# The effect of visualizing audio targets in a musical listening and performance task

Lonce Wyse[1, 2] Norikazu Mitani[2] Suranga Nanayakkara[2]

[1] Department of Communications and New Media
[2] Arts and Creativity Laboratory
National University of Singapore
{lonce.wyse, norikazu.mitani, suranga}@nus.edu.sg

## ABSTRACT

The goal of our research is to find ways of supporting and encouraging musical behavior by non-musicians in shared public performance environments. Previous studies indicated simultaneous music listening and performance is difficult for non-musicians, and that visual support for the task might be helpful. This paper presents results from a preliminary user study conducted to evaluate the effect of visual feedback on a musical tracking task. Participants generated a musical signal by manipulating a hand-held device with two dimensions of control over two parameters, pitch and density of note events, and were given the task of following a target pattern as closely as possible. The target pattern was a machine-generated musical signal comprising of variation over the same two parameters. Visual feedback provided participants with information about the control parameters of the musical signal generated by the machine. We measured the task performance under different visual feedback strategies. Results show that single parameter visualizations tend to improve the tracking performance with respect to the visualized parameter, but not the non-visualized parameter. Visualizing two independent parameters simultaneously decreases performance in both dimensions.

## Keywords

Mobile phone, Interactive music performance, Listening, Group music play, Visual support

## 1. INTRODUCTION

The last ten years or so has seen the development of many novel interactive media devices designed to support the engagement of participants through sound. Many are explicitly designed for anyone to enjoy, not only those with specific musical skills. Examples include the hyperinstruments created for the Brain Opera [15], and a variety of different table top devices such as Jam-O-Drum [1], and Reactable [10], "new media art" installations that involve sound, and musical games. Recent developments in mobile phones are also inspiring new kinds of interactive music [14,18]. Many of these devices and systems offer the potential for collective sound play that we might recognize as a form of musical improvisation. Recently, sensor-rich mobile phones have become ubiquitous computational devices providing new opportunities as public interfaces for media and musical performance environments.

Because of their computational and communicative power and their ubiquity, mobile phones hold enormous potential for supporting essentially an unlimited number of people to participate in interactive musical environments [16]. However there are still barriers to spontaneous and rewarding musical engagement for many due to a lack of musical experience.

Professional musicians exhibit a wide variety of sophisticated improvisational behaviors including conversational patterns, complimentary role-playing, coordinated transitions, etc. much more than non-musicians do. These patterns of improvisatory play may exploit technical skills, but don't seem to depend critically on their complexity. Improvisation is rather a skill that depends on the ability to synchronize one's own physical actions while simultaneously maintaining an awareness of another's musical activity [3]. Considerable attention has been devoted to providing non-musicians with instrumental interface that sounds musical without the need for technical skills [12], but less to supporting non-musicians in collaborative improvisation without having previous training in the requisite musical listening and communications skills, although there are exceptions [5][2].

Using the rotational dimensions of hand-held devices as simple instrument interfaces, we have been exploring whether and how graphical displays can be used to support non-musicians in the kind of listening and performance practices that make for engaging collaborative improvisational behavior. A preliminary study using mobile phones for collaborative music making showed that non-musicians often get lost in their attempt to listen to others while they concurrently engaged in playing their own instrument. Previous music cognition studies [11] show difference in listening skills of professional, amateur, and non-musicians. It has also been shown that non-musicians do not perform gestural imitation tasks as well as musicians [17].

The challenges faced by non-musicians in simultaneous music listening and performance motivates our exploration of visual feedback to support and encourage musical behavior for the musically untrained. This paper reports on a user study we conducted on the effect of different kinds of visual feedback to guide behavior.

## 2. RELATED WORK

It is known that real-time feedback is a useful tool for teaching singers to sing on pitch [8]. While these results are primarily about self-monitoring, target pitches are generally simultaneously visualized [9]. These systems are also oriented toward the long-term effects of learning, not just the performance during learning itself. Some of the visualizations described use multiple windows showing two parameters at once (pitch and spectral information), but there is little in this literature about the attentional issues of multiple displays of independent information. Wilson [19] found that computer-based visual feedback helped singers increase pitch accuracy, but showed differences between results for novice and advance singer. This result shows that supporting novices presents different challenges than supporting experts due to the different musical skill levels.

