An Optimization Method for Audio Experience Improvement for Audiences

Jeyanth Rajan Babu, *Masters in Computer Engineering* and Rishab Kinnerkar, *Masters in Computer Engineering*

***Abstract*—Closed gathering halls which are not specifically designed to play music through a multi-speaker system can result in an unpleasant sound experience for the audience. Based on the seating arrangement of the audience members the speaker volumes in the halls may not be set to an optimal audio intensity level. It is possible that some audience members may experience too loud volume and some others too low. Hence, we created a methodology which would make use of a wireless sensor network system responsible for detecting audience spatial position with respect to the speakers and optimizing the speaker intensities so as to ensure maximum audience satisfaction.**

***Index Terms*—wireless sensor networks, SLSQP, reverberation, acoustic noise.**

I. Introduction

AUDITORIUMS are specifically designed to ensure every audience member experience optimal sound quality and intensity. Many different aspects of auditorium acoustics have been researched on since many centuries. In 1835, Dr. D.B Reid had advised and had worked on the acoustic treatment contemporary house of commons, Westminster [1]. He was able to recognize the problem of prolonged reverberations in large rooms and from then on, his work was used in consideration by many when it came to the acoustics of a room [1]. In 1838, Scott Russell had made progress in calculation of optimal floor profiles of rooms for good hearing [1]. Today we have state-of-the-art auditoriums optimized to bring pleasant audio experiences to the audience members.

However, there are many large rooms built for multipurpose activities. It is in such rooms that there isn't enough consideration given to the potential audio acoustics. Sound intensity, reverberation and sound interference at a point are the 3 main factors which determine the audio experience of an audience member. Reverberation and the sound interference factors would be a difficult problem to solve for such rooms because these factors are determined by the dimensions and material used in constructing the room. It would be very costly to rebuild these rooms for the purpose of improving audio quality and in many cases infeasible. Thus in our project, we focus on improving the sound experience of audience by modulating the sound intensity produced by the speakers.

We focus on improving the sound experience of audience members by optimizing the speaker intensities. In a large room supposing there are 'm' number of speakers and 'n' number of people. The 'm' speaker positions would be fixed but the 'n' people's position would change as and when the large room would be used, and 'n' would change as well. Also, it is possible for the people's position to change during an in-session activity happening in the large room which involves the speakers. It is very important to get the exact location of the audience members at all time for the speaker-system to be set to an optimal volume. People could be scattered when seen from top and in such cases for an audience member to experience optimal sound intensity the speaker volumes would end up varying from each other.

Wireless sensors setup in a wireless localization network could be used to detect the position of the people with respect to the speakers. This information would be fed to a main control which would then adjust the speaker volumes after performing the optimization calculations.

II. Noise Model

Whenever a speaker or a music player outputs sound not necessarily would it be clear. This is because, only a tiny fraction of the sound output from the speaker is heard by the audience and it is referred to as direct sound. Whereas most of the sound output is lost or does not hit the intended target; but may bounce back off the walls or floor or any other object and is referred to as indirect sound. Exceptional handling of the indirect sound is required for enhanced hearing. Good acoustics deal with utilizing the indirect sound in the best possible way for improving listenability. There is always some noise or unwanted disturbances, which maybe a result of external agents such as audience, *HVAC* systems, rotating fans or creaking of chairs; or maybe a result of echo or reverberation. Noise generally blunts the listenability of sound. The general categories of noise for enhanced listening are detailed here, as explained in [4].

A. Background Noise

Background noise includes all the noise that is not generated by the sound from the speakers. It includes the whispering of

This project report was submitted on Friday, 10 May, 2019, as a part of CprE 546 course project requirement.

audience, running of air-conditioners, rotation of fans, humming of ballasts and all other similar external sources of noise. Background noise, in an auditorium, is generally considered at a level of 20 dB-A below the level of sound of speech. Ideally the background noise is at 20 dB-A when no sound is being reproduced by the speaker. The signal – to – noise ratio (*SNR*) metric gives a better understanding of the level of noise that is present in the location. Improving the SNR would be the primary objective of enhancing listenability. Reducing the background noise directly improves the SNR.

