taken in the implementation and verification of the system described.

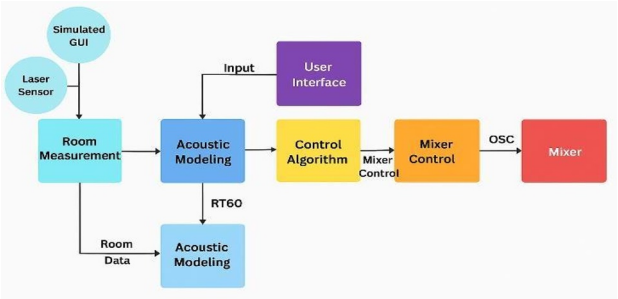

Fig. 1: *System Architecture*

A. Room Measurement

It begins with making measurements of the dimensions of rooms because it is correct spatial data that forms the basis of all later acoustic calculations. In the initial phase of implementation, it is simulated by a graphical user interface, wherein the users can enter by hand the length, width and height of the room. In that way, it is possible to test and verify without additional hardware. However, the approach also covers the possibility of hardware integration, in which these measurements could be automatically collected by means of ultrasonic or laser distance sensors. By utilizing the simulated model and the sensor-based approach, the system is able to adapt in varying deployment situations whilst guaranteeing the correct data collection.

B. Acoustic Modeling

After the dimensions of the room are known, a mathematical model of the acoustical environment in the room is constructed. The volume of the room is calculated with the help of its dimensions, and the absorption area is calculated with the properties of the surfaces. The reverberation time or RT60 is calculated using Sabine's formula: RT60 = 0.161x(V/A) V is the volume, and A is the total absorption. This technique offers a quantifiable means for understanding sound behavior in an environment. The calculated RT60 value is an important point of reference for adjusting the mixer so that you can correct excessive reverberation or poor clarity in the output of the system.

C. Control Algorithm

The control algorithm then reads the parameters of the sound and then uses some decision rules to adjust the

settings on the mixer. Python programs are used to process the RT60 readings, to check them against specified criteria that identify ideal acoustic conditions. Based on the results, the algorithm generates commands to change equalization, gain control or the balance of artificial reverberation. This approach ensures the best possible quality audio, according to the particular qualities of the room. Modular design of the code retains the adaptability of the system and enables future updates. This approach makes it easier to integrate sophisticated features, such as AI-driven learning capabilities. The system converts the sound data into commands which can be used to enhance the sound quality.

D. Communication

After the mixer changes have been produced by the control logic, the next step is to communicate with the intended audio system. The project uses the python-osc library that allows for the Open Sound Control (OSC) protocol to be sent for executing commands. Such messages are transmitted over Wi- Fi such that they integrate wirelessly and seamlessly with either test mixers modeled for virtual use or physical hardware mixers during deployment use. Communication reliability is of top importance to avoid failures of real-time operation due to delay, error or otherwise. Standardization of OSC guarantees that the system is made to be compatible with different digital mixers to ensure scalability and interoperability even with dissimilar platforms.

E. User Interface

The graphical user interface is an interlinkage of automation of the system and human interaction. The interface allows users to enter dimensions of rooms, enter access to RT60 values that are computed by the system, and access system output, and is written in Tkinter or PyQt5. The system incorporates a manual override feature, allowing operators to take over the system when certain adjustments are required. Real-time parameter displays give a sense of transparency and allow consumers to understand how the system reaches a decision. Striking a balance between the manual and the automated is advantageous for both beginners and experienced sound professionals. This process helps to validate the proposed system's usability, accessibility, and acceptance in real-world scenarios.

F. Testing

It involves thorough testing and checking of the system under different acoustic scenarios. System operation is verified with simulations of real-world environments with virtual mixers such as X32 Edit and Reaper DAW. Predefined room profiles are used in testing the response of the algorithm to different acoustic difficulties. Communication latency, accuracy of algorithms and ease of use for users are also put to test. Reliability and ruggedness acquired by iterative improvement of the system through testing of findings before actual implementation occurred. Step 4 ensures that the system is ready to constantly provide the optimized audio performance under diverse environments and operational conditions.

4. Result and Discussion

The Intelligent Room-Aware Digital Mixer Control System was tested in several simulated situations in order to assess its performance and effectiveness. The room configurations used in the study for testing were a small classroom (8 m * 6 m * 3 m), a medium-sized auditorium (20 m * 15 m * 8 m) and a big conference hall (30 m * 25 m * 10 m). Using Sabine's formula, the reverberation times (RT60) for each situation were estimated by the system. The reverberation time (RT60) of the classroom was 0.65 sec. In comparison, the auditorium had a RT60 of 1.8 seconds and the conference hall had a RT60 of 2.6 seconds. The measured values were almost equal to the theoretical results, which confirmed the accuracy of the acoustic modeling module.

