COMPUTER MODEL UTILIZATION FOR SPEECH INTELLIGIBILITY ASSESSMENT IN ENCLOSED SPACES USING SOUND SYSTEMS

C Nestoras Acoustics Research Group, FESBE, London South Bank University
S Dance Acoustics Research Group, FESBE, London South Bank University

1 INTRODUCTION

This study concentrates on the acoustical design of auditoria with reference to speech intelligibility and particular emphasis on designs using distributed sound reinforcement systems. Objective measures and computer prediction software were used in the process of evaluating existing setups with the aim to optimize the acoustical conditions and establish an accurate way of assessment in the terms used. Special consideration was given in the measurement session and the means for computer model adaptation, since erroneous actions in the latter would affect the character of any conclusions reached.

Commonly, modelling a space for which acoustical parameters are not known requires some means of model validation to ensure consistency of outcomes. Auralization is an additional factor to consider since it is directly related to the prediction quality. Model preparation in this sense is assessed. Ultimately, a test methodology for an objective verification of auralization accuracy is presented and validated.

2 THEORETICAL BACKGROUND

In designing a computer model, an assessment of the architectural details within a space is required so as to determine the level of detail necessary for an adequate computer representation. With the dimensions and shape of a space being essential, several additional parameters can be accounted for such as the fittings' dimensions, their location and density. Due to the specific use of computer modelling for this study (post-evaluation tool) and given the lack of absolute information regarding the acoustical performance of room boundaries, a validation/optimization procedure (see 3.2) was undertaken in comparison to the actual measurement results so as to ensure an accurate model performance for different setup conditions/scenarios. Optimization of room acoustical conditions could thus become feasible on an experimental basis using the model.

A validated computer model enables an accurate prediction of acoustical parameters, when compared to actual measurements. In this respect, an accurate auralization should incorporate these results so as to achieve a subjective impression that closely approximates the actual acoustical environment. In order to directly relate the prediction process to an acoustical measurement a swept sine can be used as the anechoic audio material in the auralization process. The advantage is that the auralized sample has essentially the form of a raw impulse response (having originated from a standard test signal), equivalent to the result of an actual measurement. Further processing using an open loop measurement system can be used; a comparison of the derived parameters to the predicted and/or the actual values (relating to the room) comprising a direct descriptor of auralization quality.

3 METHODOLOGY

3.1 Room acoustics measurement

Room acoustics measurements were performed in three lecture theatres using the WinMLS 2004©[1] platform and a pair of omni directional sound source and receiver. A swept sine test signal was utilized to excite the space and multiple measurements, conforming to BS ISO 3382[2], were taken for six receiver positions (see figures 1-3).

3.2 Computer model validation/adaptation

A simple procedure was used for a generic model adaptation to actual conditions. Reverberation time (T_{30}) values were the determinants of model performance and thus, a step by step evaluation was performed using the different receiver positions. Performance optimization, initially for one receiver, allowed for further sequential reference to remaining data. Differences among positions would suggest that the acoustical properties of the materials close to the source and (mainly) the receiver were not accounted for correctly. Reaching a balance among the prediction setups enabled the determination of a set of acoustical properties for the room materials, which could be assumed as correct and thus, allow for reliable predictions under the conditions necessary for each case. CATT Acoustics 8.0f[3] was used in experimentation.

3.3 Authentication of auralized room responses

The accuracy of the predicted auralizations for a lecture room was assessed by objective means using a new hybrid approach. The experimental procedure utilizes a raw test signal (swept sine wave) and an open loop measurement system.

For the validation procedure, settings within the modeling software were adjusted to correspond with the reference actual measurements. Most notably the source characteristics and the type of receiver were matched and a set of auralizations was predicted using an anechoic sample of a 12 second sine sweep.

Results (being essentially raw impulse responses) were then assigned as the input of an open loop measurement system (B&K Dirac 3.0[4]) and post processed to derive a set of room acoustics parameters. Given that the datasets match the predicted and/or actual data an auralization could be described as accurate in these terms. Furthermore, as additional measures, other than RT, should theoretically agree an assessment was made to account for possible differentiation of individual parameters. The data comparison can be seen in section 4.

