Enhancing Communication and Dramatic Impact of Online Live Performance with Cooperative Audience Control

Takuro Yonezawa  
Keio University  
Delta S213, Endou 5322, Fujisawa, Kanagawa, Japan  
takuro@ht.sfc.keio.ac.jp

Hideyuki Tokuda  
Keio University  
Delta S213, Endou 5322, Fujisawa, Kanagawa, Japan  
hxt@ht.sfc.keio.ac.jp

ABSTRACT
Recent progress in information technology enables people to easily broadcast events live on the Internet. Although the advantage of the Internet is live communication between a performer and listeners, the current mode of communication is writing comments using Twitter or Facebook, or some similar messaging network. In one type of live broadcast, musical performances, it is difficult for a musician, when playing an instrument, to communicate with listeners by writing comments. We propose a new communication mode between performers who play musical instruments, and their listeners by enabling listeners to control the performer’s camera or illumination remotely. The results of four weeks of experiment confirm the emergence of nonverbal communication between a performer and listeners, and among listeners, which increases camaraderie amongst listeners and performers. Additionally, the dramatic impact of a performance is increased by enabling listeners to control various camera actions such as zoom-in or pan in real time. The results also provide implications for design of future interactive live broadcasting services.

Author Keywords
Interactive live broadcasting, collaborative camera control, musical performance, nonverbal communication, amateur

ACM Classification Keywords
H.5.2 Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors, Communication, Measurement, Experience

INTRODUCTION
Continued progress in the expansion of the Internet and information technologies has enabled us to provide information easily worldwide. Especially, amateur users can broadcast various contents live on the Internet today, which was heretofore possible only by expert users with professional-level hardware and software. Due to the proliferation of services such as Ustream [3], YouTube Live [1], and NicoNico Live [2], performers and listeners can increasingly enjoy live broadcasting. One advantage of internet live broadcasting is the two-way real-time communication that can occur between performers and listeners. For example, Ustream enables listeners to send messages to a performer or other listeners through social network services such as Facebook or Twitter. Moreover, in NicoNico Live, the users’ comments are shown over the live streaming movie. Thereby, performers and listeners can communicate more directly.

A main component in live broadcasting is its musical instrument performance. Traditional live musical performances were performed only in concerts or halls, in the street or through TV, however live broadcasting services enable musicians to perform live in their homes. Because communications between performers and listeners are limited to sending messages, it is difficult for musicians to communicate with listeners while they are playing their instruments. Therefore, interactivity—a great advantage of live internet broadcasts—is lost in live musical performances.

To address and solve that problem, we present the Listener-Cooperative Live Production System called LCLPS, which enables listeners (1) to control the performer’s environment (e.g., camera or illumination) remotely in real time, and consequently (2) to enhance communication between the performer and the listeners. Although professionals can afford a large number of staffs to support a musician’s performance by controlling cameras or lighting during live performances, such support is too costly for amateur musicians. By enabling listeners to control cameras or other devices remotely, the LCLPS provides a high dramatic impact similar to that of a professional live performance but without the cost. Additionally, through the listeners’ control, we aim to bring nonverbal communication, which has an important impact on musical performances. The performer and listeners can send and share their intent for the musical performance through control of the performer’s devices. This paper reports our study with four weeks’ experimentation using the proposed system with thousands of listeners, and presents an evaluation of the LCLPS’ effectiveness and usability. Our results show that the LCLPS enhances the dramatic impact of a musical performance, and communication between the performer and listeners.
In summary, our work offers three main contributions:

- Creating a live broadcasting support system called LCLPS, which enhances the dramatic impact of the performances.
- Verifying the emergence of nonverbal communication between the performer and listeners, and also among the listeners, by sufficient experimentation using the proposed system.
- Presenting implications for design of future interactive live broadcasting services.

