Intelligent Agents and Networked Buttons Improve Free-Improvised Ensemble Music-Making on Touch-Screens

Charles Martin
Research School of Computer Science, The Australian National University
Canberra, Australia
charles.martin@anu.edu.au

Henry Gardner
Research School of Computer Science, The Australian National University
Canberra, Australia
henry.gardner@anu.edu.au

Ben Swift
Research School of Computer Science, The Australian National University
Canberra, Australia
ben.swift@anu.edu.au

Michael Martin
Research School of Finance, Actuarial Studies & Applied Statistics, The Australian National University
Canberra, Australia
michael.martin@anu.edu.au

ABSTRACT

We present the results of two controlled studies of free-improvised ensemble music-making on touch-screens. In our system, updates to an interface of harmonically-selected pitches are broadcast to every touch-screen in response to either a performer pressing a GUI button, or to interventions from an intelligent agent. In our first study, analysis of survey results and performance data indicated significant effects of the button on performer preference, but of the agent on performance length. In the second follow-up study, a mixed-initiative interface, where the presence of the button was interlaced with agent interventions, was developed to leverage both approaches. Comparison of this mixed-initiative interface with the always-on button-plus-agent condition of the first study demonstrated significant preferences for the former. The different approaches were found to shape the creative interactions that take place. Overall, this research offers evidence that an intelligent agent and a networked GUI both improve aspects of improvised ensemble music-making.

Author Keywords

Creativity Support Tools; Collaborative Interaction; Mobile Music; Agent; Design

ACM Classification Keywords

H.5.5. Information Interfaces and Presentation (e.g. HCI): Sound and Music Computing—*Systems*

CHI 2016, May 7–12, 2016, San Jose, California, USA.
 The definitive Version of Record was published in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 2295–2306, San Jose, California, USA, May 7–12, 2016, http://dx.doi.org/10.1145/2858036.2858269

INTRODUCTION

Novel digital music instruments involving laptops [17], mobile phones [24], tablets [20], or DIY instruments [29] frequently have musical ensembles as a focus of their performance practice. In some of these settings, intelligent agents have been used to create virtual performers that collaborate [18] or virtual conductors that direct [7, 28] a group of performers. These two models of agent interaction can be thought of as representing two ends of a continuum of responsiveness: a virtual conductor can synchronise a group of musicians to a metronome or to a pre-planned musical score; a virtual performer can listen to an ensemble and respond to the other musicians continuously and in real time.

One model of agent interaction with a musical ensemble within this continuum might be termed an "intelligent listener". In contrast to virtual performers or virtual conductors, an intelligent listener agent intervenes only occasionally in a musical performance; the agent listens continuously and occasionally nudges human performers by making suggestions via re-configurations of their interfaces. Such a system was proposed for the first time by Martin et al. [20]. This agent executes a form of real-time HCI protocol analysis on the interaction data from ensemble improvisations on touch-screens, classifying the touch commands of each performer into a finite set of interaction states on-the-fly. It then uses the frequency of transitions between these interaction states to identify when ensemble members are exploring new gestural material in their musical improvisation, and responds by occasionally broadcasting interface changes across the ensemble.

Although there may appear to be a compelling rationale for the design of such an agent, and although in [20] there appears to have been a successful arts practice associated with systems which include such an agent, the question still re-

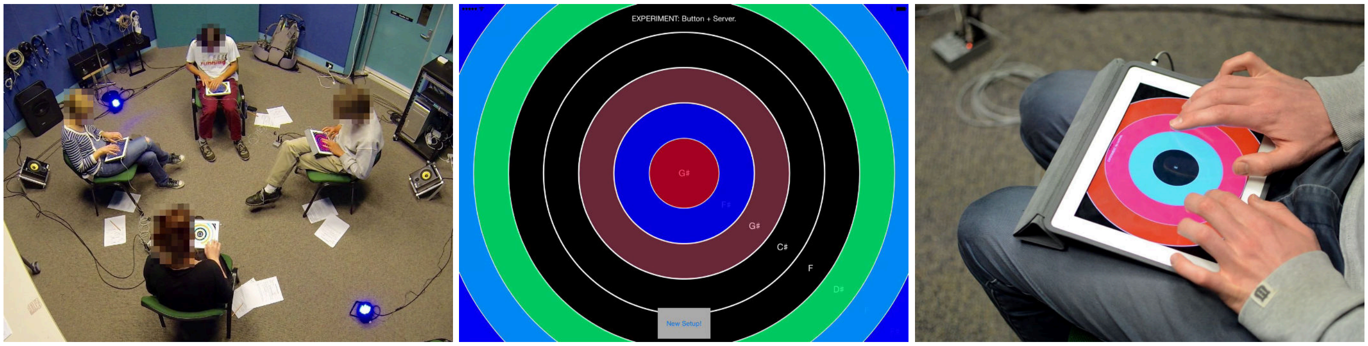

Figure 1. From left to right: study participants improvising with our iPad app during a study session; a screenshot of the app; and a performer playing the app. Tapping a ring plays a short pitched sound, swirling plays a continuous sound. Tapping the GUI button (at the bottom of the screen) updates the interface with new rings, notes and sounds on all performers' iPads.

mains whether an intelligent-listening agent really does *improve* free-improvised performance. On the negative side, the agent may broadcast UI updates to performers at inappropriate times, leading to a disruption of their musical focus and concentration. It could be that leaving users in control of the group's UI updates is preferable to agent interaction. In order to investigate this question we developed a simple digital musical instrument using Apple iPad devices that can interact with an intelligent-listener agent on a central server. Experimentally, the goal was to compare the effects of the agent (the "Agent-Control" factor), against a direct-interaction version of the same interface where a GUI button was accessible to all performers in the ensemble (the "GUI-Control" factor). These interfaces were tested individually, removed entirely (as a control condition), and combined, forming a 2×2 factorial design with four conditions: "Button", "Agent", "Nothing", and "Both".

The difficulties of systematically evaluating creativity support tools, particularly in performing arts practices, are well documented (see, for example [31]). We have followed the approach of other HCI evaluations of musical interfaces [5] in both surveying performers as well as analysing objective data from formally-structured ensemble performances. Although it has been noted that the use of measures such as time-to-completion of musical tasks [40] might not be appropriate for studies of creative interfaces [36], in our case time-to-completion was able to be used as an objective measure because of the conceptual association of a long improvisation with a deeply engaging one. In our first study, we collected data from 16 improvisational ensemble performances in an order-balanced experiment across four different ensembles, each with four performers.