It has been shown by the authors in previous work[5-6] that a closed loop measurement system should produce comparable results with an open loop system, particularly for controlled conditions as in the current experiment. Details of the systems used for this purpose are described in the associated work.

4 RESULTS

4.1 Actual (WinMLS) and prediction (CATT) data for the spaces considered

The validated computer models analyzed here (figures 1-3) achieved a satisfactory result in terms of the T_{30} measured (i.e. prediction and measurement output is comparable). Additional parameters were referenced to validate model behaviour (table 1 shows an example comparison between actual and predicted values) with the C_{50} and STI measures being of most interest due to their high correlation to speech intelligibility[7-8]. For the single source case it was found that resulting values were comparable, giving an STI (and C_{50}) marginally over the Just Noticeable Difference (JND) [7]. A somewhat altered character for the sound system assisted conditions was observed here in some instances, any differentiations however attributed to the erroneous input in terms of the source's actual sound power level (approximated here).

It was highlighted that the incorporation of a sound system requires accurate performance characteristics at hand to ensure a realistic comparison. Nonetheless, given a consistent model, experimentation can take place for alternative sound source setups, as the model could reveal to a large extend the room potential and/or limitations on a relative basis.

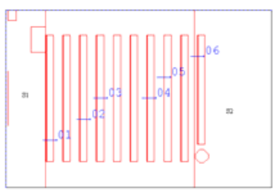

Figure 1. Lecture room B

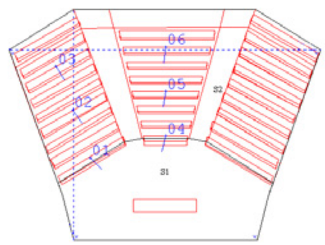

Figure 2. Nelson Haden Amphitheatre

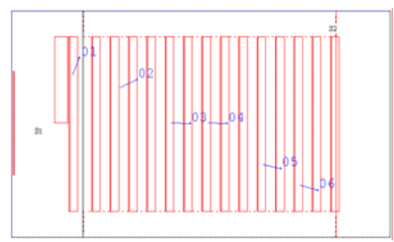

Figure 3. Manor Lecture theatre

The validation/optimization procedure overall provided confidence that the models would perform consistently under different configurations and therefore could be safely used in an investigation.

F[Hz]	125		250		500		1000		2000		4000	
	Prediction	Measurement	Pred.	Meas.	Pred.	Meas.	Pred.	Meas.	Pred.	Meas.	Pred.	Meas.
EDT[s]	0.58	0.61	0.56	0.60	0.61	0.47	0.67	0.53	0.73	0.57	0.71	0.54
T15[s]	0.58	0.61	0.55	0.55	0.61	0.57	0.7	0.64	0.76	0.68	0.72	0.67
T30[s]	0.60	0.65	0.58	0.58	0.64	0.63	0.72	0.77	0.78	0.80	0.73	0.73
C50[dB]	4.2	2.8	4.4	3.5	3.6	5.8	3	5.3	2.2	4.2	2.6	4.5
SPL[dB]	76.8	N/A	76.6	N/A	77.3	N/A	78.1	N/A	78.6	N/A	78.2	N/A
STI	0.68	0.73	Rating:		Good	Good						
STIrMal	N/A	0.73	Rating:		N/A	Good						
STIrFem	N/A	0.74	Rating:		N/A	Good						

Table 1. Mean values for single omni directional source (prediction against measurement)

4.2 Auralization accuracy

The acoustical parameters derived from Dirac (open loop system) were compared to the equivalent output from CATT to assess the auralization quality (for 3 receiver positions). The parameters used included EDT, T_{30}, and STI.

For the first assessment, a close approximation to the existing trends in terms of EDT (figure 4) was observed, with the absolute values giving a slight deviation from the expected results. A maximum difference of 13msec (two values ignored) was derived, translating to a maximum of 20% error for individual cases. It is worth mentioning that the JND for RT alterations with music stimuli has been estimated at 20-30% in a recent study[9] (5-10% for noise length). While this result is not fully supportive of the accuracy of auralizations it comprises a point of reference for the naturalness of the resulting audio.

