Audio-Based Feedback Techniques for Teaching Touchscreen Gestures

URAN OH, University of Maryland, College Park
STACY BRANHAM, University of Maryland, Baltimore County
LEAH FINDLATER, University of Maryland, College Park
SHAUN K. KANE, University of Colorado, Boulder

While sighted users may learn to perform touchscreen gestures through observation (e.g., of other users or video tutorials), such mechanisms are inaccessible for users with visual impairments. As a result, learning to perform gestures without visual feedback can be challenging. We propose and evaluate two techniques to teach touchscreen gestures to users with visual impairments: (1) *gesture sonification* to generate sound based on finger touches, creating an audio representation of a gesture; and (2) *corrective verbal feedback* that combined automatic analysis of the user's drawn gesture with speech feedback. To refine and evaluate the techniques, we conducted three controlled laboratory studies. The first study, with 12 sighted participants, compared parameters for sonifying gestures in an eyes-free scenario. We identified pitch+stereo panning as the best combination. In the second study, ten blind and low-vision participants completed gesture replication tasks for single-stroke, multistroke, and multitouch gestures using the gesture sonification feedback. We found that multistroke gestures were more difficult to understand in sonification, but that playing each finger sound serially may improve understanding. In the third study, six blind and low-vision participants completed gesture replication tasks with both the sonification and corrective verbal feedback techniques. Subjective data and preliminary performance findings indicated that the techniques offer complementary advantages: although verbal feedback was preferred overall primarily due to the precision of its instructions, almost all participants appreciated the sonification for certain situations (e.g., to convey speed). This article extends our previous publication on gesture sonification by extending these techniques to multistroke and multitouch gestures. These findings provide a foundation for nonvisual training systems for touchscreen gestures.

Categories and Subject Descriptors: H.5.2 [**User Interfaces**]: Auditory (nonspeech) Feedback; K.4.2 [**Social Issues**]: Assistive Technologies for Persons with Disabilities

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Blindness, visual impairments, sonification, touchscreen, gestures

ACM Reference Format:
Uran Oh, Stacy Branham, Leah Findlater, and Shaun K. Kane. 2015. Audio-based feedback techniques for teaching touchscreen gestures. ACM Trans. Access. Comput. 7, 3, Article 9 (November 2015), 29 pages.
DOI: http://dx.doi.org/10.1145/2764917

1. INTRODUCTION

With the widespread adoption of touchscreen devices, gestural interaction has become a primary means of computer input across a range of devices, from mobile phones

This research was supported by a Google Research Award, and was supported in part by the National Science Foundation under grant IIS-1353312.

Authors' addresses: U. Oh; email: uranoh@cs.umd.edu; S. Branham; email: sbranham@umbc.edu; L. Findlater; email: leahkf@umd.edu; S. K. Kane; email: shaun.kane@colorado.edu.

DOI: http://dx.doi.org/10.1145/2764917

to tablets to personal computers. While gesture-based user interfaces often rely upon on-screen visual representations, gesture-based user interfaces have successfully been adapted into nonvisual gestures for people with vision impairments (e.g., Azenkot et al. [2012], Bonner et al. [2010], Kane et al. [2008, 2011a]). Thus, when properly designed, gestural interaction can be an effective interaction method for users with diverse visual abilities.

Key to designing and deploying effective gesture-based user interfaces is informing users about the available gestures and teaching them to perform these gestures accurately [Bau and Mackay 2008]. Most popular gesture-based user interfaces provide a fixed set of gestures, and offer limited opportunities to customize gestures for specific users. While we may consider gestural interfaces to support a so-called “natural user interface,” touchscreen gestures obey certain conventions that must be learned and performed correctly [Norman 2010]. For example, many current gesture-based user interfaces support directional swipe gestures, in which a user drags their finger across the touchscreen in a specified direction. The system’s underlying gesture recognizer may identify this gesture based on characteristics of the gesture such as its location on screen, the speed of the user’s finger movement, the length of the gesture (i.e., how far the finger travels across the screen), and the angular direction of the gesture. Gestures that do not match the recognizer’s expected parameters may be misclassified or ignored by the system.

Learning how to perform gestures correctly may be especially difficult for users with vision impairments. Sighted users may learn to perform gestures through observing other users, in-application tutorials, or even television commercials. For visually impaired users, such observation is not accessible and, as a result, learning how to draw gestures on a screen can be challenging [Leporini et al. 2012; Plimmer et al. 2011]. Furthermore, visual feedback helps sighted users perform gestures consistently [Anthony et al. 2013]; without this feedback, accurately performing a gesture without sight may be more difficult. While recent commercial and research advances have addressed touchscreen accessibility for users with visual impairments, the learning process for performing gestures without sight has largely been ignored.

In this article, we propose the use of nonverbal and verbal audio feedback for teaching individuals with vision impairments to identify and reproduce gestures. We refer to this nonverbal and verbal feedback as *gesture sonification* and *verbal feedback*, respectively. In gesture sonification, continuous sound is generated as the finger touches the screen (e.g., increasing pitch as the finger moves up), to provide an audio representation of each possible gesture. Users can compare the sound of their own gesture to that of a reference gesture to try to match the two. Sound parameters such as pitch, stereo, volume, or timbre can be mapped to the x and y axes. The verbal feedback technique automatically analyzes a user’s gestures and provides text-to-speech feedback on how to change each gesture to make it more similar to a reference gesture (e.g., “make it longer”). While prior work has used sonification to convey visual information for blind users (e.g., graphs [Brown and Brewster 2003; Zhao et al. 2008] or geometric shapes [Harada et al. 2011; Plimmer et al. 2011]), findings from those studies are not necessarily applicable to touchscreen gestures, which can vary not only in shape, but also in location, direction, speed, size, and number of fingers. Flick and drag gestures, for example, might have the same length and direction but are differentiated by the system based on the speed of the gesture.

To refine and evaluate these gesture-training techniques, we conducted three controlled laboratory studies. First, to identify the best parameters for sonifying gestures, we compared different sonification possibilities with 12 sighted participants in an eyes-free scenario. Among the parameters tested (pitch, volume, timbre, and stereo), stereo on the x axis and pitch on the y axis were significantly more accurate at conveying gestures than any other parameter mapping. In the second study, 10 blind and low-vision participants tested the optimal gesture sonification technique (stereo+pitch) for teaching single-stroke, multistroke, and multitouch gestures. We found that multitouch gestures were more difficult to understand in sonification, but that playing each finger sound serially may improve understanding. In the third study, six blind and low-vision participants compared the two feedback techniques—sonification (using stereo+pitch) and corrective verbal feedback—while performing tap gestures, directional swipe gestures, and shape gestures. Subjective data and performance findings indicate that the two techniques offer complementary advantages. For example, although the verbal feedback was preferred overall primarily due to the precision of its instructions, almost all participants appreciated the sonification feedback for certain situations (e.g., to convey speed).

