Using Audio Cues to Support Motion Gesture Interaction on Mobile Devices

SARAH MORRISON-SMITH and MEGAN HOFMANN, Colorado State University
YANG LI, Google Research
JAIME RUIZ, Colorado State University

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. In this article, we describe and evaluate a training and feedback technique, *Glissando*, which uses audio characteristics to provide feedback on the system's interpretation of user input. This technique enables feedback by verbally confirming correct gestures and notifying users of errors in addition to providing continuous feedback by manipulating the pitch of distinct musical notes mapped to each of three dimensional axes in order to provide both spatial and temporal information.

CCS Concepts: • **Human-centered computing → Auditory feedback;**

Additional Key Words and Phrases: Motion gestures, mobile interaction, audio feedback

ACM Reference Format:

Sarah Morrison-Smith, Megan Hofmann, Yang Li, and Jaime Ruiz. 2016. Using audio cues to support motion gesture interaction on mobile devices. ACM Trans. Appl. Percept. 13, 3, Article 16 (May 2016), 19 pages.
DOI: http://dx.doi.org/10.1145/2897516

1. INTRODUCTION

The smartphone form factor limits both input and output. To allow the device to fit into a pocket or purse, screens are small and keyboards are thumb sized. On many devices, the thumb keyboard has been replaced by a soft keyboard displayed on the screen to minimize the size and weight of the device. As a result, the primary interaction with a smartphone consists of a user tapping or swiping on the device's display. Recently, Ruiz et al. [2011] proposed taking advantage of the internal motion sensors (e.g., the gyroscope and accelerometer) commonly found in mobile devices to extend the input space. Their work demonstrated how *motion gestures*, gestures performed by translating and rotating a mobile device in three-dimensional space, can be mapped to a device command, allowing interaction without the use of the touchscreen. However, beyond rotating to change screen orientation [Hinckley et al. 2000] or shaking to shuffle songs [Apple Inc. 2009], few motion gestures have been incorporated into typical users' daily lives. This disparity is surprising considering the many potential benefits

This work was funded by a Google Faculty Research Award.

Authors' addresses: S. Morrison-Smith, M. Hofmann, and J. Ruiz, Colorado State University Fort Collins, CO 80523 USA; emails: {sarahms, mkh, jgruiz}@cs.colostate.edu; Y. Li, Google, Inc., Mountain View, CA, USA; email: yangli@acm.org.

ACM 1544-3558/2016/05-ART16 $15.00
DOI: http://dx.doi.org/10.1145/2897516

granted by using motion as an input modality for mobile interaction. Recent research (e.g., Negulescu et al. [2012a, 2012b] and Ruiz and Li [2011]) has highlighted several of these possible advantages, including the potential to expand the input space for mobile phones, provide shortcuts for multistep smartphone commands, and facilitate smartphone use while distracted.

The underutilization of motion gestures for mobile input is a multifaceted problem with a variety of contributing factors. Negulescu et al. [2012b] identified several crucial barriers to the widespread adoption of motion gestures, including increasing user awareness of available gestures and providing opportunities to practice and receive feedback on gestures during the learning process. Lack of feedback is particularly problematic since it makes it difficult for users to correctly diagnose and correct errors. While these challenges exist for all gesture interfaces [Bragdon et al. 2009], feedback and training are especially difficult for motion gestures because the movement of the device is three-dimensional.

Bridging the gap between the user's input and the recognizer's expectations is typically accomplished using one of two general approaches: (1) training the device to recognize the gesture as performed by the user and (2) training the user to perform the gesture as expected by the device. Both methods require a contribution of time and effort from the user. The first method, training the device, allows users to customize gestures since it depends on the user's input rather than on a predefined template. This caters to the user's individual needs and comfort. However, during training, the user's input typically deviates from the original core gesture, sometimes significantly. In this situation, it is possible for two gestures to drift toward each other, making it difficult for the device to differentiate the gestures. In contrast, the second method of training the user prevents inadvertent gesture collisions by enforcing thresholds that limit deviation from the predefined gesture. These thresholds can be relaxed to allow the user some flexibility when learning or performing the gesture [Negulescu et al. 2012a]. Given this advantage, for this study, we focus on training users to perform gestures previously elicited from users by Ruiz et al. [2011].

Gestures consist of both timing restrictions and a path through space. The mobile device's spatial path is obviously a key characteristic of the gesture; however, the temporal component is equally important from both a theoretical and recognition standpoint. Conceptually, timing can be the primary distinction between two otherwise similar gestures. For example, for most mobile devices, a short tap on a hyperlink triggers the device to open that link in a web browser, while a long tap opens a copy/paste menu. In this case, a gesture recognizer that is flexible regarding timing would not be able to distinguish between the two gestures, which indicates that this is an issue that likely cannot simply be solved by implementing a "better" recognizer. Thus, it is vital that any training solution addresses both the spatial and temporal aspects.

To address the need of a training and feedback system for motion gestural input, we developed *Glissando*, a technique that assists in learning motion gestures by using audio characteristics to provide feedback on the system's interpretation of user input. This technique assists in training and provides feedback by (1) verbally confirming correct gestures, (2) notifying users of specific errors, and (3) providing additional continuous feedback by mapping distinct musical notes to each of three axes and manipulating audio characteristics to specify both spatial and temporal information.

The remainder of the article is organized as follows: First, we give an overview of related work in Section 2. Next, we recount the development of Glissando in Section 3. This includes a narrative of an exploration study to determine the optimal method for providing continuous feedback with Glissando (Section 3.2), a pilot study evaluating the effectiveness of continual feedback in assisting users learning the spatial component of the *DoubleFlip* gesture shown in Figure 1 [Ruiz and Li 2011] (Section 3.3), a brief exploration study to determine whether the time-dependent aspect of a motion gesture can be enforced using time limits (Section 3.4), and an assessment of the effectiveness of using continuous feedback to express temporal information about the gesture (Section 3.5). These initial studies focus on

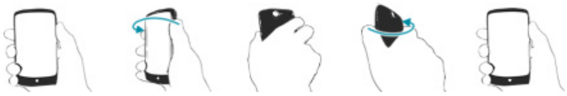

Fig. 1. The DoubleFlip gesture [Ruiz and Li 2011]. The user holds the phone in his or her right hand. The user rotates the phone along its long side so that the screen faces away, and then back.

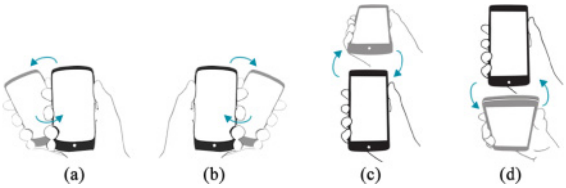

Fig. 2. Additional motion gestures influenced by Ruiz et. al [2011]. (a) FlickLeft, (b) FlickRight, (c) FlickUp, and (d) FlickDown.

use with the DoubleFlip gesture since recent work reported that users had difficulties performing the gesture when no feedback was present, despite its relative simplicity [Negulescu et al. 2012a, 2012b]. Finally, we evaluate Glissando in Section 4 by examining gesture memorability for users trained with and without the system using an expanded gesture set that, in addition to DoubleFlip, includes several gestures inspired by a previous elicitation study (shown in Figure 2) [Ruiz et al. 2011]. We close with a discussion of findings and a synopsis of future work in Sections 5 and 6.

2. RELATED WORK

2.1 Motion Gestures

Several researchers have explored various applications for motion gestures. Rekimoto [1996] was the first to demonstrate how mapping device tilt can be used for selecting menu items, interacting with scroll bars, panning or zooming around a digital workspace, and performing complex tasks such as 3D object manipulations. Harrison et al. [1998], Small and Ishii [1997], and Bartlett [2000] extended the use of tilt sensors to enable navigating through widgets on mobile devices. Additional potential applications of motion gestures have been developed, such as using tilt to allow a user to change screen orientation [Hinckley et al. 2000], navigate maps or images [Rekimoto 1996], input text [Jones et al. 2010; Partridge et al. 2002; Wigdor and Balakrishnan 2003], control a cursor [Weberg et al. 2001], access data on virtual shelves around a user [Li et al. 2009], and verify user identity [Liu et al. 2009].

