Comparison of Input Methods for Remote Audiences of Live Music Performances

Yuya Morino, Kei Miyazaki, Hiroyuki Tarumi[(✉)], and Junko Ichino

Kagawa University, Takamatsu, Japan
`tarumi@eng.kagawa-u.ac.jp`

Abstract. With the spread of the Internet broadcasting services, live music performances on the streaming service are getting popular. However, real time communication between performers and remote audiences is insufficient. We have developed a system using animation that supports communication between performers and remote audiences. With the system, remote audiences can show three types of their body actions to the performers. We have implemented three input methods – mouse devices, keyboards, and smart phones – and evaluated them. According to our evaluation, subjects who have experiences in listening to live music preferred smartphones, but some subjects liked keyboard input.

1 Introduction

Live streaming on the Internet is getting more and more popular. It enables many people, including musicians who only have tight budget, to have their own channel that reaches everyone on the net. In case of musicians, they often make use of the streaming channel to show their live music performances.

However, one of the problems in those cases is that the communication channel is basically one-way. In case of live music performances, especially in cases of rock or popular music, audiences take actions responding to the played music. Responses from the audiences are considered as important elements of the successful live performances. Live streaming on the net lacks upstream communication functions that let the musicians perceive remote audiences' responses.

Of course, live streaming tools have limited functions of communication from the audiences to the musicians: text chat. Ustream [1] and Niconico Live [2] provide text chatting interfaces to the audiences. In case of Niconico Live, especially, it has a very unique user interface that displays chatted texts overlaid on the main video contents. Nicofare [3] is a live concert hall in Roppongi, Tokyo, which has walls where text comments from remote audiences are displayed. However, they do not satisfy our requirements. We need non-verbal communication channels from the audiences to the musicians.

Yoshida and Miyashita tried to display audiences' body actions overlaid onto the video contents [4]. However, it was developed for stored video contents, not for real time live performances. Several projects are found that try to enhance communication between musicians and audiences at the live venue (e.g., [5–7]). But they are using local

communication technologies like WiFi or ultra sound (in case of [5]), which cannot be applied to remote audiences.

Our research is focusing on these problems: how to support remote audiences to give their responsive actions and how to let musicians perceive audiences' responses. It is an important key to have successful collaboration between the music players and remote audiences.

We limit the problem domain to only rock and popular music, because audiences' responses differ depending on the music genre. We also exclude "big" artists' cases, because huge budgets completely change the problem conditions. We know many young fledgling musicians want to give their music to as many audiences as possible, with limited money. They are our target users.

In this paper, we focus on the input methods for remote audiences. We have implemented three methods to input their responsive actions and compared them from the viewpoint of remote audiences. Other aspects, such as evaluations from the players' side, are not described in this paper.

2 Experimental System

2.1 Actions

According to our experiences, we have selected five typical responsive actions by audiences for rock and popular music performances [8]. They are: *sing*, *wave* (wide movement of one hand), *push-up*, *joggle* (rhythmical shakes of one hand), and *hand-clap* (Fig. 1). These actions are usually taken by audiences all together, during particular parts of songs or especially encouraged by musicians. Of course, they do not cover all kinds of actions observed at live venues. However, these five actions are common to many musicians' cases and we consider they are enough to support typical cases.

Currently, we have implemented three of them: *wave*, *push-up*, and *joggle*.
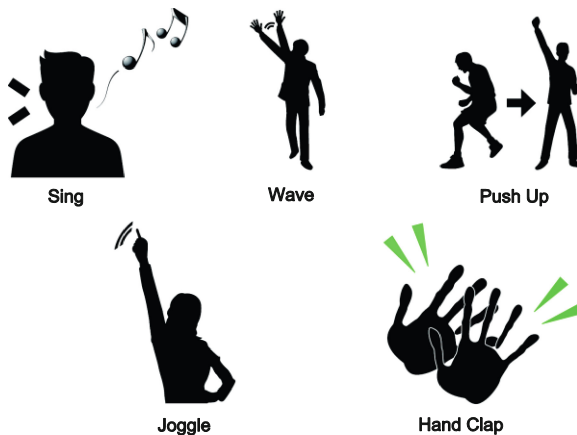