This article makes the following contributions: (1) two straightforward techniques to provide sonification and verbal feedback to blind users based on reference gestures; (2) evaluation of these sonification techniques across a variety of common touchscreen gestures; (3) empirical evidence to show that pitch+stereo is more effective for gesture sonification than alternatives that use volume or timbre on the x axis; and (4) characterization of the trade-offs between gesture sonification feedback and verbal instructions. These findings provide a foundation for the design of future gesture tutorial systems to improve initial touchscreen learnability for blind users.

This article extends our prior work in developing gesture sonification techniques [Oh et al. 2013]. In that paper, we introduced the first study presented here identifying optimal gesture sonification parameters, and the third study comparing gesture sonification to corrective verbal feedback. This extended article applies our prior work on gesture sonification to more complex touchscreen gestures, including multistroke and multitouch gestures (Study 2). We also include an extended discussion section that reflects upon the three studies and identifies opportunities for applying these gesture feedback techniques to additional use contexts.

2. BACKGROUND

2.1. Accessible Touchscreen Gestures

Adapting touchscreen gestures to support users with vision impairments consists of two major components: identifying which gestures can be performed accurately by blind people, and creating systems that combine accessible gestures to enable users to interact with applications. Relatively few studies have specifically explored gesture production by individuals with vision impairments [Kane et al. 2011b; Plimmer et al. 2011]. Most research projects, including the following, have explored how simple touchscreen gestures such as taps and swipes can be combined to create accessible user interfaces.

While early approaches for accessible gesture-based interaction for people with visual impairments combined touchscreen input with physical buttons for confirmation, more recent systems have focused on the touchscreen only [Bonner et al. 2010; Frey et al. 2011; Guerreiro et al. 2008; Kane et al. 2008; Kane et al. 2011a]. *Slide Rule* [Kane et al. 2008] was a multitouch screen reader for exploring touchscreen applications. Slide Rule enabled users with vision impairments to explore graphical applications by dragging their finger across the screen to explore available options, and enabled the use of directional swipe gestures and tap gestures to make selections and switch between application modes. *NavTouch* [Guerreiro et al. 2008], *No-Look Notes* [Bonner et al. 2010], and *BrailleTouch* [Frey et al. 2011] enabled the use of touchscreen gestures for entering text via directional swipes, a radial pie menu, and multitouch input, respectively.

Table I. Examples of Accessible Gestures Supported by Current Mobile Operating Systems

Operating System	Gesture Description	Gesture Type	Effect of Gesture
Apple iOS	Swipe left or right	Single	Browse menu items
Apple iOS	Double tap	Multistroke	Select menu item
Apple iOS	Tap with four fingers at the top of the screen	Multitouch	Move to first item of list
Apple iOS	Write letters with your finger	Multistroke	Enter text in search fields
Apple iOS	Rotate two fingers	Multitouch	Activate "rotor" menu for adjusting system settings
Google Android	Swipe up then down	Multistroke	Move to first item of list
Google Android	Swipe down then left	Multistroke	Back button

Following from these early research efforts, mainstream mobile operating systems have also incorporated accessible touchscreen gestures. For example, Apple's iOS operating system provides access to accessible gestures through its *VoiceOver* screen reader,[1] and Google's Android operating system provides access to accessible gestures through its *Eyes-Free Project*.[2] These systems typically include gestures such as single and multifinger tap, double tap, tap and hold, and directional drag and swipes. Table I presents examples of accessible gestures currently used in mainstream operating systems; however, it is important to note that these gestures are only a subset of gestures that users with vision impairments may be able to perform [Kane et al. 2011b].

While many current mobile devices provide access to accessible touchscreen gestures, users must still learn which gestures are available, and must also learn how to correctly perform these gestures. Performing a gesture correctly may involve performing a gesture in a specific location, and with a specific size, speed, or direction. If the gesture is not performed within the gesture recognizer's expected parameters, the gesture may be ignored, or may be misrecognized as a different gesture entirely. For example, on iOS, a quick finger movement from left to right is considered a horizontal swipe gesture. This gesture should be performed along the horizontal axis of the screen; if it is performed at an angle, it may be considered a diagonal or vertical swipe instead.

Learning to perform gestures accurately can be especially difficult for users with vision impairments. While a sighted person may learn how to perform a gesture by watching someone else perform that gesture, or by watching a tutorial video, this information is not always accessible to individuals with vision impairments. For example, *VoiceOver* and *Eyes-Free* provide textual descriptions of gestures (e.g., "swipe left"), but do not provide detailed information about the dynamics of the canonical gesture, such as size and speed. *VoiceOver* also provides a practice area for users to perform gestures and to hear which gestures are recognized by the system, but no feedback is provided about how to perform a specific gesture reliably. Even when a user learns to perform a gesture, they may have difficulty performing that gesture on different devices, as there is little consistency between gestures supported by different software platforms [Kane et al. 2011b]. For example, using iOS, the user navigates to the top of a list by tapping the top of the screen with four fingers, while in Android the user performs the same action by swiping down and then up. In general, iOS uses a set of multitouch swipe gestures for navigation, while Android uses both multitouch and multistroke gestures. Even when two systems share common gestures, the details of how that gesture is recognized may be different. For example, iOS and Android sometimes use different time

[1] http://www.apple.com/accessibility/iphone/vision.html.
[2] https://code.google.com/p/eyes-free/.

intervals for determining whether two consecutive taps are recognized as a double-tap gesture.[3]

2.2. Sonification of Spatial Data

Converting spatial information to nonspeech audio for blind people has been explored for many years (see Hoggan and Brewster [2012] for a general overview of nonspeech audio output). Brown and Brewster [2003] combined pitch and stereo to represent line graphs with two data series: each series was represented using pitch, while stereo position was used to separate the series. *iSonic* [Zhao et al. 2008] combined pitch and stereo panning to represent two-dimensional map and table data. Walker and Mauney [2010] explored sonification mappings for blind and sighted readers of auditory graphs, finding that both blind and sighted individuals typically (with some exceptions) applied similar mappings between audio pitch and other variables, such as size and velocity. Walker et al. [2005] found that three-dimensional spatial audio beacons could guide individuals through a map path in a virtual reality environment. While these systems have informed the selection of audio parameters for our studies, sonification to provide an understanding of spatial data is unlike gesture sonification, in that users must also be able to *reproduce* gesture accurately—with details such as location, size, speed, and direction.

2.3. Training Gestures and Shape Drawing

Sighted individuals have many opportunities to learn gestures through visual observation. For example, Apple provides video tutorials for touchpad gestures in Mac OS X,[4] while several research solutions provide continuous gesture recognition and visual guidance in real time (e.g., Bau and Mackay [2008] and Kristensson and Denby [2011]). These approaches are inaccessible to blind users. As an alternative, several projects have used sonification to teach blind users shapes (though not necessarily gestures). GUESS [Kamel et al. 2001], for example, allowed blind users to explore simple shapes using a stylus and tablet with sonification (pitch+stereo) feedback. Timbremap [Su et al. 2010] combined stereo, pitch, and spearcons to guide blind users in exploration of a touchscreen map. Harada et al. [2011] mapped vowel sounds to radial direction to enable blind people to trace shape contours. These systems enabled tracing of shapes, but were primarily optimized for slow exploration of a shape, rather than aspects such as rotation and speed of a gesture. Furthermore, these systems often focused on simple shapes and gestures, and did not consider multiple-stroke or multiple-touch gestures, as we do here.