Further research has explored the various aspects of designing gestures, including the development of systems to aid designers of systems that use motion gestures such as Exemplar [Hartmann et al. 2007] and MAGIC [Ashbrook and Starner 2010]. Ruiz et al. [2011] developed a taxonomy that described the attributes of smartphone motion gestures and the natural mappings of motion gestures onto smartphone commands.

Prior work has also examined the cognitive demands of using motion gestures. Negulescu et al. [2012b] examined the relative cognitive demands of tapping the touchscreen; performing *surface gestures*—gestures performed on display surfaces; and performing motion gestures. Results from this study showed that no significant difference in reaction time exists between the three types of input, meaning that using gestures does not result in an observable increase in cognitive cost. Additionally, it was shown that motion gestures result in significantly less time spent looking at the device screen while walking than tapping on the screen, even when the device interface is optimized for eyes-free input.

2.2 Visual Feedback Techniques

The need to provide feedback for gestural interaction is not limited to motion gestures. Surface gestures have the advantage of being readily displayed as two-dimensional diagrams, which, in addition to facilitating the communication of available gestures, facilitates the provision of continuous feedback by displaying the correct surface gesture alongside the user's input [Bartlett 2000]. OctoPocus, developed by Bau and Mackay [2008], utilizes this approach to provide continuous feedforward and feedback to learn, remember, and execute surface gestures. However, this method is difficult to apply to motion gestures due to inherent difficulties with projecting a three-dimensional gesture onto a two-dimensional surface. Additionally, the nature of motion gestures requires the user to rotate and translate the device, meaning that continuous visual feedback displayed on the device's screen is not always feasible since the screen may not be visible at all times.

Sodhi et al. [2012] presented LightGuide, a visual continuous feedback system for *midair gestures*, gestures performed in three-dimensional space without holding a device (e.g., pointing and gestures performed using the Microsoft Kinect). LightGuide projects visual cues, such as arrows and colors, onto a user's hand to guide the user in performing physical movements, such as moving his or her hand along a predetermined path. The similarity between physical movements and motion gestures suggests that LightGuide can be easily adapted to provide feedback for motion gestures. However, while LightGuide's system mitigates occlusion of visual feedback by not using the mobile device's screen, we believe that this is not a viable solution for everyday use of motion gestures due to its reliance on additional devices (a projector and a depth camera).

Recent work by Kamal et al. [2014] explored the effect of using various gesture representation systems, with and without visual feedback, on user performance of motion gestures. Methods for representing motion gestures included icons, videos displayed on the device screen, and a combination of Kinect and videos displayed on an external screen. Feedback consisted primarily of visualization of the distance between the ideal gesture and the user's attempt either through a numerical scale displayed on the device screen or by directly comparing the Kinect representations of the user's attempt and the ideal gesture. Results indicated that scaffolding techniques that rely only on the mobile device, with no additional devices or hardware, can be a feasible solution for training users to perform motion gestures.

2.3 Aural Feedback Techniques

Audio feedback may be appropriate for providing training and feedback for motion gestures since it has been successfully utilized for assisting various spatial and surface gesture tasks and does not rely on users being able to see the screen or possessing an additional device. Furthermore, concurrent auditory feedback has been shown to be more effective than visual concurrent feedback in enhancing learning of new skills [Edwards 2010].

Previous work has examined the use of audio characteristics as feedback for spatial tasks such as aiding navigation for blind users [Talbot and Cowan 2009], determining radial direction [Harada et al. 2011], expressing two-dimensional paths [Harada et al. 2011], enhancing target selection tasks [Eslambolchilar et al. 2004a; Marentakis and Brewster 2004, 2005, 2006], enhancing tilt-controlled speed-dependent automatic zooming [Eslambolchilar et al. 2004b], and replacing joint and muscle sensory information for patients who lack proprioception [Ghez et al. 2000] or are recovering from a stroke [Wallis et al. 2007; Schmitz et al. 2014].

Several researchers have explored the integration of continuous and end-of-gesture audio feedback for teaching and improving the accuracy of surface gestures [Brewster et al. 2003; Lumsden and Brewster 2003; Müller-Tomfelde and Steiner 2001; Oh et al. 2013] and tasks similar to performing surface gestures [Plimmer et al. 2011]. Additional work has focused on the combination of audio feedback and surface gestures to promote accessibility [Kane et al. 2008, 2011].

Notably, Andersen and Zhai [2008] explored application of audio feedback to pen-gesture interfaces, but concluded that it is difficult to achieve benefits with audio feedback. However, the observed negative effect of audio feedback on gesture performance is likely due to the type of feedback provided. In this study, gestures were mapped to feedback characterized by complex tones using frequency, timbre, jitter, amplitude, and displacement [Andersen and Zhai 2008], which likely provided too much information for the users to effectively utilize [Edwards 2010]. Additionally, users were only provided with a visual reference of the gesture and did not receive an audio reference that corresponded to audible feedback. Furthermore, the authors' concern regarding the efficacy of audio feedback for gestures was partially based on the idea that audio feedback is too slow to improve handwriting. However, it is unclear whether this conclusion applies to motion gestures.

Williamson and Murray-Smith [2002] developed a method for communicating high-dimensional, dynamic information to users interacting with systems via continuous audio feedback generated by asynchronous granular synthesis. This audio feedback mechanism was postulated to be applicable to surface and motion gestures and was incorporated into a framework, SIGIL, designed for developing and testing gesture recognizers [Williamson and Murray-Smith 2002, 2005]. However, there is no indication that this system is fully developed or examined in a user study. As such, we are unaware of any work implementing the use of audio as the sole feedback mechanism for training users to use motion gestures.

3. DESIGNING AN AUDIO SYSTEM FOR GESTURAL TRAINING AND FEEDBACK

In light of relevant work, we designed our gestural feedback system to meet the following design goals:

(G1) Minimize visual feedback since the device screen may not be constantly visible while performing motion gestures.
(G2) Refrain from using any external hardware or additional devices in order to promote the mainstream adoption of motion gestures.
(G3) Be compatible with current-generation smartphones to facilitate quick adoption.

3.1 Initial Concept

To address these goals, we developed a concept centered around providing *continuous concurrent feedback*, which allowed users to manipulate their input before an unsuccessful gesture has been detected, and a *reference* that users could compare their input to. To enable continuous feedback, we mapped distinct musical notes to each of three spatial axes; a change in note characteristics (e.g., pitch and/or volume) was used to specify the spatial information of rotating and/or translating the device around a specific axis. This mapped each gesture attempt to a unique audio representation with distinct characteristics. A reference consisting of the audio representation of a perfect gesture was available for the user to listen to at any time. This allowed users to directly compare the representation of their gesture attempt to the representation of a perfect gesture—any differences in the characteristics of these representations indicated differences between the ideal gesture and the performed gesture.

Additionally, upon recognition of a complete gesture or detection of an extreme error, the system informed users that the gesture was correct or identified the user's error. Error messages included identifying when a user passed a threshold of movement in an undesirable direction or failed to meet a threshold. For example, if a user attempting to perform DoubleFlip tilted the phone sufficiently toward him- or herself, the system simply said "too far up." Furthermore, if a user tried to perform a gesture that required rotating the screen (e.g., DoubleFlip) and did not rotate the phone to the required threshold, the system stated "not far enough." Finally, for our exploration study on enforcing strict time limits, the system included error messages that notified users when they took too long to complete the gesture. In this case, the system stated “not fast enough.” For clarity, error feedback was designed to be verbal rather than nonverbal.