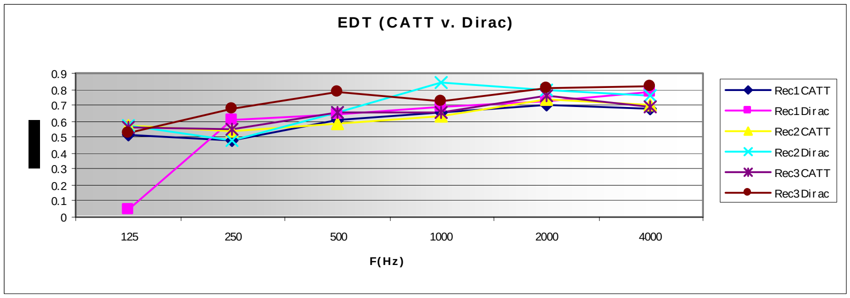

Figure 4. Comparison of EDT for CATT v. Dirac (3 receiver positions)

Examining RT values in more depth, it was found that T_{30} values were almost identical to the expected outcome (figure 5), giving only insignificant deviations (similarly for T_{20}). Being one of the main parameters for subjective impression it was established that auralizations accurately incorporated the effects of RT, in terms of T_{30}, being well within the limits of the JND for RT alterations.

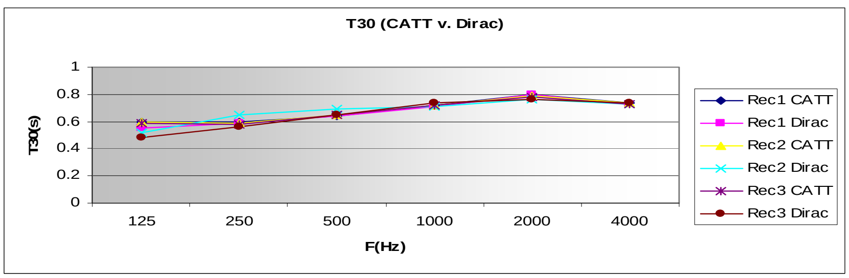

Figure 5. Comparison of T_{30} for CATT v. Dirac (3 receiver positions)

Speech intelligibility parameters gave contrasting results, with the Clarity (C_{50}) measure giving unrealistic results but STI closely following the actual conditions (see table 2). The slight differences observed for the latter measure (receiver position 3) were attributed to the erroneous estimation of individual Transfer Indices (i.e. octave band specific 'STI') relating to the EDT calculation, as the associated MTF summary suggests. An increasing error with distance to the source appeared to be the case (for EDT), implying a minimized error for multi source conditions. C_{50} values could not be realistically calculated here due to limitations relating to system setup.

Results nonetheless suggested that an accurate (within the JND for STI) assessment of speech intelligibility can be made in terms of STI, as also this measure relies largely on RT. Increasing accuracy can be expected for smaller source-receiver distances considering the effects observed in the EDT calculation.

Overall, the validation of auralizations using sine sweep impulse responses appeared to give a good indication of the subjective impression of auralized material, given also that speech intelligibility was one of the main parameters to be judged. The incorporation of additional measures relating to speech stimuli (e.g. C_{50}) in the current process could give a more complete representation of the actual conditions, highlighting the potential of the assessment method.

STI		
Receiver	**CATT**	**Dirac**
1	0.70 (Good)	0.67 (Good)
2	0.68 (Good)	0.66 (Good)
3	0.67 (Good)	0.60 (Good)

Table 2. STI derived from CATT Acoustics and Dirac (3 receiver positions)

5 DISCUSSION

5.1 Speech intelligibility optimization based on CATT predictions

In the attempt to enhance speech intelligibility ratings several conditions were examined using the models of the spaces considered. Model validation prior to experimentation gave confidence that any actions to alter the acoustical conditions in the actual spaces would be consistent with the prediction results and thus a feasible target could be set. Experimentation with the models here allowed optimal conditions of several cost efficiency levels to be highlighted, nonetheless since these are not in the immediate scope of this paper they will not be reported here.