Multimodal audio and haptic feedback have also been used to convey shapes. Crossan and Brewster [2008] combined pitch and stereo sonification with a force feedback controller to drag the user along a trajectory, and found that performance was higher with audio and haptic feedback than haptic feedback alone. *McSig* [Plimmer et al. 2011] used this same combination of sonification and force feedback to teach handwriting to blind children, while *SemFeel* [Yatani and Truong 2009], *SpaceSense* [Yatani et al. 2012], and work from Noble and Martin [Noble and Martin 2006] used primarily tactile feedback to transmit directional and shape data. These systems used custom hardware with multiple actuators, technology that is not available on most touchscreen devices. We have thus focused on training with audio feedback only.

[3] http://www.quirksmode.org/mobile/clickdelay.html.

[4] http://www.apple.com/osx/what-is/gestures.html.

3. DESIGNING EFFECTIVE NONVISUAL GESTURE FEEDBACK

As previously discussed, performing touchscreen gestures is not necessarily intuitive. Touchscreen gestures have various parameters that must be followed, including location, size, speed of finger movement, and number of fingers involved. Changing these characteristics often results in a different action taken by the system. For example, slowly swiping a finger on screen from right to left may move on-screen content, while swiping more quickly may switch to another page or application. These characteristics may be easily conveyed visually: a novice user may watch an expert's demonstration or tutorial, and may copy the visual performance to reproduce the gesture. Our prior research in this area has shown that people with vision impairments often learn about gestures through the devices' tutorial functions, or by receiving instructions from another person. In both cases, the gesture may be described verbally. Nonvisual verbal description alone, however, is not always adequate; for example, describing a gesture as a "swipe from left to right" indicates the direction of the gesture, but not the speed, size, or location.

To create flexible, widely applicable audio-based techniques that communicate touchscreen gesture characteristics, we identified the following design requirements:

—**Convey multiple gesture parameters**. We explored techniques that could teach multiple aspects of a gesture, including the *position, shape, size, speed, direction,* and *number of fingers.*

—**Provide guidance for multiple types of gestures.** While many touchscreen applications can be controlled with a simple set of taps and swipes, some applications use more complicated gestures, including complex shapes, multiple strokes, and multiple fingers. Apple's VoiceOver screen reader, for example, uses multiple-finger swipes and multiple-finger rotation gestures. Thus, our approach should support simple and complicated gestures, including multistroke and multitouch gestures.

—**Enable independent practice.** Ideally, providing better feedback about how a blind or visually impaired person is performing a gesture could enable that person to learn to use a touchscreen device independently. Thus, our approach should allow the user to practice independently, and provide the user with feedback about how their performance matches the "correct" reference gesture.

—**Use commodity hardware and software**. While touchscreen devices are increasingly common, many individuals may not have access to more complex hardware such as haptic feedback devices or multichannel audio systems. Thus, our approach should be usable on a standard touchscreen device using standard headphones.

Based on these requirements, our early iterative design resulted in two techniques that provide feedback about the user's gesture performance. First, we developed a *gesture sonification* technique that enabled a mapping between touches on the screen and continuous sound output. As the user touches the screen, the system plays a series of tones that correspond to the locations of the touches. The idea is that by learning the mapping between touch and sound, users can first listen to a gesture's sonification, then attempt to replicate that sound by articulating the gesture themselves. Second, we developed a *corrective verbal feedback* approach that could analyze a user's gesture and compare it to the reference gesture. If the user's gesture differed from the reference gesture in size, shape, speed, or location, the system could specifically instruct the user to perform the gesture differently (e.g., to make it larger).

While these two types of feedback present quite different user experiences and may provide complementary benefits, in the following studies we evaluated the approaches separately to identify their relative strengths and weaknesses. Following the presentation of the three studies, we consider how the two types of feedback may be combined.

4. STUDY 1: DETERMINING OPTIMAL EYES-FREE SONIFICATION PARAMETERS

To explore possible forms of gesture sonification feedback, we conducted a controlled laboratory study with 12 sighted participants. We tested different sound parameters (e.g., pitch, timbre) mapped to absolute (x, y) screen coordinates to assess how effectively each parameter conveyed gesture characteristics such as location, size, speed, direction, and shape. We conducted this initial study with sighted participants to achieve a larger sample than possible with blind participants alone, and to refine the sonification technique before presenting it to blind participants in Study 2 and Study 3. Perception of sound mappings has been shown to be largely consistent between blind and sighted people [Walker and Mauney 2010], so we believed that testing with sighted people would provide useful guidance for designing gesture sonification schemes for blind or sighted users.

4.1. Method

4.1.1. Pilot Testing. To identify which sound parameters would be most useful for gesture sonification, we conducted pilot testing with four sighted participants using a variety of audio filter parameters in the *Pure Data* library. We varied the following sound parameters along the x axis: pitch, volume, timbre (tone), stereo, vibrato, attack/decay (time to increase to and decrease from a peak sound), and tempo (beats per minute). Participants provided qualitative verbal feedback about their experiences using each of the sonification methods.

Following this initial testing, we conducted an additional five pilot sessions to identify distinguishable lower and upper ends of the range and step sizes for each parameter, where applicable (i.e., identifying comfortable low and high volume settings).

After pilot testing, we excluded vibrato, attack/decay, and tempo from further evaluation because they each had a temporal component that interfered with conveying gesture speed. The details of the final sonification parameters are described next.

4.1.2. Participants. Twelve sighted volunteers (five female) were recruited through campus email lists. They were on average 26.4 years old (range 20–35). All but one participant owned a touchscreen device; nine reported daily touchscreen use. No participants reported hearing difficulties. Half reported playing a musical instrument.

4.1.3. Apparatus. We used a Samsung Galaxy Nexus running Android 4.2.2 with a display resolution of 124ppcm. We also built a custom Android application, which used *Pure Data*[5] to generate real-time audio based on the (x, y) location of fingers on the screen. Study sessions took place in a quiet room and participants wore closed, supra-aural stereo headphones (Sennheiser HD 202 II). Since the Galaxy Nexus does not have a tactile edge to the screen, we created a physical overlay to demarcate a 700×700px region corresponding to the active input area in the app (Figure 1, left); the overlay also covered the experimenter's controls, preventing accidental selections by the participant. To impose eyes-free interaction, the device and hands were shielded from view inside a box (Figure 1, right). The software logged all interactions with the touchscreen.