We developed this concept for use in a training environment where the user is attempting to learn a specific, predefined gesture. This is opposed to normal, everyday use, where audio feedback would not be provided. As demonstrated by our final study, the system can be harnessed to assist a user in learning multiple gestures by tailoring implementations for each gesture in the set. For this use case, it is not necessary for the system to differentiate between multiple gestures since it is reasonable to specify which gesture will be performed.

3.2 Exploration Study: Determining Appropriate Audio Characteristics for Spatial Representation

Since this system relies on audio characteristics to represent spatial information, it is important to choose a characteristic configuration that allows the user to easily discriminate between correct and incorrect gestures. Furthermore, it is important to limit feedback to the manipulation of only a few characteristics since excessive feedback becomes an issue as feedback begins to exceed a learner’s ability to internalize and react [Edwards 2010]. Thus, the goal of this exploration study was to determine the optimum continuous feedback configuration, using DoubleFlip as an example gesture.

Our choice of using the DoubleFlip gesture in our initial studies was influenced by the findings of Negulescu et al. [2012b], who found that users had a difficult time performing the DoubleFlip gesture after the initial training session. They also note that the lack of feedback resulted in user frustration and users performing random gestures in hopes that they would be recognized as the correct gesture. Results from the initial study resulted in Neglescu et al. proposing adaptations to the recognition thresholds [Negulescu et al. 2012a] as a method to minimize user frustration. Instead of modifying recognition thresholds, potentially increasing the number of false positives, our approach is to develop better training methods to enable users to perform successful gestures.

3.2.1 *Prototypes.* Common audio characteristics include pitch, volume, timbre, tempo, and rhythm. Timbre was rejected as a potential characteristic due to concerns that the limitations of the mobile device’s [LG Nexus 4 and 5] internal speaker would make discerning between different tones exceedingly difficult for this specific application. Tempo, while easily discernible using the mobile device’s internal speaker, seemed uniquely suited to providing temporal information, such as gesture speed, and was reserved for that purpose. Rhythm, which seemed similarly suited to providing temporal information, will be of interest in future research. As such, to determine the appropriate configuration for this system, we considered the following four methods that utilized the remaining note characteristics, pitch and volume:

Additive Pitch (AP). This prototype uses changes in pitch to convey information about the user’s core movement. Feedback starts by playing only notes mapped to the axes of desired movement. Notes mapped to axes along which or around which movement is undesirable are not played. The pitches of these notes change as the phone is moved. A correct DoubleFlip gesture results in the smooth transition of these notes ranging between a low-pitched note (A^4, 69 MIDI) and a high-pitched note (C^6, 84 MIDI). Notes mapped to axes along which or around which movement is undesirable (e.g., the x- and z-axes) are not played initially. However, these notes are played once a threshold is passed, indicating error in the associated direction (e.g., 15° around either undesired axis). The pitches of these notes retain their respective distances (–4 \+3 MIDI) from the center note pitch. The y-axis is 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. AP uses a broad range of pitches to ensure that that minor movements are readily apparent. For example, see Figure 3(a).

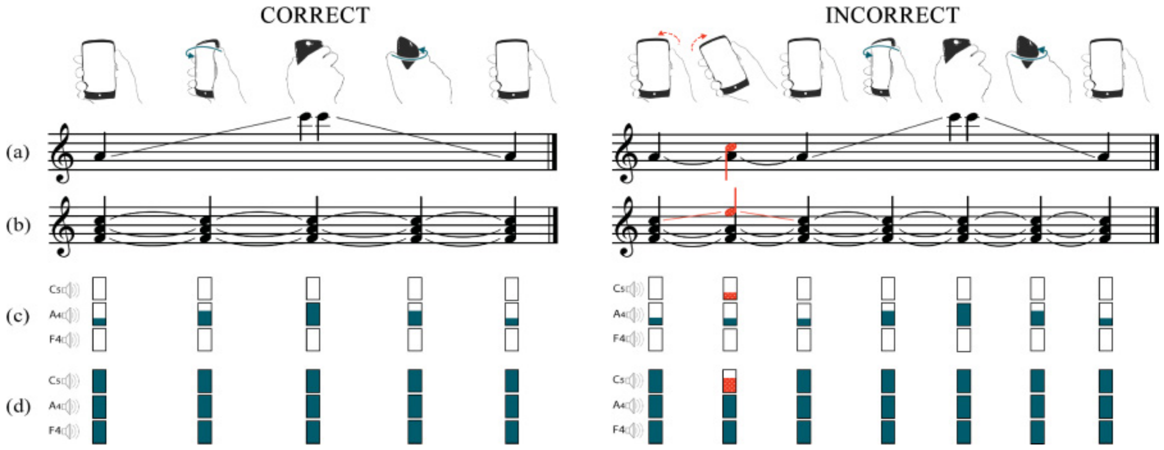

Fig. 3. Examples of feedback for a correct (left) and incorrect (right) DoubleFlip gesture for (a) Additive Pitch, (b) Wandering Pitch, (c) Additive Volume, and (d) Wandering Volume.

Wandering Pitch (WP). This prototype uses harmony and discord to give the user full awareness of his or her movement. Feedback consists of playing all notes mapped to an axis. Deviation from the reference gesture causes each note mapped to an affected direction to independently change pitch, causing discord. Correct gestures result in all notes being played continuously in harmony without pitch change. For example, see Figure 3(b).

Additive Volume (AV). This prototype uses changes in volume to convey information about the user's core movement. Feedback starts by playing only notes mapped to the axes of desired movement (e.g., C^4, 60 MIDI). The volumes of these notes change as the phone is moved. A correct DoubleFlip gesture results in the smooth transition of the note mapped to the y-axis, ranging from $\approx$ 16dB to $\approx$ 80dB. Notes mapped to axes along which or around which movement is undesirable are not played initially (e.g., the x- and z-axes). However, these notes are played once a threshold is passed, indicating error in the associated direction (e.g., 15° around either undesired axis). For example, see Figure 3(c).

Wandering Volume (WV). This prototype uses changes in volume to give the user full awareness of his or her movement. Feedback consists of playing all notes mapped to an axis (e.g., C^4, 60 MIDI; A^4, 65 MIDI; and F^4, 69 MIDI). Deviation from the reference gesture causes each note mapped to an affected direction to independently decrease in volume. Correct gestures result in all notes being played continuously without volume change. For example, see Figure 3(d).

This system maps each axis to one of three distinct notes composing a major chord that meets the requirements of all the methods mentioned earlier. 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 (C^4, 60 MIDI). A major chord was chosen because of its tendency to generate a positive effect [Cook 2007] when resolving from an error chord (i.e., the chord heard due to a deviation in one or more axes) 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.

Prototypes WP and WV were rejected during the initial design process due to difficulty discerning differences between the changes in audio feedback. Specifically, developers were unable to accurately identify the axes along which or around which undesirable movement was occurring, and were concerned that participants would be overwhelmed by the amount of feedback presented by those methods. The feasibility of options AP and AV were determined by eliciting feedback from users.

3.2.2 *Design and Procedure.* This evaluation study consisted of each participant using one of two feedback techniques (AV and AP) to perform a correct DoubleFlip gesture. Participants were randomly assigned to each technique. The number of participants in each group was counterbalanced. The study began with the participant listening to a verbal description of the gesture and explanation of the technique. Participants were encouraged to listen to the reference at least once before attempting the gesture. The reference gesture was available to be played throughout the study at the user's discretion. Each participant performed the DoubleFlip gesture while undertaking a think-aloud protocol. Since this was our first study, a think-aloud protocol was employed to present participants with an opportunity to call our attention to any additional issues with the feedback mechanism. To prevent undue frustration, participants were stopped if they could not complete a gesture within 10 minutes.