The remainder of the paper is organized as follows. Firstly, we present the motivation of our research with a summary of the current status of live broadcasting services, their strengths and weaknesses. Then, we describe the concept and implementation of the LCLPS. In the following section, we evaluate the LCLPS in an actual live broadcasting situation, in terms of its effectiveness and usability. Subsequently, we discuss its implications for designing future interactive broadcasting systems. Finally, we describe related work and our conclusion.

**MOTIVATION**

As evidenced by the spread of the Internet-based live broadcasting services [3, 1, 2], many users enjoy not only seeing live streaming of performances, but also to broadcast themselves. Additionally, the ubiquity of personal computers, smartphones and internet access enables users to see and broadcast live streaming anytime at anywhere. The contents of these broadcasts have also become increasingly diverse. Especially, musical performances are a popular type of internet live performance; many amateur musicians enjoy broadcasting their musical performance from their own homes.

Amateur musicians frequently devise new ideas to enhance the entertainment features of their performances. For example, some musicians play instruments in concert with existing music contents found on the Internet such as in music videos on YouTube, as “mashup” live session (see Fig. 1). Although music contents have faced copyright problems in the past, recent live internet services allow users to use this music in their programs by signing deals with its copyright owners such as record companies. Consequently, such mashups are expected to become increasingly popular in the future.

One advantage of live internet broadcasting is the high potential for interactivity between performers and their listeners. Through social streaming or chat space in live internet web services, they can communicate with each other in real time. Consequently, although the skills of amateur musicians are often inferior to those of professional musicians, both the performers and listeners can increase enjoyment of musical performances by communicating with each other (e.g., the listeners can request songs to be performed by the performers). Nevertheless, two main problems arise during amateur musicians’ live broadcasts. First is the intermittent communication problem. When amateur musicians play instruments, the communication between the musician and listeners is greatly decreased because the amateur musician cannot read and reply to messages from listeners while they are concentrating on playing instruments. This shortcoming in turn reduces the big advantage of the Internet live performance: interactivity. Second is the lack of dramatic impact. In live broadcasts of professional musicians, dramatic impact is enhanced through the cooperation of a large support staff including camera operators or illuminating staff. For example, professional camera operators can enhance a performance by manipulating a camera’s zoom or rotation (see Figure 2). However, because amateur musicians are not supported by such a professional staff, their broadcasts are often static in nature.

To solve these two problems simultaneously, we specifically devote attention to remote control of a musician’s broadcasting equipment by listeners. The purpose of our research is to show that such control is useful for enhancing both communication and dramatic impact. We especially expect the emergence of nonverbal communication between amateur musicians and their listeners. For example, we assume not only that the listeners control a performer’s camera in response to the musician’s performance, but also that the musician’s performance can change in response to listener’s camera control. Network-controlled pan-tilt-zoom (PTZ) cameras are commercially available. No technical difficulty exists in controlling the equipment remotely. However, our motivation arises from human factors: recognizing what will occur when remote control of those devices is used within a live streaming contexts. The main purposes of such commercially available PTZ cameras are for security, or monitoring traffic. No research has been done into giving camera control to many users during live streaming. Another important aspect of our research is manual control. Ubicomp technologies enable us to control devices automatically in a context-aware manner. There also have been many efforts to control cameras automatically in lecture rooms [4, 18] and meeting rooms [17, 9, 21]. Our research is conducted to explore the advantages of manual control: the emergence of a
new kind of communication.

LISTENER-COOPERATIVE LIVE PRODUCTION SYSTEM
This section presents a description of concepts, design and implementation of the LCLPS.

Concept
Figure 3 portrays the LCLPS concept. In the performer’s environment, equipment such as cameras and lighting is used for live broadcasts. Listeners can control such devices remotely while they are seeing the live broadcast. The performer sends their live performance through a streaming service (e.g., Ustream or NicoNico Live). The listeners can manipulate performer’s devices by sending control messages through a chat space on the live broadcast websites or Twitter. Consequently, the listeners need not install additional applications on their PC or smartphone to control the performer’s devices. Our implementation targets the control of cameras because cameras have the most dramatic impact on a broadcast.

