Using Audio Cues to Support Motion Gesture Interaction on Mobile Devices

Sarah Morrison-Smith
Colorado State University
Fort Collins, CO 80523 USA
sarahms@cs.colostate.edu

Jaime Ruiz
Colorado State University
Fort Collins, CO 80523 USA
jgruiz@cs.colostate.edu

CHI 2014, Apr 26 - May 01, 2014, Toronto, ON, Canada.
ACM 978-1-4503-2474-8/14/04.
http://dx.doi.org/10.1145/2559206.2581236

Abstract
Motion gestures are an underutilized input modality for mobile interaction, despite numerous potential advantages. Negulescu et al. found that the lack of feedback on attempted motion gestures made it difficult for participants to diagnose and correct errors, resulting in poor recognition performance and user frustration. Here, we describe and evaluate a training and feedback system consisting of two techniques that use audio characteristics to provide: (1) a spatial representation of the desired gesture and (2) feedback on the system's interpretation of user input. Results show that while both techniques provide adequate feedback, users prefer continuous feedback.

Author Keywords
Motion gestures; mobile interaction; audio feedback.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Hand motion—pointing, gesturing, grasping, shaking, tapping—is a rich channel of communication. Yet, the repertoire of hand motion is largely ignored in mobile interfaces: modern smartphone users generally hold the device stationary while tapping or swiping its

surface. Inclusion of hand motion as input is not limited by technology, as devices contain an evolving set of sensors for recognizing phone movement. However, beyond rotating to change screen orientation or shaking to shuffle songs, little has been done to enable gestural input through device motion.

Given the potential advantages granted by using motion as an input modality for mobile interaction, several researchers (e.g. [1,4,6]) have explored various aspects of *motion gestures*, gestures performed by translating and rotating the device in three-dimensional space. Barriers to widespread adoption, as identified by Neulescu et al., include increasing user awareness of available gestures, and providing opportunities to practice and receive feedback on gestures during the learning process [5]. While these challenges exist for all gesture interfaces [2], feedback and training are especially difficult for motion gestures. *Surface gestures*, gestures performed on display surfaces, are displayable as two-dimensional diagrams, facilitating communication of available gestures and provision of feedback by displaying the correct surface gesture alongside the user's input [2]. However, these methods are not applicable to motion gestures due to inherent difficulties with projecting a three-dimensional gesture onto a two-dimensional surface. In addition, continuous visual feedback is not always feasible since the screen may not be visible at all times during the gesture.

To address the need of a training and feedback system for motion gestural input, we present *Glissando* and *Silenzio*, two techniques that use audio characteristics to provide (1) a spatial representation of the desired gesture and (2) feedback on the system's interpretation of user input. Both techniques enable feedback by verbally confirming correct gestures and notifying users of errors. Glissando provides additional continuous feedback by mapping distinct musical notes to each of three axes and manipulating pitch to specify spatial information. Evaluation shows that both techniques are capable of providing feedback and training for motion gestures. Participants overwhelmingly preferred Glissando, indicating that the technique is a strong feedback mechanism for motion gestures.

Using Audio for Gesture Training & Feedback

Audio feedback is appropriate for providing training and feedback for motion gestures since it does not rely on users being able to see the screen, can be used to express temporal constraints, and has potential applications in improving accessibility. While previous work has examined the use of audio characteristics as a navigational aid [8], we are unaware of any work using audio as a feedback mechanism for motion gestures. As a result, we've designed two variants of an audio feedback mechanism, Silenzio and Glissando, which differ in that Silenzio only provides feedback at the end of a gesture attempt, while Glissando also provides continuous feedback throughout the gesture.

Silenzio

Silenzio provides minimal information to the participant upon the completion of a gesture by either stating that the gesture is correct or identifying the user's error. For example, if a user tries to perform a gesture that requires rotating the screen, e.g. the DoubleFlip gesture (Fig. 1), and does not rotate the phone to the required threshold, Silenzio will state "Not far enough." Error feedback was designed to be verbal rather than nonverbal to reduce the amount of user interpretation.