B. Acoustic Noise

Noise that is produced as a result of reflection of the sound on the walls and other objects in the location contribute to the acoustic noise, echo or reverberation. Echo is simply the reflected sound off the walls. Reverberation is when sound is reflected from different objects on to a spot lacking any sort of direction. Reverberations are omni – directional and are categorized as quiet and loud reverberations. Quiet reverberations enhance the listenability whereas loud reverberations disrupt it. Reverberations are further studied on the following three categories:

1) *Onset time delay*
   Onset time delay is the time difference between when the direct sound is heard and when the reverberation is heard. Lecture halls and auditoriums should have an onset time of $1/3^{rd}$ second ideally.
2) *Loudness of the reverberation*
   Reverberations should ideally be 10 dB-A lesser than the direct sound.
3) *Duration of the reverberation*
   Reverb time is the time it takes for the reverberations to die down to 60 dB-A. Large rooms have a reverb time of 1.5 seconds. Relatively smaller rooms can have a reverb time of 0.9 seconds. Different kinds of rooms with varied functions have different reverb times. But the above-mentioned categories are the most common ones.

Increasing the audio intensity to ensure a better signal – to – noise ratio (*SNR*) alone will not be effective as this may not improve the signal – to – acoustic noise ratio (*SANR*). These reverberation factors must be considered while working to improve the audio experience of the audience. However, the improvement of *SANR* rests solely on the acoustic principles, the basis on which the building construction is done. So, this improvement in *SANR* is beyond the scope of this project. In our project, we intend to improve *SNR* by measuring the sound intensities at the location of every person in the room and estimating the speaker power output which must be set in order to ensure the people are experiencing audio at optimal levels of hearing.

III. FORMULA

A. Sound Intensity Calculation

The sound intensity at a place is the sound power experienced per unit area. So sound intensity,

$$I = P/A \text{ (W·m-2)}$$

where P is the sound power from the source and A is the area covered. As sound waves travel in the form of sequence of waves of concentric circles, we identify the area covered to be

$$A = 4 \times \pi \times r^2,$$

where r is the distance between the source and the point at which the sound intensity is calculated. This leaves us with the sound intensity measurement as,

Sound intensity $I = P/(4 \times \pi \times r^2)$ (W·m-2).

For multiple sources, the sound intensity equation is modified to,

Sound Intensity at A,

$$I_a = I_{a1} + I_{a2}$$

$$I_a = P_1/(4 \times \pi \times r_{a1}^2) + P_2/(4 \times \pi \times r_{a2}^2) \text{ (W·m-2)}$$

Here, P_1, P_2 - speaker power output in Watts (W); r_{a1}, r_{a2} - distance of person A from speakers *S1 and S2* in meters (m) respectively.

As an example, we can see Fig.1 is a 4-speaker 3-people system. r_{a1}, r_{a2}, r_{a3} *and* r_{a4} are the distances of speaker S_1, S_2, S_3 *and* S_4 from A respectively.

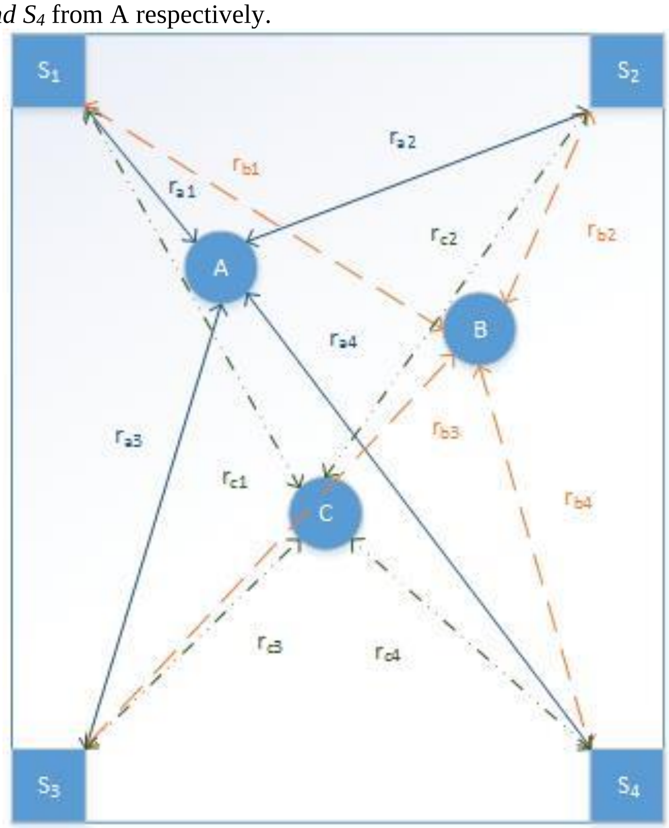

Fig. 1. The schematic showing the distribution of audience and audio speakers in a hall for a 3 people 4 speaker system.

