Biases and interaction effects in gestural acquisition of auditory targets using a hand-held device

Lonce Wyse
National University of Singapore
lonce.wyse@nus.edu.sg

Suranga Nanayakkara
Singapore University of Technology and Design
suranga@sutd.edu.sg

Norikazu Mitani
National University of Singapore
norikazu.mitani@nus.edu.sg

ABSTRACT

A user study explored bias and interaction effects in an auditory target tracking task using a hand-held gestural interface device for musical sound. Participants manipulated the physical dimensions of pitch, roll, and yaw of a hand-held device, which were mapped to the sound dimensions of musical pitch, timbre, and event density. Participants were first presented with a sound, which they then had to imitate as closely as possible by positioning the hand-held controller. Accuracy and time-to-target were influenced by specific sounds as well as pairings between controllers and sounds. Some bias effects in gestural dimensions independent of sound mappings were also found.

Author Keywords

Mobile phone, Interactive music performance, Sound cognition, Listening, Gesture control

ACM Classification Keywords

H5.2. User Interfaces (D.2.2, H.1.2, I.3.6). H5.5. Sound and Music Computing.

1. INTRODUCTION

With the ubiquity, mobility, and multimodal interactive enrichment of mobile phone devices, one of the thousands of ways in which they have been put to use is as musical instruments. Gestures detected with embedded accelerometers, compasses and other sensors, are particularly useful as a musical controllers for non-professional musicians because the basic patterns of pointing and rotating do not require a specialized learning regimen. Pitch (up/down), roll (twist), and yaw (left/right) motions seem to provide three easily separable dimensions of independent control. However, designing effective interactive instruments for non-musicians with these controllers may not be as straightforward as it seems.

In fact, musical engagement with even the simplest instrument is a demanding task. First, it generally involves listening to several different streams of activity simultaneously (whether they are individual voices or instruments, or different aspects of a single voice such as rhythm and pitch). Secondly, playing an instrument involves controlling several different physical dimensions simultaneously. It also requires understanding the mapping between control dimensions and sound dimensions, and finally it involves listening and exercising control simultaneously. Basic questions about whether some mappings between gesture and sound parameters make it easier to achieve desired musical results than others, and whether and how different dimensions in multidimensional sound control influence each other need to be understood before more complex instruments can be effectively designed and before complex musical performance data can be properly interpreted.

2. RELATED WORK

There is a strong kinship between musical instrument design and related nonmusical research in HCI (Orio, Schnell, and Wanderly, 2001). Gestural control of music has garnered attention as sound synthesis and sensor technologies have resulted in new instrument control and interaction techniques. The significance of embodied interaction has been shown to be an important aspect of a musical experience (Godøy, R. I., & Leman, M. 2009). Mapping strategies between gesture and sound have been developed (c.f. Wanderley, M. M., & Depalle, P. 2004), and a perception and intentionality perspective was presented by Van Nort (2009). Vertegaal & Eaglestone (1996) found significant differences in performance for timbral target acquisition between different interfaces, but did not systematically explore individual gesture/sound dimension pairings.

The majority of musical gesture research has been in the context of skilled instrumentalists. However, previous research about music related task performance with using gesture interface shows clear difference of novice and experienced music conductors (Lee et al., 2005). More needs to be understood about the ways physical control and listening interact in the simplest musical tasks.

Mobile phones in particular have become a popular musical interface device (Wang et al. 2008; Weinberg et al. 2009) because of their multidimensional, high-resolution responsive sensing capabilities, their increasing computational and communicative abilities, and their ubiquity among musician and non-musicians alike. However, much of this research focuses on specific tools or applications rather than on exploring the specific gestural affordances of these devices for their intrinsic capabilities and limitations from an HCI point of view.

Previous work (Wyse, Mitani, Nanayakkara 2011) also suggested that both specific sound dimensions as well as specific gestural control dimensions on hand-held devices

OZCHI '11, Nov 28 – Dec 2, 2011, Canberra, Australia

may have an effect on tracking task performance, and furthermore, there may be interaction effects between different dimensions in both sound and gesture.

3.METHOD

Twelve participants (eight females and four males) were recruited from the university student community. Their median age was 22 years ranging from 20 to 26. All reported normal hearing. Previous experience with gestural controllers is not known, but was considered irrelevant for the purpose of this audio tracking task. The study was conducted in accordance with the ethical research guidelines provided by the Internal Review Board of the National University of Singapore.