Our results showed that while performers preferred GUI control over their interface, and rated these performances more highly in terms of technical proficiency, complexity, creativity and quality, they improvised for significantly longer under the agent conditions. These findings motivated the development of a refined version of the iPad interface utilising both the agent and GUI controls in a mixed-initiative design. The new design was shown in a follow up study to support better ensemble interactions and improvisation structure, as well as higher quality and more enjoyable performances. This research offers evidence that an intelligent listener agent and a networked GUI both improve aspects of improvised music-making.

Related Work

Free-improvised ensemble music, where there are no restrictions on style and no pre-determination of the music that will be played, has a significant history of enquiry focussed on how these open-ended collaborations lead to the spontaneous creation of structured music [35]. Although related to free-jazz, "non-idiomatic" free improvisation has few boundaries and is often a process of exploration and discovery of new performance methods and sounds [1]. Such improvisations can be modelled as a sequence of non-overlapping event clusters which each contain the exploration of a particular musical idea [27]. The group interactions which lead to the emergence of a coherent structure over the performance are often considered to be a marker of skill in improvisation [30]. Borgo emphasises the concept of "group mind" [2] in ensemble improvisation, where a state of creative flow among the group leads to seemingly effortless interaction. Mazzola [22] defines free jazz performances in terms of interactions where performers must "negotiate" each musical decision and engage in a "dynamic and sophisticated game" with their ensemble. In fact, this emphasis on collaborative creativity makes free-improvisation an ideal part of musical education and has been adopted in pedagogies such as Cahn's "Creative Music Making" [6].

For novel computer-music instruments such as those demonstrated at the New Interfaces for Musical Expression conference [26], ensemble improvisation is frequently a focus of the performance practice. The Daisyphone allows a group to collaboratively compose looping phrases using a shared compositional workspace [4]. Jordà's Reactable [14] allows a group of performers to collaboratively manipulate synthesis processes with tangible objects on a tabletop interface. Rosli et al.'s feedback ensemble [29] asks performers to create their own musical device and connect to an instrumented audio feedback network.

"Laptop Orchestras" [38] pioneered the practice of large computer music ensembles with identical hardware and software and a repertoire of compositions that frequently included improvisation [33] and collaborative networked interfaces [17]. "Mobile Music" [10] performance using mobiles devices such as smartphones and tablets takes advantage of the many sensors, touch screens and small form-factors of these devices [37]. Mobile music apps are now common in professional music production [13], in educational performance settings [44], as well as in research-focussed ensembles [42, 24].

There are also many commercially available apps for mobile music performance which, like Orphion [39], Ocarina [41], and Magic Fiddle [43], frequently emphasise novel multi-touch and sensor interactions that allow performers direct control over synthesis parameters and the performance of melody. However, the majority of commercial apps ignore possible ensemble contexts of their use and do not make use of network connectivity between performers' devices. The most common approach has been to synchronise rhythmic information between sequenced music processes on multiple devices such as in Korg's Wireless Sync-Start Technology [16], which is similar to synchronisation available using the MIDI standard's timing clock [23].

In this paper, we seek to compare networked, ensemble-focussed features in a mobile music app emphasising the collaborative interaction that distinguishes free-improvised music. The experimental features in our app use direct, indirect [32], and mixed-initiative [12] interaction models and our intelligent agent sits between those that participate in [18] and merely conduct [7] ensemble performances. Such ensemble tracking agents have been theorised since the 1990s [28] but, as examples and evaluations of their use are rare, the parameters for agent-performer interactions are not yet fully defined.

Evaluating new digital musical instruments can include both quantitative and qualitative approaches [36]. Evaluations can focus on audiences, composers, or designers, but the performer is usually considered to be the most important stakeholder in the use of new interfaces [25]. Qualitative methods have been used to identify skills emerging when performers engage with very simple electronic instruments [11], to understand users' evaluations of interactive systems using machine learning [9], and to identify collaborative learning on the Reactable [46]. Bryan-Kinns and Hamilton [5] used quantitative preference surveys as well as Likert-scaled ratings to allow performers to compare multiple similar interfaces. Ordinal rating scale questionnaires have also been used in creativity and musical performance research to assess multiple facets of solo, free-improvisations on piano keyboards [8] and jazz improvisations on wind instruments [34].

SYSTEM DESIGN

Our system consists of an app for Apple's iPad platform designed for ensemble performances, and server-based software that tracks and mediates such performances. This software has been discussed in detail in previous research [20], we will describe our system briefly below and focus on the elements investigated in our experiment.

App

Our iPad app can be seen in Figure 1 and consists of an annular interface for triggering sounds using simple percussive

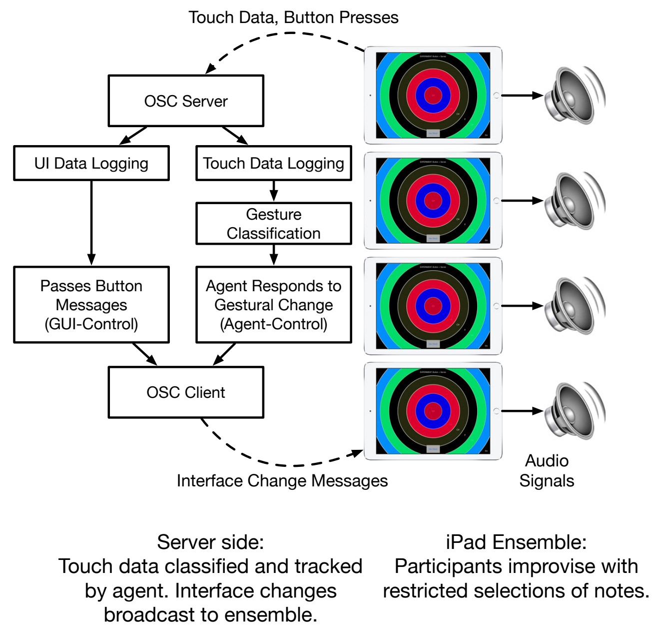

Figure 2. The architecture of our performance system. Performers use the iPad app to improvise with a selection of notes, while an agent on the server continually classifies and tracks the ensemble's gestural changes. Under the GUI-Control condition, a UI button is visible on the screen to update the interface with new notes and sounds, while under the Agent-Control condition these changes are made by the agent.

touch gestures similar to those previously shown to support a varied performance practice [19]. Each ring in the app's interface represents a single pitch, tapping the ring will trigger a short sound, while a moving touch will play a continuous sound. Volume for taps is proportional to touch-radius, and for continuous sounds is proportional to the velocity of the moving touch. In this way, the app allows direct control over basic musical concepts of pitch and rhythm in a percussive interface similar to other instruments well-suited to improvisation by beginners.