3.2.3 *Apparatus and Participants.* This system was developed in Java using the Android SDK [Google Inc. 2013] and libpd library [Create Digital Music 2012]. The study was performed using an LG Nexus 4 smartphone running Android 4.2. Eight participants aged 20 to 64 ($\mu = 31.0$, $\sigma = 14.9$, four females, one left-handed) were recruited using a departmental email list. No participants reported having any hearing impairments or prior experience using motion gestures.

3.2.4 *Results.* In one instance, a user was unable to discern correct gestures from incorrect gestures using AV due to the similarity of high-volume notes ($\approx 72dB\ to\ \approx 80dB$). Additionally, an older participant using AV reported difficulty discerning between differences in volume, especially for low volumes ($\approx 16dB\ to\ \approx 28dB$). AP did not suffer from either of these problems, and one participant using AP reported that the task "seemed very easy."

3.2.5 *Discussion.* We observed that participants had difficulty using the AV prototype, especially when the feedback was at the edges of the volume spectrum. This limits the range of audio characteristics that can be mapped to movement, which is problematic since a broad range provides more room for discernible feedback variations. In contrast, participants did not encounter similar problems using the AP prototype. As a result of this exploration study, prototype AV was discarded. Additive Pitch was used to provide continual feedback in the remaining studies outlined in Sections 3.3 through 3.5 and Section 4.

3.3 Pilot Study: Evaluating Continual Feedback

The goal of this pilot study was to evaluate the effectiveness and feasibility of using continual concurrent audio feedback assist in learning and performing motion gestures with a smartphone.

3.3.1 *Design and Procedure.* For this study, participants were asked to perform five correct DoubleFlip gestures using two implementations of our feedback technique: *Prototype*, which provided continuous feedback using Additive Pitch, and *Control*, which omitted continuous feedback and, consequently, did not provide a reference. Participants were randomly assigned to one of two groups in order to determine which technique (Prototype or Control) they would use first. The number of participants in each group was counterbalanced.

The study began with the participant listening to a verbal description of the gesture and the first technique. Participants using the audio system were encouraged to listen to the reference at least once before attempting the gesture. The reference gesture was available to be played throughout the study at these users' discretion. Participants were then asked to complete five gestures. To prevent undue frustration, participants were stopped if they could not complete a gesture within 5 minutes. Then, participants repeated the task using the second technique. Finally, users participated in a brief (5- to 10-minute) semistructured interview in which they were asked to identify the most helpful technique for learning the gesture.

3.3.2 *Apparatus and Participants.* Prototype was developed and run on the same hardware and software as our previous study. Thirty-two participants aged 18 to 55 ($\mu = 22.9$, $\sigma = 7.7$, six females, three left-handed) took part in the study. Participants were affiliated with a local university. No participants reported having any hearing impairments or prior experience using motion gestures.

3.3.3 *Results.* Two participants who initially used the Control technique were unable to correctly perform a DoubleFlip gesture, but were able to complete the required five gestures using Prototype. Both participants requested to stop their Control trial early out of frustration. One participant was unable to complete a gesture using either technique. The majority of our participants (90.63%) were able to use both techniques to accomplish the task, suggesting both provided adequate feedback.

When asked which technique they preferred, 26 out of 32 participants (81.25%) indicated a preference for Prototype, while two participants preferred the Control technique and four participants had no preference. A CHI-squared test showed that technique order had no significant effect on preference. Participants stated that Prototype was especially helpful when determining the direction and magnitude in which to rotate the phone. Several participants commented that Prototype was more helpful because it provided “more complete feedback.” Additionally, one participant reported imagining the sounds generated by Prototype while subsequently using the Control technique.

3.3.4 *Discussion.* Results from this pilot study indicated that while both Control and Prototype provide adequate feedback to users, users prefer continuous feedback. Although temporal constraints were not imposed during this study, we observed that participants attempted to move their phone in such a way as to mimic the speed of the change in pitch presented by the reference gesture, in addition to replicating the pitches themselves. Participants used the timing of the pitches’ directional shift to signal when they should change the direction of the phone’s movement. This calls into question whether or not a need to provide an explicit temporal constraint exists—our observation implies that the implicit temporal information provided by listening to the reference gesture may be sufficient. The strict enforcement of temporal constraints was investigated in the following exploration study.

3.4 Exploration Study: Enforcement of Strict Temporal Constraints Using Time Limits

Considering that motion gestures must be performed by the user in a time-dependent manner, it is important to ensure that information regarding the temporal aspect of the gestures is adequately communicated to the user. The goal of this exploration study was to investigate the potential for including temporal feedback by imposing strict time limits.

3.4.1 *Design and Procedure.* Our prototype was modified to examine hard temporal constraints during audio feedback. This “timed” version added a constraint that required the user to complete the gesture within 3 seconds. Since the provided reference gesture was 2 seconds long, 3 seconds was considered sufficient time to complete the gesture. Anything longer might result in high false-positive rates. If a user failed to complete a gesture within the allotted time, the application stated “out of time.”

Participants were asked to use this feedback technique to perform a single correct DoubleFlip gesture. The study began with the participant listening to a verbal description of the gesture and explanation of the technique. Participants were encouraged to listen to the reference at least once before attempting the gesture. The reference gesture was available to be played throughout the study at the user’s discretion. 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.

3.4.2 *Apparatus and Participants.* The prototype was developed and run on the same hardware and software as our previous studies. Four participants aged 20 to 39 ($\mu = 25.5$, $\sigma = 9.1$, two females, one left-handed) were recruited using a departmental email list. No participants reported having any hearing impairments.

3.4.3 *Results.* Participants using the timed version of the prototype overwhelmingly expressed frustration regarding not having enough time to learn the gesture. No participants were able to complete a gesture.

3.4.4 *Discussion.* As a result of the frustration expressed by participants in the exploration study, it became clear than an alternative to enforcing hard time constraints was needed to express temporal information. It is possible that all of the participants in this study found the DoubleFlip gesture to be too difficult to perform. However, we hypothesize that users in this study were unable to perform the gesture as a result of the strict enforcement of temporal constraints, given the low failure rate exhibited by our previous study (Section 3.2). Since we observed participants in the previous study attempting to match the speed of the reference gesture while using the prototype, we explored the incorporation of implied temporal information in the following pilot study.

3.5 Exploration Study: Using Tempo to Include Temporal Information

Given our observations of participants during the previous study, the goal of this exploration study was to determine whether the audio representation of a reference gesture created by Glissando provides sufficient temporal constraints to ensure that a user performs motion gestures in a time-dependent manner, without enforcing strict time limits.

3.5.1 *Design and Procedure.* Participants were asked to perform a DoubleFlip gesture five times correctly using one of four techniques: *Control* (an implementation of our prototype that both omitted continuous feedback, thus providing no implied temporal information, and meaningful error messages—instead, the Control prototype mimicked a real-world situation where the user only knows whether the gesture was recognized or not), *Slow* (implied gesture time of 4 seconds), *Medium* (implied gesture time of 2 seconds, identical to the speed of the representation in our previous pilot study), and *Fast* (implied gesture time of 0.5 seconds). The reference gesture representations for Slow and Fast were obtained by scaling the original reference gesture representation from our previous pilot study to the desired duration. The number of participants using each technique was counterbalanced.

Participants first listened to a verbal description of the DoubleFlip gesture and application use. Participants in the experimental groups were instructed to first listen to the reference and then match the sound and speed of the reference gesture. The reference gesture was available to be played throughout the study at the user's discretion. Participants in the Control group were simply asked to perform the gesture. To prevent undue frustration, participants were stopped if they could not complete a gesture within 3 minutes. Each gesture attempt was timed, in milliseconds, by the application.

3.5.2 *Apparatus and Participants.* The prototypes were developed and run on the same hardware and software as our previous studies. Sixty-eight participants aged 18 to 61 ($\mu = 25.8$, $\sigma = 9.1$, 14 females, three left-handed) took part in the study. Participants were affiliated with a local university. No participants reported having any hearing impairments.