Music has many simultaneous dimensions changing at the same time that non-musicians in particular may have trouble tracking. Our experimental task requires that participants pay attention to two auditory features simultaneously. Previous work on auditory feature conjunction has focused mainly on the detection of specific stimuli in the presence of auditory distracters [20]. Although our task is not primarily one of detection, it is based on the idea that features can be separately attended, and that attention can be modulated with visual displays.

In our experiments, features of the visual stimulus are co-modulated with auditory features creating an audio-visual object akin to audio-visual speech integration that might direct attention to some auditory features at the expense of others that are not co-modulated with visual features. Intersensory facilitation is well established as decreasing reaction times [7]. Visual speech cues matched to auditory speech can enhance the detection of speech in noise [6]. It is also possible that dual-modality presentations reduce cognitive load thereby facilitating performance improvement in learning tasks [13].

The effect of visual feedback on music performance has also studied by Brandmeyer et al. [4]. A user study was conducted comparing two different visual feedback strategies for supporting the imitation of recorded material. One strategy was to show high-level abstract feedback reflecting expressive styles of performance. Another showed low-level descriptive feedback such as timing and dynamics of individual notes. Their results showed that for musicians, the more abstract high-level visual feedback improved imitation performance better than the low-level descriptive feedback.

# 3. HYPOTHESES

Based on previous literature and our preliminary studies, indications are that it may be possible to improve the performance of non-musicians in simultaneous musical listening and performance tasks with the support of visual feedback. In the present study, a sequence of tones is generated as a target pattern by the computer with two dimensions of variation: musical pitch, and density of note events. The subject uses a mobile phone as an interface with the same dimensions of control, and is given the task of tracking the target pattern as closely as possible. The two musical dimensions change smoothly with the rotational angle of the mobile phone so that there is no need for tightly coordinated temporal gestures (such as "hitting" a note). This makes physically playing the instrument easy and allows for attention to be directed toward listening and task execution.

The task was performed under different visualization conditions that provided information about neither, one, or both of the musical dimensions of the computer-generated pattern. The tracking task was the same under all visualization conditions, and optimal task performance still always required listening.

We report here on the results of testing two hypotheses:

- *H1. Visual feedback about the target pattern would result in better performance than no visual feedback condition.*
- *H2. Visual feedback of one machine target pattern dimension would improve the participants' performance in that dimension only.*

# 4. USER STUDY

Twenty-eight participants (two male participants and 26 female participants) took part in the study. Their median age was 21 years ranging from 20 to 25 years. Eight participants had studied music as a subject in primary and secondary school level and eleven of them were entirely inexperienced with musical instruments although they reported a wide range of musical listening tastes. The participants were recruited from the university student community. Participants were given a mobile phone with embedded accelerometers running an application to control sounds by rotating the phone up/down and left/right. They were asked to follow a machine generated sound pattern under different visual feedback conditions. The study was conducted in accordance with the ethical research guidelines provided by the Internal Review Board (IRB) of the National University of Singapore and with IRB approval.

## 4.1 Apparatus

The study was carried out in a quiet room with a 42 -inch LCD display, two speakers and office chair where participants were seated during the study. The visual display was placed at a constant horizontal distance (approximately 240 cm) from the chair and constant elevation (approximately 100 cm) from the floor. Two speakers were places on each side of the screen to present audio feedback.

The participant's mobile phone interface controlled the sound of an acoustic piano with two parameters. Rotating the mobile phone up and down changed the pitch of notes from A5 (MIDI note #81) to C3 (MIDI note #48) along a pentatonic scale. Rotating the mobile phone left and right changed the density of note events from slow to fast (143 BPM to 900 BPM). The effective angles were within 45 degrees (up/down or left/right) from center (keeping the phone parallel to ground and pointing directly towards display screen). Sound was toggled on and off with a tap of the thumb on the touch screen of the mobile phone.

