The regulation of voice levels in various room acoustic conditions

Manuj Yadav, Densil Cabrera, Doheon Lee, Luis Miranda and William L Martens

Faculty of Architecture, Design and Planning, The University of Sydney, 2006, Australia

ABSTRACT

Talking-listeners show a systematic variation in their voice level due to changes in the ambient noise level (Lombard effect) and the perception of their own voice. While these effects have to be interpreted in the context of the communication task, some studies point out an additional influence of room acoustics. This pilot study investigates the changes in the voice levels of participants in various room acoustic conditions that were simulated in anechoic conditions. The participants vocalized three vowels at a comfortable level to a manikin at a distance of 5 meters, while listening to the sound of their voice in a simulated room. The results are in some agreement with recent findings that show a negative relationship between the voice level of a talking-listener and the level of the room reflections returned from their own voice (quantified as room gain). However, the overall trend was not uniform across the room acoustic conditions studied here, and current results are discussed in comparison to results from past studies.

INTRODUCTION

Over a century ago, Lombard discovered that talking-listeners exhibited a systematic increase in their voice level with an increase in the ambient noise level (Lombard, 1911). In the early days of telecommunication, researchers investigated a related phenomenon, known as sidetone compensation (Lane et al., 1961). A sidetone refers to presentation of the airborne sound of one's own voice to the ear by electronic means, while the natural airborne mouth-to-ear conduction pathway is occluded in one or both ears. In this scenario, studies have shown that talking-listeners 'automatically' compensate for a reduction in the sidetone level by increasing their voice level, and vice-versa (refer to Lane and Tranel, (1971) for a review).

Lombard posited, and it has since been justified in many studies, that the Lombard effect and sidetone compensation (or in more general terms, sidetone variation) are in fact, two sides of the same coin. In humans and many other species (refer to (Brumm and Zollinger, 2011) for a review), these effects have been shown to be complements serving the same underlying function: maintenance of a seemingly effective signal-to-noise (S/N) ratio in changing communication scenarios.

Each of these effects is associated with slopes of approximately 0.5 (voice level vs. stimulus level, both in decibels), albeit with opposite signs (Lane et al., 1961). However, two comprehensive reviews have highlighted the importance of contextualising these slopes under the requirements of the communication scenario (Lane and Tranel, 1971; Brumm and Zollinger, 2011). The evidence presented in these reviews shows that the slope magnitude is dependent on the premium placed on intelligible communication. A higher premium, implying higher emphasis being placed on the listener being able to understand the talker, who talks under varying ambient noise or sidetone levels, leads to slopes closer to 0.5. Flatter slopes, found in some studies (Lane and Tranel, 1971), have been accompanied with a low premium placed on communication, e.g., while reading a list of names not addressed to anyone in particular.

Besides ambient noise (broad- or narrow-band) and sidetone, talkers also vary their voice levels (sound power or sound pressure levels) when they perceive a change in the level of room reflection from their own voice (Black, 1950; Brunskog et al., 2009; Pelegrín-García and Brunskog, 2012). This indicates an effect of room acoustics on the voice level of a talking-listener, when the background noise remains reasonably low and steady. The sound of a room's response (in the form of reflections) to one's own voice is referred to as the *autophonic room response* in this paper. It provides the indirect airborne component of the overall perception of one's own voice (i.e., *autophonic perception*, which also includes the perception of direct airborne and body conducted sound (Pörschmann, 2001), and the term autophonic room response is introduced to highlight the influence of room acoustics. Voice level changes due to autophonic room response can be considered a variation of sidetone compensation (hence, related to Lombard effect), where the sidetone is the autophonic room response that is received binaurally, without any occlusion.

To address the influence of a communication scenario on autophonic room response, consider the case of two or more people engaged in a conversation, or a talking (or singing)-listener addressing an audience across a certain distance in a room. Other psychological factors aside (mood etc.), research indicates that voice levels (quantified as sound power or sound pressure levels) change according to the autophonic room response (Brunskog *et al.*, 2009; Pelegrín-García and Brunskog, 2012). More generally, Lane and Tranel (1971) reasoned that communication seems to be governed by interplay of an exocentric public loop (scenario dependent, such as distance to the listener, room acoustics, etc.) and an egocentric private loop (sensory mechanisms dependent). This pilot study presents the results of a study where the effect of autophonic room response on talking-listeners' voice level was measured in a simulated communication scenario. While other voice characteristics (pitch, duration, etc.) have also been shown vary with autophonic room response (Brumm and Zollinger, 2011), this paper is focussed on the variation in sound pressure levels (SPL).

