ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Architectural Acoustics
Session 3aAAa: Virtual Concert Hall Acoustics I

3aAAa4. Evaluation of stage acoustics preference for a singer using oral-binaural room impulse responses
Luis A. Miranda Jofre*, Densil Cabrera, Manuj Yadav, Anna Sygulska and William Martens

*Corresponding author's address: Faculty of Architecture, Design and Planning, University of Sydney, Sydney, 2006, NSW, Australia, lmir9852@uni.sydney.edu.au

There are two objective measurement methods in current practice that can be used to evaluate the stage acoustic conditions for singers. One is the stage support metrics (ST\textsubscript{Early} and ST\textsubscript{Late}, included in the standard ISO 3389-1), and the other is the voice support metrics proposed by Brunskog et. al. and revised by Pelegrín-García (room gain (GRG) and voice support (STV)). All of these metrics use energy integration from impulse responses to derive the acoustic descriptors. This overlooks two potentially important features of the responses: the temporal distribution of the impulse response within the evaluation period, and the directional distribution for the spatial impulse response within the evaluation period. In this paper, a method to study the effect of these features is proposed and tested. This method is based on the auralisation of one's own voice in rooms using oral-binaural room impulse responses (OBRIRs). The OBRIRs used are created by combining synthesized early reflections with a recorded reverberant tail. Results of a pilot study indicate that a wide range of on-stage acoustic quality ratings can be observed for stimuli with a similar ST\textsubscript{Early} value due to variation in the temporal and spatial distribution of reflected energy.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

In unamplified performances, the ability of a musician to perform is influenced by the acoustic conditions on stage. Over the last few decades, researchers have studied the relationship between measured acoustical values and musicians’ auditory impressions of stages (for a thorough review on the subject see Gade (2011)). Advances in stage acoustics understanding have led to metrics in the standard ISO 3382-1 (ISO, 2009) \( \text{ST}_{\text{Early}} \) and \( \text{ST}_{\text{Late}} \). The metrics included in this standard have their origins in the stage acoustics investigations conducted by Gade (1989a; 1989b), and later revisions also by Gade (1992). In these studies it was found that a common acoustical characteristic of stages as experienced by musicians was the feeling of ‘Support’ and that this parameter had a strong correlation with the early energy of impulse responses measured on stage. \( \text{ST}_{\text{Early}} \) describes the ratio of reflected energy received by a microphone positioned 1 m from the source during the time interval from 20 to 100 ms to the direct energy produced by an omnidirectional source. While these metrics provide a useful general assessment, studies by Marshall et. al. (Marshall, Gottlob, & Alrutz, 1978; Marshal & Meyer, 1985) found that the perceptual characteristics determining quality of a stage might differ for musicians playing in ensemble versus playing solo. This difference is especially important for singers performing in ensemble or solo.

While not strictly addressing musicians, the parameters proposed for relating vocal support with room acoustics by Brunskog et. al. (2009) and the revision by Pelegrín-García (2011) can provide some insight to the experience of solo singers on stage. The parameter termed voice support (\( \text{ST}_V \)) is closely related to stage support, in that it relates the energy received by the talker in the form of room reflections to the sum of the early energy produced by a talker (typically the direct level for the direct path from mouth to ear). Pelegrín-García et. al. (2011) have demonstrated that the amount of ‘amplification’ provided by the room for a speaking subject (expressed as \( \text{ST}_V \)) and the associated parameter of room gain, \( \text{G}_{\text{ac}} \) influences the vocal effort of the talker. To the knowledge of the authors, this metric has never been applied to the study of singers’ experiences on stage.

\( \text{ST}_{\text{Early}} \) and \( \text{ST}_{\text{Late}} \) have had wide acceptance since their inception, and remain the preferred method for stage acoustic measurements. \( \text{ST}_V \) appears to be a viable solution to the problem of measuring the influence of room acoustics in voice related tasks. Nonetheless, these measurements assume that only the energetic ratio of early to late parts of an impulse response is important, and disregard the spatial and temporal details.