Hardware Components
Figure 4 shows the performer-side hardware components in our implementation: three web cameras (Cameras 1-3), two keyboards with different sounds (Keyboards 1 and 2), and a PC which connects to those devices. The cameras are placed accordingly: Camera 1 captures the performer’s upper body, face, and Keyboard 1. Camera 2 captures Keyboard 2 and the performer’s hand, and Camera 3 captures the performer, Keyboard 1 and Keyboard 2 from the ceiling.

Software Components
Figure 5 shows the software architecture of our implementation. To enable many listeners to participate in the control of the performer’s devices, we claim that the system should be simple. Therefore, we designed our system to obviate installation of software on the listeners’ computers. The Comment Analyzer in Figure 5 collects and analyzes listeners’ comments on live broadcast websites using XML-Socket, and listeners’ posts on Twitter through the Twitter API. When listeners’ comments are identified as control commands, Comment Analyzer invokes Camera Controller to operate the camera. Table 1 presents each command and its operation. Listeners can control the performer’s camera by posting these commands. We implemented basic camera work using digital image processing, as illustrated in Figure 2. Additionally, we prepared an operator “;” that executes multiple commands in one message. For example, “cam3:zooming” means selecting camera 3, then performing a zoom-in effect. Figure 6 shows the sequence of camera work according to each command. We implemented the system with Java and OpenCV.

Conflict resolution of multiple commands sent by listeners simultaneously is not the focus of this study. We designed a simple priority policy as follows. The commands “cam1”, “cam2”, “cam3” are executed exclusively. The latest command is always operated. Although conflicting effects of each zoomin/out, panright/left, panup/down and rotateright/left are executed serially, effects of different types are operated simultaneously. For example, when “zoomin” and “rotateright” are requested simultaneously, then the effects overlap. These effects are canceled when a serial camera selection command is processed because we consider that such effects should be combined with each camera.

EVALUATION
This section presents a field experiment and two surveys to evaluate the LCLPS approach.

Experiment
To evaluate how the system influences dramatic impact and communication among the performer and listeners, we conducted broadcast experiments during one month (Mar. 25 – Apr. 24.) The performer was an author of the paper who has a background in playing piano. Table 3 shows the dates, time, and the length of each live broadcast in the experiment. The number of participants, all comments, and comments for control of the camera are also presented in the Table 3.
For the live broadcasting platform, we chose NicoNico Live, one of the most popular live broadcasting services in Japan. A key feature of NicoNico Live is that listeners’ comments are overlaid directly onto the streaming video. This feature motivates listeners to expose their feelings spontaneously. During the experiments, we performed on the keyboard almost in a mashup live session which was described in the previous section (see Figure 1). Through the experiments, all comments were recorded. We also administered a questionnaire survey to elicit detailed comments and a usability evaluation.

We didn’t provide any special training for listeners using the LCLPS. Instead, we presented simple instructions (see Table 2) to show how to use the LCLPS on top of the streaming video frame. This means that all of the participants learned how to control the cameras on-the-fly during the actual broadcast.

**Emergence of Communications between the Performer and Listeners**

In this section, we present what kind of communication emerged among performer and listeners using our system. We will use a live streaming program that we broadcast on Apr. 22 as an example. We acquired the largest number of listeners on this day, because our program was featured into the official program by NicoNico Live. At that time, more than half of the listeners saw the performer’s broadcast and also the LCLPS for the first time, therefore, we were able to obtain
Table 3. List of experimental dates and the number of listeners, comments and control comments.