The app displays a limited number of rings on each performer's screen. All of these notes are taken from the same musical scale, so that the ensemble can play concordant notes, but are generated independently so that performers have different sets of pitches. Performances with our app are segmented in time by moments where the performers' interfaces are simultaneously updated with a new set (and potentially a different number) of notes which may be from a different scale or have a different timbre. Individual players are free to explore the melodic space afforded by the randomly-chosen pitches displayed on the screen, but the harmonic and timbral progression of the improvisation is common to the whole group. This design was influenced by contemporary improvisation practice and was the most preferred interface in a previous study comparing three different apps in free-improvisation [21].

In our app, interface updates are triggered by two different mechanisms, roughly corresponding to direct user interface and indirect agent interaction models in HCI [32]. In the direct case, a GUI button is present in the touch interface for each performer (as it is in the screen capture in Figure 1). When one player activates this button, the app interface is updated for all performers in the group. The presence of this GUI element is a factor in our experiments (GUI-Control). In the indirect case, our intelligent-listener agent, modelled on the one implemented in previous research [20], tracks the performance of each individual player in the group and occasionally intervenes by updating all of the interfaces simultaneously.

Agent

The agent triggers an interface update when it calculates that the amount of gestural change in the ensemble has exceeded a predetermined threshold. These moments can correspond to natural "segments" of ensemble improvisations [20], and so are appropriate times to intervene in the performance. The agent performs gestural classification on each performer's touch data, once per second, to produce a sequence of gestural states. Transitions between these states are summed to calculate a transition matrix (TM) for the ensemble, calculated over a sliding 15-second window of touch data. A metric on TMs is used to determine the current rate of gestural change within the group. A sharp increase in this measure can indicate that the ensemble is shifting to a new musical "idea". In response, the agent sends a "new-idea" message which updates the interface on the iPad apps with a new set of notes or sounds.

This response is posed as "reward" to the ensemble for exploratory behaviour. Intuitively, if a group has changed musical idea, they may appreciate new sounds to work with. This interaction results in a musical experience similar to "structured improvisation" or "game pieces" except with an ensemble-tracking agent director, rather than a human. This agent, and the new-idea reward interaction, has been previously compared with a design that generated random interface updates in a study with three experienced iPad performers [21]. That study suggested that the ensemble-tracking agent was more positively received than the generative agent.

In the present system, the agent runs as a server application on a laptop on the local network, to which the iPad apps connect automatically at the start of each performance. The architecture of this system is shown in Figure 2. The agent also logs all the touch data during the improvisations, as well as interactions with the button interface and new-idea messages. The presence of the agent in the performance was a factor (Agent-Control) in our controlled experiments. The nature of the interface changes was the same under both the GUI-Control and Agent-Control factors and these systems could be switched on or off at the start of each performance. This was by design, to facilitate the two-factor experiment described in the next section where all combinations of Agent-Control and GUI-Control are compared: Button, Agent, Both and Nothing.

FIRST EXPERIMENT

The first experiment took place over four 90-minute sessions spread across two weeks. The venue was a recording studio equipped with a quadraphonic speaker system and a control room for controlling and recording the sessions and monitoring the performers. This studio setting allowed multi-track audio and video recordings to be made of the whole session.

Group	Perf. 1	Perf. 2	Perf. 3	Perf. 4
1	Agent	Both	Nothing	Button
2	Nothing	Agent	Button	Both
3	Button	Nothing	Both	Agent
4	Both	Button	Agent	Nothing
5	Both	Fade	Both	Fade
6	Fade	Both	Fade	Both

Table 1. Schedule for the experiment showing the counter-balanced ordering of conditions. Each group participated in one session including an induction, the four performances, and a debrief interview. Sessions 5 and 6 were subsequently added as a follow-up study with a revised interface.

In each experimental session, all of the participants played in four separate improvisational group performances, each with a different interface update regime: agent-controlled interface changes (Agent), button-controlled interface changes (Button), both agent- and button-controlled changes, (Both), and no changes—a static interface for the whole performance (Nothing). These conditions correspond to the combinations of the two top-level factors, GUI-Control and Agent-Control. The groups were exposed to these configurations in counter-balanced orders (shown in Table 1) chosen by taking a random order and applying Bradley's balancing procedure [3]. This was done to counteract any immediate sequential effects as well as the ensemble learning effects that are known to be exhibited by new groups of improvisers [6].

Participants

Participants were recruited for this study through posters and presentations at a university music school. 25 respondents were asked to select available times through a web-interface and 16 participants (7 female, 9 male) were invited to attend one of four sessions based on availability and order of response. 14 of these participants were university students with 11 of those studying music. Three of the participants had previously performed with our iPad apps and were distributed into different ensembles.

Procedure

Each session began with a 20 minute induction during which the participants were introduced to the app and the experimental procedure and were given the chance to try each of the experimental conditions. The participants also filled out an entry survey to record demographic information and their experience levels in electronic music and improvisation. In each of the four performances, the experimenter remotely activated all of the iPad apps to indicate the start of one of the sessions and monitored the participants from the control room.

To give each condition a fair trial and to ensure the sessions did not run over time, the participants were asked to improvise for at least seven minutes but no longer than 11 minutes in each performance. After seven minutes, the performers were free to stop when they wished and the performance was considered to be finished when all performers had stopped playing. If the performance ran longer than 11 minutes, the performance was halted by turning off the speaker system. The seven-minute mark was indicated to the participants by two remote controlled stage lights positioned in the studio. These lights faded to green to indicate the start of the performance and faded to blue at seven minutes to indicate that the participants could finish whenever they wanted. Participants were not aware that the time-to-completion was being used as a metric for this study.

Questionnaires

At the end of each improvisation, the participants were asked to fill out a survey of twenty-four Likert-style questions to record their ratings of various aspects of the performance on nine-point scales. The questionnaire was divided into five sections to evaluate aspects of the improvisations, similar to those defined by Eisenberg and Thompson [8]. The sections were: technical proficiency, musical interaction, musical structure, creativity, and performance quality:

Technical Proficiency

1. How much did you focus on particular touch gestures in that performance?
2. How much did you explore a range of touch gestures?
3. How would you rate your technical proficiency using the app in that performance?
4. How much did the app impede your performance?
5. How much did the app enhance your performance?

Musical Interaction

6. How much did you interact musically with the other performers?
7. How much did the other performers interact musically with you?
8. How well were you able to respond to the other musicians' actions?
9. How well were you able to respond to the app?
10. How would you rate the overall level of musical interaction among the ensemble?

Musical Structure

11. How would you rate the complexity of that performance?
12. How appropriate was the length of that performance?
13. How would you rate the app's influence on your own playing?
14. How would you rate the app's influence on the ensemble performance?
15. How would you rate the overall musical structure of that performance?