Earlier studies have shown the influence that the temporal and spatial characteristics of reflections can have on musicians. Marshall et. al. (Marshall et. al., 1978) conducted a study where preference of arrival time was investigated in an anechoic setup in which simulated reflections were produced using loudspeakers located to the left, right, behind and above the subject. In that study the simulated reflection playback was synchronous; in order to break up the reflections pattern the subject was positioned off center. The study reports results in terms of the onset arrival time, however, analysis of the finer spatial and temporal details is absent.

Studies have been conducted that relate the preferred timing of a single reflection to the Auto Correlation Function (ACF) of the singer’s own voice (Noson et. al., 2000) and of the musical signal being followed (Nakayama, 1984). Also, studies have been conducted that show that the arrival time of a reflection has no dependence on the arrival angle (Nakayama & Uehata, 1988). However, it is difficult to extend the applicability of these results to complex soundfields that include more than one reflection.

Perhaps the closest study to the present one examining the influence of the temporal and spatial distribution of reflections was conducted by Marshall and Meyer (Marshall & Meyer, 1985). This study was conducted in an anechoic chamber using four loudspeakers to reproduce discrete reflections from the ceiling, sidewalls and back wall, plus another three loudspeakers reproducing a reverberant field. The temporal relationship between discrete reflections in that study was based on the dimensions of physically plausible stages. The study reported the results in terms of the onset of the first reflections. Again, there was no mention of the finer temporal structure of the stimuli or the spatial characteristics of the reflection patterns.

Similarly, Dammerud et. al. (2011) point at the importance of the direction of early reflections and correlate architectural configuration to preference on stage. From these results, guidelines were derived based on the width, height and depth of stages. Although Dammerud’s result suggest that particular ratios of width, height and depth are more successful within the spaces tested, these guidelines are difficult to generalize to other spaces, which were not tested within that research. Furthermore, they offered no analysis of the discrete spatial and temporal characteristics of early reflection patterns. Therefore, we have proposed that further analysis could enable the creation of better measurement techniques to validate new and existing stages.

The study presented in this paper aims to examine some of the variables that have not been included in previous studies: namely, the influence of the spatial and temporal characteristics of early reflections on perceived quality of stages. This is done by systematically varying the temporal and spatial diffusivity of the room impulse responses.
used as experimental stimuli. Furthermore, the experimental setup used for this study allows for complete three-dimensional reproduction via (Head Related Transfer Functions (HRTFs), instead of using discrete loudspeakers, which limit the reproduction field.

**METHOD**

An experiment was designed to explore the effect that temporal and spatial diffusivity of early reflections have on singing support, while keeping ST parameters constant. The data shown in this paper were collected in a pilot study conducted using one performer as subject (author #4). The subject is a female classically trained opera singer, who also holds a PhD and a lecturer position in acoustics.

**Experimental Setup**

The audio system used for this experiment is the same system as that proposed by Yadav, et. al. (2012), and briefly described for the current paper in this section. The main feature of the system used is the ability to simulate a room acoustic environment using one’s own voice as the input. The system was conceived as a method of simulating autophony using Oral Binaural Impulse Responses (OBRIR). Autophony is described as the perception of one’s own voice in an acoustic environment (Lane, Catania, & Stevens, 1961). An OBRIR is an acoustical measurement that describes the response of a source/receiver system, with the characteristics of a human mouth as the source and the ears of a collocated head as the receivers (Cabrera, Sato, Martens, & Lee, 2009).

In order to simulate autophony, three acoustic paths need to be accounted for:
1) The direct acoustic path from the mouth to the ears.
2) Bone and body conduction from the mouth to the cochlea.
3) Room reflections arriving at the ears.

In the system used, the first path is accounted for by the unobstructed sound arriving from the subject’s mouth to the subject’s ears; the second path also requires no simulation. The third path is reproduced using ear loudspeakers fed by a convolution system that combines the subject’s voice and a set of OBRIRs.