<table>
<thead>
<tr>
<th>Dates</th>
<th>Time length (min)</th>
<th># of L</th>
<th># of AC</th>
<th># of CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar. 25</td>
<td>60</td>
<td>222</td>
<td>596</td>
<td>156</td>
</tr>
<tr>
<td>Mar. 26</td>
<td>60</td>
<td>141</td>
<td>336</td>
<td>79</td>
</tr>
<tr>
<td>Mar. 30</td>
<td>90</td>
<td>266</td>
<td>651</td>
<td>177</td>
</tr>
<tr>
<td>Apr. 1</td>
<td>90</td>
<td>210</td>
<td>403</td>
<td>138</td>
</tr>
<tr>
<td>Apr. 3</td>
<td>30</td>
<td>95</td>
<td>110</td>
<td>49</td>
</tr>
<tr>
<td>Apr. 4</td>
<td>90</td>
<td>499</td>
<td>1210</td>
<td>152</td>
</tr>
<tr>
<td>Apr. 5</td>
<td>60</td>
<td>256</td>
<td>674</td>
<td>263</td>
</tr>
<tr>
<td>Apr. 6</td>
<td>90</td>
<td>343</td>
<td>763</td>
<td>216</td>
</tr>
<tr>
<td>Apr. 9</td>
<td>90</td>
<td>328</td>
<td>836</td>
<td>181</td>
</tr>
<tr>
<td>Apr. 13</td>
<td>90</td>
<td>327</td>
<td>1226</td>
<td>194</td>
</tr>
<tr>
<td>Apr. 14</td>
<td>60</td>
<td>196</td>
<td>672</td>
<td>119</td>
</tr>
<tr>
<td>Apr. 16</td>
<td>60</td>
<td>192</td>
<td>346</td>
<td>152</td>
</tr>
<tr>
<td>Apr. 18</td>
<td>90</td>
<td>193</td>
<td>494</td>
<td>148</td>
</tr>
<tr>
<td>Apr. 20</td>
<td>60</td>
<td>181</td>
<td>234</td>
<td>89</td>
</tr>
<tr>
<td>Apr. 21</td>
<td>90</td>
<td>250</td>
<td>1148</td>
<td>212</td>
</tr>
<tr>
<td>Apr. 22</td>
<td>90</td>
<td>658</td>
<td>6323</td>
<td>560</td>
</tr>
<tr>
<td>Apr. 24</td>
<td>90</td>
<td>352</td>
<td>1436</td>
<td>311</td>
</tr>
</tbody>
</table>

# of L = number of listeners, 
# of AC = number of all comments, 
# of CC = number of comments for camera control

unbiased feedbacks from them, not just like friendly feedbacks from fans.

The live performance was broadcast for 90 min. We had 658 listeners and 6323 comments, which included control commands. All listeners were ordinary people with whom we were not acquainted. Out of 658 total users, there were 61 users (≈10%, it seems to be followed by 90-9-1 rule [15]) who sent a total of 560 control commands. The breakdown of camera controls were camera selection (168), zoom effect (374), pan effect (58), and rotate (118). The total number of camera controls (718) is greater than the number of comments for controls (560) because some comments use the “:” operator to execute multiple commands in one comment.

Figure 7 presents a timeline of two different scenes (scenes A and B) during the Apr. 22 live broadcast. The timeline shows video frames captured at five second intervals. We assigned a number to each frame (from A-1 to A-22, and from B-1 to B-29) and display comments for control commands with a balloon on the timeline. In looking at the timeline, we can see that listeners often select a camera in response to a performance. For example, when the performer starts to play keyboard 2 in A-7, a listener sent a “cam2” command to capture the performance clearly. As another example, when the performer ends playing in B-24, a listener sent a “cam1” command to capture the performer’s facial expression after the performance. We observed that listeners often try to understand the performer’s intention for performance, and control the cameras accordingly to the performance. Therefore, the performer and listeners mutually communicate through the performance.