Creativity

16. How much did you present new musical ideas to the others in the ensemble?
17. How much did you take on and develop musical ideas first presented by the others in the ensemble?
18. How would you rate your personal creativity in that performance?
19. How would you rate the other performers' creativity in that performance?
20. How would you rate the overall creativity in that performance?

Performance Quality

21. How would you rate the quality of your contribution to that performance?
22. How would you rate the quality of the other performers' contribution in that performance?
23. How would you rate the overall quality of that performance?
24. How enjoyable or unpleasant was that performance?

At the end of the entire session, participants filled out another survey in which they were asked to choose which condition they most preferred over seven aspects of the performances. Finally, their responses to this preference survey were used as a starting point for an unstructured interview/discussion with each ensemble which lasted about 10 minutes.

RESULTS

The data gathered in this study included coded survey results, performance protocols that include records of touch interactions, button presses, and agent interactions, as well as recordings of the performances and post-session interviews.

Survey Data

The survey data was analysed using an Aligned Rank Transform (ART) procedure followed by a two-way mixed-effects ANOVA to assess significance. As the survey data is ordinal, rather than interval-scaled, the assumptions of the classical within-groups factorial ANOVA procedure may be violated. The ART procedure has been recommended as the most appropriate procedure for factorial HCI studies with ordinal dependent variables, such as Likert-scaled responses [45]. Analysis was performed in `R` using the `ARTool` v0.9.5 [15] package. Post-hoc analysis was performed using Bonferroni-corrected paired t-tests.

The ANOVA procedures showed that the presence of the UI button in the touch-screen interface (GUI-Control) had a significant main effect on nine questions in the survey. A detailed overview of the ANOVA results is shown in Table 2 and the distribution of responses for these questions is shown in Figure 3. When the button was present in the interface, performers reported higher personal proficiency, that the app had a more positive influence on their personal performance, and that they presented more new ideas to the group. They rated the complexity of the performances as higher and also rated their personal creativity, the creativity of others in the group, and the overall creativity more highly. Finally, they rated the quality of their personal contribution and the overall quality of the performances as better when the button was present. The survey data did not show any significant main effect due to the Agent-Control factor.

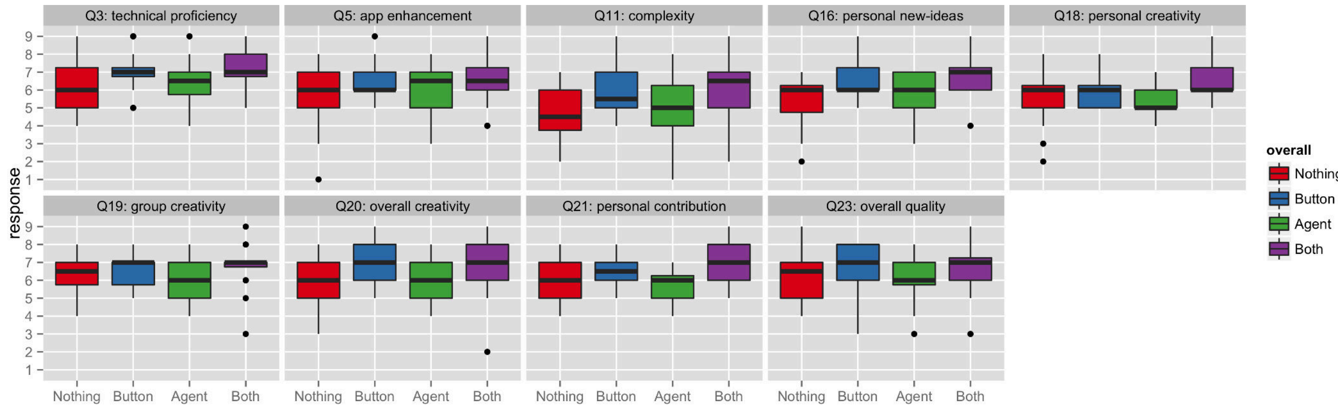

Figure 3. Distribution of results for questions where a significant main effect for GUI-Control was found. "Nothing" tended to attract the lowest ratings, and "Both" the highest.

Overall, these results suggest that the performers felt their proficiency and creativity were enhanced by the presence of the button in the interface (whether in the Button or in the Both conditions). This may have been due to the extra personal performance options, or the feeling of control that they had when interacting with the button. The main effect for GUI-Control was in spite of the potential disruption caused by one performer pressing the button and causing everyone's interface to change; as discussed below, at times there were a very large number of these button presses during a performance.

Post-hoc testing

Post-hoc Bonferroni-corrected paired t-tests confirmed that Button-only performances were rated as more complex (Q11) than Agent-only performances ($p = 0.05$). Performers thought that they presented *more new ideas* to the group (Q16) in Both performances than in Agent-only performances ($p = 0.04$) and Nothing performances ($p = 0.06$). They regarded their *personal creativity* (Q18) to be significantly higher in Both performances than Agent-only ($p = 0.02$) and they also perceived their *personal contribution* (Q21) to be better in Both than in Agent-only performances ($p = 0.05$).

Performer Preferences

After completing performances with each of the four interfaces, the performers were asked to pick one performance condition that ranked highest across seven aspects of the performances. The results shown in Figure 4 convincingly demonstrate that the Button-only condition was the preferred choice across all questions except for "most challenging" and "most creative". Chi-squared tests were used to determine how significantly the distribution of responses had deviated from chance. The questions about which condition made it easiest to perform ($\chi^2(3, N = 16) = 9, p < 0.05$) and which condition resulted in the best musical interaction ($\chi^2(3, N = 16) = 12, p < 0.01$) had significant results where performers favoured the Button-only condition. Performers rated the Nothing condition most frequently as the most challenging to use, and this question also showed a significant deviation ($\chi^2(3, N = 16) = 14, p < 0.01$).

Question	F	Significance
3. How would you rate your technical proficiency using the app in that performance?	8.288	$p < 0.01$
5. How much did the app enhance your performance?	4.201	$p < 0.05$
11. How would you rate the complexity of that performance?	11.977	$p < 0.01$
16. How much did you present new musical ideas to the others in the ensemble?	9.698	$p < 0.01$
18. How would you rate your personal creativity in that performance?	10.646	$p < 0.01$
19. How would you rate the other performers' creativity in that performance?	5.113	$p < 0.05$
20. How would you rate the overall creativity in that performance?	8.684	$p < 0.01$
21. How would you rate the quality of your contribution to that performance?	11.324	$p < 0.01$
23. How would you rate the overall quality of that performance?	5.387	$p < 0.05$

Table 2. Questions showing a significant main effect for the presence of the button in the instrument interface (GUI-Control). The performers felt their proficiency, creativity and contribution to performances was enhanced when the button was present.

