LISTENERS' EXPECTATION OF ROOM ACOUSTICAL PARAMETERS BASED ON VISUAL CUES

By

Daniel L. Valente

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Architectural Sciences

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Jonas Braasch, Thesis Adviser

Curtis Bahn, Member

Ted Krueger, Member

Pauline Oliveros, Member

Ning Xiang, Member

Rensselaer Polytechnic Institute
Troy, New York

25 June 2009
(For Graduation August 2009)

ABSTRACT

Despite many studies investigating auditory spatial impressions in rooms, few have addressed the impact of simultaneous visual cues on localization and the perception of spaciousness. The current research presents an immersive audio-visual study, in which participants are instructed to make spatial congruency and quantity judgments in dynamic cross-modal environments. The results of these psychophysical tests suggest the importance of consilient audio-visual presentation to the legibility of an auditory scene.

Several studies have looked into audio-visual interaction in room perception in recent years, but these studies rely on static images, speech signals, or photographs alone to represent the visual scene. Building on these studies, the aim is to propose a testing method that uses monochromatic compositing (blue-screen technique) to position a studio recording of a musical performance in a number of virtual acoustical environments and ask subjects to assess these environments.

In the first experiment of the study, video footage was taken from five rooms varying in physical size from a small studio to a small performance hall. Participants were asked to perceptually align two distinct acoustical parameters – early-to-late reverberant energy ratio and reverberation time – of two solo musical performances in five contrasting visual environments according to their expectations of how the room should sound given its visual appearance.

In the second experiment in the study, video footage shot from four different listening positions within a general-purpose space was coupled with sounds derived from measured binaural impulse responses (IRs). The relationship between the presented image, sound, and virtual receiver position was examined. It was found that many visual cues caused different perceived events of the acoustic environment. This included the visual attributes of the space in which the performance was located as well as the visual attributes of the performer. The addressed visual makeup of the performer included: (1) an actual video of the performance, (2) a surrogate image of the performance, for example a loudspeaker's image reproducing the performance,

(3) no visual image of the performance (empty room), or (4) a multi-source visual stimulus (actual video of the performance coupled with two images of loudspeakers positioned to the left and right of the performer). For this experiment, perceived auditory events of sound were measured in terms of two subjective spatial metrics: Listener Envelopment (LEV) and Apparent Source Width (ASW) These metrics were hypothesized to be dependent on the visual imagery of the presented performance. Data was also collected by participants matching direct and reverberant sound levels for the presented audio-visual scenes.

In the final experiment, participants judged spatial expectations of an ensemble of musicians presented in the five physical spaces from Experiment 1. Supporting data was accumulated in two stages. First, participants were given an audio-visual matching test, in which they were instructed to align the auditory width of a performing ensemble to a varying set of audio and visual cues. In the second stage, a conjoint analysis design paradigm was explored to extrapolate the relative magnitude of explored audio-visual factors in affecting three assessed response criteria: Congruency (the perceived match-up of the auditory and visual cues in the assessed performance), ASW and LEV. Results show that both auditory and visual factors affect the collected responses, and that the two sensory modalities coincide in distinct interactions.

This study reveals participant resiliency in the presence of forced auditory-visual mismatch: Participants are able to adjust the acoustic component of the cross-modal environment in a statistically similar way despite randomized starting values for the monitored parameters. Subjective results of the experiments are presented along with objective measurements for verification.