Do we hear differently? Comparing spatial hearing between East-Asian and North-American listeners

Sungyoung Kim*

ECTET, CET, RIT (Rochester Institute of Technology),
ENT-2152, 78 Lomb Memorial Drive, Rochester, NY, 14623, USA

Abstract: This paper reviews previous experimental studies on the relationship between a listener’s cultural framework and auditory perception of an enclosed space. Cultural influence on auditory perception of noise and music has been assessed through a range of studies. Is it same for spatial hearing? When we enter to a space, would a particular cultural framework influence on understanding of the corresponding auditory environment? As physical buildings and enclosures reflect architectural and visual heritage, the auditory environment of an enclosed space also represents a unique and distinct heritage where people have interacted with and shaped their culture. When two listener groups (East-Asian and North-American) compared a reproduced field, previous findings show that (1) the semantic value of a same descriptor was distinctly different for two groups, and (2) there was an inverse relationship between the area of a personal space and size of a desired (preferred) auditory environment. With the advance of virtual reality (VR) technology, listeners can enter any auditory environment ubiquitously. Therefore, researchers and developers in the field should consider multiple user groups and the role of cultural framework in virtual environments.

Keywords: Cultural psychoacoustics, Auditory environment, Spatial impression, Subjective evaluation

PACS number: 43.66.+y [doi:10.1250/ast.41.129]

1. INTRODUCTION

A number of previous studies have examined how local cultural practices shape psychological processes, some showing that North Americans and Eastern Asians appear to significantly differ for certain cognitive processes. A study by Yokosawa et al. [1] showed that Japanese and U.S. subjects were different in their color preferences, influenced by “the affective nature of people’s interactions with and beliefs about the objects in the physical and social environments within their culture” [1]. Other studies [2–4] explained that differences in cognitive style of thought between North Americans and East Asians could result in differences in perceptual judgments. Nisbett and Miyamoto [4] showed that North Americans are characterized by an analytic cognitive process: “organizing objects by emphasizing rules and categories and to focus on salient objects independently from the context,” while East Asians uses a holistic cognitive process: attending to the entire “context and to the relationship between the objects and the context.”

A recent study by Saulton et al. [5] shows that spatial cognitive process between Germans and Sound Koreans are different; Koreans tend to look around the rooms more, attending to the entire context. Korean subjects showed significantly less bias when judging a room’s rectangularity (width-to-depth ratio) and viewpoint than their German counterparts.

Cultural influence on auditory perception of noise and music has been investigated through a wide range of studies [6–9]. In particular, the author investigated a cultural influence on perceived spectral balance of music reproduced over a pair of headphone [9], showing that the Japanese group differed from the other three listener groups (United States, Korea, and the Netherlands) with a relatively reduced gain in the mid-frequency band and increased gains in both low and high frequency bands.

An important research questions that has yet to be investigated thoroughly, however, is whether cultural framework has an influence on listeners’ understanding of an auditory environment or auditory perception of an enclosed space.

The acoustic environment is defined as “the combination of all the acoustic resources within a given area—natural sounds and human-caused sounds — as modified by the environment” [10]. An auditory environment, in contrast, refers to the listener-recognized space perceived through acoustical characteristics. How we recognize and

*e-mail: sxkiee@rit.edu
comprehend the auditory environment of a space may be influenced by the acoustic characteristics of the environments most common to our daily lives. For example, according to data on the average residential floor space per capita in 2009, a Japanese person occupies 43 m² while a Canadian takes up 72 m² and an American 77 m² [11]. North American listeners live in spaces almost twice as large as the average Japanese listeners. Therefore, it would be hard to expect identical cognitive understandings from two groups when they enter to a specific room. (e.g., how big is “big enough” for each of the two groups?) Cultural psychologist, Richard Shweder, claims [12] that mind and environment are interdependent.