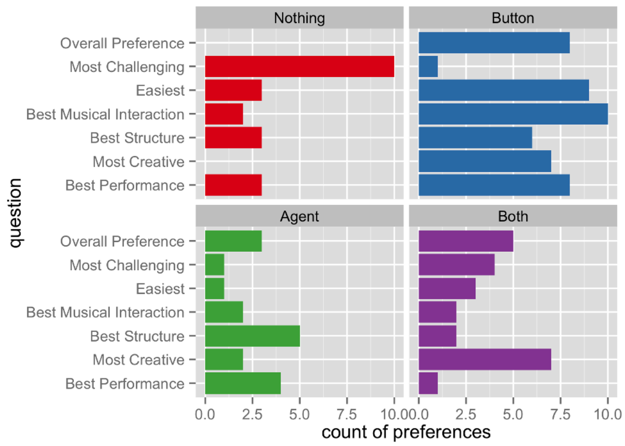

Figure 4. Distribution of the performers' overall preferences from the exit survey. The preference for the button-only condition is apparent, especially when performers considered the easiest interface and musical interaction. The Nothing condition was considered to be the most challenging.

Performance Length

As previously discussed, performances in this study had a minimum duration of seven minutes and a maximum duration of eleven minutes. The precise length of the performance was defined as the time from the first sound (green lights) to the time when all performers had ceased playing. We can also define an alternate performance length for each performer in the group, where the start is given by the first performer's touch, and the end for each performer is given by their own final touch. The distribution of these performer-lengths by experimental condition is given in Figure 5.

By treating this performer-length as a dependent variable and modelling using a two-way within-groups ANOVA procedure, we found that there was a highly significant main effect due to Agent-Control in the lengths of performances ($F(1,15) = 85.4, p < 0.001$). Post-hoc Bonferroni-corrected paired t-tests showed that the Agent, Button, and Both performances were longer than the Nothing performances ($p < 0.01$), but that Agent-only performances were not significantly longer than Button-only performances.

This result is at odds with the survey results that showed a significant main-effect for GUI-Control. It may be that even though performers stated that they preferred having the button present in the interface, in Agent and Both performances they ended up more deeply engaged in the improvisation, leading to longer performances. Observations from video recordings of the sessions (as seen in the video figure) showed that performers appeared to be deeply engaged in every single performance in this experiment and they were often unaware of the relative length of performances.

New-Idea Messages

Under the Agent and Both conditions, the iPad app interfaces would be updated with new notes and sounds in response to new-idea messages sent from our agent software. However, the agent was still actively classifying new-idea instances in the Button and Nothing conditions, even though they had no effect on the interface. This means that the number of new-idea messages can be used as a dependent variable that measures how much the performers interacted with each other in the way that the agent was designed to measure.

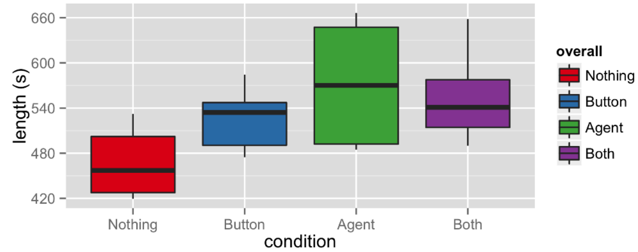

Figure 5. Performer-distinguished improvisation lengths by overall interface condition. The effect of Agent-Control was highly significant with longer performances under Agent and Both conditions.

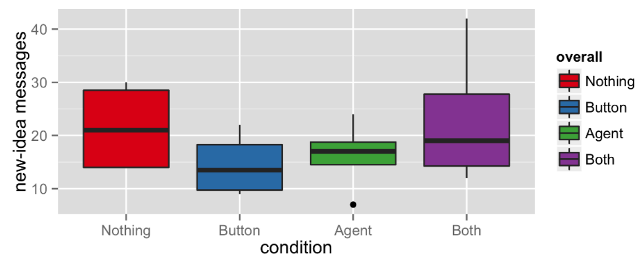

Figure 6. In all performances, the agent tracked touch-gestures of the ensemble and attempted to identify "new-idea" moments, but these were only used by the app under the Agent and Both conditions. The count of new-ideas was highest in the Nothing and Both conditions, with a significant interaction between GUI-Control and Agent-Control effects.

Figure 6 shows the number of new-idea messages sent in performances under each experimental condition. While one might expect more new-ideas to be produced when the interface was actively responding to the new-idea messages, curiously this was not always the case. The Nothing and Both conditions seem to have higher numbers of new-idea messages than either the Agent or Button only conditions. A two-way within-groups ANOVA procedure shows a significant interaction effect ($F(1,3) = 11.7, p < 0.05$) between the GUI-Control and Agent-Control factors and no significance for either main effect. It may be that under the Nothing condition, where there were no interface updates, the performers focussed more on their ensemble interactions. Conversely, in the Button-only condition, the button may have distracted from performing with varied touch gestures. The Both condition may have allowed the most natural ensemble interaction and gestural exploration, resulting in more new-idea messages than Agent-only performances.

Button Presses

During Button and Both performances, the performers were able to trigger iPad app interface updates at any time during the performances. Interaction logs recorded during the performance allowed us to see which performers pressed the button at which times. The median number of button presses in one performance per performer was 3.5, however the maximum was 457! Of the sixteen participants, three had pressed the button more than 100 times in a single performance. Observation of the performance recordings shows that these "button maniacs" had used the button not to find new notes in the iPad app interface or to segment the performance harmonically, but to create a unique sound where all performers' interfaces rapidly changed synthesis parameters and pitch.

"Button mania" was not anticipated in the design of the app, which had no mechanism to limit the frequency of button presses. For the participants, and the button maniacs in particular, the button may have represented not just one of two ways to segment performances, but a new musical device. In Button and Both performances, this device could be used to create a unique sound that was not available in Nothing and Agent performances. This interpretation may explain why performers felt more technically accomplished when the button was present—they were not only in control of interface updates, but able to create an entirely new sound.

FOLLOW-UP STUDY

The results of our first study suggested that both the GUI-Control and Agent-Control factors had an effect on different aspects of the performance. Direct group interactions, via the button, affected the performers' perceptions of the performance, while the agent may have helped them to achieve a deeper engagement resulting in longer performances. We then conducted a follow up study to compare the Both condition with a new mixed-initiative interface that interlaced the dynamics of the Agent-Control and GUI-Control conditions. Under the new "Fade" condition, the button was normally not accessible to the performers. When the agent sent a new-idea trigger, the button was faded into view and became accessible. If a performer tapped the button, it would update the interfaces as normal and then fade away. If performers did not tap the button, it would stay visible and they could use it when they wished. In this way, the Fade condition prevented the button from being activated repeatedly, while allowing the performers to have the final say over interface changes.