Black (1950) provides the results of one of the earliest investigations in studying the effect of autophonic room response on voice levels by using 8 actual rooms. The rooms used were of two volumes (4.2 m^3 and 45.3 m^3), with each volume tested with two variations in shapes (drum and rectangular) and reverberation times (0.8-1 and 0.2-0.3 s). The talking-listeners were instructed to vocalize in order to make themselves clearly audible to an experimenter facing them at a distance of 2.4 m. The results showed a significant effect of room volume and reverberation time (*RT*) on the talking-listeners' voice level and duration of vocalizations. Voice level of talking-listeners was greater in smaller, less reverberant rooms.

Apart from the relevant acoustical parameters defined in ISO 3382-1 (2009), recent studies on change in autophonic levels have used acoustical parameters that characterise the airborne sound transmission of voice from mouth to two ears in a room environment, represented by an oral-binaural room impulse response (OBRIR) (Cabrera *et al*., 2009). Two such parameters are room gain and voice support. Theoretically, room gain is the amplification (in dB) provided to one's own voice by an acoustic environment, relative to anechoic conditions (Brunskog et al., 2009). Room gain (G_{RG}) can be expressed as:

$$G_{RG} = 10\log\left(10^{ST_V/10} + 1\right) \qquad (1)$$

It is derived in Equation 1 from first calculating ST_V, or voice support, which is similar to Gade's stage support (Gade, 1989). The concept underlying voice support can be understood by categorizing an OBRIR into its direct and reflected components. The interval for the direct and reflected components is task specific. For example, Pelegrín-García et al. represented the direct component as the energetic integral of the first 5 ms of an OBRIR, and the reflected component as the energetic integral of the remainder of the OBRIR (Pelegrín-García, 2011). Voice support is then evaluated as the difference between the direct and reflected components of the OBRIR.

Brunskog *et al*. (Brunskog *et al*., 2009) showed that the voice level of talking-listeners delivering a lecture in 6 actual rooms (including a 1000 m^3 anechoic chamber) was correlated with the logarithm of the room volume (log*V*), and varied at a rate of -3.6 decibels per decibel increment in G_{RG}. The range of volumes was between 94-1900 m^3. The authors commented on the lack of correlation with reverberation times, which could have been due to a large separation in the *RT* values of rooms used.

In order to circumvent logistical limitations in conducting in-situ experiments in real rooms, recent studies have been exploring the possibilities in assembling factorial experiments by using realistic auditory simulations (Vorländer, 2008; Pelegrín-García et al., 2011; Yadav et al., 2012) (with/without a visual component). Using such auralization principles, Pelegrín-García *et al*. (2011; 2012) have studied the change in voice levels of talking-listeners from anechoic conditions to simulated acoustic conditions in a series of experiments, which are discussed briefly.

The first study used binaural impulse responses, which were derived by applying a set of exponential decays to Gaussian noise (Pelegrín-García et al., 2011). Their results showed a significant effect of various acoustic conditions' G_{RG} on the voice level (presented as sound pressure level) changes, and no influence of reverberation time. The relationship between voice level changes and G_{RG} (or equivalently ST_V) was modelled with a non-linear equation in which G_{RG} was represented as a negative exponential term. This implies that in acoustic conditions with higher G_{RG}, voice level was lower and vice-versa. They also pointed out that the auditory feedback alone could not account for the changes in voice levels; sensory and proprioceptive mechanisms, communication premiums, etc. are also likely to have an affect on autophonic room response in different acoustic conditions, which is corroborated by the findings from previous studies (Lane *et al*., 1961). Since the aim of their study was not to simulate actual room conditions, the results need to be interpreted accordingly. It could be argued that the generalization of these results to actual rooms might be limited. Actual rooms generally vary in many subtle and/or obvious ways (spectral, spatial, interaural, etc.), besides the relation between the physical dimensions and sound absorption qualities. It must also be noted that the task in their study was the talking-listener matching the level of their anechoic voice to the level of autophonic room response of simulated acoustic condition. Hence, this scenario seems to be placing a low premium on communication (or absolute lack of it, as it was a matching task with one's own voice).

In another study, Pelegrín-García and Brunskog (2012) conducted experiments under 4 different simulation settings, with a different instruction given to the talking-listeners per setting. The OBRIRs used in this study were derived from computer models of rooms (representing classroom settings) in ODEON. The original OBRIRs were subjected to similar gain changes as their previous study (Pelegrín-García et al., 2011). The findings show a variation in voice level with the communication scenario. The average values of the changes in voice levels with ST_V ranged from -0.93 dB/dB with instructions to speak freely to an imaginary audience (i.e., high premium on intelligible communication) to -0.1 dB/dB in scenarios with low premium on intelligible communication. The authors also pointed out variation in the slopes of voice level changes with ST_V changes across individual talking-listeners, within the same and different communication scenarios. This implies a more variable interaction between the talking-listeners, communication scenario and autophonic room response than their previous study (Pelegrín-García et al., 2011) where the impulse responses were not representative of actual rooms.