We also observed other nonverbal communication between the performer and listeners. When the listener wants the performer to change his playing style, the listeners can control the cameras to communicate this (e.g., switching the cameras to encourage the performer to move). In such a case, the performer can detect the listeners’ initiative and change his playing style in response to the listeners’ camera manipulations. Scene C in Figure 8 demonstrates this. After a listener changed the camera source from camera 1 to camera 2, the performer changed from keyboard 1 to keyboard 2. There was another type of camera control in a mashup session. Some listeners controlled the camera by synchronizing movements with the camera work of the existing music videos playing next to the performance.

We received many positive comments for the LCLPS while we broadcasted, such as “New way to enjoy music!”, “I can feel togetherness.”, “This program makes both the performer and listeners happy,” and “Listeners can participate directly in the performance!” Some listeners described the emergence of communication between the performer and listeners, and also among listeners themselves, such as “I enjoy communication by switching cameras according to performance, and also by switching cameras to change performance.”, “I enjoy collaboration with other listeners. When I feel how other listeners want to enhance dramatic impact, I help them by performing additional effective camera control.”, “It is very interesting to make a better program with the performer and other listeners,” and “Thanks to the system, I can feel closer to the performer and other listeners.” These positive clearly shows how nonverbal communication emerged. Moreover, we find that the communication enhanced camaraderie among the many participants in the session.

Enhancement of Dramatic Impact

This section presents a description of how dramatic impact of the program was enhanced using the LCLPS. We observed many positive comments about dramatic impact in the program. Comments such as “There is a professional camera operator definitely.”, “What wonderful camera work.”, “Good camera angle.”, and “This is just like a professional program.” are some examples. We also recognized that the LCLPS reduces the gap in the quality of the dramatic impact between professional programs and amateur programs during mashup live sessions. Scene D in Figure 9 portrays an example of listeners’ control based on a professional music clip. In scene D-1, the professional music clip shows camera shots of an overhead view. The screen capture shows that a listener changes the camera to the same overhead view perspective. Because this effect merges the professional’s and our broadcast, we actually received many comments such as “The performer appears on the same stage with professional musicians.” and “I feel like I’m in a real concert.”

We also conducted an online survey to ascertain how listeners feel about the dramatic impact of the LCLPS. We recruited participants of the survey by announcing it on our broadcasting web page in NicoNico Live. The statements and results of the online survey are presented in Table 4. Users indicated their ratings of agreement for the statement on a
1–5 scale (1=strongly disagree, 2=partly disagree, 3=neither agree nor disagree, 4=partly agree, 5=strongly agree). We also required users to explain why they selected their rating using comments. The participants were made up of 44 users who saw the broadcast, including 10–19 year olds (4), 20–29 year olds (20), 30–39 year olds (6), 40–49 year olds (12), and 50–59 year olds (2). Results show that the average score was 4.47.

Commenters who scored positively (4 or 5) included comments such as “We can see the performer in multi angle.”, “When the mood of the song was changed, camera effects...
are also changed. This is similar to the effects of professional program.” and “We can feel the existence of multiple camera operators. This makes me feel like I have joined a real concert which is created by multiple staff.” We also had some comments which described negative effects on dramatic impact. For example, some users reported conflicts of many camera control commands such as “when many comments for camera control are sent from many listeners, camera effects sometimes change very quickly. I guess this case is not good for visual effect itself.” In contrast, we confirmed that some users enjoy such conflicts “Conflicts are a very interesting phenomenon. It’s like a spice of the handmade live streaming. I can feel that there are many listeners who want to make the broadcast better through camera control.” Through our experiments, overall we confirmed that the LCLPS can enhance the dramatic impact of live streaming.