The follow-up study was conducted with an identical procedure to the first study. Two groups of four participants took part, seven of the participants had been in the first study and the last one had been a performer in earlier performances and rehearsals with the same app. The two groups performed two replications of the two conditions for a total of four improvisations in balanced order (see sessions 5 and 6 in Table 1). The same survey was used after each performance and a preference survey with the same questions was used at the end of the session. As before, each session concluded with a group interview. It turned out that two of the "button maniacs" from the first study were in one of these groups.

Again, an ART and one-way mixed effects ANOVA procedure was performed on the results of the final two performances in the follow up study. Table 3 shows questions with significant main effects for the change in interface and Figure 7 shows the distribution of survey responses for these questions. These results showed that the performers perceived an improvement in their ensemble interaction with the new Fade interface, as well as improvements in the musical structure and overall quality and enjoyment of the improvisations.

Question	F	Significance
6. How much did you interact musically with the other performers?	26.82	$p < 0.01$
7. How much did the other performers interact musically with you?	8.8407	$p < 0.05$
10. How would you rate the overall level of musical interaction among the ensemble?	10.924	$p < 0.05$
15. How would you rate the overall musical structure of that performance?	9.7391	$p < 0.05$
23. How would you rate the overall quality of that performance?	8.7576	$p < 0.05$
24. How enjoyable or unpleasant was that performance?	5.8973	$p < 0.05$

Table 3. ART ANOVA results for questions showing significance for the effect of the different interfaces in the follow up study. A one-way ART and mixed-effects ANOVA was performed on the results of the final two performances by each group.

The results of the preference survey shown in Figure 8 show the participants roughly split in preference for the two interfaces, with none of the questions leading to significant differences related to the conditions. The Both condition was seen as an easier interface than the Fade condition, where particular types of ensemble interaction were required in order for the agent to present the button to the performers.

While performance length had been an important discriminating dependent variable in the original study, the two conditions in the follow up had no significant effect on the length of improvisations, despite the Fade condition giving rise to the longest performance in each session. In the post-performance interviews, the performers reported that they were not aware of how long the improvisations had been. It is likely that, at least for these two groups, both conditions supported an optimal length of improvisation.

DISCUSSION

Over both studies, both in the qualitative survey responses and in the interviews, the participants expressed a great deal of enthusiasm for the experience. This once again demonstrates the ability of a simple iPad interface, judiciously augmented with some ensemble-wide interface dynamics, to give rise to real musical interaction. As discussed in the Related Work section, evaluating interfaces for open-ended collaborative creativity is a difficult task. It is encouraging to note that our attempts with these studies to perform systematic, controlled evaluations of different aspects of the interface did not lead to a lifeless, unmusical experience for the performers.

The results of our two studies show that no particular interface we tried has a clear-cut advantage over the others. The

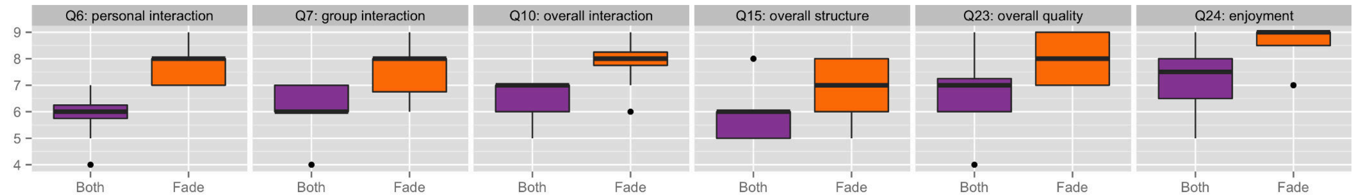

Figure 7. Distribution of results for the follow-up study where a significant effect of the interface condition was found. In all of these questions, Fade was rated more highly than Both.

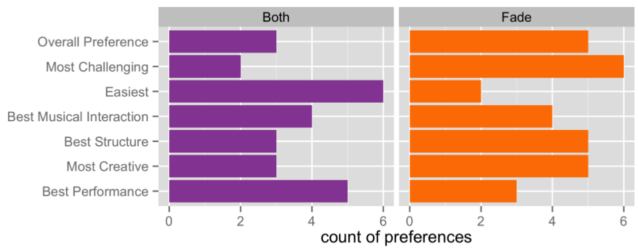

Figure 8. Preference surveys from the follow up study. The performers were roughly split on most questions, however Both was seen as easier and Fade as more challenging.

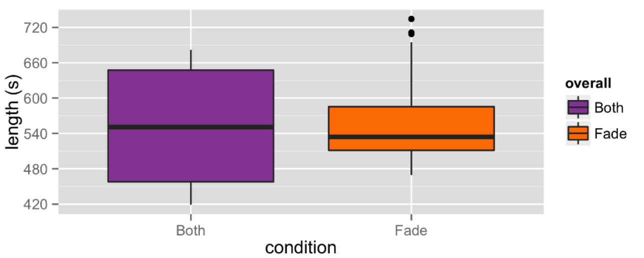

Figure 9. Length of performances with Both and Fade conditions in the follow up study. Although the longest performance was with Fade, there is not a significant effect.

only interface that was rejected by the participants was the Nothing condition, even though some commented that the limited interface had forced them to "make do" and that they enjoyed trying to find creative ways to use their setup.

It was notable that the original and follow-up studies showed significant effects in different parts of our survey. In the original study, the GUI button interface resulted in significant effects for technical proficiency, the participants' personal performance, improvisation complexity, and creativity. In the follow-up sessions, the revised, mixed-initiative interface achieved significant effects for group interaction, performance structure, and enjoyment. Both studies showed a significant effect on the question about overall performance quality.

The performers generally used the button relatively sparingly throughout performances, however the two "button maniacs" among them used the button far more than was anticipated in the design. Once they started this behaviour, they effectively dominated the performance, as the other performers' interfaces behaved in the same way. While this behaviour was chaotic, it also added a unique sonic aspect to some of the performances and demonstrates the creative exploration at play in the sessions.

In the two follow-up sessions most of the performers picked the Fade condition as their favourite. However they suggested in their interviews that both interfaces, Fade and Both, could be interesting and useful in performance situations. While the constant presence of the button allowed the finest direct control over the performance, the agent interaction in the Fade condition encouraged the performers to make the most of their current interface without rushing to new melodic options.