In the current study, a wider range of volumes, G_{RG} and *RT* are tested in an experiment where various rooms are simulated for autophonic perception by using their OBRIRs (some actual, some manipulated). The rationale behind the selection of OBRIRs, their measurement, manipulation, and simulation is presented in the methods section. The experiment and the communication task used are also presented in the methods section. This is followed by the results and conclusions.

METHOD

A total of 32 room conditions were tested in the experiment. These 32 conditions were created from 6 actual rooms that were measured with the OBRIR measurement procedure (Cabrera *et al*., 2009). This procedure derives binaural impulse response from the mouth to the two ears of a head and torso simulator (HATS, Brüel & Kjær Type 4128C). The resulting OBRIR contains the direct sound from the mouth to the two ears, generally followed by the floor and the room reflections.

The 6 real rooms that were chosen were classified into three categories based on their volumes (V), i.e., *small, medium and large* (Figure 1). Room SB is a recording booth; ENS is an ensemble music practice room; L2T is a lecture theatre; ODS is a performance space with a slanted church ceiling architecture; RCH is a music recital hall for soloists or small ensembles with 116 raked seats on the floor level; and VRB is a recital hall with 528 seats over several sections.

The OBRIRs from these 6 real rooms were then manipulated to derive the full stimulus set in the following manner:

- Each OBRIR was subjected to a RT manipulation procedure (detailed in the following subsection) to increase and decrease the duration of reverberant decay by a certain ratio.
- Each OBRIR was subjected to a G_{RG} manipulation procedure (detailed in a following subsection) to increase and decrease the level of amplification (in dB) provided by the room reflections, within realistic limits for the actual rooms.
- 2 control stimuli
- Total: 6 original + (6 × 2) RT manipulations+ (6 × 2) G_{RG} manipulations + 2 control stimuli = *32* rooms

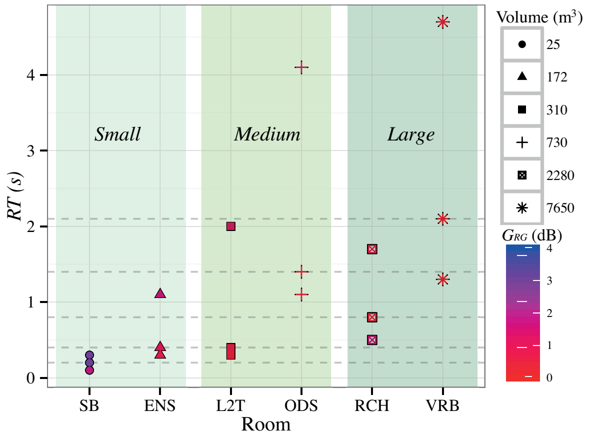


Figure 1. Three RT values for each room. Per room, the original RT is specified by a dotted line passing through its plotted point. Values above and below the original represent the increased RT decreased RT, as per manipulations described in this section. The color gradient for each value represents its G_{RG} value.

RT manipulation of real rooms' OBRIRs

The octave band RTs of the OBRIRs were increased and decreased by a factor of 2.3 and 0.6 respectively (except for room SB where a multiplicative factor of 1.1 was used). These changes were performed by multiplying the reverberant decay (in octave bands) by an exponential curve, as detailed elsewhere (Cabrera *et al.*, 2011). The reasoning behind this manipulation was twofold: to obtain a wider range of RTs from the available rooms, and to selectively group together rooms of relatively vast difference in volumes to have similar RTs.

G_{RG} manipulation of real rooms' OBRIRs

The G_{RG} manipulation involved changing the value of the original G_{RG} (calculated from its OBRIR) to a value up and a value down. This manipulation was performed at the convolution stage, which is addressed below in the experimental setup. The G_{RG} manipulations, as seen in Figure 2, were chosen individually in each case to represent realistic gain changes within the room. As a result, the gain changes were in the range of -0.9 dB to 1 dB. Even though some of the G_{RG} manipulations were less than the JND of broadband noise levels (Zwicker and Fastl, 1999), such small changes have been reported as being detectable in recent findings (Hafke, 2009; Pelegrín-García et al., 2011; Pelegrín-García and Brunskog, 2012).