Fig. 1. Five typical responsive actions

2.2   System Configuration

Figure 2 shows the system configuration. Remote audiences watching the live streaming video can input three types of responsive actions by one of the three methods: *moving mouse pointers*, *typing keyboard*, or *using smartphones*. The information of actions is sent back to the computer at the live venue. The format of action information is very simple. It does not show analog values of each motion (e.g., amplitude, acceleration, time, and speed), but shows only the type of actions taken by each audience.
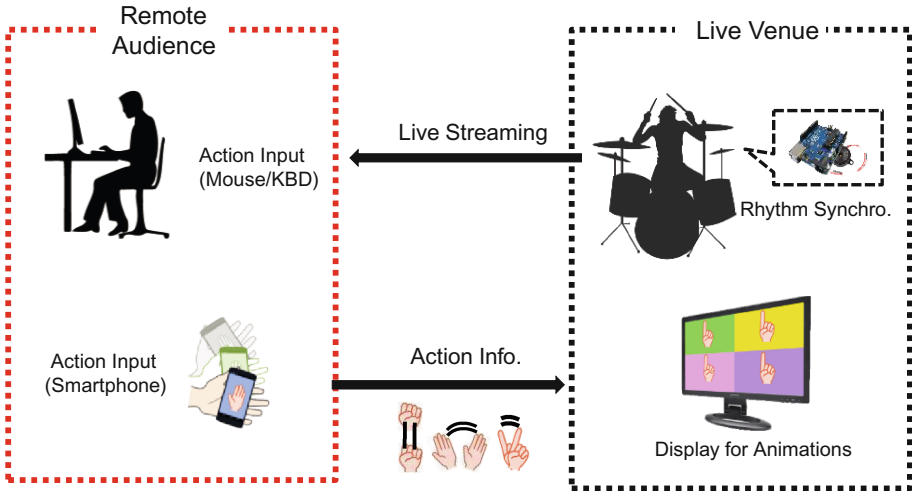

Fig. 2. System configuration

The actions taken by each audience are displayed to the musicians during their play, by animations. The animation is represented by avatars by illustrations of hands. The motions of animations are synchronized with the music played at the live venue, by detecting the drum beats with a vibration sensor. The synchronization technique is still under research to pursue a better solution, but we support a very simple synchronization technique for the current prototype.

2.3   Input Methods

We have implemented three input methods for remote audiences, to compare and evaluate them.

Keyboards. Audiences can simply type "b" key for push-up actions, "n" key for joggling actions, and "m" key for waving actions.

Mouse Devices. In this case, we have defined a "mouse event area" overlaid on the motion video (Fig. 3). It is needed to detect mouse motions. Audiences can move the mouse pointer horizontally within the "mouse event area" to input the waving actions.
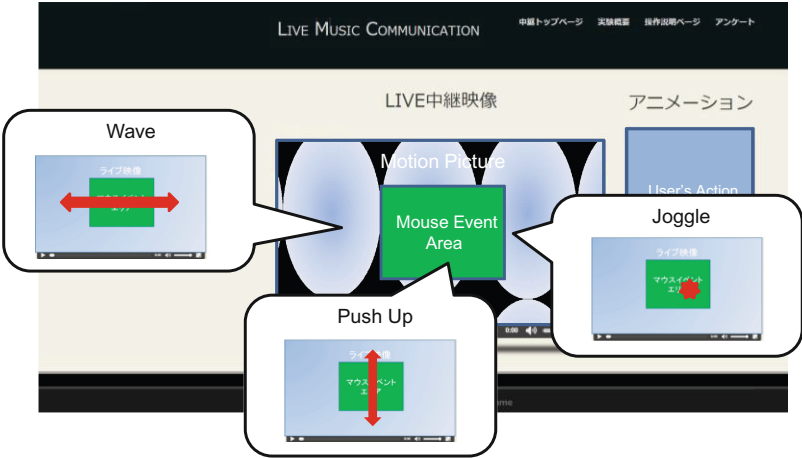
Fig. 3. Input method for a mouse

They can move it vertically to input the push-up actions. In order to input the joggling actions, they need to click the mouse within the "mouse event area."