3.1 Apparatus

The experiments were conducted in a quiet room with the participants sitting comfortably in a chair facing a 42-inch display monitor. Participants held the mobile phone device used for controlling sound in whichever hand was most comfortable for them. The device had three possible rotational dimensions for controlling musical sound parameters, each with a range of 90^o chosen for comfort of movement and to avoid potential "edge effects" if targets were placed in extreme physical positions. The physical parameters were pitch, roll, and yaw where

- Pitch (abbreviated CP for "control pitch") is the angle of the phone pointing in the up/down direction over a range of +/- 45^o,
- Roll (CR) is the angle of the phone rotated in the clockwise/counter clockwise direction over a range of +/- 45^o, and
- Yaw (CY) the angle of the phone pointing in the left/right dimension over a range of +/- 45^o.

The "center" reference position was where the phone was level and pointing straight ahead. The controller dimensions were mapped to the sound dimensions of musical pitch, event (musical note) density, and timbre, chosen because they are easy to manipulate independently, and require no training to hear.

- Pitch (abbreviated SP for "sound pitch") was continuously variable between 254 Hz and 605 Hz (approximately musical notes C4 and D5) linear on a $\log_2$ frequency scale,
- Density of events (SD) was continuously and linearly variable between 6 events per second to 18 events per second, and
- Timber (ST) which used a simple frequency modulation algorithm, with a continuously variable modulation frequency parameter in the range [1,6] expressed as a factor of the carrier frequency. The index of modulation was held constant at 100.

An amplitude envelope was imposed on each event with a 21 ms rise time, 30 ms initial decay to 24% amplitude level, and a 190 ms decay. The sounds can be auditioned at http://anclab.org/projects/gtat-user-study.

3.1 Task

The participant's task in each trial was to listen to a one-second presentation of a sound, and then match the sound as quickly and accurately as possible by manipulating the gestural dimensions of the device controlling the sound.

Following the target presentation, the participants had to touch the screen of the device and position it to within 4^o of the center reference position before the sound would start and the timing of the trial would begin. When the participant felt that they had matched the target sound, they lifted their thumb from the device, stopping the sound synthesis and the timing of the trial. The maximum amount of time for each trial was limited to 10 seconds.

The device position was measured from the sensor stream transmitted to a computer. Two measures of performance were used: (1) the time-to-target which measured the interval between the time when the participant initiated and terminated the sound, and (2) the target error measured as the angle between the position where the participant stopped the sound, and the position of the device that would have generated the same sound as the target. The final position of the device was taken as an average over the 160 ms preceding sound termination (to smooth out a motion "jerk" that frequently occurred as the participant lifted their thumb to indicate target acquisition).

Before each session, subjects were told that they were participating in an experiment about the relationship between gestures and sound making with a hand held instrument. They were told that the three dimensions were pitch roll and yaw, and that only angle mattered, not the position of the device. Before each session, they were also given time to explore and become comfortable with the device, and asked to demonstrate the maximum and minimum angle in each dimension to which the sound control was sensitive. They were informed that they would hear a target sound and that their task was to match it as quickly as possible, and that the trial would end after 10 seconds even if they had not found the target.

Each participant took part in two sessions: one in which trials mapped one control dimension to one sound parameter, and a second in which each trial mapped two control dimensions to two sound parameters. Different trials within each session used different control dimensions and mapped them to different sound dimensions. Participants were told before each trial which control dimensions were being used and which sound dimensions they mapped to. The display also showed text identifying the control dimensions and the mapping to sound dimensions.

For the 1-dimensional session, each participant matched 36 targets that corresponded to four different positions along the control dimension. Target locations were at -30 ($T4_{1D}$), -15 ($T3_{1D}$), 15 ($T2_{1D}$), and 30 ($T1_{1D}$) degrees. The sessions lasted approximately 15 minutes.

For the 2-dimensional sessions, targets were distributed on a grid spaced by 15^o as in Figure 1. This spacing in 2D generates too many targets for a single experimental sessions, so to keep the session length short, targets were distributed across subjects so that each had the same number of targets (six targets from $T1_{2D}$ to $T6_{2D}$ - labelled in increasing order of distance) at a given angle and distance. The only difference in targets between subjects was a rotation of the pattern shown in Figure 1 by 0, 90, 180, or 270 degrees. The 2-dimensional session lasted approximately 30 minutes.

In both the 1-deminsional and 2-dimensional experiments, the sound parameters for the dimension(s) not being tested were held constant at the level corresponding to the centred position of the controller(s).

