A Comparison of Spatial Listening in a Soundbooth versus an Immersive Virtual Environment

Jacob Maracle[1], Sin Tung Lau[1], Dario Coletta[1], Gurjit Singh[1, 2], M. Kathleen Pichora-Fuller[1, 2], and Jennifer Campos[2]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd N, Mississauga, Ontario, Canada, L5L 1C6
[2]Toronto Rehabilitation Institute, 550 University Ave., Toronto, Ontario, Canada, M5G 2A2

1. INTRODUCTION

The soundbooth is usually considered to be the ideal location for psychoacoustic studies. Listening studies that have involved complex or multi-talker listening tasks have also typically been conducted in the controlled acoustical space a soundbooth provides. Transposing a psychoacoustic study from its original controlled soundbooth environment to a more complex one has always been challenging. As the environment becomes more complex, testers need to be concerned about maintaining a level of control over extraneous variables. There is, however, great advantage in conducting listening studies under conditions that more closely approximate real world interactions. Such conditions would allow investigators to better understand whether the results of soundbooth studies can be generalized to real life.

1.1. Earlier Research

In earlier research, a multi-talker listening task was completed in a soundbooth. Participants were asked to identify words in a target sentence when it was presented simultaneously with two other sentences, each sentence originating from a separate spatial location (Singh, Pichora-Fuller & Schneider, 2008; see also Kidd, Arbogast, Mason & Gallun, 2005). Participants' performance was affected by the certainty with which the target was presented from a known location. When location certainty was high (i.e., 100% chance of the target being presented from a specified location), identification performance was higher than when the location was uncertain (i.e., 33% chance of it being presented from each of three possible spatial locations).

1.2. Current Experiments

The objective of our study was to see if the results from Singh et al. (2008) could be replicated in an immersive, virtual reality (VR) environment. The first part of this study was completed in the controlled environment of a soundbooth at the University of Toronto at Mississauga (UTM) campus. The second part of this study was identical to the first part, but took place in the Toronto Rehabilitation Institute's new state-of-the-art VR lab – "StreetLab".

2. METHODS

2.1. Participants

The participants were three English speaking, healthy young adults with normal hearing (pure-tone air-conduction thresholds < 25 dB HL at 0.25, 0.5, 1, 2, 3, 4, and 8 kHz).

2.2. Stimuli

A target sentence and two competing sentences drawn from the sentences recorded by four males for the Coordinate Response Measure (Bolia et al., 1999) were presented concurrently to the listeners from three different locations (either to the left, right, or in front of the listener). The sentences had the format "Ready [callsign], go to [colour] [number] now", where there were eight different callsigns and numbers, and four different colours.

2.3. Procedures

All participants completed 8 sessions in each of two conditions. The two conditions differed in terms of the test environment (soundbooth or VR). Each session consisted of 4 blocks of 30 trials. There were four location certainty conditions (1.0, 0.8, 0.6, 0.33). Within a session, each of the four location certainty conditions was tested in random order, with one block used to test each certainty condition. At the start of each trial, two pieces of information were presented visually to the participant. The first was the callsign, which the listener used to identify the target sentence. The second was one of the four possible probability specifications, which indicated the likelihood of the target being presented from the left, centre, or right locations, respectively (0-100-0; 10-80-10; 20-60-20; 33-33-33). For example, "10-80-10" indicated that there was an 80% chance of the target being presented from in front of the listener and a 10% chance of the target being presented from either the left location or the right location. The listener's task was to report the colour and number from the sentence that contained the target callsign. In both the soundbooth and VR conditions, feedback regarding the percentage of correct trials was provided at the end of each block. Each sentence was presented at 60 dB A. Participants completed two practice blocks at the 1.0 and 0.6 location certainty conditions at the beginning of each testing day.

<u>Condition 1: Soundbooth</u>

Condition 1 was a replication of the real spatial separation condition in the study of Singh et al. (2008). Participants were seated in a soundbooth and told to face forward for the duration of testing. The three loudspeakers in the soundbooth were positioned at ±54° and 0° azimuth at a distance of 1.83m from the participant.

