FIRST STEPS TOWARD DETERMINING THE ROLE OF VISUAL INFORMATION IN MUSIC COMMUNICATION

Rumi Hiraga, Nobuko Kato
Tsukuba University of Technology,
4-3-15 Amakubo, Tsukuba 305-8520, Japan
rhiraga@a.tsukuba-tech.ac.jp

ABSTRACT

We reviewed the resuts of our previous experiments on comparing the emotions recognized by recipients for several types of stimuli. Until recently, we had assumed that visual information provided simultaneously with musical performance was useful in helping hearing-impaired people to recognize the emotions conveyed by music. After our review, we became unsure whether multichannel information that supposedly helps deaf and hard-of-hearing people to recognize emotions actually works well. In this paper, on the basis of a new question that we pose about multichannel music information, we describe an experiment on comparing stimuli provided by types of music only, by music with video sequences, and by video sequences only. The results show that visual information has less effectiveness for hearing-impaired people, though visual information does have a role in supporting the recognition of emotions conveyed by music.

Keywords: music communication, hearing-impairment, multimedia, emotion.

1 INTRODUCTION

We previously believed that visual information played a role in supplementing sound information, especially for deaf and hard-of-hearing people. We also believed that this would apply to the case of deaf and hard-of-hearing people listening to music. These assumptions and our experiences with students in the Department of Industrial Technology at National University Corporation, Tsukuba University of Technology (NTUT), all of whom have hearing impairments, gave us the idea of building a music performance assistance system that uses visual information in a supplementary manner to enable people to communicate through music. After reexamining a series of experiments we previously performed to ascertain the possibility of recipients recognizing emotions after music intended to convey an emotion, we became unsure as to whether visual information can be an effective supplementary tool for enhancing musical performance appreciation. Thus, we conducted another experiment and came to conclude that visual information certainly does have a supplementary role in the recognition of emotion in music.

The background of building an assistance system goes back to a computer music class that we conducted for the abovementioned NTUT students for six years from 1997. All the students had been hearing impaired before starting elementary school and had quite limited experience of music, in terms of both playing and listening. Nevertheless, they enjoyed the class and the experience of playing new types of electronic instruments. In particular, they enjoyed playing music together to arouse mutual sympathy that gave them a certain satisfaction in the music activity.

In developing our system, we assumed that drums would be suitable instruments because they produce strong vibrations and some of these students had had the experience of playing large Japanese drums called wadaiko. Thus, we conducted several experiments on recognizing the emotions conveyed by drum performances and other stimuli that were possible musical accompaniment candidates. We used the four basic emotions—joy, fear, anger, and sadness—because they are often used in experiments on music cognition [9]. As we conducted our experiments, one concern we had was that the visual information we assumed would play a supplementary role might not support music information but would instead eliminate or replace it.

In this paper, we touch upon the background to our idea of building a musical performance assistance system, namely a computer music class for hearing-impaired college students, review previous experiments to see how effective visual information can be as a tool for recognizing emotion, explain the question we faced, and describe experiments we conducted to compare three types of stimuli (music only, music with video sequences, and video sequences only) as a means of recognizing emotion conveyed by several types of media.

2 RELATED WORK

Our work has been evolving from the idea of building a system to assist performances by deaf and hard-of-hearing people playing in ensembles. For this purpose, we needed to understand how the deaf and hard-of-hearing listen to music, especially how they recognize emotion in improvised drumming, because we had percussion ensemble music in mind for the system. We conducted experiments with deaf and hard-of-hearing people to understand the possibility of music communication with visual assistance. This research can be interpreted as an interdisciplinary approach involving musical activities by hearing-impaired people, aesthetics, multimedia, cognition, and computer systems. The studies mentioned below are related to our research from different viewpoints.

Apart from the many active music classes and music therapies for deaf and hard-of-hearing, there are many music activities done by deaf and hard-of-hearing: there are deaf professional musicians, e.g. Dame Glennie, a percussion soloist, Dr. Paul Whittaker OBE [18], an accomplished pianist, and the participants of Deaf Rave in London. In her research on music understanding by deaf and hard-of-hearing people, Darrow studied the referential meaning of pictures vis-à-vis music for hearing-impaired children [3]. That study used the musical performances that can be associated with concrete objects such as specific animals, while our study aims to pursue a more general association between audio and visual information, including abstract nuances such as the abovementioned four emotions. Ota investigated the musical abilities of young deaf and hard-of-hearing students from the viewpoint of special education [13].