The sound intensity at A is given by,

$$I_a = I_{a1} + I_{a2} + I_{a3} + I_{a4}$$

$$I_a = P_1/(4 \times \pi \times r_{a1}^2) + P_2/(4 \times \pi \times r_{a2}^2) + P_3/(4 \times \pi \times r_{a3}^2) + P_4/(4 \times \pi \times r_{a4}^2) \quad \text{(W·m}^{-2}\text{)}$$

Generally, if 'n' number of speakers are present and there are 'm' number of people, we will be having 'm' number of intensity equations, each with 'n' unknowns (speaker powers P_n), given by,

$$I_m = \sum_a^n I_{mn} = \sum_a^n {P_n}/{(4 \times \pi \times r_{mn}^2)} \text{(W·m}^{-2}\text{)} \qquad (1)$$

B. Decibel – Speaker Power correlation

We calculate the decibel levels of the sound intensity at the position of the audience by the relation,

$$db = 10 \log_{10}(P/P_0) \qquad \text{(dB)}$$

Here *db* refers to the sound intensity in decibels; *P* refers to the speaker power output in Watts and P_0 refers to the threshold of hearing, which is the minimum level of sound that human ear can hear. Typically, $P_0 = 1 \times 10^{-12}$ ($\mathrm{W{\cdot}m^{-2}}$). The intensity level of a normal conversation has been established to be at 60 dB.

Simplifying this correlation and substituting the value of P_0, we can compute the speaker power output for a decibel value as,

$$P = 10^{(db\times 0.1)-12} \qquad \text{(W)} \qquad (2)$$

We will be using (2), in our python script to verify the correctness of our estimation.

IV. *SLSQP* ALGORITHM

According to the *pyOpt* website [2] , "*SLSQP* optimizer is a sequential least squares programming algorithm which uses the Han–Powell quasi–Newton method with a *BFGS* update of the B–matrix and an L1–test function in the step–length algorithm. The optimizer uses a slightly modified version of Lawson and Hanson's *NNLS* nonlinear least-squares solver."

Here we explain the terms mentioned in the definition.

The Han-Powell method is an optimization method which solves Nonlinear Programming methods of the type [3]:

Minimize: *F(x)*

Subject to: $c_i(x) = 0$, $i = 1, 2,...,m'$

$c_i(x) >= 0$, $i = m'+1 , \dots , m$

Where *F(x)* belongs to R^1, $x \in R^n$, and c_i belongs to R^1, $i = 1, 2, \dots, m$.

The Han-Powell method falls under the category of quasi-Newton methods. Quasi-Newton methods differ from the traditional Newton optimization methods by making use of gradient information, which is stored when the algorithm visits points on the plane, to construct an approximation of the Hessian matrix. This is done because it could be difficult or time-consuming to calculate the Hessian matrix.

The Broyden–Fletcher–Goldfarb–Shanno algorithm (*BFGS*) is used in *SLSQP* to update the B-matrix. The B-matrix is the approximation to the Hessian matrix which are used in Quasi-Newton methods. The *BFGS* update is done in the following ways:

1) A point x_0 is our initial guess on the plane. The approximate B-matrix is calculated with respect to this point.
2) A search direction, p_0, is obtained by solving the equation $B_{0P0} = -\Delta f(x_0)$
3) Once the direction is obtained, we need to get an acceptable step size. Otherwise it is possible that we skip over the area we are searching for. An acceptable step size is calculated as follows:

$$\alpha_k = \arg\min f(x_k + \alpha \times p_k)$$

4) Then we set $s_k = \alpha_k \times p_k$ and update $x_{k+1} = x_k + s_k$
5) Then point y_k is calculated and the B-matrix is updated. Y_k and B-matrix are calculated as follows:

a) $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

b) $B_{k+1} = B_k + \frac{y_k y^T{}_k}{y^T{}_k s_k} - \frac{B_k s_k s^T{}_k B^T{}_k}{s^T{}_k B_k s_k}$

6) The above steps are repeated till $\|\nabla f(x_k)\|$ converges.