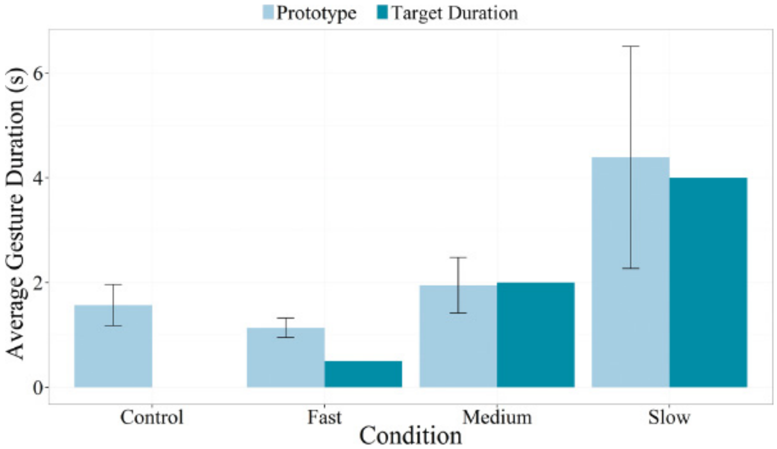

Fig. 4. Mean duration and corresponding reference gesture duration (where appropriate), in milliseconds, by condition. Whisker bars indicate one standard deviation.

3.5.3 *Results.* Figure 4 illustrates average gesture duration, in seconds, by condition. As shown in Figure 4, SlowP resulted in gestures with the longest duration ($\mu = 4.37s$, $\sigma = 2.14s$) followed by MediumP ($\mu = 1.94s$, $\sigma = 0.52s$), Control ($\mu = 1.65s$, $\sigma = 0.65s$), and FastP ($\mu = 1.14s$, $\sigma = 0.19s$).

Factorial analysis of variance (ANOVA) was performed on technique (Control, Slow, Medium, Fast), gesture attempt number (first attempt, second attempt, etc.), and the average duration of each individual's correct gestures. We observed a significant main effect for condition on gesture duration ($F_{3,314} = 136.4$, $p < 0.001$), but no significant main effect for gesture attempt number on gesture duration ($F_{19,260} = 0.748$, $p > 0.1$). Post hoc comparisons using Bonferroni correction showed a significant difference on gesture duration between all prototype conditions ($p < 0.001$ in all cases). It also showed the Control technique to be significantly faster than Slow ($p < 0.001$). However, there was no significant difference between Control and Fast ($p > 0.3$) or Medium ($p > 0.7$).

Four participants were unable to perform a correct DoubleFlip gesture within 3 minutes. However, the majority of participants (94.12%) were able to complete the task.

3.5.4 *Discussion.* Our observations regarding gesture duration in the previous study indicate that the prototype's audio representations of motion gestures significantly influenced the speed at which users attempted to perform a gesture. It is important to note that while our results show that there is no significant difference between the Control technique and Fast or Medium, this is acceptable since it is natural for users performing the gesture without being prompted for speed to achieve gestures with durations somewhere between very slow (as in Slow) and very fast (as in Fast). Additionally, our results show that the difference between Fast and its reference is larger than the differences between Medium and Slow and their respective references. This is likely because the reference gesture for Fast is exceedingly short (0.5 seconds) and therefore presumably too quick for users to reproduce accurately. The fact that the observed gestures for Fast were significantly shorter than the corresponding gestures for Medium is sufficient to indicate that the speed of the reference gesture had the desired effect.

This indicates that the speed at which participants perform motion gestures can be manipulated by changing the speed of the reference gesture, which provides a method of ensuring that motion gestures are performed in an appropriately timely manner without either enforcing strict time limits or including an additional characteristic, such as amplitude [Andersen and Zhai 2008], in the audio feedback.

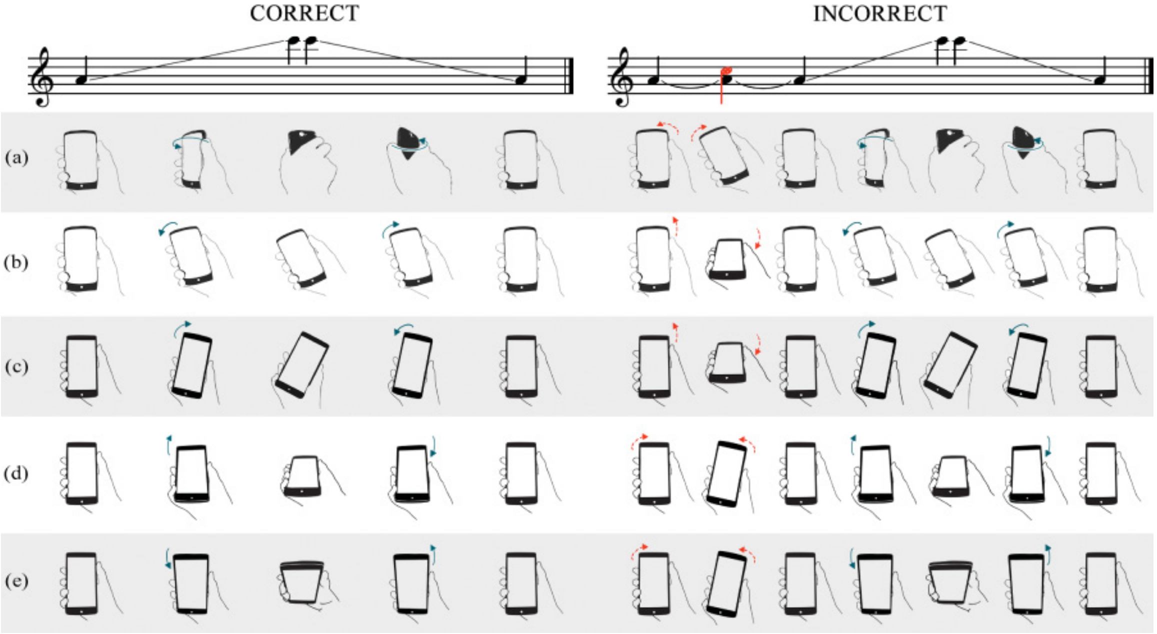

Fig. 5. Examples of feedback using final audio characteristics for a correct (left) and incorrect (right) gesture: (a) DoubleFlip, (b) FlickLeft, (c) FlickRight, (d) FlickDown, and (e) FlickUp.

4. EVALUATING GLISSANDO

We refined the prototypes used in our preliminary studies (Section 3) to create *Glissando*, which we then adapted to provide feedback for additional gestures inspired by a previous elicitation study (Figure 5) [Ruiz et al. 2011]. Glissando functions by using the Additive Pitch (Section 3.2) method to manipulate audio characteristics in order to convey temporal and spatial information. Tempo is used to imply soft temporal constraints in lieu of strict time limits (Sections 3.4 and 3.5)—users don't receive explicit feedback about the duration of their gesture. Glissando's reference was modified to include a short video demonstrating a perfect gesture being performed along with the audio representation in order to facilitate comparison to instructional videos, which, at the time this research was conducted, were the nearest thing to a training technique that did not require additional hardware.

The goal of our final user study was to evaluate Glissando by examining memorability by comparing error rates and temporal deviation of recalled gesture (defined by Equations (1) and (2)) for users trained with and without the system. Specifically, we hypothesized that:

(H1) Using Glissando would produce a lower error rate than not using Glissando, in a Control condition.
(H2) Using Glissando would result in smaller temporal deviation than the Control condition.
(H3) Participants would prefer Glissando to the Control condition.
(H4) Using Glissando would show persistent results, 1 week later.