Condition 2: Virtual Reality Environment (StreetLab)

Condition 2 was conducted in the Toronto Rehabilitation Institute's StreetLab (See Fig. 1). A 240^0 horizontal curved projection screen surrounded participants. A high-resolution virtual representation of a major city road intersection was displayed on the screen. No traffic or pedestrians were included in the simulation. Sound was presented from seven loudspeakers embedded behind the projection screen in the horizontal plane at head height.

Figure 1: Interior of StreetLab

Auditory stimuli were presented at ±90° (to the right or left) or 0° azimuth (in front). The distance from the loud-speakers to the participant was 2.14 m. Participants remained standing in one position for the duration of the experiment and the visual scene was spatially static.

3. RESULTS

3.1. Soundbooth vs. StreetLab

Overall, similar results were observed in both the soundbooth and StreetLab conditions (see Fig. 2). In both test environments, performance was best when listeners were certain about the location of the target and performance declined as location certainty decreased. Furthermore, the largest difference in performance between the two test environment conditions was only 4.6 percentage points (for the 0.8 location certainty condition). Finally, the mean results from the current study are within 1 SD of those found in the previous experiment conducted in a soundbooth (Singh et al., 2008).

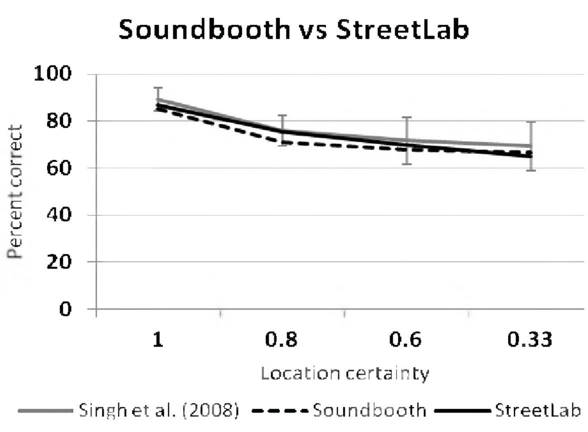

Figure 2. Percent Correct as a function of Location Certainty.

4. DISCUSSION AND CONCLUSIONS

The goal of this study was to determine whether the findings of the study of Singh et. al. (2008) could be replicated in a realistic, yet acoustically less controlled, VR laboratory environment (StreetLab). Investigators are now beginning to understand the importance of taking the knowledge that has been accumulated through traditional, highly controlled laboratory studies and moving towards evaluating the implications of these findings for real world interactions. Without understanding whether baseline performance measures in well-validated tasks under highly controlled conditions can be reproduced within highly realistic, ecologically valid conditions in simulated real world scenarios, differences in findings between the two in future experimental paradigms will be difficult to explain. In the current study, the results obtained in the soundbooth were replicated in an acoustically less-controlled space where additional non-auditory information (i.e., a complex visual scene) was presented. We were able to demonstrate that the findings are robust and likely translate to more complex situations. These results also suggest that future studies within this type of VR simulation may provide a valid testing environment for other psychoacoustic studies. Finally, this study provides a baseline for future studies in which we plan to test increasingly complex, real world challenges, such as multi-tasking activities requiring, for instance, listening and/or talking while walking with varying concurrent demands in terms of negotiating the sights and sounds of pedestrian and/or vehicle traffic.

REFERENCES

Bolia, R.S., Nelson, W.T., Ericson, M.A., Simpson, B.D. (1999). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America, 107,* 1065-1066.

Kidd, G., Arbogast, T.L., Mason, C.R., and Gallun, F.J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America, 118,* 3804-3815.

Pichora-Fuller, M.K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42,* 59-67.

Singh, G., Pichora-Fuller, M.K., and Schneider, B.A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *Journal of the Acoustical Society of America, 124,* 1294-1305.

ACKNOWLEDGEMENTS

This research was conducted by the first two authors for a 4th year undergraduate independent research course. The study was funded by an NSERC grant to author MKPF. StreetLab was partially funded by the Canadian Foundation for Innovation as part of the iDAPT (Intelligent Design for Adaptation, Participation and Technology) Centre for Rehabilitation Research; www.torontorehab.com/Research/Facilities/Labs/StreetLab.aspx). Thank you to Bruce Haycock for programming assistance.