4.1.4. Sound Parameters. Following the pilot testing, we considered four parameters: pitch, volume, timbre, and stereo:

—**Pitch.** We varied sound frequency to generate 10 pitch values that correspond to consecutive musical notes near middle C on a piano (261.63Hz). Pitch ranged from a low of B3 (246.94Hz) at the bottom of the screen to D5 (587.33Hz) at the top of

[5] http://puredata.info.

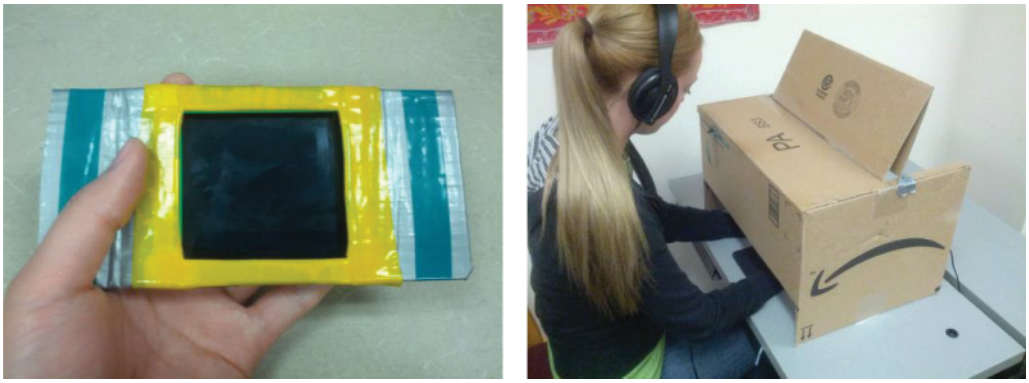

Fig. 1. Study 1 setup. (Left) Since the Samsung Galaxy Nexus screen has no tactile edge, we used a physical overlay to demarcate the active screen area. (Right) To impose eyes-free use, participants placed their hands inside of a box during tasks.

the screen. Moving a finger vertically across the screen effectively played a C major scale.

—**Volume.** To manipulate perceived volume, we adjusted the gain of the amplifier from 0.1 (0 is absolute silence) to 1 (full gain). A step corresponded to a 0.1 increase/decrease in gain.

—**Timbre.** Timbre refers to tone quality. We varied timbre from a pure sine wave (smooth) on the left side of the screen to a pure triangle wave (jagged) on the right side of the screen; we did not use sawtooth or square waves due to their relatively discomforting sounds. To transition from the triangle wave to the sine wave, we perceptually combined the two sound waves by reciprocally adjusting the gain of each—that is, the triangle wave gain (α) decreased uniformly from 1.0 on the left to 0.0 on the right and the sine wave gain correspondingly increased (always $1 - \alpha$).

—**Stereo (pan).** To create the perception of sound panning left-right as the finger moves horizontally, we adjusted the gain in the right and left channels. For a touch point on the left of the screen, no sound played (no gain) in the right channel, and the left channel gain decreased by from a high of 1.0 on the far left to 0.2 near the middle (step size of 0.2 gain); vice versa for the right side.

Of these four parameters, pitch was found to be best for all participants at conveying directionality and the start/end location of swipe gestures. We thus mapped the y axis to pitch for all conditions. Along the x axis we compared three sound parameters: volume, timbre, and stereo. Rather than using a continuous sound change, we instead divided each axis into 10 equal-sized, discrete steps, which made it easier to detect auditory changes.

4.1.5. Procedure. The procedure was designed to fit in a single two-hour session. The three different x-axis sound conditions (*volume*, *timbre*, and *stereo*) were fully counterbalanced and participants were randomly assigned to a presentation order. For each condition, participants began by freely exploring the screen while hearing sound feedback for 30 seconds, then performed several gestures as instructed by the researcher: drawing vertical, horizontal, and diagonal swipes, tapping the four corners and center of the screen, and drawing a few patterns of their choice. Following this practice activity, participants completed four tasks in the following order: *Swipe Length*, *Swipe*

Speed, *Tap Location*, and *Shape*.[6] A fifth task, *Tap Type*, was tested only once at the end of the session (not per condition) because it did not require two-dimensional (2D) sonification.

—**Swipe length.** Sixteen swipe gestures of varying direction and length: eight directions (left, right, up, down, and the four diagonals)×2 lengths (short: 315px, long: 630px). Swipe gestures were centered on screen.
—**Swipe speed.** Sixteen swipe gestures of varying direction and speed: eight directions (same as previously)×2 speeds (fast: 1/2px/ms, slow: 1/6px/ms). Swipe gestures were centered on screen. Speed was calculated as the length of the stroke divided by the time between the beginning and end of the stroke. The strokes were 630px long.
—**Tap location**. Nine tap locations, distributed one per cell across a 3×3 grid filling the entire screen. The location within a cell region was randomly chosen.
—**Shape.** Five single-stroke shapes with varying characteristics (e.g., closed vs. open, curved vs. straight): circle, diamond, small letter "e," capital letter "W," and "$\sum$." While these shape gestures do not directly correspond to existing touchscreen gestures, we included them here because we believe that shape gestures could be an important component of future accessible gesture-based interfaces.
—**Tap type.** Four tap types: single short (200ms) and long (1000ms) taps, and double and triple short taps (with a 400ms gap between taps). Since 2D location and trajectory are not necessary to communicate tap type, we tested this task with only one sound (a midrange pitch and volume).

We included two different swipe tasks (*Swipe Length* and *Swipe Speed*) because swipe gestures are quite common in touchscreen user interfaces. We wished to examine whether participants could accurately reproduce gestures with different lengths and speeds. However, to make the task somewhat easier for participants who were unfamiliar with sonification, we adjusted speed and length separately. For each of these tasks, we were interested in how accurately the participant could produce the gesture, and thus measured differences in speed and angle, and length for each of these tasks.

For each task, we first gave a description of possible gesture variations (e.g., "we will be testing direction and size of a swipe gesture") and had participants complete a small number of practice trials. Two blocks (repetitions) of the full set of gestures were then given, with trials randomized within a block. For each trial, the software played the *sound prompt* and the participant drew the corresponding gesture. A gesture was deemed to be correct if it was closer to the reference gesture in every characteristic (e.g., direction and length) than to any other gesture in the tested set. After a correct gesture, a chime sound played, while for incorrect gestures, an atonal "thunk" sound played. If the attempt was incorrect, the participant was allowed a single second attempt. The *Shape* task was an exception because its gestures were the most complex: thus, the sound prompt played twice per attempt, participants were *required* to complete two attempts per trial, and no audio feedback on correctness was provided. For all tasks, participants held the device inside a box so that it was shielded from view (Figure 1, right). Questionnaires were given after each task and at the end of the study.

4.1.6. Experiment Design and Analysis. This experiment used a within-subjects design that examined the effects of *Sound Parameter* (levels: *Volume*, *Timbre*, *Stereo*). The main measures for *Swipe Length* and *Swipe Speed* were angular difference (in degrees, calculated as the angle between the start and end points), speed difference, and length difference, respectively, between the reference gesture and drawn gesture. We

[6] The Swipe Length and Swipe Speed tasks were named Line Length and Line Speed in our previous publication [Oh et al. 2013], but were renamed here to improve consistency across studies.