The latency of the convolution system determines the limit of the arrival time for early reflections. If we think of this in terms of sound in space, the minimum latency will determine the closest surface from which we can obtain a reflection. The latency of the system has been measured at 7 ms with a sampling frequency of 48 kHz, resulting in a minimum distance of 2.04 m for the earliest reflection. The system is setup in an anechoic chamber, therefore the only reflections that the subject hears are the reflections simulated with the convolution system, apart from the reflections from a plywood floor installed to simulate the stage floor.

The convolution system is based on the SIR2 convolution plugin, hosted in the Max/MSP software running on a Windows platform. The AD/DA converter is an RME ADI-8 QS. The system’s input is provided by a DPA 4066 microphone. The ear loudspeakers are a pair of AKG K1000. The inclusion of the ear loudspeakers has been shown to have a negligible effect on the direct sound path (Yadav et al., 2012).

**Stimuli**

The aim of the experiment was to examine the influence of spatial and temporal diffusivity of early reflections on singing support, and these parameters were varied to create the early part of the stimuli. Temporal diffusivity
refers to the temporal spread of the reflections in the 20 to 100 ms interval. A low temporal diffusivity would yield stimuli for which the reflections would arrive at a single temporal moment. A high temporal diffusivity would yield stimuli for which the reflections would arrive spread across the entire period of interest. Spatial diffusivity refers to the spatial spread of the reflections in the 20 to 100 ms interval. A low spatial diffusivity would yield reflections arriving from the same direction, regardless of their timing. A high spatial diffusivity would yield reflections distributed so as to arrive with near to equal probability from all angles over a sphere around the listener. Two additional variables describe the temporal and spatial central tendencies. The temporal centroid was defined as the moment in time where the middle reflection in a sequentially ordered train of reflections would arrive. The spatial centroid was defined as the mean spatial direction for the reflections arriving at the listener position.

The OBRIRs were created by concatenating synthetic early reflections (prepared by systematically varying the spatial and temporal characteristics) and a reverberant tail recorded using the method outlined by Cabrera, et al. (2009). The temporal distinction between early and late reflections follows the ST Early Convention, with a threshold of 100 ms between early and late reflections.

![FIGURE 2. Stimulus structure. The three temporal diffusivity conditions used for the experiment are shown: one reflection shown in red, three reflections shown in blue and ten reflections shown in green. It should be noted that temporal jitter was added to the reflections to prevent sound coloration.](image)

To create the early part of the impulse response, Head Related Impulse Responses (HRIRs) measured using a Head And Torso Simulator (B&K 4128C) were used. Each discrete reflection was modeled as a single HRIR. The pattern for early reflections was created by concatenating these HRIRs. With regards to the timing of reflections, temporal jitter was introduced in order to break up the reflection pattern, which could create comb filtering and unwanted excessive coloration. The direction of arrival of reflections for stimuli with higher spatial diffusivities was chosen based on the calculated great-circle distance. Distances to the spatial centroid were calculated and directions were chosen by selecting the reflections that would yield an equal sampling of the sphere based on the desired angle of coverage and number of reflections.

![FIGURE 3. Example of distribution of reflection directions for 10 reflections and three diffusivity parameters. Single direction diffusivity shown in red, half a sphere diffusivity shown in blue and full sphere diffusivity shown in green.](image)
The reverberant tail used was captured in a 610 m\(^3\) lecture theatre, the measurement of which is described by Cabrera et. al. (2010). The reason for using this reverberant tail was its high signal-to-noise ratio. The reverberant tail was adjusted to have a 1 second reverberant tail across all frequency bands, adjusted using the method described by Cabrera et. al. (2011). The reverberant tail was constant across all stimuli, yielding an ST\(_{\text{Late}}\) of -16 dB.