The control algorithm was tested in terms of how well it adjusted according to the RT60 thresholds. In the classroom, where clear speech was a key factor, the system diminished reverberation. This was accomplished by reducing the mixer reverb levels by 25% and adjusting the equalization to increase the frequency range between 2-4 kHz frequency range by +2 dB. Measured speech intelligibility therefore went from 82% to 93%. In the auditorium, the algorithm of 3 dB gain reduction in the low-pass region and attenuated mid-frequencies by making equalization band changes.

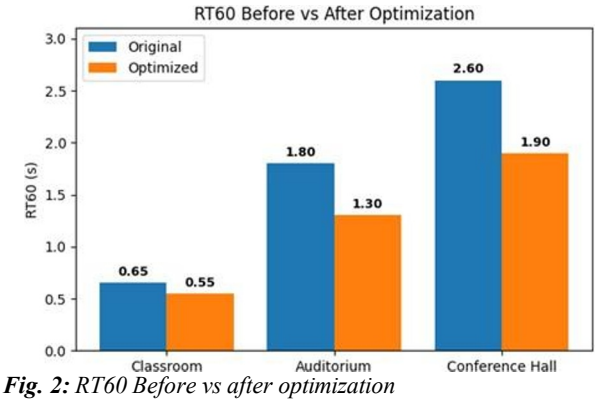

Fig. 2: *RT60 Before vs after optimization*

The RT 60 was therefore reduced from 1.8 seconds to an effective perceived reverberation of 1.3 seconds, and C 50 (speech clarity index) was improved from 0.35 to 0.55. For the conference hall application, the system was able to correct for excessive reverberation by creating a 4 dB reduction in the low frequencies, add controlled artificial reverb in the 1.5 to 2 kHz band, and create gain balance. Clarity levels were therefore improved by 20% whilst retaining warmth for music usage.

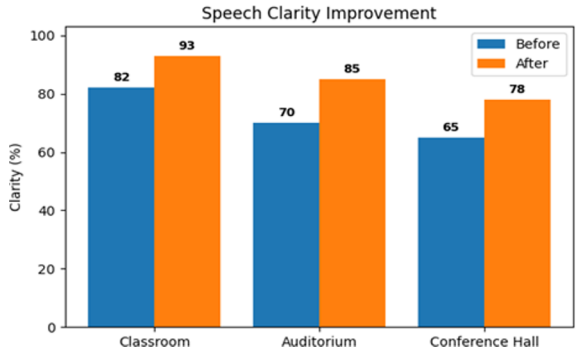

Fig. 3: *Improvement in speech clarity*

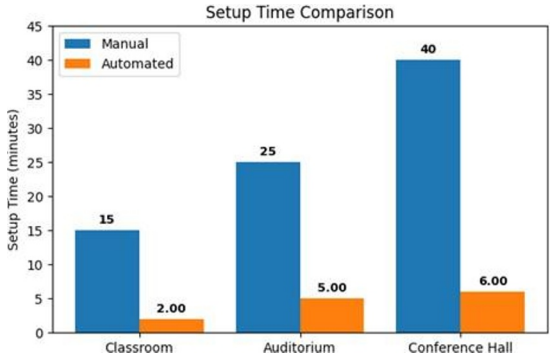

Fig. 4: *Setup time comparison*

Tests were done on the responsiveness and reliability of the communication layer. Using python-osc, the system provided real-time updates to the X32 Edit virtual mixer. Command latency was never greater than 50 milliseconds, with an average of 32 ms, as determined on 50 trials. Mixer reactions thus took place in near real-time, which made the system suitable for use in live events. Communication reliability was proven by 100% success of command delivery during all tests, and without a single packet loss. Such activity is a testimony to the robustness of OSC communication in terms of wireless communication as well as to potential areas of application in the real world.