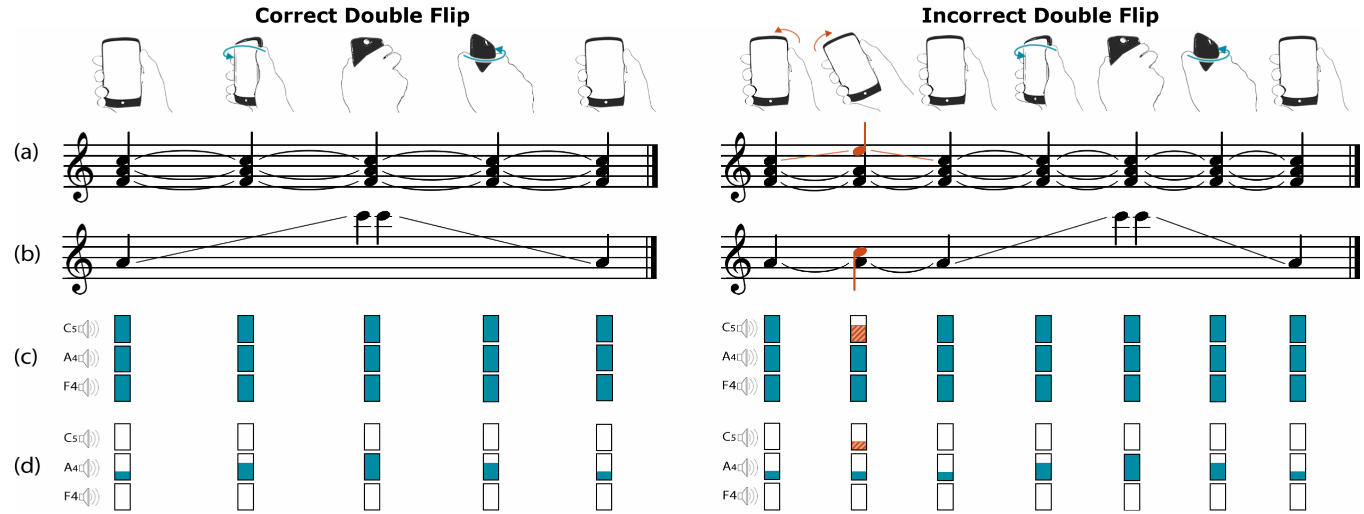

Figure 1: Examples of feedback for a correct (left) and incorrect (right) DoubleFlip gesture for (a) Wandering Pitch, (b) Additive Pitch, (c) Wandering Volume, and (d) Additive Volume.

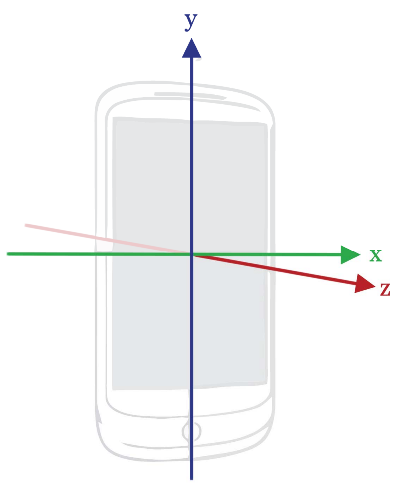

Figure 2: Coordinate system used by Android Sensor API [9].

Glissando

Glissando builds on Silenzio by adding continuous feedback as the user performs the gesture. This allows users to manipulate their input before an unsuccessful gesture has been detected. To enable continuous feedback in Glissando, we mapped distinct musical notes to each of three axes (see Fig. 2); a change in note characteristics (e.g. pitch and/or volume) specified the rotation and/or translation of the device around a specific axis. The resulting audio representation of the reference (ideal) gesture was available to be played to the user, as well as the representation of the most recent gesture attempt. Any differences in these representations indicated differences between the reference gesture and the user's gesture[1].