The relationship between sound and visual information is an interesting research area and researchers from several fields have worked on it, e.g., Yamasaki as a psychologist [19] and Parke et al. as computer scientists [14]. They have analyzed how visual information supports the understanding of music (Yamasaki) and how music supports movies (Parke et al.). Levin and Lieberman [10], as media artists, tried to generate sound from pictures.

Emotion in musical research is a significant area of interest for both performers and listeners. Juslin introduced research methods for analyzing emotion in music [9]. Senju, a professional violinist, played the violin by herself conveying some emotions and investigated the possibility of emotional communication in music [16]. Schubert and Fabian analyzed performances of piano music into a Chernoff face [15]. Bresin and Friberg's system automatically generated music performances with emotions [2].

Our focus in the past was on determining whether there is a difference between people with and without hearing disabilities in recognizing emotion in several types of media (e.g., [5][6]). Our results demonstrated that there were no significant differences in most cases at the 5% level between deaf and hard-of-hearing and people with hearing abilities in recognizing emotion when the emotion was conveyed through musical performances, several types of visual media, or musical performances accompanied by visual information. The most difficult emotion to recognize was fear, regardless of the medium was.

3 COMPUTER MUSIC CLASS

During the period 1997–2002, we conducted a computer music class for hearing-impaired students at NTUT [1]. Most of the attending students were majoring in electronics. All of them were hearing impaired, though to different degrees. Their hearing difficulties were discovered before they began attending school, and their educational backgrounds were either special schools or general schools. In either case, their musical experience was generally more limited than others in their age group who had no hearing problems.

Some of those with lesser impairment enjoyed listening to music via MP3 players, but we were not sure whether they enjoyed listening to music in other cases, e.g., when playing Nintendo video games. In terms of sound, speaking and listening to a language is more relevant them than playing and listening to music. Thus, it was difficult for us to choose topics that would both maintain their interest and cater to their individual skills, needs, and desires. If we had organized the same class for students with hearing abilities, we could have taught them how to use desktop music software systems, such as those for sequence or music notation, or software systems for digital signal processing such as Max/MSP [1]. Since these topics were unsuitable for the students we actually had, we chose to give our students the chance to use new types of instruments in an attempt to let them experience the joy of music.

In the class we used the wearable instrument "Miburi" and the electronic drum set DD55 (developed and sold by Yamaha Corp). Both

[1] NTUT was then Tsukuba College of Technology, a three-year institute.

Figure 1: A student playing Miburi instrument.

instruments generate audio data along with MIDI (musical instruments digital interface) data. The image movement of the visualization software product Visisounder (developed and sold by NEC) is controlled by MIDI. We demonstrated these instruments and software to the students and showed them how to use them. Instead of using any published musical scores, we allowed the students to decide for themselves what they would play and how they would play it, including the chosen sound color. At the end of the semester they gave performances to show what they had learned. A photograph of a student playing the Miburi as if she were playing the wadaiko is shown in Figure 1. In one year, we suggested three rhythm patterns to students with which they could play batucada, a kind of samba music played with only percussion instruments. They found this interesting; in particular, playing together enabled them to gain an appreciation of music performance.

In the final year, we conducted our first experiment [7], which was an attempt to see how visual information could be useful in playing a simple rhythm. We offered four students three types of visual information that would guide a rhythm and patting sound only information. The results showed that none of the visual information was helpful to them in this respect; using the guiding sound only helped them improve their ability to follow the rhythm pattern. These results did not particularly surprise us because all four of the students had been accustomed to playing instruments before the start of the class.

4 MUSIC RECOGNITION

Since the last year of the class, we have conducted several experiments to investigate how deaf and hard-of-hearing students listen to music, focusing on the possibility of music communication with emotions. Since we believed that visual information can assist one in listening to music, we used several types of stimuli to understand the recognition of emotions that music can elicit. Besides single-medium stimuli of “music only” and “drawings only”, we used the following as multimedia stimuli.

- music and drawings,
- music and its performance scene,
- music and video sequences intended to convey no specific emotion,
- music and video sequences intended to convey the same emotion as the musical performances.

