euronoise

Comparison of Japanese and English language descriptions of piano performances captured using popular multichannel microphone arrays

William Martens[a], Sungyoung Kim[b] and Atsushi Marui[c]

[a]McGill University, Schulich School of Music, 555 Sherbrooke Street West, Montreal, QC, Canada H3A 1E3
[b]Yamaha, 203 Matsunokijima, 438-0192 Iwata, Japan
[c]Tokyo University of the Arts, 1-25-1 Senju, Adachi, 120-0034 Tokyo, Japan
wlm@music.mcgill.ca

In a cross-cultural comparison of musical sound evaluations, the way in which bipolar adjective pairs are used by native speakers of Japanese and English language was studied via a subjective rating task. These ratings were collected in response to eight solo piano performances that had been captured using four popular multichannel microphone arrays, reproduced via a standard 5-channel loudspeaker array, re-recorded binaurally, and finally presented via headphones. This allowed nearly identical stimuli to be presented to all listeners, without any modulation of the loudspeaker signals via listener head movements. Mean ratings for the 32 stimuli on associated Japanese and English bipolar adjective scales were compared for the following English terms: clear-muddy, dark-bright, and compact-wide. The obtained mean ratings showed a similar pattern of correlations within each language, and were correlated across language as well, suggesting that the selected terms were used in a similar manner by native speakers of Japanese and English language in the context of this study.

1 Introduction

Although it may seem like a straightforward task to translate descriptive terms from one language into a second language, such from English to Japanese, there is quite a bit of uncertainty involved in selecting a set of descriptive terms from one language that will be used in a similar fashion to supposedly associated terms in a second language. However, when the context is tightly constrained, as they are when listeners attempt to characterize the perceptual differences between reproduced musical performances, it is possible to determine empirically whether such terms will be used in similar ways or not. The work reported in this paper was designed to test whether bipolar adjective pairs generated by native speakers of Japanese and English language correspond equally well to the perceptual distinctions that could be made between a set of 32 stimuli about which a good deal of data was already available from previous work. The stimuli in question were binaural reproductions of a set of four solo piano performances that had been simultaneously captured using four different multichannel microphone techniques. These stimuli were presented via headphones (i.e., binaurally) in order to allow nearly identical stimuli to be presented to all listeners, without any modulation of the loudspeaker signals via listener head movements. Since data was already available for native speakers of English, the current study involved a collection of similar ratings from a relatively large number of native speakers of Japanese language (73 undergraduate students registered at the University of Aizu).

Eight native speakers of the English language had previously completed a series of descriptive analysis (DA) sessions on the stimuli employed in this study (see Martens & Kim [1]). These English speakers also made ratings on the stimuli during two experimental test sessions separated in time by 6 months using the bipolar adjective scales that resulted from those DA sessions, and the retest showed that 5 of the 8 could made highly consistent ratings on three of the bipolar adjective scales that described the perceptual differences within a set of 32 stimuli comprising solo piano performances captured using four different multichannel microphone techniques. Furthermore, the mean ratings of these 5 listeners could be predicted from physical measures made on the binaurally-recorded stimuli presented in that study. These data provided a basis of comparison for the larger number of Japanese listeners who participated in the study reported herein.

2 Methods

In the study to be described in this paper, the experimental variable that was under direct control was the multichannel microphone technique that was used to record a selection of solo piano performances. Another important factor here was the selection of musical program material to be used in evaluating the results of using the microphone techniques to be evaluated. Previous reports on this project have already given more in-depth introduction to these issues (e.g., Kim, et al. [2]), and only details relevant to this particular cross-cultural study will be provided here.

Two short excerpts of each of four solo piano pieces composed in the European concert musical tradition were chosen for this study: works by Bach, Schubert, Brahms, and a contemporary improvisation by Plaunt. It was hypothesized that some microphone techniques might be preferred for certain musical selections within the performance space, which was the 600-seat Pollack Concert Hall located at McGill University, and expert advisors agreed that the following four surround microphone arrays were appropriate selections for comparison: Fukada Tree, Polyhymnia Pentagon, Optimized Cardioid Triangle combined with a Hamasaki Square, and a SoundField microphone. All musical excerpts were performed in the same concert hall by a single musician, and played on a single piano. Each performance was recorded simultaneously using the four selected microphone arrays, so that no difference between performances could confound the differences between microphone techniques. The particular short excerpts of the recordings that were presented had been selected in order to cover a relatively wide range of values on the physical measurements that were found to predict variation in preferences (see [2, 3] for more details).