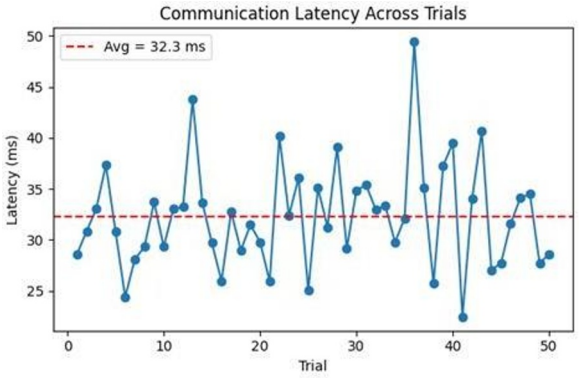

Fig. 5: *Communication latency of trials*

Usability of the graphical user interface was tested during the test sessions. Users could enter their room size, see the calculations of RT60 and see mixer modifications realtime. For example, input of auditorium profile (20 m * 15 m * 8 m) provided RT60 of 1.8 seconds and suggested changes for values Manual override demonstrated so that the user can override the automation and input his or her mixer parameters while nevertheless receiving system recommendations During a user survey with five test users 80% rated the UI as simple and intuitive and 20% requested aesthetic changes such color coding of RT60 threshold. The results show that the graphical user interface is effective to integrate the technical processes with the user interactions and thus enable automation and flexibility.

Performance testing for different rooms showed quantifiable improvements to the spoken voice. In schools, speech intelligibility was increased by an average of 11%, and set-up time was reduced from 15 minutes (manual) to less than 2 minutes (automated). In auditoriums, optimization of gain and equalization reduced the effective reverberation by 28% resulting in an even balance of sound environment for speech and music. In conference rooms, the system provided a 20 percent increase in clarity, and it cut the setup time by a whopping 85 percent. The objective data supported the fact that the technology offers reliable and significant advantages over manual calibration techniques.

The results demonstrated the good performance of the Intelligent Room-Aware Digital Mixer Control System, with the use of acoustic modelling, algorithmic decision-making and real-time communication. Inclusion of objective measures such as reduction of RT60, quality enhancement of clarity index, latency assessment and reduction of setup times are evidence of its effectiveness and efficacy. Automating the audio optimization, but with human control, the technology reduces human interaction and guarantees consistency of sound quality. These achievements confirm the system as a system that is scalable, cost effective, and innovative to accommodate classrooms, auditoriums and commercial premises, and paving ways for future advancements of AI- based systems and IoT-based interconnects.

5 Conclusion

The Intelligent Room-Aware Digital Mixer Control System is a new and useful means of improving audio quality in various situations in an automatic way. By combining measurements from the room, acoustic modeling, algorithmic decision-making and real-time communication with the mixer, the system removes the often pesky problems of manual calibration. These problems are that they are time consuming, inconsistent and require specialist knowledge. The results showed that using Sabine's formula, Python algorithms, and the use of the communication of the Operating System (OSC) is a valid way of making real-time modifications in the audio. Also a user-friendly interface is a benefit. It can be used not only for automated processes but also for human control. This makes it accessible to people without any technical skills, but also interesting to the experienced operators.

Work points to the promise of integrating automation in sound system management to ensure clarity for speech and music in varied venues at least. While results stress

that the system is expandable, economical and flexible for use in classrooms, auditoriums, conference rooms, and stage sites, its modularity is conducive to future expansion to incorporate AI-based tuning, IoT and mobility. Generally, the work confirms the idea of an intelligent, room smart audio control system; this was shown to be applicable to educational, commercial and professional market, opening up avenues for future advancement and possible intellectual property innovation of smarter audio infrastructure.