A probability of less than 5% was considered a significant difference in all the experiments. In these experiments, our interest was mainly on understanding whether there were more differences or similarities in recognizing emotions elicited from several types of stimuli between deaf and hard-of-hearing people and people with hearing abilities.

.

4.1 Single-medium stimuli

We started with experiments using two types of single-medium stimuli: “music only” and “drawings only”.

4.1.1 Music only

We used performances given by (1) NTUT students, (2) amateurs with hearing abilities, and (3) professional percussionists. The number of performances in each set and the number of subjects who listened to each set was varied. The numbers are given in Table 7 in the Appendix, with the number of performances given in parentheses.

The two subject groups (deaf and hard-of-hearing subjects and subjects with hearing abilities) showed no significant differences in recognizing emotion in music when they listened to performances played by hearing-impaired students and amateurs. On the other hand, there was a large difference when they listened to performances played by professional percussionists. Subjects with hearing abilities recognized emotion in this music significantly more than subjects with hearing impairments [8]. One possible reason is the difference in sound source. Professional percussionists played acoustic instruments, while other groups played the MIDI drum set and the performances were replayed from the MIDI data.

4.1.2 Drawings only

We then conducted an experiment where we used drawings only as stimuli. Experiments on emotional cognition from drawings have been conducted in the area of psychology [17]. We used three sets of drawings that were intended to

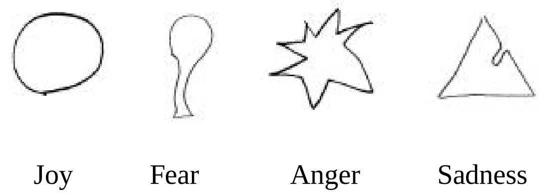


Figure 2: Drawing samples.

elicit emotions from the subjects, drawn respectively by students with hearing abilities who were design majors, hearing-impaired students who were electronics majors, and hearing-impaired students who were design majors. The number of subjects differed for each set of drawings. The numbers are given in Table 8, with the number of drawings given in parentheses.

Differences between subject groups were obtained when subjects looked at two of the three drawing sets. The exception was the set drawn by hearing-impaired students who were design majors [6].

4.2 Multimedia Stimuli

Below, we present our results for investigating whether there were significant differences in recognizing emotion in two types of stimuli: (1) music only and (2) music with one of the following types of visual information: drawings, performance scenes, video sequences intended to convey no emotion, and video sequences intended to convey the same emotion as the musical performance. We present results for both deaf and hard-of-hearing subjects and for subjects with hearing abilities. For all experiments with stimuli, whatever the type, we used drum improvisation performances.

4.2.1 Music and drawings

For the first multimedia stimulus type, we provided drawings intended to convey the same emotion as the musical performance during the first half of the performance. We used drawings based on our experiment on recognizing emotion conveyed by drawings (Section 4.1.2). They were provided along with the musical performances used in our previous experiments (Section 4.1.1). Some of the drawings we used are shown in Figure 2 [4]. Eleven hearing-impaired subjects (three males and eight females, aged 18–22) and 15 subjects with hearing abilities (13 males and two females, aged 20–24) participated in the experiment.

The results showed that there were significant differences between the stimulus types of music only

Figure 3: A professional drummer drumming.

and music accompanied by a drawing intended to convey the same emotion for both subject groups.

4.2.2 Music and video sequences

In attempting to provide a visual assistance tool, we expected a video sequence accompanying a musical performance to be helpful in enabling deaf and hard-of-hearing people to recognize emotion. We used the following three types of video sequences.

- Performance scene. Since it was difficult to gather subjects at a live performance, we used replays of musical performances from our past experiments. We thought that if showing videos of the performance scenes improved recognition rate, then we could assume that players in an ensemble performance would recognize emotion for better music communication.
- Video sequences not intended to convey any particular emotion. Since it seemed preferable for deaf and hard-of-hearing students to obtain added information from visual media, we thought that video sequences, even those not intended to convey any particular emotion, might be helpful in enabling them to recognize emotion in musical performance.
- Video sequences intended to convey the same emotion as the musical performance. We assumed that these sequences would improve the emotion recognition rate the most for deaf and hard-of-hearing subjects.

In experiments using these three stimulus types, we used different sets of musical performances and different subject groups.