In a pilot study reported by Kim, et al., [4], a comparison was made between descriptive terms used by native speakers of Japanese and those used by native speakers of English in describing differences in the perception of multichannel reproductions of 4 solo piano performances. These two groups of listeners were presented with versions of these performances captured using four different multichannel microphone techniques. After an examination of all the elicited adjective pairings, three bipolar anchoring adjective pairs were selected by the experimenters both from their relative frequency of occurrence in descriptive analysis sessions, and also through informal discussions between bilingual speakers of Japanese and English (confirmed authoritatively by co-author Dr. Atsushi Marui

of the Tokyo National University of Fine Arts and Music). The contextually-informed translations of these English terms into Japanese (written here using italicized roman characters) are as follows:

1 (*hakkiri shita*) clear ⇔ muddy (*fumeiryou na*)

2 (*kurai*) dark ⇔ bright (*akarui*)

3 (*matomatta*) compact ⇔ wide (*hirogatta*)

On three separate days, a group of 73 undergraduate students registered at the University of Aizu (average age of 21) participated in a short listening session at the beginning of a class period in a listening lab within which students were provided with matched pairs of Sony HS-90 headphones. In a single session, all 73 students were binaurally presented with the 32 piano performance excerpts, all with the same random order. The stimulus order was different on each day, and the listeners were asked to rate all 32 stimuli on only one of the three bipolar adjective scales during a single session. For each stimulus, one of five responses could be recorded using a paper form (PARSCORE) that was later automatically read and scored, preserving the students anonymity. The participants were given a short introduction describing the task each day, and viewed a projected slide show throughout the session indicating which adjective was associated with each extreme of the rating scale, the five levels of which were labelled "A" through "E." All participants completed all trials and all sessions.

4 Results

To begin with, the ratings of the 73 native speakers of Japanese will be examined. The graphs shown in Figure 1 plot the mean ratings obtained for each of the 32 stimuli, with a common plotting symbol used for each of the 8 stimuli that were recorded using the same microphone technique (i.e., the same plotting symbol was used for each of the 8 short musical excerpts in order to reveal any tendency for clustering of mean ratings according to microphone technique). The individual ratings that contributed to the plotted means were scored on the 5-point bipolar adjective scales as follows: The extreme responses "A" and "E" were scored as a -2 and +2, respectively. The moderate responses "B" and "D" were scored as a -1 and +1, respectively. If the listener gave the "C" response, that was scored as a 0, indicating that the rated stimulus was midway between the two extremes on the bipolar adjective scale in question. Before calculating the mean response over the 73 listeners, however, these raw scores were separately standardized within each set of ratings from each listener, so that all would have the same central tendency and utilize roughly the same range of the 5-point scale.

The upper panel of Figure 1. shows a scatterplot of mean ratings for scales identified as "Width" and "Brightness" (though Japanese listeners were never provided with these English terms, and were making ratings anchored by the terms *hirogatta* and *akarui*). The lower panel of Figure 1. shows a scatterplot of mean ratings for scales identified as "Width" and "Muddiness" (the latter scale being anchored by the Japanese term *fumeiryou na*). The plotting symbols used are explained in the figure caption.

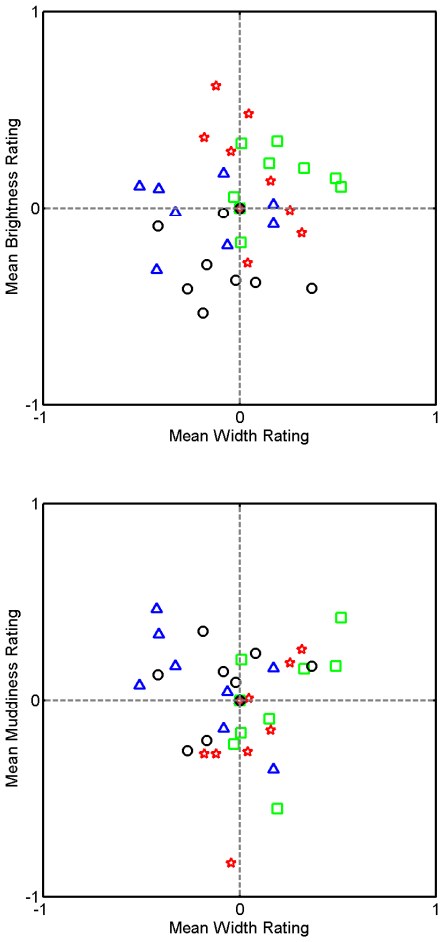

Figure 1. Upper panel: Scatterplot of Width vs. Brighness showing mean ratings on these two attributes made by 73 listeners for each of 32 stimuli. The plotting symbols used here made no distinction between listeners or the 8 musical programs to which they listened; rather the symbol shape codes only which microphone technique was employed for each rated stimulus: blue triangle for Fukada Tree, red pentacle for Polyhymnia Pentagon, green square for Optimized Cardioid Triangle with Hamasaki Square, and black circle for SoundField MKV. Lower panel: Scatterplot of Width vs. Muddiness, again showing mean ratings on these two attributes.