Sound controlled by the participant was played from the speaker on the left-hand side of the screen. The computer-generated pattern was made with the sound of a marimba, and was played from the speaker on the right-hand side of the screen. The assignment of different instrument timbres (piano and marimba), and different stereo channels for playing the two source patterns was designed to make it easy for users to discriminate between the sounds patterns generated by the computer and themselves. A follow-up questionnaire confirmed that none of the participants had difficulty differentiating between the two simultaneous instrument sources.

Parameters for the pitch and the density-of-note events for the computer pattern were read from the same look-up table for each session to insure that the difficulty of the tracking task for each participant was exactly the same, but they were initiated from random starting points in the tables to minimize the possibility that patterns could be learned. The system outline is shown in Figure 1.

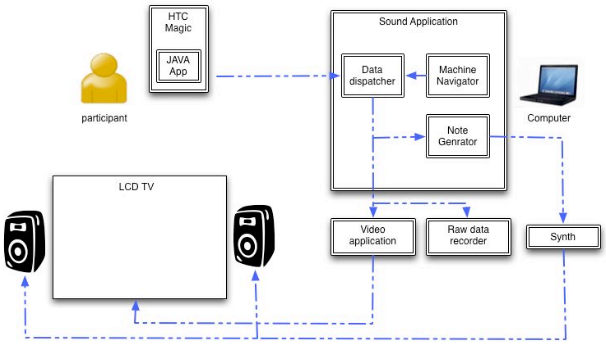


**Figure 1. System outline**

## 4.2 Procedure

In each trial, both participants and the computer played their instruments simultaneously for a period of 45 seconds during which time the participant engaged in the task of following the musical pattern of the computer as closely as possible.

Each of the trials was accompanied by visualizations tracking different control parameters generated by the subject and/or the computer instrument except for one segment where there was no visual feedback. Participants were informed about which controls would be visually tracked at the beginning of each trial. In addition, a color-coded legend appeared in the right hand side of the display to highlight what was being visualized. Visualizations mapped parameters in the vertical dimension (high pitch and high density were plotted higher in the vertical axis), and the graph scrolled to show a history of the parameter value (see Figure 2).

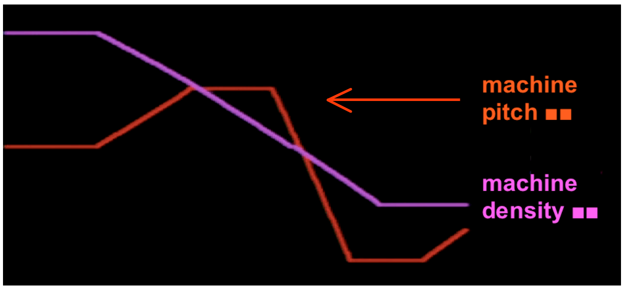


**Figure 2. A typical visualization tracking both pitch and density of events (PD condition). Real-time measurements appear on the right as the visualization scrolls to the left to show a history of the parameter values. [Note: Arrow pointer was not a part of the visualization; but has been added for black and white viewing]**

We tested ten conditions in total, but here we will consider only (a) visualization tracking the pitch of the computer's sound, P; (b) visualization tracking density of event of the computer's sound, D; (c) visualization that tacks both pitch and density of events of the computer's sound, PD; (d) no visual feedback, N; to address the hypothesis presented in this paper.

Before starting the study, each participant was told that the purpose of the experiment was to study the relationship between visualizations and making sound with a hand-held instrument. In addition, they were given the chance to become comfortable with the mobile phone instrument. They were also shown four examples of visualizations tracking pitch or density of events of their own instrument and pitch or density of the computer generated pattern. Once the participant was ready, they were reminded that their task was to follow the sound pattern of the computer as closely as possible. Trials were presented in a random order across subjects.

Data from the movement of the phone in the two control dimensions as well as the parameters generating the target pattern were recorded in a log file. After all the trials, the participants were asked to share their experience by answering a questionnaire. Each subject took approximately 30 minutes to complete the experiment session. It took two days to collect responses from 28 participants.