4.2.3 Performance scene

We asked a professional percussionist to play a drum set in an improvisational style to convey emotions. He played four sets of performances (i.e., a total of 16 performances), using a bass drum, snare, set of concert toms consisting of five different sizes, Chinese gong, and suspended cymbal with drumsticks, mallets, brushes, and other items (Figure 3). Eleven deaf and hard-of-hearing subjects (eight males and three females, aged 19–24) and 10 subjects with hearing abilities (three males and seven females, aged 19–47) viewed a 19-inch display screen while listening to the musical performances from a speaker set [5].

The results showed that there were no significant differences between the two stimulus types of music only and music with its performance scene for both subject groups.

4.2.4 Video sequences intended to convey no emotion

We used the visual effects "Amoeba" and "Fountain" in Windows Media Player 10 [4]. These effects were controlled by sound data and the resulting animated scenes were not intended to convey any emotion by themselves. We chose Amoeba because its figure looked a little like some of the drawings we used in the drawings category. We chose Fountain because it looked quite different in shape and movement from Amoeba and used fewer colors than other effects. The number of deaf and hard-of-hearing subjects who watched Amoeba was the same as in the experiment using musical performance and drawings (Section 4.2.1), while eight subjects (two males and six females, aged 18–22) watched the Fountain scene. Fifteen subjects with hearing abilities (13 males and 2 females, aged 20–24) participated in the experiment.

In this experiment too, the results showed that there were no significant differences between the two stimulus types of music only and music with video sequences for both subject groups.

4.2.5 Video sequences intended to convey emotion

We wrote a software program called "Music with Motion Picture" (MPM) that generates a motion picture scene from sound data and a drawing by using Max/MSP and DIPS (digital image processing with sound) [12] on Mac OSX. This program provides a set of effects for modifying the shapes of drawings to make video sequences. Each effect is given a set of preset parameter values for the shaders of GLSL (OpenGL shading language) and those that Apple Core Image supports. The most important parameters for modifying image shapes are strength, depth, and speed. Each of these corresponds to and affects how much, how smoothly, and how quickly a shape varies. The set of preset effects provided by MPM is given in Table 1.

Table 1: Effects of MPM.

Effect names	Traits
Soft	Gentle movement; shallow depth
Wave	Wave-like movement; fine movement with more depth and less strength
Notch	More strength; sharper notch with greater speed and depth values
Collapsing	More strength and large speed values
Jell-o	Tremulous movement; less strength, more speed, and greater depth values
Bump	Bold shivering movement; using bump distortion
Torus	Sight through lens; using torus distortion
Blur	Smoothing movement, creating afterimages, and some other blurring
Disappearance	Presentation or deletion of objects with amplitude
Vortex	Shape modification obtained by calculating coordinates

Sixteen deaf and hard-of-hearing students (five males and 11 females, aged 20–21) and twelve subjects with hearing abilities (three males and nine females, aged 22–48) participated in the experiment. We used one of the four sets of performances used in the experiment described in Section 4.2.3. The set in which the professional percussionist played a snare drum was the one that elicited the lowest emotion recognition rate among the four sets. For each musical performance, we used MPM to create two video sequences comprising a musical performance and a seed drawing intended to convey the same emotion as the performance. Thus, we used a total of eight video sequences accompanying the musical performances. The seed drawing, its effect on generating a motion picture scene, and a sample frame from each scene are given in Table 3. Although we provided two media with the same intended emotion, some subjects with hearing abilities pointed out a mismatch between the two media. No deaf and hard-of-hearing subjects mentioned this.

The results we obtained in this experiment were unexpected. The recognition rates for subjects with hearing abilities improved more with the addition of video sequences than those for the deaf and hard-of-hearing subjects did. Moreover, there were significant differences between the two types of stimuli (music only and music with video sequences) for subjects with hearing abilities, but none for deaf and hard-of-hearing subjects.

4.3 Comparison of results

The results of experiments on multimedia stimuli (Sections 4.2.1, 4.2.3–4.2.5) are summarized in Table 2. We conducted three experiments to obtain the results. In the first experiment we used "Drawings", "Amoeba", and "Fountain"[2] and in the other two we used "Performance Scene" and "Video Sequence with Emotion". Since the musical performance sets and subject groups were different, we thought that comparing the numbers between them would be a valid way to ascertain certain tendencies. Our findings in this respect are summarized below.