L-1 test function refers to the usage of Euclidean distance in the Han-Powell method.

In general, *SLSQP* can solve problems of the form:

Minimize: *f(x)*
Subject to equality and inequality constraints:

$G_j(x)=0 \qquad j = 1,...,m_e$

$G_j(x) >= 0 \qquad j = m_e +1,..., m$

Subject to lower and upper bounds variables:

$L_i <= x_i <= U_i \qquad i = 1,..., m$

V. PROBLEM STATEMENT AND ALGORITHM

A. Problem Statement

Minimize: $\sum_1^n(|60 - I_n|)$
Subject to: $(60 - k) < I_1, I_2, \dots, I_n < (60 + k)$

Where I_i stands for the intensity experienced by the i^{th} person in decibels.

B. Algorithm:

1) Since we are minimizing the Euclidean distance of the intensities with respect to our optimal value of 60 dB we start by finding keeping the step value k equal to 0. This would make the upper and lower bounds 60 dB. Then we increase the k value by an arbitrary value(0.1 in our case) and try to find a solution. If no solution is found, then the k value is doubled from its present value till a solution is found.
2) Once a solution is found at k_i then the value is decreased by calculating the mean of k_i and k_{i-1} and setting it as the new k value
3) This value of diverging and converging k is repeated till we find very little change in the outputs of k_n and k_{n-1}

VI. EXPERIMENTAL ANALYSIS

We implemented our python script to estimate the speaker power outputs for optimal sound intensity for every person. We executed the python script for 4 varied scenarios in a hall of dimensions 20m x 40m, which we will discuss below:

A. *Scenario 1 – Equal number of speakers and people*

1) *Two people Two speaker system*

We assume a system with speaker coordinates at (0, 0) and (20, 0); and people at (10, 5) and (15, 5). The following results were obtained:

#k - steps: 0
Final k step value: 0
Speaker power outputs:
P_1 = 1.178097244941505 mW
P_2 = 0.39269908182798563 mW
Sound intensities at people location:
I_a = 59.999999999929756 dB
I_b = 60.00000000067964 dB

2) *Three people Three speaker system*

We assume a system with speaker coordinates at (0, 0), (20, 0) and (10, 40); and people at (5, 5), (15, 5) and (10, 20). The following results were obtained:

#k - steps: 1
Final k step value: 0.1
Speaker power outputs:
P_1 = 0.5335417114722429 mW
P_2 = 0.5335417114722425 mW
P_3 = 3.768902780867989 mW
Sound intensities at people location:
I_a = 60.99999999988689 dB
I_b = 60.99999999988689 dB
I_c = 59.63613595922614 dB

B. *Scenario 2 - #speaker > #people (3 speaker 2 people system)*

We assume a system with speaker coordinates at (0, 0), (20, 0) and (10, 40); and people at (5, 5) and (10, 20). The following results were obtained:

#k - steps: 0
Final k step value: 0
Speaker power outputs:
P_1 = 0.6415926654481205 mW
P_2 = 0.6415926654481205 mW
P_3 = 4.0 mW
Sound intensities at people location:
I_a = 60.32113250262753 dB
I_b = 60.000000016392974 dB

C. *Scenario 3 - #speaker < #people (2 speaker 3 people system)*

We assume a system with speaker coordinates at (0, 0) and (10, 40); and people at (5, 5), (15, 5) and (10, 20). The following results were obtained:

#k - steps: 3
Final k step value: 0.3
Speaker power outputs:
P_1 = 1.0983030013720687 mW
P_2 = 3.8839321201087738 mW
Sound intensities at people location:
I_a = 63.000000000132545 dB
I_b = 57.758721119183896 dB
I_c = 59.76571928902232 dB

The coordinates of the speakers and people are in meters. The above results are tabulated in Table I.

