Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Interaction of Auditory and Visual Information in the Gidayu-bushi Performer's Expression of Gender and Age

Satoshi Kawase

Graduate School of Human Sciences, Osaka University, Japan

kawase@hus.osaka-u.ac.jp

ABSTRACT

Gidayu-bushi is a traditional Japanese performing art accompanied by *shamisen* music. The performer plays several characters. This research focuses on communication in performing arts where one performer plays various characters. The aim of this study is to explore the interaction of auditory and visual information in communicating the Gidayu-bushi performer's expressions of gender and age. The object of the study was a performance in which two Gidayu-bushi performers played various age and gender roles and which featured 81 stimuli consisting of 27 sound-only modes, 27 vision-only modes, and 27 sound-and-vision modes. Participants were asked to determine the gender and age of the persons being played by the performer. First, the results of statistical analysis suggested that participants could correctly determine both gender and age in sound-and-vision mode. Second, it was shown that it was most difficult for the participants to identify gender when the performer transmitted expression of gender in vision-only mode. Participants were able to determine the age groups of the respective roles most clearly in sound-and-vision mode, second most clearly in sound-only and least clearly in vision-only. Finally, social skills affected determination by the participants. This study showed the importance of auditory cues rather than visual information in communication. It can also be inferred that on such occasions people use methods which differ from those applied in daily communication, where social skills are important.

INTRODUCTION

Communication is the elemental character of music. And, in the case of musical performance, communication between the performer and the audience is particularly important. As many studies show, performers can convey their intentions to the audience only through sound (e.g., Laukka & Gabrielsson, 2000; Kendall & Carterette, 1990; Senju & Ohgushi, 1987; Nakamura, 1987). The acoustical cues which relate to such conveyance of intentions are also analyzed. For example, expression of happiness is accompanied by a fast mean tempo, high sound level and bright timbre.

On the other hand, it is known that visual information plays important roles in live performance where the audience can see the performers. Sakuma & Ohgishi (1994) and Ogushi (2005) analyzed how presence or absence of visual information affected audience comprehension of the intentions of the musicians in performances by a pianist and by a percussionist, and suggested that visual information enabled the audience to understand the intentions of the performers. Visual information comprises various channels, including facial expression and gaze. Davidson (1993) recorded body movements of pianists, using a point light technique. The performers played in three different modes: 'Deadpan', 'Projected', 'Exaggerated' and the raters observed their behaviors. The results showed that the raters could distinguish the differences among these three modes via the movement of markers which were attached to the performers' joints. The researcher also analyzed the movements of each body part and suggested that head movement is particularly significant in providing cues for understanding the intention of the performance (Davidson, 1994). Davidson (2001) also coded the body movements of a solo vocalist on the basis of the Ekman & Friesen (1969) classification of non-verbal behavior, and analyzed the body movements used during singing. Moreover, Kurosawa & Davidson (2005) and Davidson (2005) coded the non-verbal behavior of a vocalist who is a member of the Irish pop band, *The Corrs*, and analyzed how frequently and at which points it occurred.

There have been many previous studies on emotional expression or magnitude of expression during music performance, in which emotional expression is pivotal. However, there are some genres of the performing arts, such as musical theater and opera, in which one performer may play more than one role. In such expressive forms, the performers have to accurately convey their respective roles to the audience while still projecting emotional expression. The major channels which enable audience members to understand the roles being played by the performers are costume and/or make-up; however, factors besides individual performances, such as sets and interaction between co-performers can also help the audience to understand the performance.

That said, if a performer is required to play two or more roles during a performance, he or she may not be able to express the different roles through the mediums of costume or make-up. Therefore, in addition to explanation ('a character talks this way', etc.), acoustical and/or visual expression becomes important, from one act to the next. In other words, the performer has to help the audience understand his or her roles by changing acoustical expression, facial expression or body movement.