Smartphones. In this case, each audience needs an additional device, a smartphone, other than the computer to watch the live streaming. We used the acceleration sensors (x, y) and a gyro sensor (z) to detect the hand motions. Figure 4 illustrates how we have defined the movements of smartphones corresponding to the audiences' actions. The current implementation is tuned for ASUS ZenFone2.
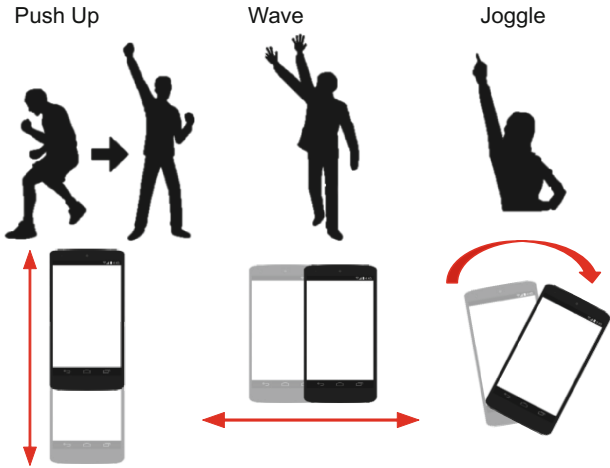

Fig. 4. Motion sensing with smartphones

3   Evaluation

3.1   Evaluation Outline

This time, the purpose of experiment is to evaluate the input methods and acquire general comments from the test audiences. Hence we recruited test users and let them experience the system in the environment of remote audiences. After the experience with three different input methods, we gave them questionnaire and interviews.

From January to February, 2016, we had nine test users. All of them were students and had experiences of attending real live concerts.

To make the evaluation setting simple, stable, and even, we did not give real live performance by human players, but adopted recorded live performance of a professional rock band. Figure 5 is a picture of testing scene in case of the input method with a keyboard. The test user was watching the recorded music performance on the left LCD. He was giving his responsive actions using the keyboard. His action was confirmed by the animation of hand motion at the left LCD. The right LCD was showing a virtual display that should have been shown to the musicians at the live venue. The test user's action was shown at the upper-leftmost sub-window. Other seven sub-windows were dummies, representing other audiences. The reason we set the right LCD was to let the test user know the total system concept.
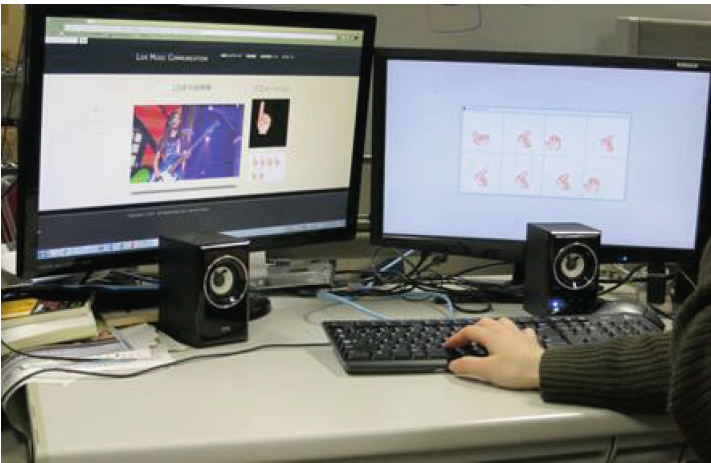

Fig. 5. Testing environment

The process of experiment was as follows. First, we gave explanations of the research background and the system concept. Next, we gave instructions of the system operation to test users, for all input methods. The users had some test practices. After the practices, all users experienced all input methods, each of which took four minutes. After the experience, test users answered the questionnaire on the web. We also gave interviews with them.