- Our results show that providing still-image drawings along with a musical performance can improve the emotion recognition rate. In spite of this, we feel that still images are inappropriate as a means of supplementing music because music is dynamic.
- Except for the "Performance Scene" experiment, we can see from the p-values in Table 2 that visual information is more likely to benefit subjects with hearing abilities than it is to benefit deaf and hard-of-hearing subjects.

The results of the experiment described in Section 4.2.5 and the abovementioned tendencies led us to question whether deaf and hard-of-hearing people could make use of visual information in listening to music. Namely, we wondered if music from multiple channels might not necessarily improve the ability of deaf and hard-of-hearing people to recognize the emotion conveyed by musical performances.

5 VISUAL INFORMATION

To determine whether deaf and hard-of-hearing people could make use of visual information, we conducted an experiment that compared subjects' ability to recognize emotion conveyed by music accompanied by visual information with their ability to recognize emotion from visual information only. This experiment extended the "video sequence only" stimulus to the experiment described in Section 4.2.5. Prior to the experiment, we thought that if any differences were observed between the recognition rates for "music only" and "music with video sequences", then we could conclude that video sequences could be a valuable tool for subjects in enabling them to recognizing emotions. In that case, though, if the recognition rates for music with video sequences and video sequences only were the same, then we could pose the question that subjects might not make use of sound information.

5.1 Subjects

Seven deaf and hard-of-hearing subjects (aged 20–21, 3 males, 4 females) participated in the experiment. Their auditory capacity was over 100 dB, except one (80 dB). They wore their hearing aids (turned on), except for one whose auditory capacity was 110 dB, though they did not know their own auditory capacities with the hearing aids. Since the number of stimuli was small, we asked them to participate in the experiment three times, once every week. The order in which the stimuli were given to them differed each time. In each case, they listened to four drum performances, hearing and watching stimuli of drum performances accompanied by video sequences, and watching video sequences without music. We used two-way analysis of variance (ANOVA) where the effects were sessions in the experiment and types of stimuli. The p-value for three experiment sessions was 0.40, where the respective mean values were 0.66, 0.61, and 0.55. Thus we used all data (seven subjects participating in the same experiment three times).

As a control group, ten subjects with hearing abilities (aged 21–50, 3 males, 7 females) participated in the experiment.

5.2 Material

Three types of stimuli were used in this experiment: (1) music only, (2) musical performances with video sequences where the emotions conveyed by both stimulus types coincided, and (3) video sequences only. They were the same as we used in the experiment described in Section 4.2.5—four musical performances and eight video sequences. The video sequences used in this experiment are summarized in Table 3.

5.3 Results

The recognition rates for all stimuli by subject groups are shown in Table 4. Drawings were shown to subjects who were different from those participating in this experiment. A total of 21 (three males, 18 females, aged 20–21) were asked to recognize emotions from drawings.

For both deaf and hard-of-hearing subjects and subjects with hearing abilities, the recognition rates for music paired with video sequences increased from that for only musical performance. The exception was the recognition of sadness (both S1 and S2) by deaf and hard-of-hearing and anger (A1) by subjects with hearing abilities.

[2] This means that the recognition rates for music for "Drawings", "Amoeba", and "Fountain" were the same.

Table 2: P-values obtained by comparing the recognition rates for music stimuli and for stimuli of music accompanied by visual information.
* denotes visual information (one out of Drawings, Amoeba, Fountain, Performance Scene, and Video sequences with emotion).

	Hearing-impaired subjects			Subjects with hearing abilities		
		Mean			Mean	
	p-value	Music only	Music with *	p-value	Music only	Music with *
Drawings	1.2e-004	0.46	0.66	9.0e-005	0.47	0.67
Amoeba	0.93	0.46	0.46	0.26	0.47	0.52
Fountain	1.00	0.46	0.45	0.53	0.47	0.49
Performance Scenes	0.11	0.54	0.65	0.83	0.74	0.73
Video sequences with emotion	0.18	0.55	0.68	0.01	0.50	0.77

Table 3: Seed drawings, effects in generating video sequences, and sample frames from scenes.