It must be noted here that only the 6 original OBRIRs mentioned in Figure 1 (i.e., the symbols intersected by the dotted lines) were subjected to gain manipulation, not the OBRIRs derived from RT manipulation. However, the RT manipulation of each OBRIR, which either extends or shortens the reverberant decay, also changes the ratio between the direct and reverberant energies, quantified here as G_{RG}. Although it is possible to have a change in RT without changing G_{RG} values, such a manipulation would involve changing the energy pattern of the reflections, which was not intended in this study. Hence, there are more sample points in Figure 2, which depicts the G_{RG} values of the 30 room conditions that were simulated, excluding the 2 control conditions that are described in the following subsection.

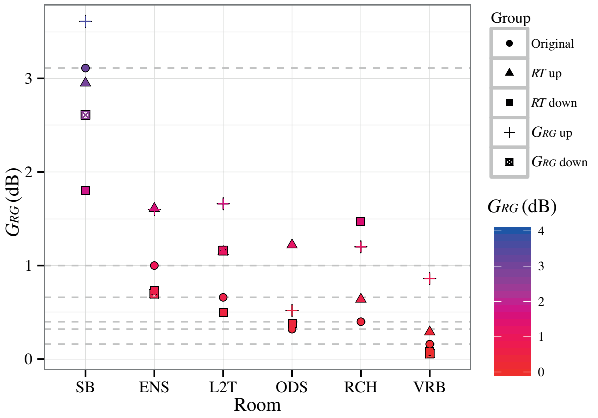


Figure 2. Five G_{RG} values for each room. Per room, the original G_{RG} is specified by a dotted line passing through its plotted point, and G_{RG} manipulations are denoted according to the group they belong to. The color gradient for each value represents it G_{RG} value.

Control conditions

There were 2 control conditions that were introduced to be used as reliability checks in later analysis. The room RCH was presented twice with gain turned up by 0.8 dB, and twice with gain turned down by 0.6 dB.

Real-time simulation for autophonic perception

The room acoustical simulation system used in the current paper has been used in similar studies on autophonic perception (Yadav *et al.*, 2011; Yadav *et al.*, 2012; Miranda *et al.*, 2013). In the system, the voice of a talking-listener is input with a headset microphone (DPA 4066) positioned 7 cm from the centre of the lips, convolved in real-time with an OBRIR, and output to a pair of head-worn ear-speakers (AKG K1000). A similar system was used in recent studies of autophonic perception by other authors (Pelegrín-García et al., 2011).

The electroacoustic latency of the system (less than 8 ms) is accounted for in the simulation by cropping the direct sound and first floor reflection from each OBRIR. As a result, the system outputs room reflections delayed by the right amount after the direct sound (provided by a talking-listener's own voice) and floor reflection (provided by placing a wooden board in an anechoic room) to accurately recreate conditions for autophonic perception in a room acoustic environment represented by an OBRIR. The system also allows changing gain levels and room conditions in real-time. Figure 3 provides an overview of the system.

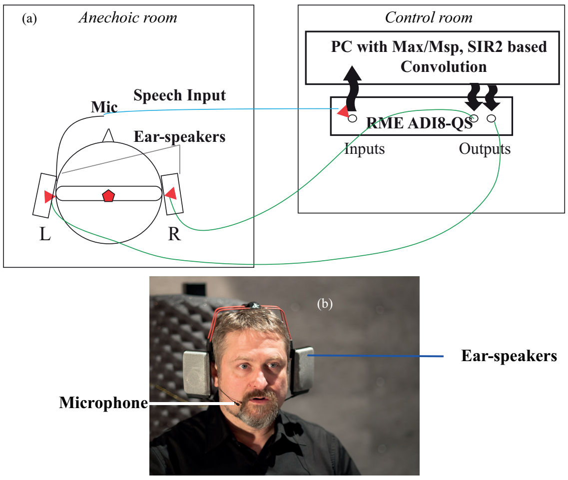


Figure 3 (a) provides an overview of the simulation system and (b) shows a seated talking-listener doing the experiment in the anechoic room.

Participants, and Experimental setup

24 participants (17 male, 7 female) were recruited for the experiment on a volunteer basis. 15 of the participants were either students or academics in the field of audio and acoustics. For the experiment, the participants were seated on a wooden chair placed on a wooden board on the floor of an anechoic chamber. The wooden board was placed to provide the floor reflections.