4.1 Design and Procedure

For this study, participants were trained to perform each of the five gestures shown in Figure 5 five times correctly while using one of two techniques: Glissando and *Control*. In this case, the Control technique was an implementation of Glissando that omitted continuous feedback, and as a consequence omitted temporal information, and replaced detailed verbal feedback with either *correct* or *incorrect* to better approximate performing the gestures in a real-world scenario where users only know whether or not their input was accepted. The Control technique was designed in this way since, at the time this research was conducted, there were no other training techniques for motion gestures that did not require additional hardware (such as a Kinect [Kamal et al. 2014]).

Glissando's reference was modified to include a short video demonstrating a perfect gesture being performed along with the audio representation. The reference provided by the Control technique displayed the same videos as Glissando, but without the corresponding audio representations. The training session was separated into five tasks, one for each gesture, with a corresponding implementation of Glissando or Control that was tailored to that specific gesture, including a corresponding reference. References were available to be played throughout the study at the user's discretion. Participants first listened to a verbal description of application use and then were asked to perform each task.

After completion of the training session, participants in the Control group were asked to rate the helpfulness of the video demonstration in learning the movement and timing of the gesture. To do this, participants answered six Likert-type questions using a visual analog scale ranging from 0 to 10, with 0 being "strongly disagree" and 10 being "strongly agree." Participants in the Glissando group were given an additional six questions to rate the audio feedback. Both groups were asked to rate the likeliness that they would use the technique to help them learn motion gestures.

Participants were then asked to return 7 days later and again perform each of the five gestures illustrated in Figure 2 five times correctly, in the same order. This return task was required in order to assess how well the gestures had been put into long-term memory. For this task, all participants were given a version of the Control technique that did not provide a reference in order to best approximate performing the gesture in a real-world scenario. The return session was separated into five tasks, one for each gesture, with a corresponding implementation of Control that was tailored to that specific gesture.

After completion of the return session, participants were asked to rate the helpfulness of the training session in learning the gestures, the easiness of learning the gestures, and the easiness of performing the gestures by answering four Likert-type questions using the same visual analog scale from the initial questions.

Participants were randomly assigned to each technique. The number of participants using each technique was counterbalanced. As this was a between-subjects design, participants performed each gesture in the same order for both the training and return session, FlickLeft, FlickUp, DoubleFlip, FlickRight, and then FlickDown, so that potential learning effects would average out. To prevent undue frustration, participants were stopped if they could not complete a gesture within 5 minutes. Each gesture attempt was timed, in milliseconds, by the application.

4.2 Apparatus and Participants

Glissando was developed using the same software as our previous studies. The study was performed using an LG Nexus 5 smartphone running Android 4.4. Thirty-eight participants aged 18 to 40 ($\mu = 21.66$, $\sigma = 4.8$, 10 females, three left-handed) took part in the study. Participants were affiliated with a local university. No participants reported having any hearing impairments.

4.3 Results

For each gesture, we calculated the error rate (ER) as

$$ER = \frac{number\_of\_incorrect\_gestures}{number\_of\_attempts}. \tag{1}$$

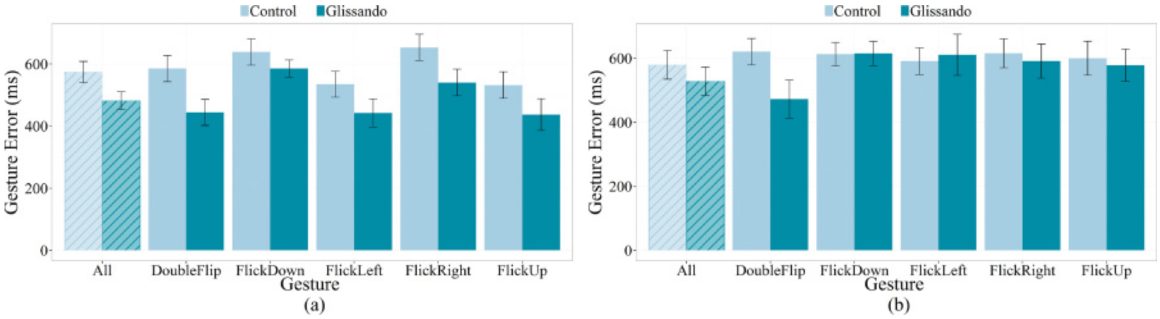

Fig. 6. Temporal deviation (TD) for (a) training session and (b) return session, in milliseconds, by condition and gesture. Error bars represent standard error.

We also calculated the temporal deviation of recalled gesture (TD) as

$$TD = |(user\_gesture\_length) - (ideal\_gesture\_length)|. \quad (2)$$

4.3.1 *Training Session Quantitative.* We observed a mean ER of 11.7% ($\sigma = 13.4\%$) for the Control group and 9.0% ($\sigma = 10.1\%$) for Glissando. We did not observe a significant effect for condition or gesture on error rate. Figure 6 illustrates TD (in milliseconds) by condition and gesture for the training session. As shown in the figure, use of the Glissando technique resulted in gestures with smaller temporal deviation from the reference gestures. ANOVA performed on TD indicated a significant main effect for condition on ($F_{1,36} = 21.03$, $p < 0.001$). We did not observe a main effect for gesture performed on TD ($F_{4,144} = 0.37$, $p > 0.8$). We found no correlation between number of attempts and temporal deviation.

4.3.2 *Return Session Quantitative.* The ER for the Control group ($\mu = 9.7\%$, $\sigma = 8.0\%$) and Glissando group ($\mu = 9.6\%$, $\sigma = 7.0\%$) were nearly identical. Similar to the training session, use of the Glissando technique resulted in gestures with smaller temporal deviation from the reference gestures for the return session (shown in Figure 6). ANOVA performed on TD indicated a significant main effect for condition on TD ($F_{1,36} = 6.78$, $p < 0.05$). Pairwise comparisons using post hoc analysis by condition*gesture indicated that the effect is spread across all gestures/conditions ($p > .95$ in all cases). Again, we did not observe a main effect for gesture performed on TD ($F_{4,144} = 0.13$, $p = 1.0$).

4.3.3 *Qualitative.* Participants favorably rated the helpfulness of the training session (mean = 7.34 for Glissando, mean = 7.24 for Control), as shown in Figure 7. In addition, participants indicated their likelihood of using the application in the future as neutral to moderately positive (mean = 6.46 for Glissano, mean = 4.80 for Control). We found no differences between conditions in participant ratings of technique helpfulness ($T_{33.28} = -0.18$, $p = 0.91$), easiness of learning the gestures ($T_{35.96} = -0.07$, $p = 0.91$), easiness of performing the gestures ($T_{31.70} = -1.81$, $p = 0.08$), or likelihood of future use ($T_{34.84} = -0.79$, $p = 0.44$). However, as shown in Figure 7, we collected a large range of responses from participants in the Control group with regard to likelihood of future use, including some negative responses, while responses from participants using Glissando were generally positive.

4.3.4 *Discussion.* Although our results fail to support H1, that Glissando would produce a lower error rate than the Control condition, using Glissando produced a smaller temporal deviation than the Control condition, supporting H2 that Glissando would result in smaller temporal deviation than the Control condition. Additionally, while our results failed to support H3, that participants would prefer Glissando to the Control condition, responses from participants using Glissando were more

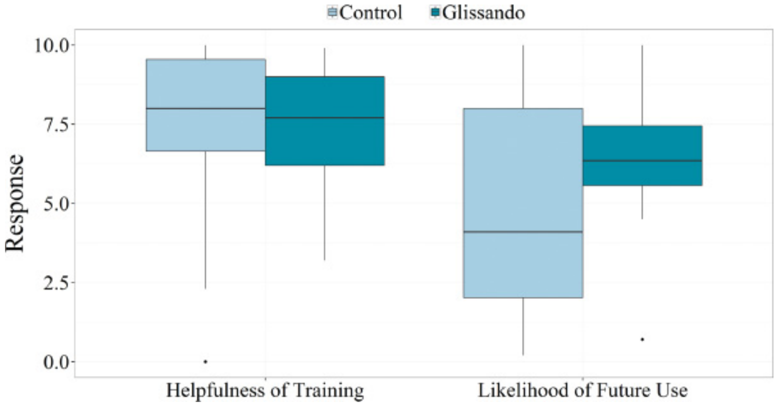

Fig. 7. Helpfulness of the training session and likelihood of future use by condition.

consistently positive, as shown in Figure 7. Since Glissando resulted in gestures with smaller temporal deviation during both the training and return sessions, our results support H4, that using Glissando would show persistent results, 1 week later.