Table II. Mean Angular Difference in Degrees for the *Swipe Length* and *Swipe Speed* Tasks. *Stereo* was Significantly More Accurate than *Volume* and *Timbre*. ($N = 12$)

	Sound Parameter		
Task	Volume	Stereo	Timbre
Line Length	30.2 (SD = 13.4)	8.6* (SD = 6.4)	21.2 (SD = 13.8)
Line Speed	22.8 (SD = 11.4)	4.2* (SD = 4.2)	19.8 (SD = 17.1)

simplified the analysis for these tasks by calculating average measures for horizontal, vertical, and diagonal directions rather than analyzing all eight directions individually. We then ran separate three-way repeated measures Analysis of Variance tests (ANOVAs) (*Sound*×*Direction*×*Speed* or *Sound*×*Direction*×*Length*) for each of the dependent measures. For the Tap Location task, we ran separate two-way repeated measures ANOVAs (*Sound*×*Location*) with horizontal and vertical difference between the reference and drawn gestures as dependent measures. All post hoc pairwise comparisons were protected against Type I error with Holm's Sequential Bonferroni adjustments. Where degrees of freedom are not whole numbers, a Greenhouse-Geisser adjustment was applied to account for violations of the sphericity assumption (tested using Mauchly's test for sphericity).

To capture more experienced behavior, we focused our analysis on the second block of trials within each task and the final attempt within a trial (a second attempt only occurred if the first attempt was incorrect). The same pattern of results occurs if we examine the first attempt only, largely because *Stereo* required fewer repeat attempts than the other sounds, meaning that the advantages of *Stereo* seen in the next section would likely be magnified by examining the first attempt alone. On average there were 1.24 attempts per trial ($SD = 0.14$) for *Stereo*, followed by 1.45 ($SD = 0.16$) for *Timbre*, and 1.54 ($SD = 0.13$) for *Volume* across *Swipe Length, Swipe Speed*, and *Tap Location* tasks. For the *Shape* task, participants had to complete all attempts regardless of accuracy.

4.2. Findings

In the interest of readability, we report only significant ($p<0.05$) main or interaction effects involving *Sound*, our primary factor of interest.

4.2.1. Reproducing Direction. *Stereo* was most effective at conveying direction among the three sound parameters and, as shown in Table II, resulted in the lowest angular difference in the *Swipe Length* and *Swipe Speed* tasks. For both tasks, there was a significant main effect of *Sound* on angular difference (*Length* task: $\mathrm{F}_{2,22} = 13.07, p < 0.001, \eta^2 = 0.54$; *Speed* task: $\mathrm{F}_{2,22} = 16.94, p < 0.001, \eta^2 = 0.61$). Post hoc pairwise comparisons in each case showed that *Stereo* was significantly more accurate than both *Volume* and *Timbre* (all $p < 0.05$). Although we did not directly compare the Swipe Length and Swipe Speed tasks, angular difference may be lower in the *Swipe Speed* task because participants had more practice at that point.

The positive effects of *Stereo* in the *Swipe Length* task were strongest for horizontal and diagonal swipes, which is not surprising given that those directions rely on x-axis sonification (Figure 2). This result was seen in a significant interaction effect of *Sound*×*Direction* on angular difference ($\mathrm{F}_{4,44} = 4.21$, $p = 0.006$, $\eta^2 = 0.277$). Post hoc pairwise comparisons showed that *Stereo* was more accurate than *Volume* for horizontal and diagonal swipes ($p < 0.05$). Additionally, a three-way interaction effect between *Sound*×*Length*×*Direction* ($\mathrm{F}_{4,44} = 5.880$, $p = 0.001$, $\eta^2 = 0.348$) was found. Post hoc pairwise comparisons were inconclusive. Section 7.6 provides some discussion regarding what level of accuracy may be "good enough" for teaching users to perform gestures.

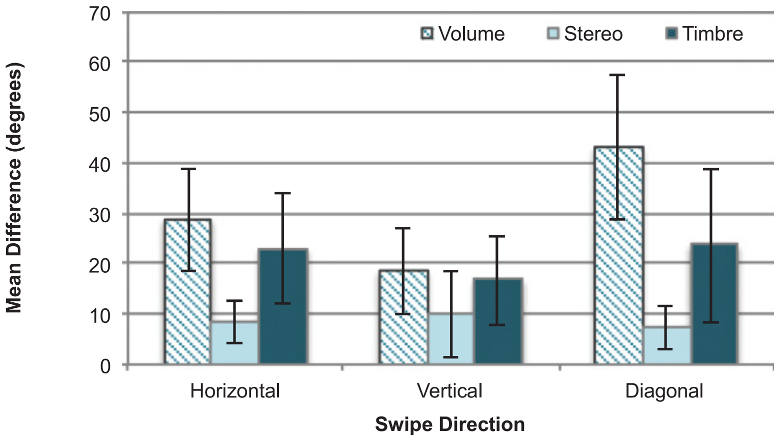

Fig. 2. Angular difference for the three sound types for swipe gestures in horizontal, vertical, and diagonal directions in the *Swipe Length* task; lower is better. *Stereo* resulted in the lowest angular difference (N = 12; error bars: 95% confidence intervals).

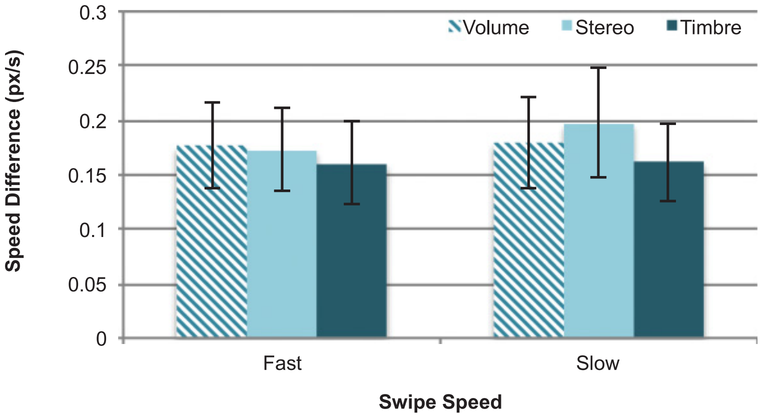

Fig. 3. Speed difference for the three sound types in the *Swipe Speed* task; lower is better. *Stereo* resulted in the lowest angular difference (N = 12; error bars: 95% confidence intervals).

4.2.2. Swipe Length. Participants were able to reproduce swipes of each length. On average across all three sound parameters, the drawn gesture lengths were off by 101.7px (*SD* = 17.3px) from the reference gesture, a much smaller amount than the difference between the short and long reference gestures themselves (315px). No significant main or interaction effects were found on length difference (Figure 3).