TABLE I

Scenario	#Speakers	Speaker coordinates	#People	People coordinates	#k steps	Speaker power output (mW)	Sound intensity at people location (decibels)
2 people 2 speaker system	2	0, 0 10, 40	2	10, 5 10, 20	0	1.17809724494150 5 0.3926990818279 8563	59.999999999929 756 60.000000000679 64
3 people 3 speaker system	3	0, 0 20, 0 10, 40	3	5, 5 15, 5 10, 20	1	0.5335417114722429 0.5335417114722425 3.76890278086 7989	60.9999999998689 60.9999999998689 59.6361359592261 4
2 people 3 speaker system	3	0, 0 20, 0 10, 40	2	5, 5 10, 20	0	0.6415926654481205 0.6415926654481205 4.0	60.3211325026275 3 60.000000016392974
3 people 2 speaker system	2	0, 0 10, 40	3	5, 5 15, 5 10, 20	3	1.0983030013720687 3.8839321201087738	63.000000000132545 57.758721119183896 59.76571928902232

Table I tabulates the different scenarios against which the Python script was executed, and the sound intensities and the power output are showcased.

As seen from the results our program was successfully able to simulate and optimize the input scenarios.

VII. FUTURE WORK

Our model was fast in finding optimal solutions. However it does not present the real situation response where data would be communicated by using a wireless sensor network and that there would be delays which would depend on the sensor specification and the overall communication architecture.

The involvement of a multi-speaker and multi-person system gives rise to a lot of different possibilities which need to be handled differently. We put our best efforts in trying to remedy whenever an exception arose but there is room for further improvement. Our algorithm is fast but it finds a sub-optimal solution. Essentially in our problem statement we decrease the values of k uniformly for all the constraints. The binary search approach taken by us to find the sub-optimal solution happens with a runtime order of O(nlog(n)) where n is the number of inequalities we are optimizes. Ideally, the brute force approach would be to check for all combinations of k for all the inequalities which would give the optimal solution but would happen in the order of O(n!). The brute force approach could be feasible when there are less speakers and people and we could incorporate a mechanism to switch to the brute force when there are less number of people. We had an optimal value of 60 dB and symmetrically worked around it to find solutions as close to it as possible. As an example, this translates to 50 dB being on the same audio satisfaction intensity as 70 db. In real life this is not true. There has not been any research done on this sound level intensity equivalency. Perhaps, in the future through surveys this data could be gathered. In any case our model would be still able to accommodate such changes.

Through this project we learned many new concepts. We started our study by understanding the nature of sound in-depth. Then we moved on to study the real life implementations of such a system after which we were confident about the project's feasibility via the use of wireless sensors. There was not much information publicly available pertaining to our project so we had to come up with the model by ourselves. In doing so we learned about different optimization techniques especially pertaining to constrained nonlinear optimization. After the equation optimization, another area where optimization was necessary was in the setting up of the equations. While doing so we ideated and researched upon many different approaches and overall it improved our knowledge in data structure and algorithms. Up until here we were testing our program in parts. Thus the last phase of our project was to completely code our system. There were many new programming concepts which we had to learn in order to implement this system. Overall it was a good experience to be working on the project.

VIII. CONCLUSION

Our simulation was a simplification which considered only the optimization of sound intensity when it came to improving the audience satisfaction. We were successfully able to break down our problem in a mathematical way and optimize it. The results which we got were accurate and the optimization was fast for even higher number of people and speakers. There is scope for improvement in the future.

ACKNOWLEDGMENT

We would like to acknowledge and express our gratitude to Professor Daji Qiao, for providing us guidance and valuable insights throughout the course of this project. The feedback we received from him throughout the course of our project ultimately motivated us to research and ideate upon different aspects pertaining to our project.

REFERENCES

[1] Barron, Michael. Auditorium acoustics and architectural design. Routledge, 2009.

[2] http://www.pyopt.org/reference/optimizers.slsqp.html

[3] Chen, Hern-Shann, and Mark A. Stadtherr. "Enhancements of the Han—Powell method for successive quadratic programming." Computers & chemical engineering 8.3-4 (1984): 229-234.

[4] https://www.acousticsciences.com/media/articles/auditorium-acoustics-101-quieter-better

[5] Avriel, Mordecai (2003), Nonlinear Programming: Analysis and Methods, Dover Publishing,