Therefore, this research seeks to examine the interaction between visual information and acoustical information in a situation where a performer plays multiple roles. Focusing on a Gidayu-bushi performance, this study also explores social skills because they appear to affect encoding and decoding of non-verbal communication. Gidayu-bushi is a traditional Japanese performing arts accompanied by *shamisen* music. The performer narrates based on a concept called *on-zukai* and plays several different characters.

The aim of this study is to explore the interaction of auditory information and visual information in communication of a Gidayu-bushi performer's expressions of gender and age. It considered the following hypotheses:

1. Standard performance accurately conveys the expressions of gender and age to participants.

2. Visual information affects transmission of a performer's expressions of gender and age.

3. Audience members' social skills affect performer-to-audience communication.

METHOD

Participants

Eleven college students, three males and eight females, participated in the survey. The average age of participants was 21.5 years and the standard deviation was 1.29. Four participants were majoring in music, the other seven were not. Participants' experience of playing a musical instrument averaged 8.4 years (SD = 7.47). Only three of the participants actively engaged in appreciation of Japanese traditional arts; two of them about once a year and the other about once every five years.

Materials

As stimuli, the study used DVD footage of two Japanese living national treasures of Gidayu performing the same monologue, "kaede irodzuku yama no asa wa", in different ways, in an anechoic room. (*Japanese Voices: A Video Archive of Singing Styles and Techniques in the Japanese Language*, ISBN: 9784990336066). The performers were Takemoto Sumitayu and Toyotake Shimatayu. Three kinds of stimuli were prepared: sound-only mode, vision-only mode and sound-and-vision mode, in which they conveyed the same utterances in different roles. These files were processed using Adobe Premier (Adobe Systems, Inc.) to extract 19 patterns performed by Takemoto and 8 patterns performed by Toyotake. Therefore, 27 stimuli in each mode were used, for a total of 81 stimuli. This experiment was conducted during a college class. The video images were projected on a screen at the front of the room and sound was conveyed through fixed audio speakers.

Appendix 1 indicates the stimuli provided by each performer. The age groupings for each stimulus were divided into 6 categories: child, young, adult, mature, old, and unclear according to notes provided with the DVD. Some stimuli were not classified as to gender because they were not adequately specified in the notes (e.g., 'a child').

Procedure

The experiment was conducted in the form of a college class. Students were free to participate in the experiment. Participants rated the gender and age for each stimulus out of two choices: 'I think it is a male' and 'I think it is a female' using a 5-point scale ranging from 'not true at all' to 'very true'. They freely rated the age in figures.

Following the auditory experiment, the participants were also asked to check boxes indicating the music genres to which they frequently listen, whether or not they were familiar with the performers or the performing arts used as stimuli in this research, and to rate items for the purpose of measuring their social skills (ENDE2; Horige, 1994, see Appendix 2). Participants were given sufficient time to answer the questionnaires. The total experiment took about one hour.

RESULTS

Level of familiarity with stimuli

Table 1 indicates the kinds of music participants usually listened to. This result shows that many of the participants were familiar with both popular music and classical music but seldom engaged in appreciation of traditional Japanese performing arts such as the genre used in this research.

Table 1. Genres of music to which participants usually listen

Musical style	Number of people
Pop	8
Rock	8
Jazz	6
Classic	5
R&B	3
Trance	3
Healing	3
Techno	3
Fusion	1
Latin	1
Reggae	1
Tango	1
Hip Hop	1
Blues	1
Easy Listening	1
Sound tracks	1
Heavy Metal	1
Musicals	1

Differences between performers and differences between styles

In this research, two performers played various gender and age roles according to two different performance styles: *jidaimono* and *sewamono*. Gidayu-bushi repertory can be classified into two types: *sewamono* or domestic plays, and *jidaimono* or historical plays that deal with common people's lives and can include love stories. This study analysed the two performers and the differences between their performance styles.