Blesser and Salter [13] assert that a listener “selects specific aural attributes of a space based on what is desirable in a particular cultural framework to describe” a given auditory environment. Physical buildings and enclosures reflect architectural and visual heritage; the auditory environment of an enclosed space also represents a unique and distinct aural heritage, where people have interacted with and shaped their culture.

The author has investigated cultural influence on listeners’ perception of reproduced sound fields for the last 12 years, focusing on a specific research question: when listeners enter to an auditory environment rendered using modern spatial audio capture and reproduction techniques, how significantly do cultural backgrounds influence their experiences? This paper will introduce and summarize results from two case-studies [14,15] investigating cultural influence on listeners’ understanding of an auditory environment. A definite conclusion will require more supporting data both from behavioral and cognitive psychology: plans and caveats for future research will also be discussed.

2. EXPERIMENT 1: COMPARING JAPANESE AND ENGLISH ADJECTIVES DESCRIBING MULTICHANNEL-REPRODUCED MUSIC

In 2006, the author conducted a listening test to discover salient attributes in both English and Japanese associated with multichannel-reproduced solo piano music as a part of his doctoral research [16]. As part of his doctoral research, the author undertook a series of experimental investigations into how multichannel reproduction of solo piano music changes depending on the microphone technique used to capture the sound field. The question was examined using both physical and perceptual measurements. During these investigations, the author had a chance to work with a Japanese participant, which lead to the question: can a researcher assure that directly translated adjectives refer to common percepts for two listener groups who have different linguistic and cultural backgrounds? The author has personally experienced that direct translation of a word often causes a misunderstanding because “meaning and culture” are mutually intertwined. A social value associated with a specific idiom is dependent on a society’s cultural framework. Consequently, a subsequent experiment was conducted to compare two listener groups from Canada and Japan.

2.1. Methods

Four solo piano pieces were captured using the following four surround microphone arrays: Fukada Tree, Polyhymnia Pentagon, Optimized Cardioid Triangle combined with a Hamasaki Square, as well as a SoundField microphone (an 1st order Ambisonic microphone array) as shown in Fig. 1. Full details of the microphone arrays can be found in [17].

Stimuli were presented through five full-range loudspeakers arranged as per ITU-Recommendation BS.775-2 [18]. Two listener groups participated in the experiment: one group comprised of eight native English speakers (all males), and a second comprised of five native Japanese speakers (2 female and 3 male). All listeners were music students with majors in either performance or sound recording, aged from 20 to 31. The listeners were asked to elicit descriptive adjectives most associated with perceptual characteristics of four sound fields captured with four microphone techniques. A triadic comparison method was used [19], in which participants first compared the auditory imagery of three randomly selected stimuli and subsequently selected the most perceptually different stimulus. Each participant was then asked to describe how the selected sound field was different, and how the rest two were similar. By repeating this process, the method established a set of bipolar adjectives associated with salient perceptual characteristics of the stimuli.

2.2. Result 1

Collected adjectives were subsequently categorized on
the basis of the number of frequencies similar constructs were used by subjects within each group. Results from both groups are shown in Table 1 in descending order of frequency for each group (so that within a given row of the table, English terms in the first column do not correspond to the Japanese terms in the second column of the table).

The most frequently used bipolar adjective pairs for English listeners were: Wide ↔ Narrow, Distant ↔ Close, Focused ↔ Diffused, Sharp ↔ Dull, and Tight-bass ↔ Muddy-bass. The most frequently used bipolar adjective pairs for Japanese listeners were: はっきりした ↔ ふめいりょうな, あかるい ↔ くらしい, ひろがった ↔ まとまった, ふかい ↔ あさい, and はくりょくのある ↔ かるい, the direct translations of which are: Clear ↔ Obscure, Bright ↔ Dark, Wide ↔ Compact, Deep ↔ Shallow, and Powerful ↔ Powerless.