3.2   Results

Comparison of the Three Input Methods.   Table 1 shows the results of questionnaire comparing the three input methods.

We had supposed that smartphone should be the best input method because it was most similar to the body actions given at the real concert. However, the result was a little different. From the viewpoint of concentration on the music, keyboards seemed to be the best solution. However, smartphones were best preferred totally and they had the feeling of attending the concert better than other devices. Mouse devices did not have any advantage.

Table 1. Comparison of three input methods.

Question	Device	Strongly agree	Slightly agree	Neutral	Slightly disagree	Strongly disagree
I was able to concentrate my attention on the music	Mouse	1	2	0	4	2
	Keyboard	3	5	1	0	0
	Smartphone	1	4	0	2	2
I felt like I really attended the concert	Mouse	0	4	1	3	1
	Keyboard	1	2	2	3	1
	Smartphone	2	6	0	1	0
		Yes				
I like to use this device (Multiple selections are allowed)	Mouse	2				
	Keyboard	3				
	Smartphone	6				

We have given more consideration based on the comments from the test users given in the interviews. The merit of keyboard was that it was physically easy and the input was stable. In case of smartphones, weight of the device was a physical burden. We also had a problem that the gesture recognition was not always perfect, which might be a psychological burden for the users.

It was interesting that some users who did not like keyboard commented that it was like a "task" to hit keys, but some users who liked keyboard commented that they felt like playing action games when hitting keys with the music.

Other Comments.   One test user requested us to introduce virtual penlights for the responsive actions. He said that some live venues inhibited audiences from using penlights due to the safety reasons, but remote audiences did not have safety problems with other audiences.

This comment is important. We have two approaches to such kind of remote systems. One is to make the remote environment as similar to the local environment as possible, and to give remote users same kind of experiences with local users. The other approach is to determine the difference between the remote and local environments and exploit

the difference to give advantages to remote users. Hitting keys instead of waving or joggling by their own hands is an example of the latter approach.

If it is a business system, practical advantages (e.g., efficiency) would be almost always accepted. However, it is a community support system. We should consider psychological and social issues. The question is, how the music players or local audiences think about it.

4   Conclusion

We are developing a system to support remote audiences to give their responsive actions during the live music performance to the players at the live venue. In this paper, we compared three input methods for the remote users: mouse devices, keyboards, and smartphones. We have conducted experiments with test users and found that smartphones are the best device to give users experiences feeling like attending the real concerts. However, keyboards were also preferred by some users, because of the low physical burden and a different kind of fun like gaming.

For the future work, we will evaluate the system totally with music players, local audiences and also remote audiences. Evaluations that should be done with musicians will include, for example, animation representation, animation synchronization mechanism, etc.

As stated in the last section, it is an important problem how the remote environment should be designed. Should the remote audiences have experiences similar to the local audiences? Or would it be accepted that they take advantage of the remote environment? We will still consider this problem with evaluating other aspects of this system.

References

1. Ustream. http://www.ustream.tv/
2. Niconico live. http://live.nicodivdeo.jp/
3. Nicofare. http://nicofare.jp/
4. Yoshida, A., Miyashita, H.: Video sharing system that overlays body movements for the sense of unity. In: Proceedings of Interaction 2012 (IPSJ), pp. 187–188 (2012). (in Japanese)
5. Hirabayashi, M., Eshima, K.: Sense of space: the audience participation music performance with high-frequency sound ID. In: Proceedings of the International Conference on New Interfaces for Musical Expression (NIME2015), pp. 58–60 (2015)
6. Lightwave. http://www.lightwave.io/
7. FreFlow. http://freflow.com/
8. Tarumi, H., Akazawa, K., Ono, M., Kagawa, E., Hayashi, T., Yaegashi, R.: Awareness support for remote music performance. In: Nijholt, A., Romao, T., Reidsma, D. (eds.) ACE 2012. LNCS, vol. 7624, pp. 573–576. Springer, Heidelberg (2012)