First, the correlation coefficients of the performances in which two performers expressed gender and age characteristics in the same performance style (sewamono maiden, sewamono old woman, sewamono suave young man) were evaluated using the average rate for all modes. The results showed that gender was $r = .731 (p < .05)$ and age was $r = .484$ ($p = .186$).

Second, the correlation coefficients of the rating scale values of stimuli in which each performer played the same gender and age in different performance styles were evaluated in order to examine the differences between performance styles. The results showed that the rating scale value for gender was $r = .790 (p < .01)$ and that for age was $r = 707 (p < .01)$.

These results suggested that participants tended to respond consistently regardless of the performers or the performance style. For this reason, this study analyzed the results without discerning between performers and performance styles (*jidaimono* or *sewamono*).

Determination of gender

Figure 1 indicates the gender ratings. The rating scale value shows the absolute values for the differences between "I think it is male" and "I think it is female". The error bar indicates the standard error. A higher rating means that participants regarded the character as "more masculine" or "more feminine". The rating scale value for each gender was average, regardless of age.

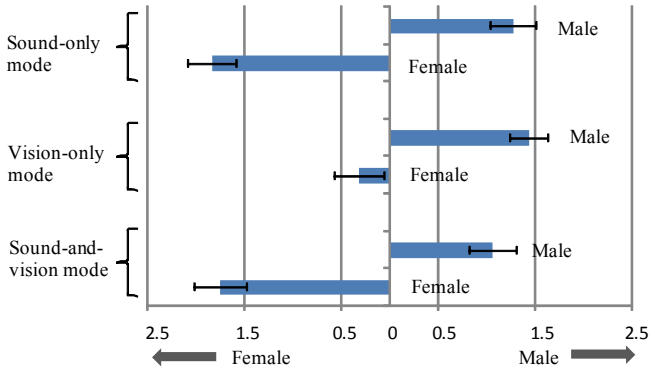

Figure 1. Expression of gender

In all modes, participants regarded those performances which expressed masculine characteristics as "male" and those which expressed feminine characteristics as "female". The most significant differences between male and female characterization occurred in sound-only mode. In other words, participants could clearly distinguish male from female through voice delivery. Sound-and-vision mode was second most effective in terms of conveying gender differences, while the vision-only mode was least effective.

Analysis of variance on two factors was conducted in order to examine mode and gender differences. The results showed significant differences in mode-gender interaction ($F(2, 20) = 12.525$, $p<.01$). This suggests that two factors, both mode and gender, affected rating scale values.

First, referring to each factor of simple main effect, there were significant differences regarding gender in all modes ($p<.01$). Therefore, participants could distinguish male from female under all conditions.

Second, in gender factors, there were no differences of stimuli which expressed "male" among the modes ($p<.05$, Bonferroni). The results show that participants regarded stimuli in which the performer played a masculine role as "male" regardless of the conditions.

The results also suggest that the rating scale values for the vision-only mode were significantly lower than those for the other two modes in the case of stimuli in which the performers played female roles ($p<.01$, Bonferroni). There were no significant differences between sound-only mode and sound-and-vision mode. Thus, it is more difficult to discern "female" in vision mode than it is in the other two modes.

Determination of age

Figure 2 indicates the results for age rating. The horizontal axis represents age classifications for the characters played by the performer, while the vertical axis shows the average rating by of participants. The error bar indicates the standard error.

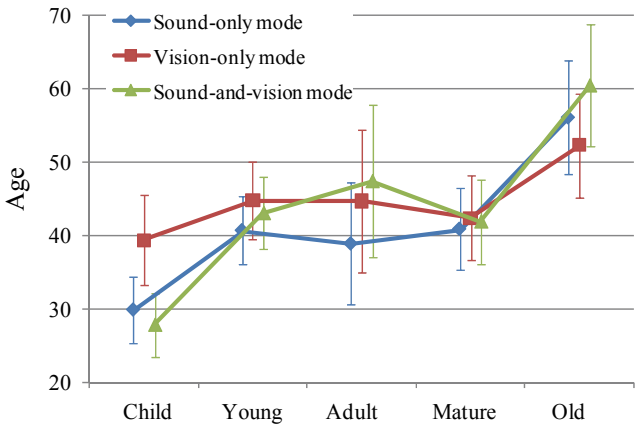

Figure 2. Expression of age

Child was rated "youngest" and Old was rated "oldest" in all modes. From this, it can be inferred that participants almost correctly rated the ages of the respective characters portrayed by the performers.