**Usability Evaluation**

This section evaluates the LCLPS’s usability with five indicators [11]: Learnability, Efficiency, Memorability, Error Handling and User Satisfaction through the online survey. We believed that some users might encounter difficulty in controlling cameras by text commands. The survey was administered on the Internet by recruiting participants through an announcement on our broadcasting web page. We had 37 participants who had experience with controlling the camera in our program. The participants included 10–19 year olds (8), 20–29 year olds (15), 30–39 year olds (7), 40–49 year olds (5), and 50–59 year olds (2). Table 5 presents statements and results obtained from the questionnaire. The respective ratings of agreement are 1=strongly disagree, 2=partly disagree, 3=neither agree nor disagree, 4=partly agree, and 5=strongly agree. Note that all questionnaires are in the form of positive statements which has a possibility to cause positive bias. Thus, we focus more on their remarks.

**Learnability**

The average rating of the statement “It was very easy to learn how to use the system.” was 3.78. We confirmed that a learning curve was present for some users learning how to use the system. We obtained both positive and negative comments for learnability from the questionnaire such as “I can use the system through some experiences.”, “Commands are easy to understand. I can control the camera intuitively.”, “It was difficult for me to understand pan effects. I sometimes confused pan right or left commands because it’s difficult to understand right or left means which direction of the camera movement.”, “It was very easy for me to learn how to use the system because I’m experienced with CUL.”, and “I was able to use the system after the first trial.” Overall, text-based control is apparently tolerated by most users.

**Efficiency**

The average rating of the statement “The system is very pleasant to work with” was 3.65. In the comments, some users described the “:” operator that connects two or more camera control commands such as “Efficiency was enhanced with multi-command using :” operator.” One user indicated the problem of text-based control, “I made an effort to imagine how my command affects camera shot. I think it will be easy to select a camera graphically like a camera switcher of TV station”. We confirmed that the text-based camera control worked well with users, but more efficient methods should also be investigated. We discuss this point in the next section.

**Memorability**

The average rating of the statement “It was very memorable how to use the system.” was 3.46. In a comments section, there were opinions given such as “I can remember the command if there are not only English commands but also Japanese commands.” In contrast, a user who seems to have much experience at controlling cameras in our program described that “I will not forget how to use the system even after one year.” The grades of memorability of the system varied widely compared with those of other statements. This also implies the necessity of intuitive interfaces.

**Error Handling**

We asked users to report their experiences of control failure in the comments. As described in the Learnability section, some users remarked that “I often confused direction of pan right and pan left, or rotate right and rotate left.” We often encountered this error when we broadcasted. Another user reported errors resulting from conflicts with other users, “Though I just thought my command was wrong, in fact another user sent another command shortly after my command.” We also observed other comments that described the latency of the live broadcasting. According to the load condition of NicoNico Live’s live streaming server, the streaming video is broadcast with latency from 3 sec in minimum to 10 sec in maximum. Consequently, in such cases, users who want to control a camera effectively must predict the performance in 3–10 sec. Some users seemed to feel that this prediction is difficult. Meanwhile, some users enjoy this prediction like a game such as “To predict the performer’s action is both difficult and interesting. When my command succeeds in making effective control, I’m very excited”.

**User Satisfaction**

The average rating of the statement “It was very fun to use the system.” was 4.27. Therefore, we conclude that our system is satisfying for many users. In the comments, we observed many positive remarks such as “interesting system”, “I enjoyed”, “excited”, “I’m very impressed when my control realizes very good effect. I’m also very happy with other listener’s feedback.” and “I enjoy communication with the performer or other listeners.” These responses indicate that communication through camera control with other listeners contributes to a high rating of listeners’ satisfaction.

Some users expressed concerns about conflicts of control requirements or mischievous users (trolls), such as “I worry about conflicts. When many and many users want to control the cameras, will bad conflicts occur?” and “I wonder if trolls would appear.” Although we did not encounter any trolls in our experiments, some function to prevent trolls might be necessary in general. As for conflicts, we observed
an interesting phenomenon by which listeners try to avoid conflicts or collaborate to create good effects by themselves. In other domains employing shared camera control, there was evidence that users wanted to fulfill their own needs, not those of others. The designers of these systems attempted to address conflicts in camera control [20, 12]. However, in our system, each user wanted to show movies with high dramatic impact for all listeners. Consequently, the users can collaborate naturally to solve conflicts and create good broadcast programs. Moreover, the users seemed to enjoy solving the conflicts, for example, “I sometimes feel that the other users seem to wait my or others’ control to avoid conflicts or to keep equal opportunities for controlling. So, I sometimes behaved in the same way. In such a case, I can recognize human kindness.”