## 4.3 Analysis

Data was analyzed by comparing the two dimensional movement of the phone by the participant with the parameters controlling the synthesized target pattern. Data collected from four participants were discarded from the analysis since had accidently toggled off the sound during some of the trials. In each trial, it took a few seconds for the participants to get used to the task of following the computer-generated sound. Therefore, a 30 second interval from 13th second to 43rd second was chosen to evaluate the performance of the sound-following task. Pitch and density parameter values from both the participant and computer controlled patterns were sampled at 25 samples per second to calculate three performance indicators as follows:

Error in pitch dimension, $E_p = \sum_{t=13}^{43} \frac{|p_u - p_m|}{n}$

Error in density dimension, $E_d = \sum_{t=13}^{43} \frac{|d_u - d_m|}{n}$

Total error, $E_t = \sum_{t=13}^{43} \sqrt{E_p^2 + E_d^2}$

Where $p_u$ is user's pitch parameter, $p_m$ is computer's pitch parameter, $d_u$ is user's pitch parameter, $d_m$ is computer's pitch parameter, and $n$ is the number of number of data points. Figure 3 shows the error rates across all experimental conditions.

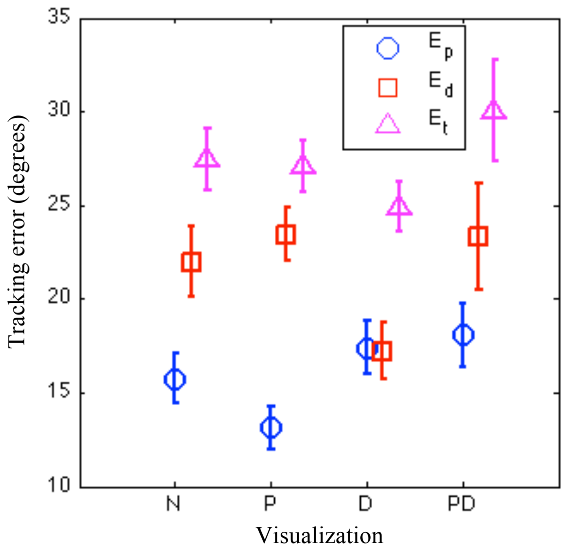


**Figure 3: Plot of error rates in the pitch dimension (Ep), density dimension (Ed) and total error (Et) with 95% confidence intervals for four different experimental conditions [N–no visual feedback, P–visualization tracking computer's pitch, D–visualization tracking computer's density of events, Visualization tracking both P and D]**

As seen in Figure 3, error in the pitch dimension is at its lowest when the visualization tracks the pitch of the computer sound. Similarly, error in the density dimension is lowest when the visualization tracks the density of events of the computer pattern. These results support the hypothesis that visualization of a target pattern control parameter reduces the tracking error in that dimension.

In general, error in the density dimension is significantly higher than the error in pitch dimension except in condition D, where the visualization tracks density parameter of the computer sound. However, this bias does not affect the cross-condition comparisons upon which we base the evaluation of our hypotheses.

Compared to the no visualization condition (D), total error in the tracking performance is slightly reduced when a single parameter is visualized (conditions P and D), however the improvement in total performance is less than the improvement in the visualized parameter. In both single-parameter tracking cases (P and D), the performance on the non-visualized parameter was slightly worse than in the no visualization condition (N). It would appear that the increased performance in the visualized parameter is at the expense of some degree of performance in the non-visualized parameter.

Given the improvement in performance for each of the individual parameter visualization conditions, one might predict that visualizing both parameters would result in a further performance improvement. However, when both parameters were visualized (condition PD), total performance error was significantly greater than all other conditions. Participants were sensitive to the difficulty of the task in this condition, and reported in the questionnaire following the experiment that trying to keep track of two visual parameters distracted them from the task of following the sound of the computer.

# 5. CONCLUSION AND FUTURE WORK

This study explored the effect of providing visual feedback with information about parameters of a target audio pattern of control parameters on a music-following task. We developed a mobile phone based interface to be simple to understand and play without any training. The results support both our hypotheses (*H1* and *H2*) concerning performance with single-parameter target visualization.

The data and the responses to the questionnaire following the experiment show that the density parameter was harder to follow than pitch. However, it is impossible to draw conclusions about any inherent difference in tracking these parameters outside of the specific ranges and mapping strategies used in our experiments. We expect that the particular characteristics of the mapping and visualization strategies have a significant effect on performance, and this warrants further study. Finally, our present study focused only on showing target parameters, although it seems likely that visual feedback of user activity would also influence listening and performance behavior.