5.2 RT as a reference for model performance

The validation procedure (see 3.2) appeared as a reliable way to adapt the models to the actual conditions. The process was performed in terms of the RT (T_{30}) values at the receiver positions thus confirming, to some extent, that the specific measure largely incorporates the general room characteristics for computer modeling purposes.

Smaller differences, found for T_{15} and EDT in the data comparison moreover, suggested that optimization relying solely on T_{30} is principally a simplified though efficient version of the process. Given a more detailed approach, a need for further references (e.g. EDT, SPL) might aid the computer model to perform at the level of required accuracy. The validation process in this case could highlight the need for alterations in the construction of the model so as to allow for better matching datasets between predictions and measurements.

5.3 Objective verification of auralization accuracy

The auralized room responses could be described as of good quality in terms of T_{30} and STI as demonstrated by the verification process. Thus, an accurate space assessment in the context of speech intelligibility could be expected under analogous conditions.

Considering the smaller error margins that were suggested for shorter source-receiver distances it appeared that modelling a distributed sound system installation could result in a more consistent outcome when compared to single source conditions for a large space. For the case of classrooms or small auditoria in particular the specific outcome supports the use of the method since speech intelligibility is the main prerequisite for a space to be characterized as well performing.

A subjective evaluation relating to the naturalness of sound complemented the objective method for Lecture Room B. Some differences were revealed between the predicted and actual case, although STI was accurately predicted. A more detailed model validation (see 5.2) could potentially enhance naturalness in the auralization, while however it could be rendered unnecessary for speech applications.

6 CONCLUSIONS

T_{30} values proved to be an efficient reference for computer model validation while also allowing for a satisfactory, in terms of T_{30} and STI, accuracy level in the auralization part of modelling.

The optimization process was enabled as an option, given the models' consistency for the three case studies in the terms described. A number of configurations were examined and optimal conditions of several cost efficiency categories could be highlighted.

The auralization quality could be objectively verified with the use of sine sweep test signals and an open loop measurement system. The overall method, using T_{30} for model validation, proved to be highly efficient within the context of speech intelligibility.

It was established that naturalness of audio and speech intelligibility performance in terms of an auralization quality assessment can comprise two individual tasks, not necessarily related with each other.

7 FURTHER WORK

It is intended to expand the investigation in the model construction and setup process so as to establish the level of accuracy that could be expected for a given setup. Verification of performance for different source setups in particular will enhance confidence for intelligibility assessments. Naturalness of audio is an additional factor to consider on the basis of the actions required in model construction and validation. Objective and subjective verification will be further referenced in this sense.

8 REFERENCES

1. WinMLS 2004, Professional Measurement Software for PC and Soundcard, User's Manual, www.winmls.com
2. BS EN ISO 3382- 2000, Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameters
3. CATT Acoustics 8.0f, Acoustics prediction/auralization software, http://www.catt.se/CATT-Acoustic.htm
4. DIRAC 3.1, Room Acoustics Software Type 7841, User's Manual, http://www.bkhome.com
5. Nestoras C. et al, Speech intelligibility measurements in a diffuse space using open and closed loop systems, 19th International Congress on Acoustics, Madrid, 2-7 September 2007
6. Gomez L. et al, Speech intelligibility measurements in a non diffuse space using open and closed loop systems, 19th International Congress on Acoustics, Madrid, 2-7 September 2007
7. J. S. Bradley et al, A just noticeable difference in C50 for speech, Applied Acoustics 58 (1999) 99-108
8. J. S. Bradley, Predictors of speech intelligibility in rooms, Journal of the Acoustical Society of America 80 (3), September 1986
9. Zihou Meng, Fengjie Zhao, and Mu He, The Just Noticeable Difference of noise length and reverberation perception, International Symposium on Communications and Information Technologies, ISCIT 2006