Table 1 shows that the relative frequency of elicitation of the terms was different for the North American and Japanese groups. In addition, the way in which Japanese subjects chose antonyms showed inter-group differences. While the North American group described “Narrow” as an antonym for “Wide,” the Japanese group used “まとまった” (Compact). The author asked a participant the reason and he explained that “Compact” was a more appropriate adjective to describe a relatively small volume of a space, while “Narrow” sounded negative. Similarly, the Japanese group frequently used “ふかい (Deep)” while the North American group used “Distant” more. The author believes that both groups tried to describe the depth of the spatial image of the instrument itself from the subject position but chose distinct descriptors. There is a chance that the adjective “Distant” may unconsciously reflect a negative value in it unconsciously among Japanese listeners, which is a topic for future investigation.

3. EXPERIMENT 2: INVESTIGATING CULTURAL INFLUENCE ON UNDERSTANDING OF AN IMMERSIVE SOUND FIELD

Another comparison between North American and Japanese listeners was made in the context of immersive three-dimensional (3D) sound field rendering, the results of which were published in Virtual Reality journal [15]. In this experiment, the author focused more on the global hedonic response of listeners and its relevance to perceptual attributes.

3.1. Methods

Two listener groups (North American and Japanese) were asked to listen to 3D music presentations with either concert hall or a church-like acoustics, and described the perceptual characteristics of the sound field in their native language. As with Experiment 1, a triadic comparison method was used. The auditory environments were rendered using a set of impulse responses (IRs) measured at two venues, convolved with anechoic recordings, and reproduced through nine loudspeakers.

Participants ranked a total of eight distinct loudspeaker configurations based on their perceived overall sound quality. To switch different configuration simultaneously, we installed total of twelve loudspeakers in the height layer (with the elevation angle of +30°). Among those twelve loudspeakers, eight configurations each of which uses four loudspeakers were used as the stimuli of the study.
Azimuth angles of four loudspeakers of the corresponding eight configurations are: (1) $\pm 30^\circ$ and $\pm 90^\circ$; (2) $\pm 30^\circ$ and $\pm 110^\circ$; (3) $\pm 30^\circ$ and $\pm 130^\circ$; (4) $\pm 50^\circ$ and $\pm 90^\circ$; (5) $\pm 50^\circ$ and $\pm 110^\circ$; (6) $\pm 50^\circ$ and $\pm 130^\circ$; (7) $\pm 70^\circ$ and $\pm 110^\circ$; and (8) $\pm 70^\circ$ and $\pm 130^\circ$. The configuration 3 (C3) is a shape of long rectangular; C8 is a wide rectangular; and C1 and C4 are similar to a trapezoid. Some listeners found it hard to define the sound quality and in such cases, the author asked them to judge how well one configuration appropriately integrated and immersed the music with the listener.

The North American listener group consisted of eleven listeners from McGill University and twelve students from Rochester Institute of Technology (RIT). The Japanese group consisted of fourteen listeners from Tokyo University of the Arts. Although the author did not conduct a formal audiometry test, none of the listeners had difficulties in either everyday listening or critical listening for music production.

3.2. Result 2

The overall sound quality perceived by two listener groups are illustrated in Fig. 2, showing distinct difference between North American (solid line with circle symbol) and Japanese group (dotted line with triangle symbol); two groups were different on their selection of configurations for an appropriate 3D auditory environment.

Was the difference because one listener group perceived idiosyncratically different characteristics from the same configuration due to the reproduction condition differences (such as room acoustics or loudspeakers used)? Or was it because of the group differences from their inherent (training and culturally associated) differences even though they perceived similar percepts for configurations? To better understand listeners’ perceptual foundations of Fig. 2 result (and to answer the previous question), the author analyzed the collected descriptors associated with each group’s high- and low-rank configurations.

The top six frequent descriptors elicited from the North American listener group included “frontal,” “wide,” “narrow,” “spacious,” “surround,” and “full” (together about 50% of the total number of descriptors). In contrast, “enveloping,” “clear,” “wide,” “reverberant,” “frontal,” and “narrow” were chosen by the Japanese group (57%). Those five representative descriptors were selected through a group discussion by the author and local collaborators.