The communication task involved the talking-listeners vocalizing the three corner vowels (Raphael *et al.*, 2007) /a/ (the “ah” sound in rather), /i/ (the “ee” sound in deep) and /u/ (the “oo” sound in boot) to a manikin placed 5 m in front of them. They were asked to address the manikin at a comfortable and steady level for 3 s. The reasoning behind this communication task was as follows:

- Vocalizing three vowels represented a parsimonious approach for a task that can be performed over the 32 room conditions without much vocal exertion.
- By averaging the contributions from these vowels, a first-order approximation of the contribution of body conducted sound could be derived (Reinfeldt *et al.*, 2010; Pelegrín-García *et al.*, 2011)

The vowel vocalizations were a part of an experiment where the participants also rated the auditory room size (Yadav *et al.*, 2013). Hence, the participants spent some time to familiarise themselves with the autophonic room response represented by each simulated room condition, prior to attempting the vowel vocalizations. The three vowels were recorded with the 7 cm microphone position for each room condition.

Vowel data extraction

The selection criteria for recordings to be included in further analysis involved two stages, as follows:

1. Vowel stage
- The recording per room was processed to extract the signals representing the three vowels.
- Each signal was analysed in 200 ms non-overlapping windows. If the standard deviation of the five consecutive 200 ms windows was less than 3 dB, only then this 1 s duration was used.

2. Participant stage
- The 1 s vowel signal level (in dB) was compared for the two control conditions for each participant. If the levels for the control conditions (room RCH, presented twice at two gain levels) were within 1 dB, only then the data for a participant was used.

Following these two stages, data for 11 (8 male, 3 female) participants was found to be reliable.

Statistical analysis

The software R (2008) was used for statistical analysis. For each participant, the level per vowel in each room was normalized ($\Delta L_{i,vwl}$ in dB for $i = 1, 2, \ldots,$ □30 and vwl = /a/,/i/,/u/) with respect to the lowest level ($Lmin,_{i,vwl}$) amongst the rooms, as follows:

$$\Delta L_{i,vwl} = L_{i,vwl} - (Lmin_{i,vwl} + 3\text{ dB}) \qquad (2)$$

The 2 levels of gender, 3 levels of vowels, and 30 levels of rooms led to a 2 × 3 × 30 factorial design. Levene’s test for homogeneity of variance showed that the variance in ΔL was homogenous for all the factor levels and their interactions ($F(177,496) = 0.9$, $p = 0.7$). A three-way (2 × 3 × 30) independent ANOVA was conducted to assess the variation in $\Delta L_{i,vwl}$ for each independent factor and the results are presented in Table 1.

For further analysis, the vowel data was combined from all the participants (ΔL_Z), by averaging the results for the three vowels. ΔL_Z was analysed against two physical acoustical parameters: RT, G_{RG}; and one subjective parameter: auditory room size ratings for the simulated rooms obtained from the same experiment. The room size ratings are discussed in a separate study (Yadav *et al.*, 2013).

RESULTS AND DISCUSSION

The three-way ANOVA results in Table 1 show that most of the variance (76%) in ΔL values is explained by gender as the main effect ($F(1,496) = 195$, $p < 0.001$). The Bonferroni *post-hoc* test revealed that female vocalizations registered significantly more SPL than male. The two-way interaction between gender and vowel accounted for 5.6% of the variance ($F(2,496) = 14.4$, $p < 0.001$). Both the room condition ($F(29,496) = 1.6$, $p = 0.03$) and the two-way interaction between the gender and room condition ($F(1,496) = 1.6$, $p = 0.03$) accounted for 0.6% of the variance each.

Table 1. Three-way analysis of variance with main effects and interactions (indicated by an asterisk between the factors) applied to ΔL. Significance (at $p < 0.05$ level) is highlighted in bold face.

	F-value	*p-value*	% Explained variance
Main effects			
Gender	195	**$<10^{-16}$**	76
Vowel	1.7	0.19	—
Room	1.6	**0.03**	0.6
Two-way interactions			
Gender*Vowel	14.4	**$<10^{-6}$**	5.6
Gender*Room	1.6	**0.02**	0.6
Vowel*Room	0.7	0.9	—
Three-way interactions			
Gender*Vowel*Room	0.9	0.7	—

In Pelegrín-García et al. (2011), room conditions had accounted for most of the variance. However, the impulse responses used in that study did not represent real room conditions and spanned a much wider range of G_{RG} values, some even representative of electronic amplification.

Figure 4 shows the correlation matrix for the combined vowel SPLs (ΔL_Z), RT, G_{RG}, and subjective room size ratings. The Spearman correlation coefficients and the corresponding confidence intervals were derived using bootstrapping methods (Wilcox, 2011) with the function boot() in R.