Although the participants in the Glissando group rated the audio feedback neutrally ($\mu = 6.36$, $\sigma = 2.48$ for "I found the audio feedback helpful"), technique seemed to have an unconscious significant effect on users' ability to match the timing of the gestures. This indicates that adding audio feedback conveys temporal information better than visual demonstration alone. This is significant because motion gestures heavily rely on temporal information to discriminate noise from input.

5. DISCUSSION

5.1 Implications for Designing Audio Feedback for Motion Gestures on Mobile Devices

In this article, we presented several user studies that examined appropriate audio characteristics for spatial representation, effectiveness of continual audio feedback, effect of enforcing strict temporal constraints, incorporation of implied temporal information, and effectiveness of audio feedback in assisting memorability. Together, the findings of these studies presented in this article provide insight into what developers need to consider when designing an audio feedback system for training users to use motion gestures on mobile devices:

(1) **Feedback should be designed with the limitations of current-generation smartphones in mind since distortion can interfere with the user's ability to receive feedback.** This was exemplified during the initial design process of Glissando, when differences in audio characteristics could not be discerned for Wandering Pitch and Wandering Volume due to the quality of the device's internal speaker. Furthermore, observations during the initial exploration study indicate that users become frustrated when they can't hear or understand feedback and want to quit attempting to learn the gesture.

(2) **Feedback should avoid excessive use of volume, as users may have difficulty hearing or discerning between volumes at the edges of the spectrum.** Results from the initial exploration study showed that two users had severe difficulty discerning between differences in very high and very low volumes. It is therefore important to control the use of volume since overuse will likely lead to user frustration and inhibit the adoption of motion gestures.

(3) **Developers should refrain from imposing strict time limits on users without providing additional assistance in learning the gesture.** Our second exploration study demonstrated that users became overwhelmingly frustrated with strict time limits when attempting to learn the gesture for the first time. Furthermore, participants in this study were unable to complete gestures while strict time limits were imposed. It was observed that, in part, users appeared to have difficulty with the time limits because they were still trying to learn the spatial aspect of the gesture. For this reason, we highly recommend that developers avoid imposing strict time limits on users who are unfamiliar with the gesture in question.

(4) **Developers should consider providing continual feedback when teaching motion gestures as users strongly prefer the inclusion of continual feedback to receiving feedback only after making an attempt.** We believe that this is particularly important when teaching gestures such as DoubleFlip that require users to meet a specific threshold before changing direction. It was observed during our evaluation of continual feedback and incorporation of temporal information that users frequently were unable to tell when they had rotated the phone far enough without continual feedback. Furthermore, we observed that users who were unfamiliar with the gesture often used Glissando's continuous feedback to determine in which direction they should begin movement. Additionally, there were instances where users were unable to perform a gesture without continual feedback, but could perform the gesture with continual feedback. Finally, our final studies indicated that temporal information could be imparted through the use of continual feedback.

Our preliminary evaluations indicate that this system is a strong technique for providing feedback and assisting users in learning motion gestures. Furthermore, since this project's feedback relies only on the smartphone and all provided instructions can be easily recorded and stored on the device for playback by the user, our system is suitable for use outside of a research laboratory. Although initial prototypes and evaluations were performed using only the DoubleFlip gesture, our final evaluation demonstrates that Glissando can easily be applied to other gestures. In light of this, we hypothesize that this system has the potential to help benefit millions of smartphone users by promoting the mainstream adoption of motion gestures.

5.2 Limitations

Although our final studies indicate that continuous feedback can be successfully used to convey temporal information, strict temporal constraints were not imposed. Further research will need to be done to determine whether continuous feedback can be used in conjunction with other techniques to teach users to perform gestures that meet specific time requirements.

Additionally, considering that Glissando is primarily an auditory feedback system, the potential exists for difficulty during use in environments with high levels of surrounding noise. Since Glissando is designed for use in a user-selected training environment as opposed to everyday use, it is not strictly necessary that the system be able to operate flawlessly in highly imperfect situations. Still, it is beneficial to evaluate its effectiveness in a variety of common environments.

6. CONCLUSION AND FUTURE WORK

In this article, we explored the use of audio characteristics to provide spatial and temporal feedback to users performing motion gestures. We described and evaluated a technique for motion gesture input, Glissando, which used audio to provide feedback on the system's interpretation of user input. This technique enables feedback by verbally confirming correct gestures and notifying users of errors, in addition to providing continuous feedback by mapping distinct musical notes to each of three axes and manipulating pitch to specify both spatial and temporal information. Extra effort was used to support all design decisions on how to present audio feedback for motion gestures on mobile devices through experimentation. Results from our first pilot study demonstrated that Glissando provided adequate feedback to users both with and without continuous feedback, though provision of continuous feedback is more preferred. Our second exploration study and pilot study show that while users have difficulty with strict time limits, temporal information can be provided via Glissando's continual audio feedback by manipulating the tempo of the reference gesture. Our final study shows that adding audio feedback conveys temporal information better than visual demonstration alone.

6.1 Future Work

Further work includes evaluating Glissando by comparing user performance during ideal and distracted use (e.g., walking) after using Glissando and other scaffolding techniques. In addition, we intend to identify the expected duration of Glissando's effect on users' performance of motion gestures and its effect on long-term accurate retention. Moreover, we plan to extend Glissando to assist users learning to use other input modalities, such as midair gestures. 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.

REFERENCES

Tue Haste Andersen and Shumin Zhai. 2008. "Writing with music": Exploring the use of auditory feedback in gesture interfaces. *ACM Transactions on Applied Perception* 7, 3, Article 17 (June 2008), 24 pages. DOI:http://dx.doi.org/10.1145/1773965.1773968

Daniel Ashbrook and Thad Starner. 2010. MAGIC: A motion gesture design tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM, New York, NY, 2159–2168. DOI:http://dx.doi.org/10.1145/1753326.1753653

Joel F. Bartlett. 2000. Rock 'n' scroll is here to stay [user interface]. *IEEE Computer Graphics and Applications* 20, 3 (May 2000), 40–45. DOI:http://dx.doi.org/10.1109/38.844371

Olivier Bau and Wendy E. Mackay. 2008. OctoPocus: A dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST'08)*. ACM, New York, NY, 37–46. DOI:http://dx.doi.org/10.1145/1449715.1449724

Andrew Bragdon, Robert Zeleznik, Brian Williamson, Timothy Miller, and Joseph J. LaViola, Jr. 2009. GestureBar: Improving the approachability of gesture-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, New York, NY, 2269–2278. DOI:http://dx.doi.org/10.1145/1518701.1519050

Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*. ACM, New York, NY, 473–480. DOI:http://dx.doi.org/10.1145/642611.642694

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

Create Digital Music. 2012. libpd. Retrieved from http://libpd.cc/.

William Edwards. 2010. *Motor Learning and Control: From Theory to Practice*. Cengage Learning, Belmont, CA.

Parisa Eslambolchilar, Andrew Crossan, and Roderick Murray-Smith. 2004a. Model-based target sonification on mobile devices. In *Proceedings of the International Workshop on Interactive Sonification (ISon'04)*. Interactive Sonification Organisation.

Parisa Eslambolchilar, John Williamson, and Rod Murray-Smith. 2004b. Multimodal feedback for tilt controlled speed dependent automatic zooming. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST'04)*. ACM.