The level of the early and late parts of the stimuli was adjusted based on the ST\(_{\text{Early}}\) and ST\(_{\text{Late}}\) definitions by Gade (1992). The main objective of the experiment presented here was to understand the influence that spatial and temporal diffusivity have on the perception of support while maintaining similar values of ST. All stimuli were adjusted for ST\(_{\text{Early}}\) to have a value of -12 dB (and ST\(_{\text{Late}}\) to have a value of -16 dB). The levels chosen provide realistic levels of support to the subject by conforming to typical values of ST\(_{\text{Early}}\) (Beranek, 2003; Hidaka & Nishihara, 2004). A level adjustment was made to take into account the fact that the ST parameters are usually measured 1 m from the source, while the microphone sits 7 cm from the singer’s mouth.

Table 1 enumerates the different parameters for the stimuli created for the test.

<table>
<thead>
<tr>
<th>Variables Under Test</th>
<th>Temporal Diffusivity</th>
<th>Spatial Diffusivity</th>
<th>Temporal Centroid</th>
<th>Spatial Centroid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units</td>
<td>Number of Reflections</td>
<td>Percentage of Coverage</td>
<td>Milliseconds</td>
<td>Angle Pair [Azimuth, Elevation]</td>
</tr>
<tr>
<td>Values Tested</td>
<td>0 1 3 10</td>
<td>0 1 50 100</td>
<td>40 50 60 80</td>
<td>NA [0,90] [0,180] [90,0] NA</td>
</tr>
</tbody>
</table>

**Experiment Test**

The test consisted of presenting an autophonic stimulus to the subject in the form of a ‘virtual room’ in which the subject would sing the same musical excerpt and provide a rating for a set of categories. The musical excerpt was the aria ‘Habanera’ (or ‘L’amour est un oiseau rebelle’) from the opera ‘Carmen’ by George Bizet. The subject was provided with a piano accompaniment to help her with timing. The accompaniment was recorded using MIDI with the subject and accompanying pianist in the same room prior to the experiment. The recorded accompaniment was presented via a small loudspeaker (Fostex PM0.4n) in the anechoic chamber and its level adjusted by the subject prior to the actual test. The level of the accompaniment was not varied throughout the test and it was not convolved with any room impulse response.

**FIGURE 4.** Graphical interface presented to the subject.
The rating questionnaire included the following scales: Support Quality, Timbre, Reverberance and Overall Quality. The scale ‘Room Size’ was included for another project. These categories were chosen based on common subjective characteristics elicited in previous studies (Gade, 1989a). The subject was presented with an iPad interface with sliders on a seven point Likert scale (shown in Figure 4); the output of the slider being a five digit precision number from 0 to 1. The ratings were made after the singer had completed the musical excerpt.

Given the large amount of possible parameter combinations, the test was reduced to a practice set (12 stimuli) and an experiment run (15 stimuli). The practice set was used to allow the singer to familiarize herself with the system. Only the results from the experiment run are reported.

RESULTS

The results are summarized in Figure 5. The recorded four category ratings are presented as scatter plots with the rating result on the y-axis and the four variables under test on the x-axis. All the results presented in Figure 5 are presented on a linear scale, except for the spatial centroid variable where the centroid is expressed as a category. For the temporal diffusivity variable, the results are presented as the number of discrete reflections within the first 100 ms. For the spatial diffusivity, the results are presented using the percentage of the sphere covered by the arriving reflections. A special red-colored case was included for clarity to distinguish the 0% case, where there are no reflections arriving within the first 100 ms, and the 1% case where reflections are arriving from only one direction. The temporal centroid is expressed in milliseconds, and in the case in which there were no reflections in the first 100 ms, the rating appears as >100 ms. The spatial centroid is expressed as categories comprising the azimuth and elevation angles. In two of the sphere coverage cases, the ‘no reflections’ case and the ‘100% sphere coverage’ case, the situation was noted as Non Applicable.