4.2.3. Swipe Speed. In the *Swipe Speed* task, participants were generally able to reproduce gestures at each speed. On average across all three sound parameters, the drawn gesture speeds were off by 0.16px/ms (*SD* = 0.22) from the reference gesture, less than the 0.33px/ms between the speeds of the short and fast reference gestures. No significant main or interaction effects on the measure of speed accuracy were found (Figure 4).

4.2.4. Tap Location. *Stereo* again performed well in the Tap Location task as compared to the other two sounds. In the horizontal direction, taps in the *Stereo* condition were only off by an average of 4.7mm, or 58.4px (*SD* = 22.6), while *Volume* and *Timbre* were

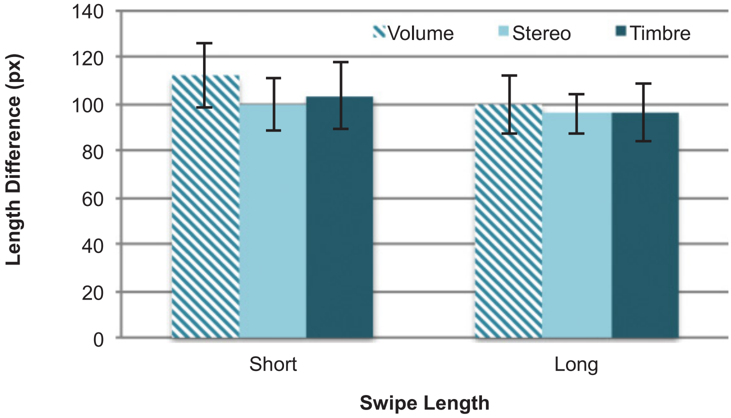

Fig. 4. Length difference for the three sound types in the *Swipe Length* task; lower is better ($N = 12$; error bars: 95% confidence intervals).

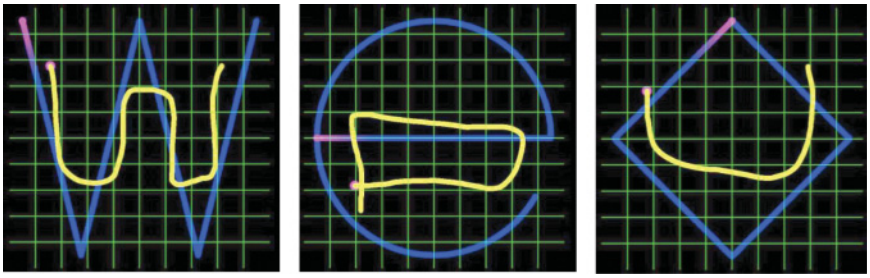

Fig. 5. Example shape gestures from one participant (P8) in the *Stereo* condition, demonstrating both the sporadic success (the "W" shape here) and common difficulties seen with all sound types. Reference gesture in blue; drawn gesture in yellow.

off by 103.5px ($SD = 31.0$) and 82.5px ($SD = 29.8$), respectively. A main effect of *Sound* on the x-direction difference was significant ($F_{1.36,14.96} = 13.97$, $p = 0.001$, $\eta^2 = 0.56$), with post hoc pairwise comparisons revealing that *Stereo* was more accurate than both *Volume* and *Timbre* ($p < 0.05$). For vertical difference, where pitch was always used on the y axis, no significant main effects were found. There was a significant interaction of *Sound*×*Location* on vertical difference ($F_{16,176} = 1.865$, $p = 0.026$, $\eta^2 = 0.145$), although no post hoc pairwise comparisons were significant.

4.2.5. Tap Type. Participants found the *Tap Type* task to be easy, and they were 100% accurate in identifying the number of taps that were played (recall that we did not compare the three sound parameters for this task). The only errors were in identifying taps of different lengths (short vs. long single tap), where participants sometimes underestimated the length of the long tap. The average duration of short single taps was 155.5ms ($SD = 94.5$), while the duration for long single taps was 740.5ms ($SD = 565.5$). These lengths were shorter than the short and long tap durations of the reference gestures (200 and 1000ms).

4.2.6. Shape. For the *Shape* task, we visually inspected the drawn shapes. There were no conclusive differences among the three sounds, with participants exhibiting difficulties in completing shapes regardless of sound type. See Figure 5 for example shape gestures.

4.2.7. Subjective Preference. When asked to rank the three sound conditions, all 12 participants ranked *Stereo* as their favorite, often citing the ease with which it conveyed horizontal differences. For second place ranking, *Volume* and *Timbre* were roughly equally split (seven and five votes, respectively).

4.2.8. Summary. The pitch+stereo combination was the most easily discernible mapping, and was preferred by all participants. It improved angular accuracy in the swipe tasks and horizontal location accuracy in the Tap Location task compared to the other sound combinations. Thus, we carry on the pitch+stereo mapping in Study 2 and Study 3. For the shape gestures, visual inspection of the drawn shapes suggested that users needed more than only sonification to reliably reproduce a shape. Thus, in Study 2 and Study 3 we use pitch+stereo within a more realistic tutorial procedure, where participants are given a verbal description to help them understand the shape that they were drawing.

5. STUDY 2: SONIFICATION FOR MULTISTROKE AND MULTITOUCH GESTURES

Our first study explored parameters for creating gesture sonifications, and findings from sighted users indicated that a combination of pitch (for representing movement along the y axis) and stereo pan (for the x axis) was preferred to other mappings. Study 1 was limited in that it did not include participants with vision impairments, whereas both Study 2 and Study 3 included participants with vision impairments.

Study 1 explored the question of whether individuals could use sonification to accurately reproduce various aspects of a gesture (direction, length, speed, and location). However, while Study 1 primarily focused on single-stroke gestures, many common accessible touchscreen gestures involve multiple strokes or multiple simultaneous touches (see Table I). These more complex gestures may be significantly more difficult to understand when sonified. As a result, Study 2 was primarily focused on participants' ability to identify sonifications of multistroke and multitouch gestures using the sonification methods established in Study 1, and to reproduce those gestures. Because we anticipated that it would be difficult for participants to identify the sonifications of more complex gestures, and because it was more difficult to precisely verify how the participants' gestures matched the reference gesture, we primarily measured participants' ability to describe and reproduce the general structure of the gesture (direction and number of strokes or touches).

5.1. Method

5.1.1. Participants. Ten visually impaired volunteers (seven female) participated in this study. The average age was 41.8 ($SD = 16.0$; range 19–67). All participants were primarily screen reader users and had little functional vision. Participants were asked if they had any auditory impairment; none reported such impairment. Five participants owned and regularly used a touchscreen device. Four participants were musically trained or practiced music as a hobby. One participant in this study also participated in Study 3; however, over 12 months separated their participation in the two studies, and we did not believe that they had any significant advantage because of their experience participating in the other study.

5.1.2. Apparatus. We used the same device as in Study 1, but did not use a box to shield the device from view, since the participants in this study were all blind or visually impaired. We replaced the cardboard overlay with a laser-cut plastic overlay of the same size, to reduce slippage. For this study only, we used a pair of Sony MDR ZX-100 over-the-ear headphones, but found in informal testing that these performed similarly to the Sennheiser headphones.