Since the performer in the experiment was one of the authors of the paper, it is hard to describe the performer’s experience without biases. However, we should still report on the performer’s experience to provide a more complete understanding of the impacts of LCLPS. Overall, we were satisfied with our performances with the LCLSP because we felt that both dramatic impact and communication with listeners are enhanced as listeners remarked. Especially, as a performer, we were very excited when our performance was directed effectively by the listeners. We also felt comfortable with performance changes initiated by listeners’ camera control. It is hard to devise ways of performance by ourselves in all the time of broadcasting. Therefore, entrusting our performance to the listeners made us easy to organize the broadcast program. We were also encouraged to devise new ideas for performance by the listeners’ camera control. For example, we were sometimes encouraged to play two keyboards simultaneously (right hand for the keyboard 1 and left hand for the keyboard 2) when our performance was captured by the overview camera. This performance style had not been appeared when we performed without the LCLPS. On the contrary, we consider that other musicians sometimes feel uncomfortable with controlled performance by listeners if the musicians’ intents for performance are greatly different from listeners’ intents. We have two approaches to solve such the case. One is that when we continue to convey our intents to listeners, the listeners changed their intents in response to our performance. Second is a negative approach, but we prepare a control interface which terminates a function to control the cameras by listeners. Though we didn’t encounter such a situation that we wanted to use the controller, it is also useful for trolls.

**DISCUSSION AND IMPLICATIONS FOR DESIGN**

The experimental analysis of our system indicates possible areas of improvement for the design of current live broadcasting services, and implications for the design of future services.

**More Effective Performances**

We focused only on camera control in the current implementation, but the LCLPS can control other devices such as illumination and special effects such as smoke generation. Our experiments prompted some comments regarding future development such as “If I could control other devices such as a mirror ball, it would be very exciting for me.” This will enable us to create more professional live performances with other users. We also consider that great potential remains for camera devices. If listeners were able to control cameras with a movable crane or flying helicopter cameras, it would provide more dynamic and attractive movies.

We also anticipate more effective interfaces to control devices. For example, we consider that it is worth designing a multi-touch and gesture-based camera control interface to enhance intuitiveness (see Figure 10). The challenge to solve latency problems is also important. This enables many users to control devices remotely with appropriate timing. From the point of view of enhancing communication and a great sense of camaraderie, we consider the importance of identifying the existence of contributors who enhance broadcast programs. From our experiments, we observed many comments which applauded the listeners who participated in the program by controlling cameras. We consider that maintaining this kind of positive communication is an important priority for maintaining a collaboration system.

**Adapting the Concept to Other Areas**

Though we focused on live streaming of musical performances in this paper, our system can be adapted to other categories. Art, dance, magic, juggling, board gaming, and various sports are examples of live performances that can be supported by our system. In the ubicomp research area,
Balancing between Manual and Automatic Control

Through our experiments, we studied the importance of balancing between manual and automatic control. If we were able to take a fully automated approach, we would have no new communication between performers and listeners, or among listeners. On the other hand, we implemented several automatic limitations for camera control. For example, the magnification percentage of zoom-in effects were limited to 150%. Even if listeners do not send a “zoomstop” command, the zoom-in effect is stopped automatically to maintain an appropriate presentation.