Analysis of variance was conducted on two factors in order to examine the mode and age differences. The results show the significant difference in mode-age interaction ($F(8, 652) = 2.362$, $p<.05$). These results suggest that both factors, mode and age, affected rating scale values.

First, there was a difference of modes in Child and Old. The vision-only mode rated more significantly than the other two modes in Child ($p<.05$, Bonferroni). There were no significant differences between sound-only mode and sound-and-vision mode. The sound-and-vision mode rated more significantly than the vision-only mode in Old ($p<.05$).

Second, Table 2 shows the results for each mode. The figures in the table represent differences of average rate in each

Table 2. Differences in rating scale value in each mode. Horizontal axis numbers subtracted from those on the vertical axis.

Age	Sound-only mode				Vision-only mode				Sound-and-vision mode			
	C	Y	A	M	C	Y	A	M	C	Y	A	M
1. Child	-				-				-			
2. Young	10.9**	-			5.3	-			15.5**	-		
3. Adult	9.2	-1.7	-		5.4	-5.3	-		19.6**	4.2	-	
4. Mature	11.3**	0.4	2.1	-	3.4	-1.9	-2.0	-	14.2**	-1.3	-5.4	-
5. Old	26.1**	15.2**	16.9**	14.8**	13.0*	7.6	7.6	9.6*	32.5**	17.0**	12.9*	18.3**

**$p<.01$, *$p<.05$

mode. These results show that there were more significant differences between Child and Old in sound-and-vision mode than in all other ratings. At the same time, there were significant differences among rating scale values other than between Child and Adult in sound-only mode. Moreover, there were significant differences only between Child and Old, and Old and Mature in vision-only mode. There were no significant differences among Young, Adult and Mature in all modes.

These results revealed that participants were most likely to understand the performers' gender expressions of age in both sound-and-vision mode and sound-only mode. Meanwhile, participants were much less likely to accurately understand the performers' gender expressions of age in vision-only mode.

Influence of social skills on determination of gender

The participants were divided into two groups: those with high decoding skills (five participants, hereinafter called the "high group") and those with lower social skills (six participants, hereinafter called the "low group") based on the results of examination of participants' social skills evaluated by ENDE2. Then the rating scale values of these two groups' determination of gender in each mode were evaluated (Figure 3).

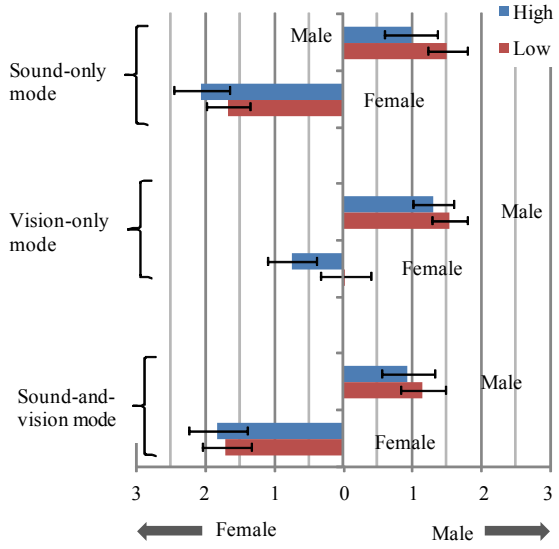

Figure 3. Decoding of expression of gender relative to social skills

The high group rated gender as "more female" compared to the low group in all modes. That is, the low group rated it as "more male". There were significant differences among each mode and gender: male ($t = 2.036$, $df = 13$, $p<.1$) of sound-only mode and female ($t = 2.213$, $df = 10$, $p<.1$) of vision-only mode.