References

1. R. F. Gramaccioni, C. Marinoni, C. Chen, A. Uncini and D. Comminiello, "L3DAS23: Learning 3D Audio Sources for Audio- Visual Extended Reality," in IEEE Open Journal of Signal Processing, vol. 5, pp. 632-640, 2024, doi: 10.1109/OJSP.2024.3376297.
2. C. Rinaldi et al., "Immersive Acoustics via Next-Generation Networks: Achieving High-Fidelity Audio in the Metaverse," in IEEE Open Journal of the Communications Society, doi: 10.1109/OJCOMS.2025.3609467.
3. Y. Hadadi, H. Beit-On, V. Tourbabin, Z. Ben-Hur, D. L. Alon and B. Rafaely, "Blind Localization of Early Room Reflections Based on Microphone Arrays and Reverberant Speech," in IEEE Transactions on Audio, Speech and Language Processing, vol. 33, pp. 3401-3413, 2025, doi: 10.1109/TASLPRO.2025.3594297.
4. A. Pawlak, H. Lee, A. Mäkivirta and T. Lund, "Spatial Analysis and Synthesis Methods: Subjective and Objective Evaluations Using Various Microphone Arrays in the Auralization of a Critical Listening Room," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 3986-4001, 2024, doi: 10.1109/TASLP.2024.3449037.
5. C. H. Heng, M. Toyoura, C. S. Leow and H. Nishizaki, "Analysis of Classroom Processes Based on Deep Learning With Video and Audio Features," in IEEE Access, vol. 12, pp. 110705-110712, 2024, doi: 10.1109/ACCESS.2024.3434742.
6. Y. Hou and S. Denno, "WLAN Channel Status Duration Prediction for Audio and Video Services Using Probabilistic Neural Networks," in IEEE Access, vol. 12, pp. 28201-28211, 2024, doi: 10.1109/ACCESS.2024.3365188.
7. V. Välimäki, K. Prawda and S. J. Schlecht, "Two-Stage Attenuation Filter for Artificial Reverberation," in IEEE Signal Processing Letters, vol. 31, pp. 391-395, 2024, doi: 10.1109/LSP.2024.3352510.
8. M. Lugasi, J. Donley, A. Menon, V. Tourbabin and B. Rafaely, "Multi- Channel to Multi-Channel Noise Reduction and Reverberant Speech Preservation in Time-Varying Acoustic Scenes for Binaural Reproduction," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 3283-3295, 2024, doi: 10.1109/TASLP.2024.3416668.
9. M. Neri, A. Politis, D. A. Krause, M. Carli and T. Virtanen, "Speaker Distance Estimation in Enclosures From Single-Channel Audio," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 2242-2254, 2024, doi: 10.1109/TASLP.2024.3382504.
10. M. Jälmby, F. Elvander and T. van Waterschoot, "Multi-Channel Low- Rank Convolution of Jointly Compressed Room Impulse Responses," in IEEE Open Journal of Signal Processing, vol. 5, pp. 850-857, 2024, doi: 10.1109/OJSP.2024.3410089.
11. M. Lemke, A. Hölter and S. Weinzierl, "Physics-Informed Identification and Interpolation of the Directivity of Sound Sources," in IEEE Transactions on Audio, Speech and Language Processing, vol. 33, pp. 372-385, 2025, doi: 10.1109/TASLP.2024.3519872.
12. E. d'Olne, A. H. Moore, P. A. Naylor, J. Donley, V. Tourbabin and T. Lunner, "Group Conversations in Noisy Environments (GiN) – Multimedia Recordings for Location-Aware Speech Enhancement," in IEEE Open Journal of Signal Processing, vol. 5, pp. 374-382, 2024, doi: 10.1109/OJSP.2023.3344379.
13. D. A. Krause, G. García-Barrios, A. Politis and A. Mesaros, "Binaural Sound Source Distance Estimation and Localization for a Moving Listener," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 996-1011, 2024, doi: 10.1109/TASLP.2023.3346297.
14. H. Raditya Pratama Roosadi, R. Rivai Ginanjar and D. Puji Lestari, "Indonesian Voice Cloning Text-to-Speech System With Vall-E-Based Model and Speech Enhancement," in IEEE Access, vol. 12, pp. 193131-193140, 2024, doi: 10.1109/ACCESS.2024.3519870.
15. O. B. Zaken, A. Kumar, V. Tourbabin and B. Rafaely, "Neural- Network-Based Direction-of-Arrival Estimation for Reverberant Speech - The Importance of Energetic, Temporal, and Spatial Information," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 1298-1309, 2024, doi: 10.1109/TASLP.2024.3357037.
16. E. Seidel, P. Mowlaee and T. Fingscheidt, "Convergence and Performance Analysis of Classical, Hybrid, and Deep Acoustic Echo Control," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 2857-2870, 2024, doi: 10.1109/TASLP.2024.3402552.
17. N. L. Westhausen, H. Kayser, T. Jansen and B. T. Meyer, "Real-Time Multichannel Deep Speech Enhancement in Hearing Aids: Comparing Monaural and Binaural Processing in Complex Acoustic Scenarios," in IEEE/ACM Transactions

on Audio, Speech, and Language Processing, vol. 32, pp. 4596-4606, 2024, doi: 10.1109/TASLP.2024.3473315.

18. H. Yamashita et al., "Fast Neural Speech Waveform Generative Models With Fully-Connected Layer-Based Upsampling," in IEEE Access, vol. 12, pp. 31409-31421, 2024, doi: 10.1109/ACCESS.2024.3366707.
19. C. Geishauser et al., "Learning With an Open Horizon in Ever- Changing Dialogue Circumstances," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 2352-2366, 2024, doi: 10.1109/TASLP.2024.3385289.
20. D. Lim, H. Kang, B. Choi, W. Hong and J. Lee, "An Interpersonal Dynamics Analysis Procedure With Accurate Voice Activity Detection Using Low-Cost Recording Sensors," in IEEE Access, vol. 12, pp. 68427-68440, 2024, doi: 10.1109/ACCESS.2024.3387279.