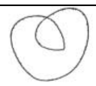
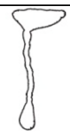
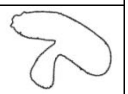
Emotions	Joy		Fear		Anger		Sadness	
Name	J1	J2	F1	F2	A1	A2	S1	S2
Drawings								
Effects	Jell-o	Bump	Wave	Jell-o	Collapse	Bump	Wave	Jell-o
Sample frames								

Table 4: Recognition rates for three types of stimuli.
DHH: deaf and hard-of-hearing subjects, HA: subjects with hearing abilities.
VS: video sequence, M&VS: music with video sequences, Music: music only.

Name		J1	J2	F1	F2	A1	A2	S1	S2
Drawings		1.00	1.00	0.95	0.76	0.38	0.43	0.43	0.43
DHH	VS	0.86	0.52	0.43	0.62	1.00	0.67	0.47	0.62
	M&VS	0.71	0.52	0.48	0.52	0.81	0.95	0.52	0.33
	Music	0.43		0.48		0.76		0.62	
HA	VS	0.90	0.80	0.90	0.90	1.00	1.00	0.60	0.90
	M&VS	0.80	0.80	0.50	0.60.	0.60	0.90	0.60	1.00
	Music	0.40		0.10		0.70		0.50	

Table 5: P-values obtained by comparing three types of stimuli (music only, music accompanied by video sequences, and video sequences only) and their mean values.

	p-value	Mean		
		Music	M&VS	VS
DHH	0.62	0.57	0.61	0.65
HA	2.64-e04	0.42	0.73	0.88

We used two-way ANOVA to analyze recognition rates, where the effects were subject groups and stimulus types. The p-values showed that there were significant differences between stimulus types (p=6.0-e004) but no differences between subject groups (p=0.25). Note that difference between stimulus types were caused mainly by subject groups with hearing abilities. The p-values and mean values for the three types of stimuli used in recognizing emotion are shown in Table 5. There were no significant differences between stimulus types for deaf and hard-of-hearing subjects, while there were differences for subjects with hearing abilities. A multiple comparison for the results of the subject group of hearing abilities showed that there were significant differences between stimulus types except for video sequences only and music with video sequences.

We also analyzed the recognition rates for each emotion for the three types of stimuli using one-way ANOVA. The p-values showed that there were no significant differences in recognition rate of any of the emotions for any stimulus type. The p-values and mean recognition rates for all the emotions are shown in Table 6.

6 DISCUSSION

6.1 Role of visual information

The experiments showed that there were no significant differences in recognizing emotions between two types of subject groups. This corresponds to the results of past experiments and suggests to us that musical communication may be possible between deaf and hard-of-hearing people and people with hearing abilities.

From Table 5 and the p-value obtained by comparing three types of stimuli, we can see that providing video sequences did not significantly improve the recognition rates for deaf and hard-of-hearing people. On the other hand, Table 5 shows the same recognition tendency between two subjects groups: providing subjects with the stimuli of video sequences accompanying musical performances helped their recognition of emotions. Table 6 also supports the increase in recognition rate for each emotion when video sequences were accompanied by musical performances.

Table 6: P-values obtained by comparing four emotions and their mean values.

	p-value	Emotions			
		Joy	Fear	Anger	Sadness
VS	0.24	0.78	0.71	0.92	0.65
M&VS	0.14	0.72	0.53	0.82	0.63
Music	0.10	0.40	0.28	0.74	0.57

In these experiments, our concern was the difference in recognition of different emotions. In most of the past experiment, fear was the least recognized emotion and the present experiments showed the similar results (Table 4). Table 6 also shows a similar tendency for fear to have the lowest recognition. Although there were no significant differences in recognizing emotions in terms of the p-values, the spontaneous remarks made by subjects revealed the difficulties in differentiating joy from anger and fear from sadness in listening to musical performances only. Though the improvement was insufficient, if a video sequence accompanies a musical performance, then it will be effective for the recognition of fear.

6.2 Video sequences as supplementary information for musical performances

The increase in recognition rates with video sequences encouraged us to use video sequences in our musical performance assistance system. On the other hand, the improvement in recognition with video sequences from musical performances only was small and also insufficient for differentiating emotions, so we must seek better video sequences for augmenting musical recognition. In seeking such video sequences, we should be careful not to forget that they are intended for recognizing musical performance and not for use by themselves.

