Interaction between auditory and visual information in conveyance of players’ intentions

Kengo Ohgushi

Professor Emeritus, Kyoto City University of Arts,
1–6–10 Oheyama, Goryo, Nishikyo-ku, Kyoto, 610–1102 Japan

Abstract: Music performers usually play music with some intentions. They play musical notes elongated or shortened and loud or soft for their expressive performances. Furthermore, they move their bodies consciously or unconsciously to enhance expression. Singers often show specific emotions on their faces during singing before audiences. Visual information from performers plays an important role for listeners to understand the performers’ intentions. Our studies in this field and a pioneering work by Davidson are reviewed.

Keywords: Expressive performance, Music cognition, Visual effect on music performance

PACS number: 43.75.Cd [doi:10.1250/ast.27.336]

This article contains the supplementary media files. Underlined file names in the article correspond to the supplementary files. For more information, see http://www.asj.gr.jp/2006/data/ast2706.html.

Supplementary files

1. Conveyance of expressiveness in piano performances
   s1v1.mpg, s1v2.mpg, s1v3.mpg, s2v1.mpg, s2v2.mpg, s2v3.mpg, s3v1.mpg, s3v2.mpg, s3v3.mpg

2. Conveyance of emotion in vocal performances
   sira_anger.mpg, sira_fear.mpg, sira_joy.mpg, sira_neutral.mpg, sira_sorrow.mpg,
   hat_anger.mpg, hat_fear.mpg, hat_joy.mpg, hat_neutral.mpg, hat_sorrow.mpg

1. INTRODUCTION

Performing musicians usually play music with some intentions. If music is performed exactly by a computer as described by a musical score, the performance is mechanical and gives an unnatural and strange impression to listeners. Performers often play musical notes elongated or shortened, louder or softer, and higher or lower compared to notes as written in the score for expressive performances. These tendencies were pointed out by Seashore [1].

Furthermore, performers usually move their bodies consciously or unconsciously to express their intentions. Singers often show specific emotions on their faces during singing before an audience. Visual information from the performer plays an important role for listeners to understand the performer’s intention.

Davidson [2] did pioneering work exploring how an observer’s impression of expressiveness (deadpan–projected–exaggerated) of recorded performances changed as the performances were presented in three modes: the visual stimuli alone, the sound stimuli alone, and the combined visual and sound stimuli. Four violinists wore tight-fitting black tracksuits with strips of reflective tape attached around the head, elbows, wrists, knees and ankles and on each hip and shoulder. In point-light conditions, they played a musical excerpt of Bach, Mozart, Mendelssohn or Kreisler in three performance manners: deadpan, projected, and exaggerated. These performances were recorded on video tape with sound. Twenty-one observers (music students) were presented the excerpts played by the four violinists in all three manners in sound-alone mode, vision-alone mode, and combined sound-and-vision mode. The observers rated each stimulus by marking the appropriate number on a seven-point scale (deadpan–exaggerated). The result revealed that all three modes showed differences between the three manners. However, the result suggested that vision most strongly conveyed the differences between the three performance manners. This means that vision is the most effective indicator of manner. She carried out a similar experiment using piano performances. In this case, the combined sound-and-vision mode most strongly conveyed the differences between the three manners. Moreover, for sound-alone mode, projected and exaggerated manners were rated similarly. These findings suggest that vision is more informative than sound in forming an observer’s understanding of the performer’s expressive intentions.
2. COMMUNICATION OF EXPRESSIVENESS THROUGH PIANO PERFORMANCE

We also did an experiment to investigate how visual and auditory information establish the observer’s understanding of the performer’s expressive intentions [3]. In the experiment, a female pianist was instructed to play a short musical excerpt with three different expressive levels. The excerpt was the beginning four bars of Nocturne op. 9, no. 2 by Chopin. The instructions for the three expressive levels were the following:

- expressive level 1 — performance without any expression (deadpan)
- expressive level 2 — performance with artistic expression (projected)
- expressive level 3 — performance with exaggerated expression (exaggerated)

First, the pianist played the excerpt with the three manners. The three performances were recorded on video tape with sound. To record combinations of different expressive levels of sound and vision, the pianist moved her body while listening to the sound reproduction of her own performances as if she were playing. Thus, six combinations of body movements and performance sounds were recorded on video tape. Finally, the nine performances (combinations of three expressive levels for performance sound and body movement) were prepared for stimuli.

The upper half of the pianist’s body was recorded but her wrists and fingers were not included in the pictures. An acoustical analysis revealed that tempo fluctuation increased with increasing expressive level.