5.1.3. Gesture Sonification Feedback. Participants in this study used a modified form of the pitch+stereo method explored in Study 1. The *y* position of each finger was mapped to pitch exactly as in Study 1. We began with the same mapping of *x* position to pan as in Study 1, but found in pilot testing that this mapping caused the sound to become quiet when the finger touched the horizontal center of the screen, making it difficult to identify when gestures started or ended in the center. Thus, we modified the stereo mapping such that for a touch point on the left of the screen, the left channel gain decreased by from a high of 1.0 on the far left to 0.4 near the middle (step size of 0.4 gain), while the right channel played 1.0—(the left volume); vice versa for the right side. The gain was evenly split (0.5,0.5) in the center. This stereo mapping presented a departure from the other two studies, but we felt that this change was necessary to enable participants to reliably identify gestures performed in the horizontal center of the screen.

For multitouch gestures, we used timbre to indicate multiple fingers touching the screen. For a single finger, a sine wave was played, as in Study 1. If a second finger touched the screen, a square wave was played with pitch and stereo pan based on its position. A sawtooth wave was played when a third finger touched the screen.

We also compared two methods for presenting multitouch gestures: *Parallel* and *Serial*. In the *Parallel* method, the sonifications for all fingers used in the gesture were played at once. In initial testing, however, we found that it was sometimes difficult to understand each finger's movement in the sonification of a multitouch gesture if the movements started at the exact same time or if they overlapped spatially or temporally, as in a two-finger swipe gesture. Thus, in *Parallel* mode, each finger's movement was offset from the previous finger by 50ms, and was shifted by 50px along the *x* axis and 50px along the *y* axis. In pilot testing, we found that these offsets improved the comprehensibility of the gestures. In the *Serial* method, each finger of the gesture was played in sequence, as in a multistroke gesture (with a 300ms delay in between strokes), but participants were instructed to perform all of the finger movements at once.

5.1.4. Procedure. Participants completed four blocks of gestures with the previous gesture sonification settings: a block of 8 single-stroke gestures, a block of 18 multistroke gestures, and two blocks of 20 multitouch gestures. The gestures were as follows:

—**Single stroke (eight gestures).** These consisted of eight directional swipe gestures (up, down, left, right, and the four diagonals). Each gesture was a straight swipe 630px long, through the center, and was drawn at 1/2px/ms.

—**Multistroke (18 gestures).** These consisted of directional double swipe gestures (e.g., *up twice*), swipe-and-back gestures (e.g., *up and down*), and cross gestures (e.g., *right then down*). For each swipe, swipe-and-back, and cross gesture, participants performed six gestures: up, down, left, right, and two randomly selected diagonal directions. Each gesture was a straight-line swipe 630px long through the center of the screen, and was drawn at 1/2px/ms. There was a 300ms delay between each stroke.

—**Multitouch (20 gestures).** These consisted of directional two-finger swipe gestures (e.g., *two-finger swipe left*) and directional three-finger swipe gestures (e.g., *three-finger swipe down*). For each type of directional gesture, participants performed six gestures in six directions: up, down, left, right, and two randomly selected diagonal directions. Participants also performed three pinch-out gestures (e.g., *from top and bottom to center*), and three pinch-in gestures (e.g., *from center to top and bottom*). For these subsets, participants performed gestures in three directions: vertical,

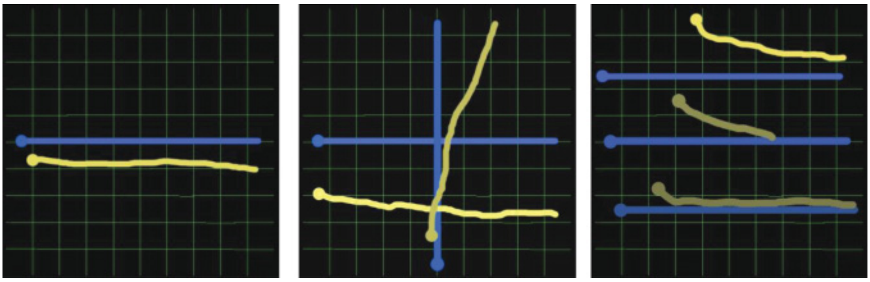

Fig. 6. Example gestures from one participant (P4): Single Stroke, Multistroke, and Multitouch (**Parallel**). Reference gesture in blue; drawn gesture in yellow.

horizontal, and a randomly selected diagonal direction. Finally, participants completed one two-finger rotate clockwise and one two-finger rotate counterclockwise gesture. Each gesture was drawn at 1/2px/ms.

Participants first completed the single-stroke gesture block, then the multistroke gesture block, and then the two multitouch gesture blocks. The order of gestures within each block was randomized. The two multitouch blocks each contained the same gestures, but one block was presented using the *Serial* presentation technique and the other was presented using the *Parallel* presentation technique. Half of the participants completed the *Serial* block first, and the other half completed the *Parallel* block first. At the start of each block, participants were told that they would be testing gestures with single strokes, multiple strokes, or multiple simultaneous fingers, but were not told which gestures they would specifically perform. Examples of performed gestures are shown in Figure 6.

Before each block, participants were told about the sonification method and asked to touch the screen to explore how that method worked. Then, for each gesture in the block, the participant attempted to describe the gesture and articulated it three times: (1) the gesture sonification played; (2) the participant described the gesture verbally (and could ask for the sonification to be repeated up to once); (3) the participant performed the gesture and received sonified feedback while doing so; (4) the textual description (e.g., “a line from left to right drawn with two fingers”) *and* the sonification played; (5) the participant performed the gesture a second time and received sonified feedback while doing so; (6) the sonification played again; (7) the participant performed the gesture a third time and received sonified feedback while doing so. Because the experimenter provided a description of the correct gestures, and because it was not possible to automatically assess whether some of the gestures were correct (specifically the multitouch gestures), participants did not hear audio feedback (chimes or “thunks”) during this experiment.