These results imply that, to some degree, the participants with low decoding skills were influenced by the performer's gender (male), compared to the participants with high decoding skills. This suggests that the high group were less influenced by the performer's gender than the low group and were better able to understand the roles. This is supported by the fact that it was difficult for the low group to recognize a character as female in the vision-only mode, whereas the high group was able to do this to some degree.

From the above results, it is suggested that the decoding skills helped participants to correctly understand the characters' genders.

Influence of social skills on determination of age

Figure 4 indicates both groups' rating scale values for each age expressed by the performer.

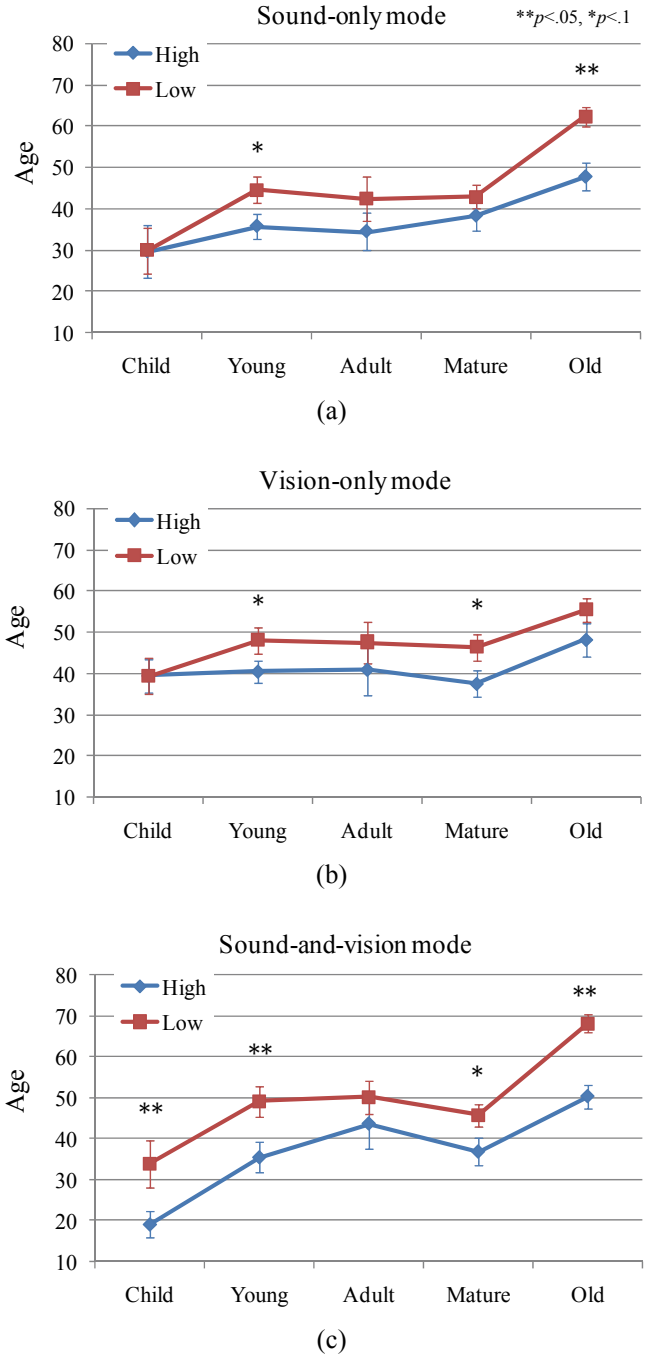

Figure 4. Decoding of expression of age relative to social skills. (a) (b) (c) indicate the results for each mode. Asterisks indicate significant differences between groups. (*t*-test)

The low group tended to rate age as older than the high group in all modes. The low group had a coherent tendency regardless of modes. (Old was the highest, followed by Young, Adult, Mature and Child).