Parasuraman et al. [16] proposed an automation framework that has ten levels (see Table 6). Although we can design and implement a system that enhances dramatic impact at any level, it is important to examine the tradeoff between efficiency and the possibility of communication among users. Another important implication for collaborative system design is related to simplicity. Our system was very simple to use, which in turn allowed listeners to devise various means of control. At the end of the experiment, some users seemed to have developed a great ability to control cameras. They combined several commands effectively and made expressive performances. Some other listeners called such skilled listeners “craftsmen” with praise. In ordinary live broadcasting programs, only the performers are spotlighted. In contrast, our system enabled some listeners to be spotlighted as important participants in the program. In such a system, simplicity plays an important role in emphasizing the existence, participation and ability of humans.

RELATED WORK

Some research has been conducted to create interactive broadcasting. Interactive television [10] is a research area that provides audiences with interactive functions to share experiences. However, most such systems only use verbal communication such as text chat functions or voice chat functions [8, 23, 5]. In contrast, our system focuses on emergent nonverbal communication via control devices. Moreover, although text chatting or voice chatting are only sharing experiences, users in our system influence broadcast content more directly. Consequently, users can also feel togetherness in becoming part of the program. There are studies related to camera control in live broadcasting. Saito et al. [19] proposes an interface that enables listeners to tell a performer to change camera effects such as zoom. However, in this case, camera control is executed by the performer. Our experiments are the first reported case of simultaneous camera control by many listeners in actual live broadcasting.

Great interest has arisen in semi-automated or fully automated camera control to create effective videos in Human Computer Interaction (HCI) research. As a semi-automated example, cameras are controlled automatically with prepared scenarios written by TVML [13], which is a markup language to express TV programs. FlySPEC [12] uses a PTZ camera on top of a panoramic camera to support camera controls by users. In the system, users overview an entire situation of a room using a panoramic camera and select regions to view more detail. Then, PTZ cameras are controlled automatically to fulfill the participants’ requests to the greatest degree possible. For fully automated cameras, some prior studies have examined lecture rooms [4, 18] and meeting rooms [17, 9]. Ranjan et al. [17] provide a system for automatic multi-camera control in meetings. The system uses a motion-tracking system and microphones, and control cameras automatically based on television production principles. Although these automatic camera control systems can provide benefits for efficient video recording/broadcasting, we consider that no communication will occur among participants. Meanwhile, we specifically examine the advantages of manual control. In our experiences, manual control plays a significant role in creating effective videos.
an important role for listeners to share experiences, produce new communication, togetherness, and share accomplishments.

Our work is also related to research in Computer Supported Collaborative Work (CSCW). Recent reports of studies investigating collaborative video production [6, 7, 22] present findings about human factors when multiple users collaborate in video creation, and also provide design implications for collaborative systems. Engström et al. [6] analyzed dance clubs’ VJ work and presented the SwarmCam prototype, which enables VJs to create collaborative videos with audiences. Engström et al. [7] also studied how live video can be produced as a group activity by amateurs using a tool called Instant Broadcasting System. From our paper’s point of view, these studies are interesting because they provide ethnographic insights regarding interaction between participants in a collaborative content-production setting. The differences between our experiments and these works are that we developed collaborative production for enhancing not only dramatic impact, but also nonverbal communication, and provide experimental results in a public broadcasting service with numerous simultaneous participants.

CONCLUSION
This paper presents the Listener-Cooperative Live Production System, which enhances communication and dramatic impact during live broadcasts. The LCLPS enables listeners to control cameras using methods such as switching other cameras or various effects (zoom, pan and rotate effects). Through one month of experiments with thousands of users, we confirmed the emergence of nonverbal communication and enhancement of dramatic impact. Especially, nonverbal communication occurred not only between the performer and the listeners, but also among the listeners themselves. Through these experiments, we anticipate the great possibilities of manual control for human communications, and its implications for the future design of interactive broadcasting systems.

ACKNOWLEDGEMENTS
We would like to thank the anonymous reviewers, and especially our shepherd William G. Griswold for their insightful comments. We also thank all of participants in our live broadcasting. This research was partly supported by National Institute of Information and Communications Technology (NICT).

REFERENCES