Overall, 24 performances were recorded in the two studies encompassing a wide variety of musical explorations (some of which can be seen in the video figure) despite all using the same minimal app. Some groups tended towards the ambient and arrhythmic, while others developed a strong pulse and explored metric ideas and melodic motifs. In the interviews, the musicians told us that they had enjoyed the improvisations, the collaborative interactions that took place, and using the iPad app. This view is confirmed in the survey results which were mostly in the upper half of the rating scale, even for the Nothing condition. The positive reception to our app, and wide range of stylistic possibilities the participants discovered, suggests that future artistic performances with any of the conditions could be rewarding.

Since the first study was conducted, the four interface variants—Button, Agent, Both and Nothing—have been used in an educational context in a high-school music class as the stimulus for engaging with different styles of free-improvisation. The ability of our app to operate under different interaction paradigms while keeping the same simple performance interface makes it very useful for such a setting and future work will examine its educational utility.

CONCLUSION

The results of our two studies lead us to conclude that networked interfaces with direct manipulation of the group performance space, via a GUI button, and indirect manipulation via an intelligent agent, give rise to different styles of interaction with touch-screen musical instruments and improve free improvisation in a musical ensemble. We identified precisely which aspects of performance were impacted by the experimental conditions. In our first study we found that performers rated direct interactions more highly when considering their *technical proficiency* and the *complexity*, *creativity*, and *overall quality* of performances. However, the indirect, agent interaction, resulted in significantly *longer* improvisations, which could indicate a higher level of creative engagement with the app, and the ensemble. This was a particularly notable, and surprising, result, as the length of performances was not easily perceived by the performers, but detected with a quantitative measure. The simplest condition, where no changes occurred in the interface, was broadly rejected by the participants while the performances using both button and agent were broadly supported (together with the button-only condition).

The results of our first study led us to design a new, mixed-initiative interface that interlaced the behaviour of the button with the agent. In this new condition, the agent was able to enable a GUI button at appropriate moments in the performance. Performances with this new condition received significantly higher ratings for *structure*, *group interaction*, *enjoyment* and *overall quality*, than the simpler Both condition. However, the performers acknowledged that both of these interfaces had the capacity to direct the improvisation by encouraging particular ways of playing.

Overall we can conclude that exploiting the networked capabilities of mobile touch-screen devices in apps for musical performance can have significant effects on how these performances are perceived and carried out, and that they really can improve group interaction, creativity, and length of engagement in improvisations. Our methodology of controlled, quantitative studies of free-improvisation, including measuring session length, has been shown to assist with digital musical instrument design leading to a refined interface. This research suggests that further designs for intelligent listener agents, networked user interfaces, and instruments that emphasise particular ensemble interactions could be useful in musical training, recreational music making and on the concert stage.

REFERENCES

1. Derek Bailey. 1993. *Improvisation: It's Nature and Practice in Music*. Da Capo, Cambridge, MA.

2. David Borgo. 2005. *Sync or Swarm: Improvising Music in a Complex Age*. Continuum, New York, NY, USA.

3. James V. Bradley. 1958. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528. DOI: http://dx.doi.org/10.1080/01621459.1958.10501456

4. Nick Bryan-Kinns. 2004. Daisyphone: The Design and Impact of a Novel Environment for Remote Group Music Improvisation. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '04)*. ACM, New York, NY, USA, 135–144. DOI: http://dx.doi.org/10.1145/1013115.1013135

5. Nick Bryan-Kinns and Fraser Hamilton. 2012. Identifying Mutual Engagement. *Behaviour & Information Technology* 31, 2 (2012), 101–125. DOI: http://dx.doi.org/10.1080/01449290903377103

6. William L. Cahn. 2005. *Creative Music Making*. Routledge, New York, NY.

7. Arne Eigenfeldt and Ajay Kapur. 2008. An Agent-based System for Robotic Musical Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Genoa, Italy, 144–149. http://www.nime.org/proceedings/2008/nime2008_144.pdf

8. Jacob Eisenberg and William Forde Thompson. 2003. A Matter of Taste: Evaluating Improvised Music. *Creativity Research Journal* 15, 2-3 (2003), 287–296. DOI: http://dx.doi.org/10.1080/10400419.2003.9651421

9. Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. 2011. Human Model Evaluation in Interactive Supervised Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 147–156. DOI: http://dx.doi.org/10.1145/1978942.1978965

10. Lalya Gaye, Lars Erik Holmquist, Frauke Behrendt, and Atau Tanaka. 2006. Mobile Music Technology: Report on an Emerging Community. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression (NIME '06)*. IRCAM & Centre Pompidou, Paris, France, 22–25. http://dl.acm.org/citation.cfm?id=1142215.1142219

11. Michael Gurevich, Adnan Marquez-Borbon, and Paul Stapleton. 2012. Playing with Constraints: Stylistic Variation with a Simple Electronic Instrument. *Computer Music Journal* 36, 1 (2012), 23–41. DOI: http://dx.doi.org/10.1162/COMJ_a_00103

12. Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 159–166. DOI: http://dx.doi.org/10.1145/302979.303030

13. Mark Jenkins. 2012. *iPad Music: In the Studio and on Stage*. Taylor & Francis, Burlington, MA, USA.

14. Sergi Jordà. 2010. The Reactable: Tangible and Tabletop Music Performance. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 2989–2994. DOI: http://dx.doi.org/10.1145/1753846.1753903

15. Matthew Kay and Jacob O. Wobbrock. 2015. ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. R package version 0.9.5. Available online at https://github.com/mjskay/ARTool. (2015).

16. Korg Inc. 2011. Korg WIST SDK: Wireless Sync-Start Technology for iOS Music Apps. https://code.google.com/p/korg-wist-sdk/. (2011).

17. Sang Won Lee, Jason Freeman, and Andrew Collela. 2012. Real-Time Music Notation, Collaborative Improvisation, and Laptop Ensembles. In *Proceedings*

of the International Conference on New Interfaces for Musical Expression. University of Michigan, Ann Arbor, Michigan. http://www.nime.org/proceedings/2012/nime2012_62.pdf

18. Aengus Martin, Craig T. Jin, and Oliver Bown. 2011. A Toolkit for Designing Interactive Musical Agents. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference (OzCHI '11)*. ACM, New York, NY, USA, 194–197. DOI: http://dx.doi.org/10.1145/2071536.2071567

19. Charles Martin, Henry Gardner, and Ben Swift. 2014. Exploring Percussive Gesture on iPads with Ensemble Metatone. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1025–1028. DOI: http://dx.doi.org/10.1145/2556288.2557226

20. Charles Martin, Henry Gardner, and Ben Swift. 2015a. Tracking Ensemble Performance on Touch-Screens with Gesture Classification and Transition Matrices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Edgar Berdahl and Jesse Allison (Eds.). Louisiana State University, Baton Rouge, Louisiana, USA, 359–364. http://www.nime.org/proceedings/2015/nime2015_242.pdf

21. Charles Martin, Henry Gardner, Ben Swift, and Michael Martin. 2015b. Music of 18 Performances: Evaluating Apps and Agents with Free Improvisation. In *Proceedings of the 2015 Conference of the Australasian Computer Music Association*, Ian Stevenson (Ed.). Australasian Computer Music Association, Fitzroy, Australia, 85–94. http://hdl.handle.net/1885/95205

22. Guerino Mazzola and Paul B. Cherlin. 2009. *Flow, Gesture, and Spaces in Free Jazz*. Springer, Berlin, Germany.

23. MIDI Manufacturers Association. 1996. *The Complete MIDI 1.0 Detailed Specification: Incorporating All Recommended Practices*. MIDI Manufacturers Association, La Habra, California, USA.