Claude Ghez, Thanassis Rikakis, R. Luke Dubois, and Perry R. Cook. 2000. An auditory display system for aiding interjoint coordination. In *Proceedings of the International Conference on Auditory Displays (ICAD'00)*.

Google Inc. 2013. Android Open Source Project. Retrieved from http://www.android.com.

Susumu Harada, Hironobu Takagi, and Chieko Asakawa. 2011. On the audio representation of radial direction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, 2779–2788. DOI:http://dx.doi.org/10.1145/1978942.1979354

Beverly L. Harrison, Kenneth P. Fishkin, Anuj Gujar, Carlos Mochon, and Roy Want. 1998. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, 17–24. DOI:http://dx.doi.org/10.1145/274644.274647

Björn Hartmann, Leith Abdulla, Manas Mittal, and Scott R. Klemmer. 2007. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, NY, 145–154. DOI:http://dx.doi.org/10.1145/1240624.1240646

Ken Hinckley, Jeff Pierce, Mike Sinclair, and Eric Horvitz. 2000. Sensing techniques for mobile interaction. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST'00)*. ACM, New York, NY, 91–100. DOI:http://dx.doi.org/10.1145/354401.354417

Apple Inc. 2009. iPhone User Guide For iPhone OS 3.1 Software. (2009). https://manuals.info.apple.com/MANUALS/0/MA616/en_US/iPhone_iOS3.1_User_Guide.pdf.

Eleanor Jones, Jason Alexander, Andreas Andreou, Pourang Irani, and Sriram Subramanian. 2010. GesText: Accelerometer-based gestural text-entry systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM, New York, NY, 2173–2182. DOI:http://dx.doi.org/10.1145/1753326.1753655

Ankit Kamal, Yang Li, and Edward Lank. 2014. Teaching motion gestures via recognizer feedback. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI'14)*. ACM, New York, NY, 73–82. DOI:http://dx.doi.org/10.1145/2557500.2557521

Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'08)*. ACM, New York, NY, 73–80. DOI:http://dx.doi.org/10.1145/1414471.1414487

Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Access overlays: Improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST'11)*. ACM, New York, NY, 273–282. DOI:http://dx.doi.org/10.1145/2047196.2047232

Frank Chun Yat Li, David Dearman, and Khai N. Truong. 2009. Virtual shelves: Interactions with orientation aware devices. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology (UIST'09)*. ACM, New York, NY, 125–128. DOI:http://dx.doi.org/10.1145/1622176.1622200

Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. 2009. User evaluation of lightweight user authentication with a single tri-axis accelerometer. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'09)*. ACM, New York, NY, Article 15, 10 pages. DOI:http://dx.doi.org/10.1145/1613858.1613878

Joanna Lumsden and Stephen Brewster. 2003. A paradigm shift: Alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research (CASCON'03)*. IBM Press, 197–210.

Georgios Marentakis and Stephen A. Brewster. 2005. A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices &Amp; Services (MobileHCI'05)*. ACM, New York, NY, 55–62. DOI:http://dx.doi.org/10.1145/1085777.1085787

Georgios N. Marentakis and Stephen A. Brewster. 2004. A study on gestural interaction with a 3d audio display. In *Proceedings of the 6th International Symposium on Mobile Human-Computer Interaction (Mobile HCI'04)*. 180–191. DOI:http://dx.doi.org/10.1007/978-3-540-28637-0_16

Georgios N. Marentakis and Stephen A. Brewster. 2006. Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*. ACM, New York, NY, 359–368. DOI:http://dx.doi.org/10.1145/1124772.1124826

Christian Müller-Tomfelde and Sascha Steiner. 2001. Audio-enhanced collaboration at an interactive electronic whiteboard. In *Proceedings of the 7th International Conference on Auditory Display (ICAD'01)*. 267–271.

Matei Negulescu, Jaime Ruiz, and Edward Lank. 2012a. A recognition safety net: Bi-level threshold recognition for mobile motion gestures. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'12)*. ACM, New York, NY, 147–150. DOI:http://dx.doi.org/10.1145/2371574.2371598

Matei Negulescu, Jaime Ruiz, Yang Li, and Edward Lank. 2012b. Tap, swipe, or move: Attentional demands for distracted smartphone input. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI'12)*. ACM, New York, NY, 173–180. DOI:http://dx.doi.org/10.1145/2254556.2254589

Uran Oh, Shaun K. Kane, and Leah Findlater. 2013. Follow that sound: Using sonification and corrective verbal feedback to teach touchscreen gestures. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'13)*. ACM, New York, NY, Article 13, 8 pages. DOI:http://dx.doi.org/10.1145/2513383.2513455

Kurt Partridge, Saurav Chatterjee, Vibha Sazawal, Gaetano Borriello, and Roy Want. 2002. TiltType: Accelerometer-supported text entry for very small devices. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology (UIST'02)*. ACM, New York, NY, 201–204. DOI:http://dx.doi.org/10.1145/571985.572013

Beryl Plimmer, Peter Reid, Rachel Blagojevic, Andrew Crossan, and Stephen Brewster. 2011. Signing on the tactile line: A multimodal system for teaching handwriting to blind children. *ACM Transactions on Computer-Human Interaction* 18, 3, Article 17 (Aug. 2011), 29 pages. DOI:http://dx.doi.org/10.1145/1993060.1993067

Jun Rekimoto. 1996. Tilting operations for small screen interfaces. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (UIST'96)*. ACM, New York, NY, 167–168. DOI:http://dx.doi.org/10.1145/237091.237115

Jaime Ruiz and Yang Li. 2011. DoubleFlip: A motion gesture delimiter for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, 2717–2720. DOI:http://dx.doi.org/10.1145/1978942.1979341

Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, 197–206. DOI:http://dx.doi.org/10.1145/1978942.1978971

Gerd Schmitz, Daniela Kroeger, and Alfred O. Effenberg. 2014. A mobile sonification system for stroke rehabilitation. In *The 20th International Conference on Auditory Display (ICAD'14)*. New York, NY, USA.

David Small and Hiroshi Ishii. 1997. Design of spatially aware graspable displays. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems (CHI EA'97)*. ACM, New York, NY, 367–368. DOI:http://dx.doi.org/10.1145/1120212.1120437

Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: Projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. ACM, New York, NY, 179–188. DOI:http://dx.doi.org/10.1145/2207676.2207702

Martin Talbot and William Cowan. 2009. On the audio representation of distance for blind users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, New York, NY, 1839–1848. DOI:http://dx.doi.org/10.1145/1518701.1518984

Isaac Wallis, Todd Ingalls, Thanassis Rikakis, Loren Olsen, Yinpeng Chen, Weiwei Xu, and Hari Sundaram. 2007. Real-time sonification of movement for an immersive stroke rehabilitation environment. In *Proceedings of the 13th International Conference on Auditory Display*. 497–503.

Lars Weberg, Torbjörn Brange, and Åsa Wendelbo Hansson. 2001. A piece of butter on the PDA display. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems (CHI EA'01)*. ACM, New York, NY, 435–436. DOI:http://dx.doi.org/10.1145/634067.634320

Daniel Wigdor and Ravin Balakrishnan. 2003. TiltText: Using tilt for text input to mobile phones. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST'03)*. ACM, New York, NY, 81–90. DOI:http://dx.doi.org/10.1145/964696.964705

John Williamson and Roderick Murray-Smith. 2002. *Audio feedback for gesture recognition*. Number TR-2002-127 in DCS Tech Report. Department of Computing Science, University of Glasgow.

John Williamson and Roderick Murray-Smith. 2005. Sonification of probabilistic feedback through granular synthesis. *IEEE Multimedia* 12, 2 (2005), 45–52. DOI:http://dx.doi.org/10.1109/MMUL.2005.37

Received July 2015; revised January 2016; accepted February 2016