On the other hand, the high group could distinguish age in the sound-only mode, but across a narrower range than that of the low group. There were no differences among determinations other than Old in the vision-only mode, while the high group could distinguish the ages of characters in the sound-and-vision mode.

In determination of Child, there were no differences attributable to decoding skills between the sound-only mode and the vision-only mode. However, the high group rated age as younger in sound-and-vision mode while the low group rated it as older in the sound-and-vision mode than in the sound-only mode. Thus, it can be inferred that the high group rated ages as close to the characters' ages in Child because both sound and vision were provided at the same time, whereas the sound-and-vision presentation did not provide a synergy effect among the low group. The same phenomenon occurred in the low group in Old. The low group's rating scale values for Old were highest in the sound-and-vision mode, followed by sound-only mode and then vision-only mode, although the change was not so great as that for Child among the high group.

DISCUSSION

First, the results of this research show that age and gender expressed by performers were accurately conveyed to participants under normal circumstances, i.e., under conditions in which both visual and acoustical information was provided. Therefore, it is suggested that audience members are able to construe character roles from voice and visual information even if they are not given context or explanation during performance. However, because the adult, existing between Child and Old was not clearly discerned, it would appear that participants distinguished the expression of age by means of other factors (e.g., speech content, manner of speaking or gestures) which this research did not analyze.

Second, the results also show that acoustical information is superior to visual information with regard to participants' understanding of expression because it was most difficult for participants to determine gender and age from only visual information whereas they came closer to accurately determining gender and age through only acoustical information.

In a normal Gidayu performance, the artists perform in formal clothing, which implies that acoustical cues are very important. The results of this research reflect this characteristic of Gidayu. Previous studies on musical communication have revealed the following facts: First, music (i.e., acoustical information) is able to convey expression of emotion quite accurately to an audience. Second, when the performer is expressing magnitude of performance, body movement included in visual information, is able to convey the intention of the performance without the assistance of acoustical information (Davidson, 1993). Third, visual information, including facial expressions, can convey emotion (Ohgishi, 2005). However, the results of this study suggest that visual information is the "least useful" channel for transmission of gender and age cues. From these results, it can be inferred that both suitable and unsuitable types of information are included in visual information (body movements or facial expressions) and acoustical information used for communication with audiences.

For example, acoustical information is suitable for conveyance of gender and age cues, while body movement included in visual information is suitable for expression of magnitude of performance.

This is indicated in Figure 5. The intention of a performer as an encoder (I_e) includes such factors as emotion (E_e), magnitude (M_e), age (A_e) and other elements (X_{ne}). Each factor affected subcategories or channels of communication included in visual channels and sound channels. There are weighting differences between each factor of intention and each channel.

For instance, emotion transmitted by the performer may affect the body movement channel ($\omega_{e,\,E,\,BM}$) less than facial expression ($\omega_{e,\,E,\,FE}$). The audience decodes factors of intention from each channel's information. Finally the audience