Overall, ΔL_Z exhibited an increase with increasing RT (at a 90% confidence level), a decrease with increase in G_{RG} (at a 95% confidence level), and an increase with increase in the subjective auditory room size ratings (at a 70% confidence level). The overall trend in ΔL_Z with G_{RG} seen in this study is consistent with previous studies (Brunskog *et al.*, 2009; Pelegrín-García *et al.*, 2011). However, the same studies did not find any correlation between ΔL_Z with RT.

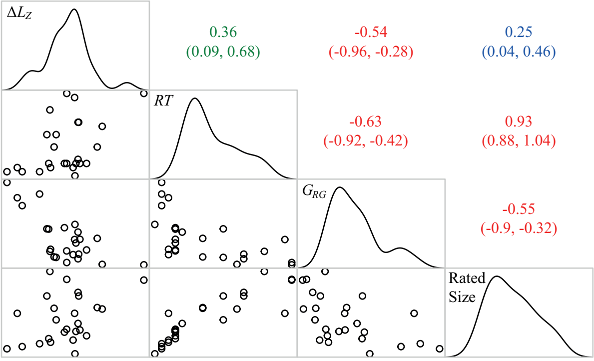


Figure 4. Correlation matrix for ΔL_Z, RT, G_{RG}, and subjective room size ratings. The diagonal panels show the density plot for each variable. Panels below the diagonal show the scatterplot of the corresponding variables on the diagonal. Panels above the diagonal show the Spearman correlation coefficients with the confidence interval for the corresponding variables on the diagonal. Red, blue and green represent the 95%, 90%, and 70% confidence intervals, respectively.

Past research in voice level changes in relation to room acoustics and conjectures from sidetone compensation studies would suggest that there would be a monotonic decrease in ΔL_Z with an increase in G_{RG}, when the communication scenario is held constant (Lane and Tranel, 1971; Pelegrín-García and Brunskog, 2012). By inspecting Figures 5 and 6 that chart the changes in ΔL_Z with RT and G_{RG} manipulations, respectively, it is clear that the current results do not completely agree with previous research.

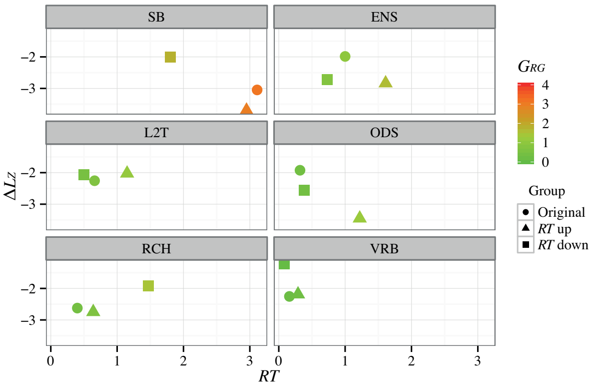


Figure 5. Change in ΔL_Z per room, with RT manipulation, which also changes G_{RG}. Original refers to the actual OBRIR.

Figure 5 shows the ΔL_Z changes with a change in both RT and G_{RG}. Apart from the room condition SB, which shows ΔL_Z changes consistent with past research, there is no clear trend in the rest of the rooms.

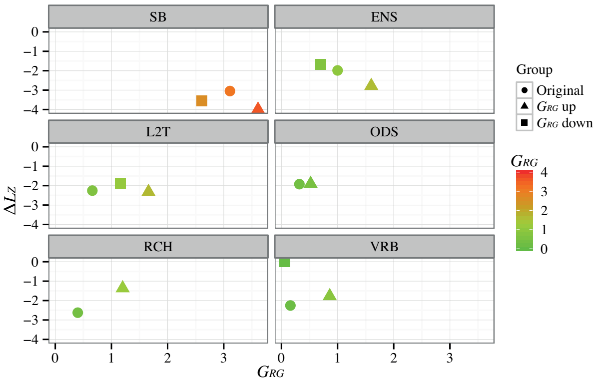
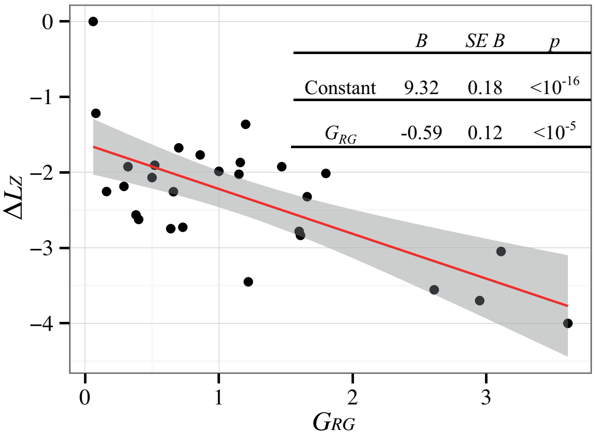


Figure 6. Change in ΔL_Z per room with G_{RG} manipulation. Rooms ODS and RCH have one data point missing each.