As Glissando relies on audio characteristics to represent spatial information, it is important to choose a characteristic configuration that allows the user to easily discriminate between different gestures. To determine the appropriate configuration for Glissando, we considered the following four methods:

WANDERING PITCH (WP)

Feedback consists of playing all notes mapped to each axis. Deviation from reference gesture causes independent pitch changes for each note mapped to a deviating direction. Correct gestures result in all notes

[1] A demonstration of these techniques is available at: http://hci.cs.colostate.edu/audio-cues/chiwip2014.mov

being played continuously without pitch change. For example, see Fig. 1 (a).

ADDITIVE PITCH (AP)
Feedback starts by playing all notes mapped to the axes of desired movement. A correct gesture results in the smooth transition of these notes ranging between a low-pitched note (C4, 61 MIDI) and a high-pitched note (C6, 84 MIDI). The notes mapped to axes associated with undesirable movement are added once a threshold is passed indicating error in the associated direction. For example, see Fig. 1 (b).

WANDERING VOLUME (WV)
Feedback consists of playing all notes mapped to each axis. Deviation from reference gesture causes independent volume reduction for each note mapped to a deviating direction. Correct gestures result in all notes being played continuously without volume change. For example, see Fig. 1 (c).

ADDITIVE VOLUME (AV)
Feedback starts by playing all notes mapped to the axes of desired movement. A correct gesture results in the smooth transition of these notes ranging from 20% to 100% volume. The notes mapped to axes associated with undesirable movement are added once a threshold is passed indicating error in the associated direction. For example, see Fig. 1 (d).

Glissando maps each axis to one of three distinct notes comprising a major chord acceptable for use in all of the methods mentioned above. For example, an audible and undistorted adequate pitch range was required for AP, while AV and WV required all notes to remain above the lowest note that could be played at discernibly different volumes (C4, 61 MIDI). A major chord was chosen because of its tendency to generate a positive effect [3] when resolving from an error chord (i.e., the chord heard due to a deviation in one or more axis) to the original chord in the WP and WV conditions. The use of the mobile device's internal speaker reduced the range of notes that could be played without distortion.

Options WP and WV were rejected during the initial design process due to difficulty discerning differences between the changes in audio characteristics. The feasibility of options AP and AV were determined by the following pilot study.

Pilot Study – Determining Appropriate Audio Characteristics for Spatial Representation

The goal of this pilot study is to determine the optimum continuous feedback configuration for Glissando. Although Glissando can be applied to a variety of gestures, this study focused on use with the DoubleFlip gesture [7], since recent work reported that users had difficulties performing it when no feedback was present, despite its relative simplicity [5].

Conditions
As the DoubleFlip gesture comprises solely of rotation around the Y axis, AP was implemented such that a correct gesture resulted in the center note smoothly transitioning from A4 (69 MIDI) to C6 (84 MIDI) and back, as shown in Fig. 1(b). Notes mapped to the Z and X axes were added once the gesture deviated ±15° around either axis. The Y axis was mapped to the center note of the chord and the X and Z axes to the highest and lowest notes, respectively, to assist users in determining which direction needed correction.

AV was implemented so that a correct gesture resulted in only the center note smoothly transitioning from

20% volume to 100% and back, as shown in Fig. 1(d). Notes mapped to the Z and X axes were added once the gesture deviated ±15° around either axis. As described above, deviation from the reference gesture resulted in independent volume changes for each note.

Procedure

This pilot study consisted of each participant using one of two feedback techniques (AV and AP) to perform a single correct DoubleFlip gesture. Participants were randomly assigned to each technique. The number of participants in each group was counter-balanced. The study began by the participant listening to a verbal description of the gesture and explanation of the technique. Each participant performed the DoubleFlip gesture while undertaking a think-aloud protocol. To prevent undue frustration, participants were stopped if they could not complete a gesture within 10 minutes.

Apparatus and Participants

Glissando was developed in Java using the Android SDK [9] and libpd library [10]. The study was performed using a LG Nexus 4 smartphone running Android 4.2. Eight participants aged 19-64 (mean = 31.0, S.D = 14.9, 4 females, 1 left handed) we recruited using a departmental email list.