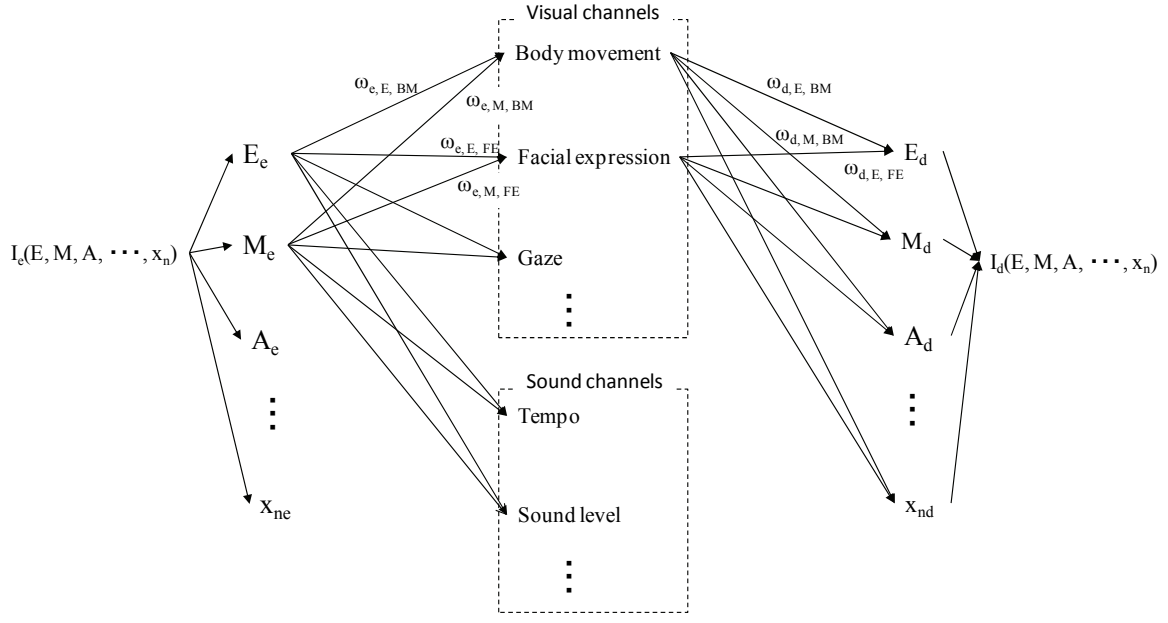

Figure 5. Communication process from performer to audience. All intervals between factors and channels are connected to each other. However, in this figure, some connections are omitted for simplicity.

members, as decoders, construe the performers' intention (I_d). Thus, the difference between I_e and I_d becomes the index of accuracy of communication. The characteristics of channels, individual musical ability and personality factors such as the encoder's social skills and those of the decoder determine the weight matrix for both encoder and decoder.

However, further analysis is needed because this model has not yet shown the hierarchical structure as precisely as did the extended lens model of Juslin and Scherer (2005).

Finally, social skills appear to influence the way in which an audience decodes the performer's roles. This means that musical communication is a kind of non-verbal communication. There is a procedure whereby senders (performers or speakers) encode intention as information and receivers (audience members or listeners) decode it, in both musical communication and non-verbal communication. Viewed in this light, it can be interpreted that audience members' decoding skills affect their understanding of the performers' intention in musical communication.

However, although the high decoding skills group was able to more accurately discern expression of gender, coherent results for age expression could not be obtained. Therefore, further research is needed.

CONCLUSIONS

The results of this study indicate the superiority of acoustical information in expression of gender and age. However, previous studies suggest that visual information is also an effective channel for expressing emotion and magnitude of performance (Davidson, 1993; Ohgushi, 2005). Therefore, it can be inferred that visual information and acoustical information include information which is suitable for conveyance. In other words, important channels differ in terms of information type. For example, when the male performer expresses 'happy female', it appears that the conveyance of 'happy' is influenced by facial expression, while communication of 'female' is influenced by acoustical information. That said, further research using minimized influence by certain channels, such as clothing and body movement, should be undertaken.

Ultimately, however, the degree of conveyance of the performer's intention depends greatly on decoding skills, which are part of the social skills of audience members. Accurate comprehension of information from each channel may affect the appreciation or understanding of music. (This ability also applies to everyday communication.) For these reasons, it is necessary to explore the relationship between various social skills and communication of music in a future study.

ACKNOWLEDGMENTS

I wish to thank Kei Eguchi for the excellent support and comments.

This presentation is funded by the Hayao Nakayama Foundation for Science & Technology and Culture.