![Figure 5](image-url)
The results presented in Figure 5 do not exhibit a clear pattern, and so it is hard to distinguish a case in which one of the variables clearly influences the subjective parameters. Given how sparse the data collected are, and the fact that only one subject was tested, the usual inferential statistical tests are not suitable. Instead, the authors have broken the ratings down into three categories: low, medium and high ratings. Then the results could be re-evaluated using the rating categories. For this analysis, ratings distributed over the whole quality metric were grouped so that each ‘grade category’ had just five observations. The results of this grouping are presented in Figure 6 as histograms representing the number of occurrences of a given variable for each grade category.

Some observations can be made with the data organized in this manner. If we look at temporal diffusivity, we can see that there are no observations in the lower quality rating when there are 10 reflections, and in general lower number of reflections are included in the lower category. Furthermore, the trial where there were no early reflections is included in the group with the lower rating.

The results for the spatial distribution are harder to interpret than those for temporal distribution, and so further tests will be completed before any tentative conclusions on the influence of spatial distribution will be suggested. On the other hand, the categorically coded results for the temporal centroid show that reflections arriving towards the later part of the initial 100 ms are included in a greater proportion in the lower ratings category. The highest rating category includes the largest amount of trials where the temporal centroid was 40 ms. This result supports a tentative conclusion that the most important reflections arrive in the earlier part of the 20 to 100 ms interval.

![Temporal Diffusivity](image1)

![Spatial Diffusivity](image2)

![Temporal Centroid](image3)

![Spatial Centroid](image4)

**FIGURE 6.** Histograms of variable occurrences plotted against overall quality categories.

Additionally, we can analyze the extreme cases individually, where the subject reported the lowest and highest values for overall quality. The three lowest and highest overall quality trials are summarized in Table 2. The most obvious pattern again is the tendency of the subject to judge as low quality the stimuli exhibiting later temporal centroids. The common factor in the three lowest rated stimuli is the 80 ms temporal centroid. For the highest rated stimuli, there are no occurrences of stimuli with temporal centroids lower than 60 ms. Also, we can see that all the highest rated stimuli have a temporal diffusivity of three reflections.

Parallel to the scale rating, the subject was asked to provide comments on the reasoning behind the ratings. It was a common description for the lowest rated cases that the room included a very prominent echo that was too
distracting for the performance. It is worth mentioning that the performer could not complete the lowest two rated cases due to the strain that was required to perform. Also, a common response from the performer was that very prominent reflections from the sidewall were too distracting. This can be observed in two of the cases included in the low rating bracket.

<table>
<thead>
<tr>
<th>TABLE 2. Summary of highest and lowest rated stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Diffusivity (number of reflections)</td>
</tr>
<tr>
<td>Lowest</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Highest</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION

It has been suggested in the past that singers, and especially solo singers, might have a different perception of what characteristics provide good support in stage acoustics compared to musicians playing in an ensemble. A method to test different conditions on stage while providing precise control of the parameters tested is proposed in this paper. The results obtained from a pilot study, while inconclusive, point at the viability of the method proposed to obtain valuable data towards the understanding of stage support quality perception for singers.

It is important to notice that all of the stimuli presented to the subject would yield the exact same values of ST<br>. It is fair to say that for a given ST<br> value different quality ratings of stage support can be experienced depending on the temporal and spatial reflection pattern. A better understanding of the preferred temporal and spatial characteristics of reflections patterns could in turn lead to better design criteria for stages.

While inconclusive in terms of providing preferred spatial and temporal parameters for stages, the pilot study presents an encouraging case in favor of further studies that could provide valuable information in the field of stage acoustics, by methods sensitive to the temporal and spatial distribution of the early sound field.

ACKNOWLEDGEMENTS

This research has also been supported by an Australian Postgraduate Award for Manuj Yadav; and an Australian Postgraduate Award and a CONACYT (a Mexican government initiative) scholarship from the Becas al Extranjero program for Luis Miranda. This research has also been supported by an Australian Acoustical Society NSW Division Travel Award to Luis Miranda. This research was supported under the Australian Research Council’s Discovery Projects funding scheme (project DP120100484).

References