# 6. ACKNOWLEDGEMENTS

We would like to express our gratitude to Angela Khoo for her input on aspects of the experimental design. This work was supported by project grant NRF2007IDM-IDM002-069 from the Interactive and Digital Media Project Office, Media Development Authority, Singapore, and the NUS AcRF project, "Listening Strategies for New Media; Experience and Expectation".

# 7. REFERENCES

[1] Blaine, T. and Perkis, T. The Jam-O-Drum interactive music system: a study in interaction design. *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, (2000), 165-173.

[2] Blaine, T. and Fels, S. Contexts of Collaborative Musical Experiences. *Proceedings of the International Conference on New Interfaces for Musical Expression*, (2003).

[3] Borgo, D. and Goguen, Sync or Swarm: Group Dynamics in Musical Free Improvisation. In R. Parncutt, A. Kessler, and F. Zimmer (eds.), Proceedings, Conference on Interdisciplinary Musicology, pages 52-53. Dept. Musicology, Graz (2004).

[4] Brandmeyer, A., Timmers, R., Sadakata, M., and Desain, P. Learning expressive percussion performance under different visual feedback conditions. *Psychological Research*, (2010).

[5] Weinberg, G. and Driscoll, S. *iltur:* Connecting Novices and Experts Through Collaborative Improvisation, *Proceedings of the International Conference on New Interfaces for Musical Expression* (2005), 17-22.

[6] Grant, K.W. and Seitz, P.F. The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America 108*, (2000), 1197.

[7] Hershenson, M. Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology 63*, 3 (1962), 289-293.

[8] Hoppe, D., Sadakata, M., and Desain, P. Development of real-time visual feedback assistance in singing training: a review. *Journal of Computer Assisted Learning 22*, 4 (2006), 308-316.

[9] Howard, D.M. and Welch, G.F. Visual displays for the assessment of vocal pitch matching development. *Applied Acoustics 39*, 4 (1993), 235-252.

[10] Jordà, S., Geiger, G., Alonso, M., and Kaltenbrunner, M. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, (2007), 139-146.

[11] Kreutz, G., Schubert, E., and Mitchell, L.A. Cognitive Styles of Music Listening. *Music Perception: An Interdisciplinary Journal 26*, 1 (2008), 57-73.

[12] Machover, T. Shaping minds musically. *BT Technology Journal 22*, 4 (2004), 171-179.

[13] Mousavi, S.Y., Low, R., Sweller, J. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology 87*, 2 (1995), 319-334.

[14] Oh, J., Herrera, J., Bryan, N.J., Dahl, L., and Wang, G. Evolving The Mobile Phone Orchestra. *Proceedings of the International Conference on New Interfaces for Musical Expression*, (2010).

[15] Paradiso, J. The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance. *Journal of New Music Research 28*, 2 (1999), 130-149.

[16] Scheible, J. and Ojala, T. MobiLenin combining a multi-track music video, personal mobile phones and a public display into multi-user interactive entertainment. *Proceedings of the 13th annual ACM Iinternational Conference on Multimedia, November*, (2005), 6-11.

[17] Spilka, M.J., Steele, C.J., and Penhune, V.B. Gesture imitation in musicians and non-musicians. *Experimental Brain Research 204*, 4 (2010), 549-558.

[18] Weinberg, G., Beck, A., and Godfrey, M. ZooZBeat: a Gesture-based Mobile Music Studio. *Proceedings of the 9th International Conference on New Interfaces of Musical Expression*, (2009), 312-315.

[19] Wilson, P.H., Lee, K., Callaghan, J., and Thorpe, C.W. Learning to sing in tune: Does real-time visual feedback help? *CIM07: 3rd Conference on Interdisciplinary Musicology,* Tallinn, Estonia, (2007), 15-19.

[20] Woods, D.L. and Alain, C. Conjoining three auditory features: an event-related brain potential study. *Journal of Cognitive Neuroscience 13*, 4 (2001), 492-509.