REFERENCES

Davidson J. W. (1993). Visual perception of performance manner in the movement of solo musicians. *Psychology of Music*, 21, 103-113.

Davidson J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, 6, 279-301.

Davidson J. W. (2001). The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. *Musicae Scientiae*, 5(2), 235-256.

Davidson J. W. (2005). Bodily communication in musical performance. In D. Miell,. D.J. Hargreaves, & R. Macdonald (Eds.), Musical Communication (pp. 215-237). New York, MA: Oxford University Press.

Ekman, P. & Friesen V. W. (1969). The Repertoire of Nonvebal Behavior: Categories, Origins, Usage, and Coding, *SEMIOTICA*. 1(1), 49-98.

Horige, K. (1994). The refinement of heterosocial skills through the development and deterioration in close interpersonal relationships. The Japanese Journal of Experimental Social Psychology, 34(2), 116-128. (in Japanese)

Juslin, P. N. and Scherer, K. R. (2005). Vocal expression of affect. In Harrigan, J. A., Rosenthal, R. and Scherer, K. R. (Eds.), The new handbook of methods in nonverbal behavior research (pp.65-135). New York, MA: Oxford University Press.

Laukka, P. and Gabrielsson, A. (2000). Emotional Expression in Drumming Performance, *Psychology of Music*, 28, 181-189.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception and Psychophysics*, 41(6), 523-533.

Kendall, R.A. and Carterette, E. C. (1990). The communication of musical expression. *Music Perception*, 8(2), 129-163.

Kurosawa, K and Davidson, J. W. (2005). Nonverbal behaviours in popular music performance: A case study of The Corrs. *Musicae Scientiae*, 9(1), 111-136.

Ohgushi, K. (2005). Interaction between auditory and visual information in conveyance of players' intentions. *Journal of Acoustical Society of Japan*, 61(5), 284-288. (in Japanese)

Sakuma, M. and Ohgushi, K. (1994). Conveyance of the player's ideas for performance of percussion – Interaction between auditory and visual processing-. *Journal of Acoustical Society of Japan*, 50(8), 613-622. (in Japanese)

Senju, M. and Ohgushi, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception*, 4(4), 311-323.

Appendix 1

List of stimuli used in this study.

Name of performer_Style of performance_Role	
Takemoto_jidaimono_nobleman	Toyotake_sewamono_old woman
Takemoto_sewamono_villain	Toyotake_jidaimono_child
Takemoto_jidaimono_mature man	Toyotake_sewamono_housewife
Takemoto_jidaimono_maiden	Toyotake_sewamono_suave young man
Takemoto_sewamono_average woman	Toyotake_jidaimono_male warrior
Takemoto_jidaimono_old man	Toyotake_sewamono_elderly woman
Takemoto_sewamono_old man	Toyotake_sewamono_maiden
Takemoto_jidaimono_commoner's child	Toyotake_sewamono_fat man
Takemoto_sewamono_little boy	
Takemoto_jidaimono_elderly woman	
Takemoto_sewamono_mature man	
Takemoto_jidaimono_young nobleman	
Takemoto_sewamono_maiden	
Takemoto_jidaimono_villain	
Takemoto_jidaimono_young man	
Takemoto_sewamono_young man	
Takemoto_jidaimono_old woman	
Takemoto_sewamono_little girl	
Takemoto_sewamono_old woman	

Appendix 2

Rating items for decoding skills in ENDE2. ENDE2 consists of 15 items. This research used the scores for the 5 items below

	Not true at all				Very true
Q. I can read another person's feelings from his/her behavior	1	2	3	4	5
Q. I can sense a slight change in another person's emotions during conversation.	1	2	3	4	5
Q. I can sort of understand what another person is trying to say without the need for words.	1	2	3	4	5
Q. I can detect a lie.	1	2	3	4	5
Q. I can read what another person is thinking about me.	1	2	3	4	5