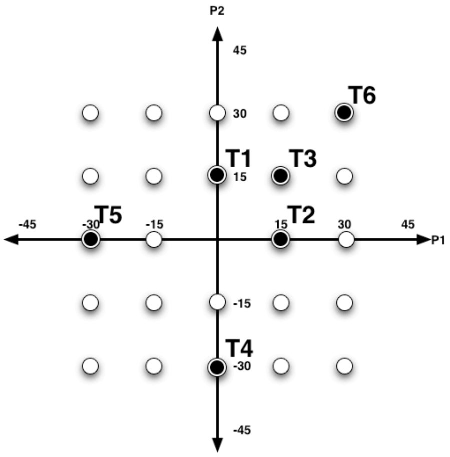


Figure 1. Target position for one subject in the 2 dimensional space of control parameters in black.

4.RESULTS

In the 1-D case, we found no main effect on error for control parameter - the performance in each control dimension, averaged across all sound mappings, was equivalent. There were also no significant interaction effects between control and sound dimensions in terms of accuracy.

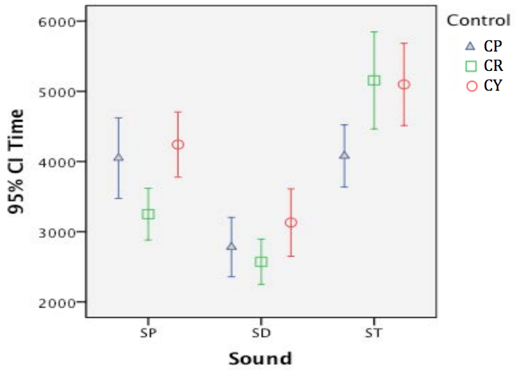
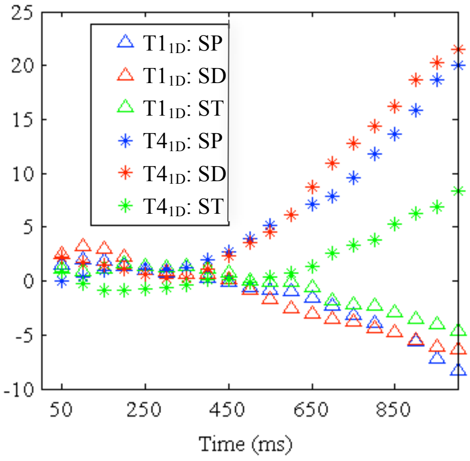


Figure 2. Error as a function of sound dimensions.

There was a main effect for sound: performance error was greater for timber when compared with either event density (by an average of 6^o) or sound pitch (by an average of 8.75^o). This result holds at the 95% significance level when error was averaged across all controller pairings. The relatively poor performance in the timbre dimension held as a trend for each control parameter paired individually, as well (Figure 2).

In terms of time-to-target, there was a main effect of sound parameter with density being the fastest to target, followed by pitch, and then timbre (when averaged over all control parameter mappings). Thus timbre showed both the worst performance accuracy and took the longest time to target.

In observing the experimental sessions, we noticed that participants seemed to explore the CP dimension with a bias toward the upward direction. Analyzing the data, we found that there was indeed a significant difference between the time-to-target for positive angles compared to negative, with upward targets being located a full second faster than downward targets (an average of 4.2 seconds vs. 3.2 seconds). This pattern is visible even during the first second of target searching. On average, when targets were located at $+30^o$ ($T4_{1D}$), participants had tracked up approximately 20^o after the first one second of searching, whereas when the targets were located at -30^o ($T1_{1D}$), tracking had only tracked down to -9^o over the first second. Figure 3 shows the bias over time in each sound dimension separately. The other control dimensions (yaw and roll), showed no significant bias.



Figre 3. Time course of target tracking for each of 3 sound parameters (color coded) mapped to the CP (up/down) controler dimension. Targets were located at + 30^o ('*') or – 30^o ('△')

Also in two dimensions, we can see a consistent picture emerge of the effect of the distance of the target from center. For each of the 3 control pairings, performance trends worse as target distance increases, with significant statistical significance for larger distance differences (Figure 4). We also found some interaction effects between different parameters in the two dimensional experiments. To summarize the salient findings for sound parameter pairings (averaged over all control parameters), we found that compared to the 1-D case,

- sound pitch performance becomes worse when paired with timber, but remains unchanged when paired with density,
- density performance deteriorates slightly and to the same extent whether paired with timbre or pitch,
- timbre trended toward worse performance when paired with pitch, but showed better performance when paired with density.

A similar analysis in control parameter pairings yielded only one significant effect: performance in the roll dimension deteriorated when paired with yaw. Although it does not reach significance overall, the pitch/yaw pairing performance was slightly better than others (see Figure 4).