Subsequent analysis showed that although the two groups provided similar descriptors, they used them differently. For example, the North American group highly favored an environment described as “frontal” and “narrow” (Configuration 1 and 4–C1 and C4–in Fig. 2) over configurations with perceptual characteristics of “wide,” “spacious,” and “surrounding” (C3 and C6). In contrast, Japanese listeners preferred “enveloping” and “wide” environments to those described as “frontal” and “narrow.” This revealed a culturally specific preference for characteristics of the auditory environment.

The result may suggest the hypothesis of an inverse relationship between the area of a personal space and size of a desired space in a reproduced auditory environment. In other words, Japanese listeners might prefer an auditory environment that is “wide” and “spacious,” while North American listeners do not need to expand the size of auditory circumstance but rather prefer a space with a full frontal image. This result may implicate that the North American group endorse the value of independence and the Japanese group interdependence even for understanding an auditory environment.

4. FUTURE WORKS

The author has compared a North American and Japanese listener groups to find a supporting evidence of culture-induced cognitive difference(s) for an auditory environment, discovering some socio-linguistic differences between two groups. However, when compared with visual studies of cultural psychology, current supporting data for auditory space cognition is still inadequate.

A new series of experiments, therefore, should be undertaken to (1) discover physiological and/or biological changes including electroencephalogram (EEG) signals and (2) investigate the independent (and analytic) nature of North American listeners, and the interdependent (and holistic) nature of Japanese listeners within the understanding of an auditory environment as they were for visual stimuli.

Auditory spatial understanding will be even more important for future virtual reality (VR) technologies that will require a degree of realism and enhanced user...
engagement through audio-visual integration. Auditory information is strongly associated with listeners’ emotions and moods [20] and as such will significantly influence listeners’ understanding of a VR environment. Future research should focus on the influence of auditory information to our understanding of virtual environments.

Meanwhile, an advanced psychophysical model should be established to minimize culture-dependent differences to allow for commonality of experience, communication and collaboration in new VR technologies.

ACKNOWLEDGMENT

The author appreciates Dr. William L. Marten (University of Sydney), Dr. Toru Kamekawa and Dr. Atsushi Marui (Tokyo University of the Arts), Masahiro Ikeda (Yamaha Corporation), Ron Bakker (Yamaha Europe), and Prof. Richard King (McGill University) for their sincere contribution to the experiment designs and data analyses introduced in this paper. Part of this work was carried out under the Cooperative Research Project Program of the Research Institute of Electrical Communication (RIEC), Tohoku University.

REFERENCES

[1] K. Yokosawa, K. B. Schloss, M. Asano and S. E. Palmer, “Ecological effects in cross-cultural differences between U.S. and Japanese color preferences,” Cogn. Sci., 40, 1590–1616 (2016), https://doi.org/10.1111/cogs.12291.
[2] A. Krishna, R. Zhou and S. Zhang, “The effect of self-construal on spatial judgments,” J. Consum. Res., 35, 337–348 (2008), https://doi.org/10.1086/588686.
[3] M. E. W. Varnum, I. Grossmann, S. Kitayama and R. E. Nisbett, “The origin of cultural differences in cognition: Evidence for the social orientation hypothesis,” Curr. Dir. Psychol. Sci., 19, 9–13 (2010).
[4] R. E. Nisbett and Y. Miyamoto, “The influence of culture: Holistic versus analytic perception,” Trends Cogn. Sci., 9, 467–473. https://doi.org/10.1016/j.tics.2005.08.004.
[5] A. Saulton, H. H. Bülhoff, S. de la Rosa and T. J. Dodds, “Cultural differences in room size perception,” PLoS ONE, 12(4), p. e0176115 (2017). https://doi.org/10.1371/journal.pone.0176115.
[6] S. Iwamiya and M. Zhan, “A comparison between Japanese and Chinese adjectives which express auditory impressions,” J. Acoust. Soc. Jpn. (E), 18, 319–323 (1997). doi: 10.1250/aat.18.319.