Figure 6 shows the ΔL_Z changes, when the G_{RG} values are manipulated. This manipulation is closer to the one reported in the previous study by Pelegrín-García *et al.*, (2011), where G_{RG} changes are applied to a single binaural impulse response over a large range of values (0.07 to 8.63 dB) for a different communication task. Again, the trends vary across the room conditions. One room condition (ENS, for G_{RG} changes of +0.6 dB and -0.3 dB from an original 1 dB) shows agreement with past research by showing a decrease in ΔL_Z value with an increase in G_{RG} and vice-versa. The room condition L2T is a medium sized lecture theatre, and has sound absorption treatment with an actual G_{RG} of 0.66 dB. Since this room represents relatively dry speaking conditions with its original OBRIR, the G_{RG} manipulations were chosen to only increase the G_{RG} by +0.5 dB and + 1 dB. With the 0.5 dB increment, ΔL_Z is increased, and then decreased slightly relative to the original with the 1 dB increment. This could be seen as an example of a room condition where appreciable G_{RG} changes may be required to reduce the ΔL_Z. Room conditions ODS and RCH had one G_{RG} manipulation removed each from the experimental data due to the G_{RG} values being too similar to the respective original values. In the room condition VRB, which represents a large recital hall with its original OBRIR, ΔL_Z values exhibited a rise with both the G_{RG} (and both RT) manipulations.



	B	$SE\ B$	p
Constant	9.32	0.18	$<10^{-16}$
G_{RG}	-0.59	0.12	$<10^{-5}$

Figure 7. Normalized voice levels (ΔL_Z) as a function of G_{RG}. The figure shows the linear regression line plotted with the 95% confidence interval in gray shade and the linear model in tabular form.

Due to the lack of a consistent trend in most of the rooms, and a relatively low percentage of variance explained in the ANOVA model by the various room conditions overall, fitting a linear model to the with G_{RG} as the independent variable was not expected to explain much of the variance in the ΔL_Z values. The model (R^2=0.44, F=22.65, $p<10^{-5}$) is presented in Figure 7.

This study did not measure any other sensory or proprioceptive mechanisms that are important in autophonic perception.

CONCLUSIONS

The current study points out that the changes in ΔL_Z (normalized voice SPL values) with changing autophonic room response may not follow a consistent trend that was exhibited in previous studies (Brunskog *et al*., 2009; Pelegrín-García *et al*., 2011; Pelegrín-García and Brunskog, 2012). Another finding, which differs from past research, is that RT in the current study was correlated with ΔL_Z values. RT, however, was not a strong predictor of ΔL_Z, when compared to G_{RG}. But the linear regression model with G_{RG} an the independent variable did not explain much of the variance in ΔL_Z, as noticed previously (Pelegrín-García and Brunskog, 2012).

Two obvious limitations of the current study were the somewhat rigid communication scenario and a limited range of G_{RG} values tested per room condition. However, similar communication task was used in a previous study where a consistent negative relationship between ΔL_Z values with G_{RG} was reported for simulated classroom conditions (Pelegrín-García and Brunskog, 2012). Hence, it could be argued that at least a similar trend should be noticeable within the current room conditions, regardless of the limited range of G_{RG} values.

Participants in the current study had reported that the autophonic room responses from the various room conditions were quite similar to the experience of listening while talking in actual rooms. Since the room conditions were chosen to represent a variety of actual speaking and singing environments, the results suggest that the ΔL_Z changes across rooms could vary even with similar communication scenario. In speaking and singing (where a very strict adherence to maintaining pitch and level are imposed) while also listening to the autophonic room response, the communication scenario and its relation to the room environment may in fact be more important than previously noticed (Lane and Tranel, 1971; Garnier et al., 2010) and needs to be studied further in appropriate contexts.

ACKNOWLEDGEMENTS

Manuj Yadav and Luis Miranda were supported by an Australian Postgraduate Award. MY was further supported by Jean and Andrew Wong Research Scholarship. LM was further supported by CONACYT (a Mexican government initiative) scholarship from the Becas al Extranjero program.

This research was supported under the Australian Research Council's *Discovery Projects* funding scheme (project DP120100484).

Travel to the conference for MY was supported by a grant from the NSW division of the Australian Acoustical Society.

REFERENCES

Black, JW 1950, The Effect of Room Characteristics upon Vocal Intensity and Rate. *J. Acoust. Soc. Am*. 22, 174.