Results

In one instance, a user was unable to discern correct gestures from incorrect gestures using AV Timed due to similarity of high volume notes. Also, an older participant using AV Untimed reported difficulty discerning between differences in volume, especially for low volumes. AV was discarded due to these drawbacks.

Final Design

As a result of this pilot study, further implementations of Glissando represent spatial information with Additive Pitch.

Pilot Study – Evaluation of Continual Feedback

We conducted a pilot study asking participants to perform five correct DoubleFlip gestures using both of our feedback techniques, Silenzio and Glissando. Participants were separated into two groups that were defined by the initial technique to be learned. The number of participants in each group was counter-balanced.

The study began with the participant listening to a verbal description of the gesture and the first technique. Participants were then asked to complete five gestures. To prevent undue frustration, participants were stopped if they could not complete a gesture within five minutes. Then, participants repeated the task using the second technique. Finally, participants were asked to identify which technique they preferred.

Apparatus and Participants

Silenzio and Glissando were developed and run on the same hardware and software as our first pilot study. Thirty-two participants affiliated with a local university, aged 18 - 55 (mean = 22.9, S.D. = 7.7, 6 females, 3 left handed), took part in the study.

Results

Two participants who initially failed to correctly perform a DoubleFlip gesture with Silenzio were able to complete the required five gestures using Glissando. Both participants requested to stop their Silenzio trial before the five minutes had elapsed out of frustration. One participant was unable to complete a gesture using either technique. The majority of our participants (29) were able to use both techniques to accomplish the task, suggesting both provide adequate feedback.

When asked which technique they preferred, 28 out of 32 participants (80%) indicated a preference for Glissando, while two participants preferred Silenzio, and another two participants had no preference. Participants stated that Glissando was especially helpful when determining the direction and magnitude to rotate the phone. Additionally, one participant reported imagining the sounds generated by Glissando while subsequently using Silenzio. Although temporal constraints were not imposed during this study, we observed that participants attempted to match the speed of the reference gesture while using Glissando.

Conclusion and Future Work

We described and evaluated two techniques for motion gesture input that use audio to provide (1) a spatial representation of the desired gesture and (2) feedback on the system's interpretation of user input. Results from our user study demonstrated that while both Silenzio and Glissando provide adequate feedback to users, users prefer continuous feedback.

While our initial prototypes and evaluations were performed using the DoubleFlip gesture, since Glissando supports movement around several axes through the use of major chords, our techniques can be easily adapted to other gestures. Further work includes incorporating additional gestures, facilitated by creating a framework for mapping motion gestures to audio feedback. Furthermore, additional investigation is needed to impose temporal constraints on gestures in an educative way. Finally, given the nature of motion gestures and our use of audio feedback, we plan on exploring the use of motion gestures and Glissando to support mobile interaction for vision disabled users.

Acknowledgements

Funding for this work was provided by a Google Faculty Research Award.

References

[1] Ashbrook, D. and Starner, T. MAGIC: A Motion Gesture Design Tool. Proc of CHI ’10, ACM (2010), 2159–2168.

[2] Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., and LaViola,Jr., J.J. GestureBar: improving the approachability of gesture-based interfaces. Proc. CHI‘09, ACM (2009), 2269–2278.

[3] Cook, N.D. The sound symbolism of major and minor harmonies. Music Perception 24, 3 (2007), 315–319.

[4] Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E. Sensing techniques for mobile interaction. UIST’00, ACM (2000), 91–100.

[5] Negulescu, M., Ruiz, J., Li, Y., and Lank, E. Tap, swipe, or move: attentional demands for distracted smartphone input. Proc. of AVI, ACM (2012), 173–180.

[6] Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. Proc. of CHI’11, ACM (2011), 197–206.

[7] Ruiz, J. and Li, Y. DoubleFlip: a motion gesture delimiter for mobile interaction. CHI’11, ACM (2011), 2717–2720.

[8] Talbot, M. and Cowan, W. On the audio representation of distance for blind users. CHI’09, ACM (2009), 1839–1848.

[9] Android Open Source Project. Google Inc. libpd. http://libpd.cc/