Table 4 reveals several interesting results. For deaf and hard-of-hearing subjects, in the case of A2, a multimedia stimulus increased the recognition rates for both cases of a single medium (musical performance only and video sequences only). On the other hand, multimedia stimulus S2 decreased the recognition for both a music-only stimulus and a video-sequence-only stimulus. These results indicate that multimedia stimuli can be both useful and detrimental, though these examples do not apply to hearing subjects.

By comparing recognition rates for multimedia stimuli where the recognition rates of the two video sequence stimuli were the same, we found that the multimedia effect can describe the difference in video effects. This happens in the case of fear and anger for hearing subjects. Recognition rates for video sequences only for both F1 and F2 were 0.90 and those for music only were 0.10. Similarly, the rates for video sequences for A1 and A2 were 1.00 and those for music only were 0.70. While the recognition rates for multimedia stimuli for F1 and F2 do not differ much, those for A1 and A2 were 0.60 and 0.90, respectively. From this, we can predict that the Bump effect might be the better effect to use in generating video sequences of anger for our purpose. Since the recognition rates for multimedia for A1 and A2 for deaf and hard-of-hearing subjects were 0.81 and 0.95, respectively, this prediction is likely to be true.

6.3 Music beyond hearing impairment

Among the seven deaf and hard-of-hearing subjects, three were from schools for special education. Regardless of the school type, they had some experience of musical instruments at school. One of them currently belongs to a street-dance circle, one is a member of a band, and four have their own favorite music.

We feel that the auditory capacity and music recognition are not necessarily related. In our past experiments, one subject enjoyed playing wadaiko despite having auditory capacity off the scale (over 120 dB). We found by enquiry that, in general, those with greater hearing losses listen to music less than those with less hearing impairment. On the other hand, we found that the recognition rate for music only for the subject who did not wear his hearing aid was the second best in this experiment.

6.4 Future work

Our next step in building a performance assisting system will be to find better video sequences as “visual assistance”. To build better video sequences for multimedia use, we should be careful in using MPM. Namely, we should not necessarily make better video sequences that can convey emotions by themselves, but should seek video sequences that supplement music information. To get several other video scenes for each emotion, we should first analyze the relationship between the movement (effect) in video sequences and emotions, in order to avoid arbitrariness in generating video sequences. When we regenerate them, we could then repeat the experiment described in the previous section.

We will analyze musical sound data and image data quantitatively so that our system will be able to judge emotion in music and generate video sequence candidates automatically.

The question of how valuable visual information is for music includes how to activate residual hearing abilities to enable better use to be made of information from several media types, as well as how to obtain better visual information for supplementary use in recognizing emotions in musical performances. This work will include developing other types of system besides our musical performance assistant system. Another, but as yet untouched issue, includes searching for a third medium that can convey the information contained in music, such as haptic information, which is treated in Miura’s work on vibration devices [11].

7 CONCLUSION

We described the background of our plan to build a musical performance assistance system for deaf and hard-of-hearing people to enable them to share in the joys of music. Although we previously assumed that the supplementary visual information provided by the system would act as an effective tool for this purpose, our experimental results raised doubts in our mind as to how useful visual information actually is in facilitating musical communication. Thus, we reviewed our past experiments and compared the emotion recognition rates obtained through musical performance only and through musical performance accompanied by several types of visual information. Since we were not persuaded that visual information is useful for music recognition, we conducted an experiment that compared the recognition of emotion in music accompanied by video sequences and in video sequences only. Though not strong enough, the results do support the role of visual information in music communication.

ACKNOWLEDGMENTS

This research was supported by the Kayamori Foundation of Information Science Advancement and the Special Research Expenses by NTUT.

8 REFERENCES

[1] F. Blum: Digital Interactive Installations: Programming interactive installations using the software package Max/MSP/Jitter, VDM Verlag (2000).

[2] R. Bresin and A. Friberg: Emotional coloring of computer-controlled music performances, Computer Music Journal, MIT Press, 24(4), pp. 44-63 (2000).

[3] A.-A. Darrow and J. Novak: The effect of vision and hearing loss on listeners’ perception of referential meaning in music, Journal of Music Therapy, American Music Therapy Association, XLIV(1), pp. 57-73 (2007).

[4] R. Hiraga and N. Kato: Understanding emotion through multimedia—comparison between hearing-impaired people and people with hearing abilities, Proc. of ASSETS, ACM, pp. 141-148 (2006).