There was a fairly high negative correlation found between muddiness and brightness ratings (r = -.352), suggesting the simple observation that stimuli with darker tone coloration were also likely to be rated as muddier than stimuli with brighter tone coloration. The remaining sets of ratings did not show significant correlations: brightness and width ratings were unrelated (r = .125), as were the obtained muddiness and width ratings (r = .023).

With regard to the correlations between ratings on the three associated adjective scales by native English speakers, brightness and width ratings were also unrelated (r = .066), as were muddiness and width ratings (r = .171). Just as for the Japanese listeners, the only significant (and negative) correlation was found between the muddiness and brightness ratings (r = -.451).

The correlations between ratings made by native English speakers and those made by native Japanese speakers on the three putatively associated adjective scales were all found to be significant. Ranking these associations from highest to lowest correlation, the obtain r values were .432 for width ratings, .374 for brightness ratings, and .297 for muddiness ratings. Although such correlations provide no definitive evidence that the perceptual distinctions made by these terms are the same, the results do support the hypothesis that the way in which these adjective scales are used by native speakers of English and Japanese may be quite similar.

5 Discussion

What's in a name?
that which we call a rose
by any other name
would smell as sweet

From Shakespeare's "Romeo and Juliet," 1594

A fundamental assumption underlying many current methods used in sensory evaluation of reproduced sound is that listeners are able to analyze their complex auditory percepts in terms of separable attributes. While this assumption may not always be well supported by experimental data, it is in fact generally accepted. Even when physical measures on stimuli are available that correlate well with ratings on attributes purportedly associated with those measures, there is a continuing problem regarding the language that is used to describe the attributes that experimental research has identified. In particular, most textbooks on sensory evaluation techniques (e.g., [5]) distinguish between the way in which sensory attributes are perceived, and the terms which may be associated with them. The problem is amplified when a comparison is to be made between terms used by native speakers of different language, such as the English and Japanese language terms examined in the current study.

In a study similar to that reported here, an attempt was made to relate multilingual semantic scales to a common timbre space [6]. That study first showed that individuals within two groups differing in their native language seemed to share similar perceptual responses to a set of musical stimuli, as revealed by INDSCAL analysis of dissimilarity judgments (relatively unaffected by the listener's native language). However, when listeners from these groups made ratings on 13 bipolar adjective scales constructed in their natives languages (Japanese and Sinhala, a language of Sri Lanka), a principle components analysis showed that the that the ratings on the adjective scales from these two related differently than expected to the dimensions of their shared timbre space. Therefore, caution is advised in the application of the results obtained in the current study. This is despite the fact that no evidence was found here to reject the hypothesis that the group of Japanese speakers used differently the bipolar adjective scales constructed to correspond to the English pairs of terms clear-muddy, dark-bright, and compact-wide.

Acknowledgments

This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Canada Foundation for Innovation (CFI) New Opportunities Program, with additional support from McGill University's Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Special thanks are due to Dr. Stephen Lambacher who organized the listening sessions at the University of Aizu for this study.

References

[1] Martens, W. L., and Kim, S. "Verbal Elicitation and Scale Construction for Evaluating Perceptual Differences between Four Multichannel Microphone Techniques." In Proc. Audio Engineering Society 122nd Int. Conv., Vienna, Austria (2007)

[2] Kim, S., Martens, W. L., Marui, A. and Walker, K. "Predicting Listener Preferences for Surround Microphone Technique through Binaural Signal Analysis of Loudspeaker-Reproduced Piano Performances." In Proc. Audio Engineering Society 121st Int. Conv., San Francisco, CA (2006)

[3] Kim, S., & Martens, W. L. "Deriving Physical Predictors for Auditory Attribute Ratings Made in Response to Multichannel Music Reproductions." In Proc. Audio Engineering Society 123rd Int. Conv., New York, NY (2007)

[4] Kim, S., Walker, K. & Martens, W. L. "Cross-cultural Descriptive Analysis of Multichannel Auditory Imagery: A Comparison of Japanese and English Adjectives." In Proceedings of the 13th Regional Convention of the Audio Engineering Society, Tokyo, Japan (2007)

[5] Meilgaard, M ,. Civille, G. V., & Carr, B. T., "Sensory Evaluation Techniques (3rd ed.)" CRC Press, Boca Raton (1999)

[6] Martens, W. L., & Giragama, C. N. W., "Relating Multilingual Semantic Scales to a Common Timbre Space." In Proceedings of the 113th Convention of the Audio Engineering Society, Los Angeles, CA (2002)