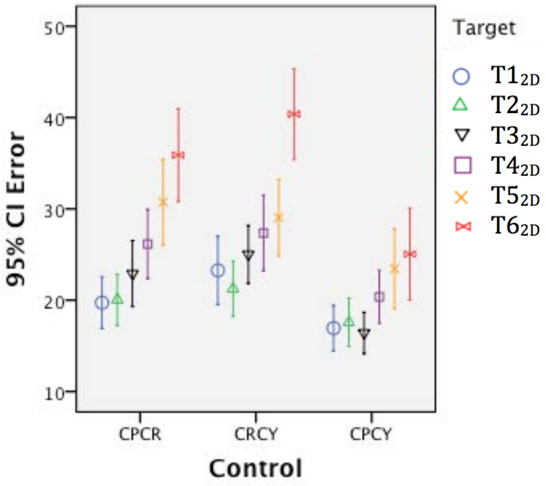


Figure 4. Accuracy for each target under the three control dimension pairings.

Finally, we noticed that the time it took for centering the device following the target presentation was variable, ranging up to 5 seconds, with the majority being fairly evenly distributed between 0 and 3 seconds. Average target error increased from 10^{o} to 15^{o} as the time-to-center increased from 0 to 3 seconds. Interestingly, the average time-to-target decreased over the same interval, from 3.8 to 3.2 seconds. One possible explanation for increased error combined with shorter time-to-target is that subjects may have felt an urgency to find the target after taking longer to center, and that the shorter time-to-targets are partly responsible for the decreasing accuracy.

5.DISCUSSION

For the purpose of design principles for gestural control of musical sound with pitch, roll, and yaw sensors on mobile devices, it is reassuring to have found that target acquisition performance was approximately equivalent in each of the control dimensions. It is interesting to note, however, that each control dimension, in particular the up/down dimension (CP) showed asymmetries in terms of how fast targets were acquired. This could be due to bodily kinetic asymmetries around our particular choice for the center reference point – level, and pointing straight ahead.

It was also not surprising to find that different sound parameters result in different performance accuracies, even when tested in a single dimension. For our selection of sound parameters - musical pitch, event density, and timbre - timbre was the worst performer in terms of time-to-target as well as accuracy, and had the most complex interaction with the other dimensions, particularly pitch.

Of course, timbre is difficult to define (an infamous attempt is found in Harpers Dictionary of Music which defines it as “the characteristic quality of a sound independent of pitch and loudness”). Our timbral dimension was created over a range of modulation frequency factors in a classic frequency modulation synthesis algorithm. Manipulating this parameter shifts components of the complex tone in the frequency domain where pitch is also expressed, which could have caused the performance interaction we found between these two dimensions.

We expected performance to deteriorate in each individual dimension when paired with others. This expected pattern was seen most clearly with event density. However, the interaction of timbre with the other dimensions was the most complex. Pitch and timbre both performed significantly worse when paired with each other than when they were paired with density. Performance for timbre was even slightly better when paired with density than when manipulated as the sole dimension.

Although we did find that performance accuracy decreased with target distance, there was no clear relationship between target distance and time-to-target. Thus, we found no results comparable to the Fitts Law (Fitts 1994). Further studies would be required to determine exactly why auditory and visual targets differ in this respect.

ACKNOWLEDGMENTS

This work was supported by project grant NRF2007IDM-IDM002-069 and the GAMBIT program from the Interactive and Digital Media Project Office, Media Development Authority, Singapore.

REFERENCES

Godøy, R. I., & Leman, M. Musical Gestures: Sound, Movement, and Meaning. Routledge (2009).

Lee, E., Wolf, M., & Borchers, J. Improving orchestral conducting systems in public spaces: examining the temporal characteristics and conceptual models of conducting gestures. In Proc. CHI’05, (2005) 731-740.

Orio, N., Schnell, N., & Wanderly, M.M. Input Devices for Musical Expression: Borrowing Tools from HCI. In Proc. NIME’01 (2001).

Van Nort, D. Instrumental Listening: Sonic Gesture as Design Principle. Organised Sound, 14(2), (2009) 177-187.

Vertegaal, R. & Eaglestone, B. Comparison of Input Devices in an ISEE Direct Timbre Manipulation Task. Interactive with Computers. 8(1), (1996) 113-130.

Wanderley, M. M., & Depalle, P. Gestural control of sound synthesis. Proceedings of the IEEE, 92(4), (2004) 632-644.

Weinberg, G., Beck, A., & Godfrey, M. ZooZBeat: a Gesture-based Mobile Music Studio. In Proc. NIME’09 (2009), 312-315.

Wyse, L., Mitani, N., Nanayakkara,S. The effect of visualizing audio targets in a musical listening and performance task. In Proc. NIME’11 (2011) 304-307.

Wang, G., Essl. G., Penttinen, H. Do Mobile Phones Dream of Electric Orchestras? In Proc. ICMC, Belfast, (2008).