[5] R. Hiraga, N. Kato, and N. Matsuda: Effect of visual representation in recognizing emotion

expressed in a musical performance, Proc. of ICSMC, IEEE, pp. 131-136 (2008).
[6] R. Hiraga, N. Kato, and T. Yamasaki: Understanding emotion through drawings—comparison between hearing-impaired people and people with hearing abilities, Proc. of ICSMC, IEEE, pp. 103-108 (2006).
[7] R. Hiraga and M. Kawashima: Performance visualization for hearing-impaired students, Journal of Systemics, Cybernetics and Informatics, Int'l Institute of Informatics and Cybernetics, 3(5), pp. 24-32 (2006).
[8] R. Hiraga, T. Yamasaki, and N. Kato: The cognition of intended emotions for a drum performance: Differences and similarities between hearing-impaired people and people with hearing abilities, Proc. of ICMPC, pp. 219-224 (2006).
[9] P. N. Juslin: Communicating emotion in music performance: a review and a theoretical framework, In P. N. Juslin and J. A. Sloboda, editors, Music and Emotion, theory and research, Oxford University Press, pp. 309-340 (2004).
[10] G. Levin and Z. Lieberman: Sounds from shapes: Audiovisual performance with hand silhouette contours in the manual input sessions. Proc. of NIME, pp. 115-120 (2005).
[11] S. Miura and M. Sugimoto: Supporting children's rhythm learning using vibration devices, Proc. of CHI, ACM, pp. 1127-1132 (2006).
[12] C. Miyama, T. Rai, S. Matsuda, and D. Ando: Introduction of DIPS programming technique, Proc. of ICMC, ICMA, pp. 459-462 (2003).
[13] Y. Ota and Y. Kato: Musical education at elementary school department for the deaf, Education for the Deaf in Japan, 43(2) (2001).
[14] R. Parke, E. Chew, and C. Kyriakakis: Quantitative and visual analysis of the impact of music on perceived emotion of film, ACM Computers in Entertainment, 5(3) (2007).
[15] E. Schubert and D. Fabian: An experimental investigation of emotional character portrayed by piano versus harpsichord performances of a J.S. Bach excerpt, In E. Mackinlay, editor, Aesthetics and Experience in Music Performance, Cambridge Scholars Press, pp. 77-94 (2005).
[16] M. Senju and K. Ogushi: How are the player's ideas conveyed to the audience? Music Perception, University of California Press, 4(4), pp. 311-323 (1987).
[17] Y. Wada, K. Tsuzuki, H. Yamada, and T. Oyama: Perceptual attributes determining affective meanings of abstract form drawings. Proc. 17th Congress of International Association of Empirical Aesthetic, pp. 131-134 (2002).
[18] P. Whittaker: Musical potential in the profoundly deaf, Music and the Deaf (1986).
[19] T. Yamasaki: Emotional communication mediated by two different expression forms: Drawings and music performances, Proc. of ICMPC, pp. 153-154 (2006).

9 APPENDIX

We varied the number of stimuli and the number of subjects participating in experiments with single-medium stimuli, for both the music-only and drawings-only categories. Table 7 shows the number of improvisational style drum performances used in the experiment referred to in Section 4.1.1. Table 8 shows the number of drawings used in the experiment referred to in Section 4.1.2.

Table 7: Number of performances (in parentheses) and number of subjects for the experiment on recognizing emotion conveyed by musical performances.
HI: hearing-impaired. Aged 20–22, 12 males, 3 females.
HA: hearing abilities. Aged 21–26, 20 males, 13 females.

	Performances by		
	HI	HA	
Subjects	(11)	Amateurs (5)	Professionals (2)
HI	10	15	15
HA	33	33	33

Table 8: Number of drawings (in parentheses) and number of subjects for the experiment on recognizing emotion conveyed by drawings.
HI EM: hearing-impaired, electronics majors. Aged 20–22, 16 males, 3 females.
HI DM: hearing-impaired, design majors. Aged 20–22, 2 males, 8 females.
HA DM: hearing abilities, design majors. Aged 21–26, 26 males, 8 females.

	Drawings by		
	HI		HA
Subjects	EM (14)	DM (11)	DM (7)
HI EM	19	10	19
HI DM	10	9	10
HA	34	11	34