24. Jieun Oh, Jorge Herrera, Nicholas J. Bryan, Luke Dahl, and Ge Wang. 2010. Evolving the Mobile Phone Orchestra. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. University of Technology Sydney, Sydney, Australia, 82–87. http://www.nime.org/proceedings/2010/nime2010_082.pdf

25. Sile O'Modhrain. 2011. A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal* 35, 1 (2011), 28–42. DOI: http://dx.doi.org/10.1162/COMJ_a_00038

26. Ivan Poupyrev, Michael J. Lyons, Sidney Fels, and Tina Blaine (Bean). 2001. New Interfaces for Musical Expression. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*. ACM, New York, NY, USA, 491–492. DOI: http://dx.doi.org/10.1145/634067.634348

27. Jeff Pressing. 1988. Improvisation: Methods and Models. In *Generative Processes in Music*, John Sloboda (Ed.). Oxford University Press, Oxford, UK. DOI: http://dx.doi.org/10.1093/acprof:oso/9780198508465.003.0007

28. Jeff Pressing. 1990. Cybernetic Issues in Interactive Performance Systems. *Computer Music Journal* 14, 1 (1990), pp. 12–25. DOI: http://dx.doi.org/10.2307/3680113

29. Muhammad Hafiz Wan Rosli, Karl Yerkes, Matthew Wright, Timothy Wood, Hannah Wolfe, Charlie Roberts, Anis Haron, and Fernando Rincon Estrada. 2015. Ensemble Feedback Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Edgar Berdahl and Jesse Allison (Eds.). Louisiana State University, Baton Rouge, Louisiana, USA, 144–149. http://www.nime.org/proceedings/2015/nime2015_329.pdf

30. R. Keith Sawyer. 2006. Group Creativity: Musical Performance and Collaboration. *Psychology of Music* 34, 2 (2006), 148–165. DOI: http://dx.doi.org/10.1177/0305735606061850

31. Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. DOI: http://dx.doi.org/10.1145/1323688.1323689

32. Ben Shneiderman and Pattie Maes. 1997. Direct Manipulation vs. Interface Agents. *Interactions* 4, 6 (Nov. 1997), 42–61. DOI: http://dx.doi.org/10.1145/267505.267514

33. Scott Smallwood, Dan Trueman, Perry R. Cook, and Ge Wang. 2008. Composing for Laptop Orchestra. *Computer Music Journal* 32, 1 (2008), pp. 9–25. DOI: http://dx.doi.org/10.1162/comj.2008.32.1.9

34. Derek T. Smith. 2009. Development and Validation of a Rating Scale for Wind Jazz Improvisation Performance. *Journal of Research in Music Education* 57, 3 (2009), 217–235. DOI: http://dx.doi.org/10.1177/0022429409343549

35. Harald Stenström. 2009. *Free Ensemble Improvisation*. Number 13 in ArtMonitor. Konstnärliga fakultetskansliet, University of Gothenburg, Gothenburg, Sweden. http://hdl.handle.net/2077/20293

36. Dan Stowell, Andrew Robertson, Nick Bryan-Kinns, and Mark D. Plumbley. 2009. Evaluation of Live Human-computer Music-Making: Quantitative and Qualitative Approaches. *Int. J. Hum.-Comput. Stud.* 67, 11 (Nov. 2009), 960–975. DOI: http://dx.doi.org/10.1016/j.ijhcs.2009.05.007

37. Atau Tanaka. 2010. Mapping Out Instruments, Affordances, and Mobiles. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Kirsty Beilharz, Andrew Johnston, Sam

Ferguson, and Amy Yi-Chun Chen (Eds.). University of Technology Sydney, Sydney, Australia, 88–93. http://www.nime.org/proceedings/2010/nime2010_088.pdf

38. Dan Trueman. 2007. Why a Laptop Orchestra? *Organised Sound* 12 (2007), 171–179. Issue 2. DOI: http://dx.doi.org/10.1017/S135577180700180X

39. Sebastian Trump and Jamie Bullock. 2014. Orphion: A Gestural Multi-Touch Instrument for the iPad. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Goldsmiths, University of London, London, United Kingdom, 159–162. http://www.nime.org/proceedings/2014/nime2014_277.pdf

40. Marcelo Mortensen Wanderley and Nicola Orio. 2002. Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal* 26, 3 (2002), 62–76. DOI: http://dx.doi.org/10.1162/014892602320582981

41. Ge Wang. 2014. Ocarina: Designing the iPhone's Magic Flute. *Computer Music Journal* 38, 2 (2014), 8–21. DOI: http://dx.doi.org/10.1162/COMJ_a_00236

42. Ge Wang, Georg Essl, and Henri Penttinen. 2008. Do Mobile Phones Dream of Electric Orchestras?. In *Routes/Roots: Proceedings of the International Computer Music Conference*. MPublishing, University of Michigan Library, Ann Arbor, MI. http://hdl.handle.net/2027/spo.bbp2372.2008.039

43. Ge Wang, Jieun Oh, and Tom Lieber. 2011. Designing for the iPad : Magic Fiddle. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Oslo, Norway, 197–202. http://www.nime.org/proceedings/2011/nime2011_197.pdf

44. David A. Williams. 2014. Another Perspective: The iPad Is a REAL Musical Instrument. *Music Educators Journal* 101, 1 (2014), 93–98. DOI: http://dx.doi.org/10.1177/0027432114540476

45. Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143–146. DOI: http://dx.doi.org/10.1145/1978942.1978963

46. Anna Xambó, Eva Hornecker, Paul Marshall, Sergi Jordà, Chris Dobbyn, and Robin Laney. 2013. Let's Jam the Reactable: Peer Learning During Musical Improvisation with a Tabletop Tangible Interface. *ACM Trans. Comput.-Hum. Interact.* 20, 6, Article 36 (Dec. 2013), 34 pages. DOI: http://dx.doi.org/10.1145/2530541