Observers were 13 music students and 10 fine arts students. First, sound-alone stimuli and vision-alone stimuli were presented to the observers to check the reliability of expressiveness of the pianist. Observers were asked to rate the expressiveness on a five-point scale (1: deadpan–5: exaggerated) for both modes. The average ratings are shown in Fig. 1. This result shows that the performer’s intentions on expressiveness of sound production and body movement were satisfactorily realized. Next, the observers were presented the nine stimuli with the combined sound-and-vision mode and rated expressiveness on the five-point scale. The results are shown in Fig. 2. The ordinate gives the average rated value and the S-axis and V-axis correspond to the expressive level of sound and vision, respectively. The sound and the pictures are presented as video files. The names of the files of expressive level 1 with the three levels of body movements are s1v1.mpg, s1v2.mpg, s1v3.mpg, respectively. The names of the files of expressive level 2 with the three levels of body movements are s2v1.mpg, s2v2.mpg, s2v3.mpg, respectively. Expressive level 3 files are s3v1.mpg, s3v2.mpg, s3v3.mpg, respectively. The experimental results are summarized as follows: (1) Visual information influenced the ratings by both music and fine arts students significantly, (2) Auditory information had a significant influence on the ratings by music students, but did not have a significant influence on fine arts students, (3) Visual information was more influential than auditory information for both music and fine arts students.

In this experiment, “expressiveness of performance” was rated. However, there might be some individual...
differences in interpretation of “expressiveness of performance.” The difference between the results for music students and fine arts students might occur due to such individual differences in interpretation. Therefore, the same experiment was carried out again with different instructions, asking observers to pay attention to “sound expressiveness of performances” in the combined sound-and-vision mode [4]. Observers were 35 music students and 46 psychology students. For the sound-alone mode and the vision-alone mode, there were significant differences between different expressive levels for both the observer groups. In the case of the combined sound-and-vision mode, the experimental results are shown in Fig. 3. These results are summarized as follows: (1) Visual information presented with auditory information had much influence on the understanding of “sound expressiveness of performances” by both music and psychology students. (2) Music students were able to understand the intention of performers in most cases in spite of some contrary influences from vision, but psychology students were not able to understand performers’ intentions due to strong influences of vision. (3) Psychology students were more disturbed by contradictory visual information than music students were.

3. EMOTIONAL COMMUNICATION THROUGH VOCAL PERFORMANCE

It is very important for singers to communicate their intended expression to listeners, especially basic human emotions such as “joy,” “sorrow,” “anger,” and “fear.” Singers usually sing facing their audiences. Therefore, visual information such as facial expression is likely to be an important factor in communicating intended emotion to listeners. We carried out an experiment in which we investigated the influence of visual information on emotional communication between singers and listeners [5].

Three excellent female singers were selected from among the graduate students of the Department of Music, Kyoto City University of Arts. They were instructed to perform the “Vocalise” by Faure with different emotional expressions, such as “joy,” “sorrow,” “anger,” “fear” and “neutral.” These performances were recorded on video tape with a high-quality stereophonic sound track.

Performances of two singers are provided as supplementary media files. The performances of Ms. Shirakawa are named as sira_joy.mpg, sira_sorrow.mpg, sira_anger.mpg, sira_fear.mpg, sira_neutral.mpg. The performances of Ms. Hattori are hat_joy.mpg, hat_sorrow.mpg, hat_anger.mpg, hat_fear.mpg, hat_neutral.mpg.

Fifteen performances (3 singers × 5 performances) were presented to the observers (ten music students) through two loudspeakers (sound), through video display (video) and through loudspeakers and video displays. The observers rated each reproduction with regard to their impression as “joy,” “sorrow,” “anger” and “fear.” The judgments were made on a scale from 4 to 0, where 4 designated the maximum, and 0 the minimum, of the respective attribute. This result was analyzed by Kruskal’s Multidimensional Scaling technique (MDS). In the case of the sound-alone mode, “sorrow” and “fear” were mixed together and “joy” and “neutral” were mixed together in the two-dimensional psychological space. In the case of the vision-alone mode, all five emotions were separated. In the case of the combined sound-and-vision reproduction mode, however, the five emotions were more clearly separated from each other and “neutral” was placed at the center as shown in Fig. 4. This suggests that visual information was more important than auditory information in communicating emotions by singing.

4. PROFILES OF PERFORMERS

Pianist Yuko Naiki: When recorded in 1994, she was a graduate student majoring in piano performance at Kyoto City University of Arts. She is currently an active pianist in Yokohama, Japan.

Singer Miyuki Shirakawa: When recorded in 1995, she
was a graduate student majoring in vocal music at Kyoto City University of Arts. She is currently an active singer in Germany.

Singer Manami Hattori: When recorded in 1995, she was a graduate student majoring in vocal music at Kyoto City University of Arts. She is currently a singer in the USA.

ACKNOWLEDGMENTS

The author would like to thank Professor A. Nishimura for helping me in producing the media files. Thanks are also due to Professor W. M. Hartmann who read an earlier version and provided helpful comments on English expression.

REFERENCES