Brumm, H & Zollinger, A 2011, "The evolution of the Lombard effect: 100 years of psychoacoustic research," *Behaviour* 148, 1173.

Brunskog, J, Gade, AC, Bellester, GP & Calbo, LR 2009, "Increase in voice level and speaker comfort in lecture rooms", *J. Acoust. Soc. Am*. 125, 2072–2082.

Cabrera, D, Lee, D, Yadav, M & Martens, WL 2011, "Decay Envelope Manipulation of Room Impulse Responses: Techniques for Auralization and Sonification", *Proceedings of ACOUSTICS 2011*, Gold Coast, Australia.

Cabrera, D, Sato, H, Martens & WL, Lee, D 2009, "Binaural measurement and simulation of the room acoustical response from a person's mouth to their ears", *Acoust. Aust*. 37, 98–103.

Gade, AC 1989, "Investigations of Musicians' Room Acoustic Conditions in Concert Halls. Part I: Methods and Laboratory Experiments", *Acustica* 69, 193–203.

Garnier, M, Henrich, N & Dubois, D 2010, "Influence of Sound Immersion and Communicative Interaction on the Lombard Effect", *J. Speech. Lang. Hear. Res*. 53, 588–608.

Hafke, H 2009, "Nonconscious control of voice intensity during vocalization", *Arch. Acoust*. 34, 407–414.

ISO 3382-1 Acoustics – measurement of room acoustics parameters – Part 1: Performance spaces, 2009.

Lane, H & Tranel, B 1971, "The Lombard sign and the role of hearing in speech," *J. Speech Lang. Hear. Res*. 14, 677.

Lane, HL, Catania, AC & Stevens, SS, 1961, "Voice Level: Autophonic Scale, Perceived Loudness, and Effects of Sidetone" *J. Acoust. Soc. Am*. 33, 160–167.

Lombard, E 1911, "Le signe de l'elevation de la voix", *Ann Mal. Oreille Larynx Nez Pharynx* 37, 25.

Miranda, L, Cabrera, D, Yadav, M, Sygulska, A & Martens, WL 2013, "Evaluation of stage acoustics preference for a singer using oral-binaural room impulse responses", *Proceedings of ICA 2013*, Montréal, Canada.

Pelegrín-García, D 2011, “Comment on “Increase in voice level and speaker comfort in lecture rooms” [J. Acoust. Soc. Am. 125, 2072–2082 (2009)] (L)”, *J. Acoust. Soc. Am.* 129, 1161.

Pelegrín-García, D, Fuentes-Mendizábal, O, Brunskog, J & Jeong, C-H, 2011, “Equal autophonic level curves under different room acoustics conditions”, *J. Acoust. Soc. Am.* 130, 228–238.

Pelegrín-García, D & Brunskog, J 2012, “Speakers’ comfort and voice level variation in classrooms: Laboratory research”, *J. Acoust. Soc. Am.* 132, 249.

Pörschmann, C 2001, “One’s Own Voice in Auditory Virtual Environments”, *Acta Acust. United Acust.* 87, 378–388.

Raphael, L, Borden, G & Harris, K 2007, *Speech science primer: Physiology, acoustics, and perception of speech*, 5th ed. Lippincott Williams & Wilkins, Baltimore and Philadelphia.

Reinfeldt, S, Östli, P, Håkansson, B & Stenfelt, S 2010, “Hearing one’s own voice during phoneme vocalization—Transmission by air and bone conduction”, *J. Acoust. Soc. Am.* 128, 751.

R Development Core Team 2008, *R: A language and environment for statistical computing*. Vienna Austria R Found. Stat. Comput. 1–1731.

Vorländer, M 2008, *Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, 1st ed. Springer-Verlag, Berlin.

Wilcox, RR 2011, *Modern statistics for the social and behavioral sciences: A practical introduction*. CRC Press.

Yadav, M, Cabrera, D & Martens, WL 2011, “Auditory room size perceived from a room acoustic simulation with autophonic stimuli”, *Acoust. Aust.* 39, 101–105.

Yadav, M, Cabrera, D, Martens & WL, 2012, “A system for simulating room acoustical environments for one’s own voice”, *Appl. Acoust.* 73, 409–414.

Yadav, M, Cabrera, D, Miranda, L, Martens, WL, Lee, D & Collins, R 2013, “Investigating auditory room size perception with autophonic stimuli”, *Proceedings of the 135th Convention of the Audio Engineering Society*, New York, NY, USA.

Zwicker, E & Fastl, H 1999, *Psychoacoustics: Facts and Models*, 2nd ed. Springer-Verlag, Berlin.