As we were interested in participants’ ability to identify and reproduce the sonified gesture, in this study we focused on whether or not the participant could identify the primary characteristics of the gesture (the direction of the gesture, the number of strokes, and the number of fingers). A gesture was considered correct if the participant used the correct number of strokes, the correct number of fingers, and moved their fingers in the correct direction. During each trial, the participant first described the gesture verbally, and then performed it three times. We asked participants to draw the *same* gesture three times for data collection purposes (and to make sure that we had an accurate recording of their intended gesture), but considered the first drawing of the gesture as the canonical gesture. We instructed participants to draw the same gesture

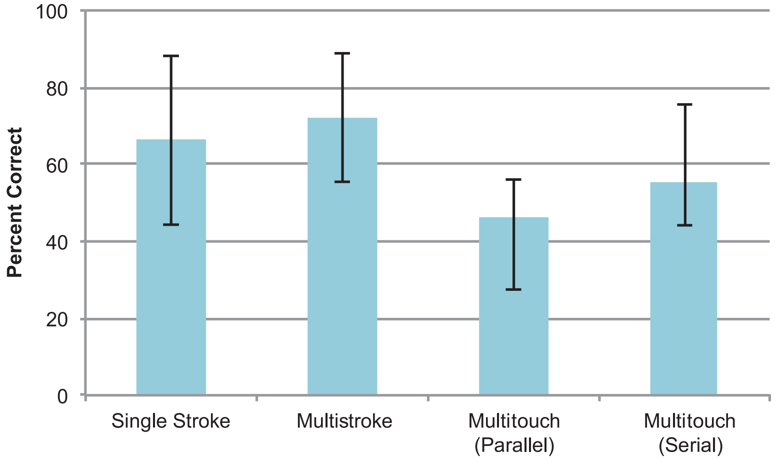

Fig. 7. Percent of gestures correctly guessed in each of the four blocks, averaged across participants ($N = 10$; error bars: 95% confidence intervals). Correct guesses were significantly lower in the *Multitouch (Parallel)* condition than in the *Multistroke* ($p < 0.01$) and *Single-Stroke* ($p < 0.05$) conditions.

each of the three times, even if they believed it to be incorrect. We measured correctness based on how the first gesture was drawn; in several instances, participants verbally swapped the words "left" and "right," but performed the correct gesture.

After completing the four study blocks, participants rated each block and their accompanying feedback on a 7-point scale (1: "I like it very much" to 7: "I don't like it at all"). The complete procedure took between 60 and 90 minutes.

5.2. Results

5.2.1. Recognizing and Reproducing the Correct Gesture. Here we explored whether the mapping could extend to more complex multistroke and multitouch gestures. Figure 7 shows the average number of gestures correctly guessed in each block.[7]

We compared the percentage correct across each of the blocks. Percentage correct scores were not normally distributed, as measured by Shapiro-Wilk's test ($p < 0.05$). We thus used a Friedman test to compare the percentage correct across the four blocks. We found a significant effect of study block on percentage correct ($\chi^2(3) = 13.99$, $p < 0.01$). We conducted a pairwise post hoc comparison using a Wilcoxon signed-rank test with Bonferroni correction, and found significant differences between the *Single-Stroke* and *Parallel* conditions ($p < 0.05$) and between the *Multistroke* and *Parallel* conditions ($p < 0.01$); no other pairwise comparisons were significant.

We note that the single-stroke gestures had a lower percentage correct than the multistroke gestures (single stroke; $M = 66.25\%$, $SD = 30.65\%$; multistroke: $M = 72.22\%$, $SD = 23.42\%$), although the difference was not statistically significant. This may be due to two factors: first, this was the first block that participants completed, and thus participants had less experience using the technique. Second, there were only eight gestures in the single-stroke gesture set (vs. 18 in the multistroke gesture set), so that each mistake had a greater effect on the percentage correct.

We also found that individual differences seemed to affect performance. Table III shows the number of gestures correctly guessed by each participant, and in each

[7] One participant (P1) only completed 18 gestures in the Multitouch Parallel condition, due to a data collection error. We calculated percentage correct based on the number of gestures each participant completed in each block, so this does not affect the calculation of the correct percentage.

Table III. Percent of Gestures Correctly Recognized and Reproduced in Each of the Four Blocks for Each Participant. The Total Column Counts the Number of Total Gestures Correctly Recognized and Reproduced by that Participant

Participant	Single (8)	Multistroke (18)	Multitouch (Parallel) (20)	Multitouch (Serial) (20)	Total (66)
1	75%	61%	28%7	70%	56%
2	88%	94%	65%	80%	80%
3	25%	78%	35%	65%	55%
4	63%	67%	60%	50%	59%
5	100%	94%	50%	75%	76%
6	25%	33%	10%	5%	17%
7	25%	50%	15%	50%	36%
8	88%	94%	65%	60%	74%
9	75%	50%	35%	80%	71%
10	100%	100%	55%	65%	76%
Mean (SD)	66% (31%)	72% (23%)	42% (20%)	60% (22%)	60% (20%)

Table IV. Participants Rated Each of the Four Blocks on a 7-Point Scale (1: “I like it very much” to 7: “I don’t like it at all”). Lower is better. Post hoc Comparison Revealed a Significant Pairwise Difference between the *Single-Stroke* and *Parallel* Conditions ($p < 0.05$)

Participant	Single Stroke	Multistroke	Multitouch (Parallel)	Multitouch (Serial)
1	2	5	7	2
2	2	4	5	2
3	1	1	3	3
4	2	2	7	3
5	1	2	4	2
6	5	6	7	4
7	5	2	1	3
8	2	2	4	6
9	2	2	5	1
10	1	2	7	3
Mean (SD)	2.3 (1.5)	2.8 (1.6)	5 (2.1)	2.9 (1.4)

category. Looking across participants, we see signs that some participants had significant difficulty in recognizing any of the gestures: 2 of 10 participants guessed correctly less than 40% of the time. These participants were among the older participants in this study; we discuss the potential implications of this in the discussion section. In testing for correlation, we found that age was correlated with the total number of correct guesses (Pearson $r = -0.702$, $N = 10$, $p < 0.05$). There were no significant correlations between performance on study tasks and either participants’ prior use of touchscreens or prior musical experience.

5.2.2. Preferred Feedback Modes. We asked participants to rate each of the four study blocks (single-stroke, multistroke, multitouch *Parallel,* and multitouch *Serial).* Table IV summarizes their ratings.

Participants tended to rate single-stroke and multistroke gestures more preferentially than multitouch gestures. A Friedman test found that the study block had a significant effect on participants’ ratings ($\chi^2(3) = 10.86$, $p < 0.05$). Post hoc comparison revealed a significant pairwise difference between the *Single-Stroke* and *Parallel* conditions ($p < 0.05$). Within multitouch gestures, three participants verbally commented that the *Parallel* task was especially difficult. Seven of the ten participants preferred *Serial* over *Parallel,* two preferred *Parallel,* and one rated the two equally. Eight of the ten participants gave *Parallel* their lowest rating.

5.2.3. Subjective Feedback. We also asked participants what they liked and did not like about each presentation technique, and which gestures were easiest or most difficult to perform. Nine participants mentioned that the multitouch gestures were especially difficult, specifically those gestures that involved fingers moving in different directions, that is, the pinch and rotation gestures. Almost all of the participants (9 of 10) stated that the straight-line, single-finger gestures were the easiest to learn. One participant said that the three-finger gestures in the *Parallel* condition were easiest to learn because it was easy to identify that there were three fingers based on the volume of the sound.

In general, participants were able to learn the sound mappings for the gestures. However, they sometimes had difficulty directly duplicating the gesture sonification sound directly, and commented that they wished they were able to. Duplicating the gesture sound was especially difficult for multitouch gestures, as participants needed to position multiple fingers precisely, and because the order in which their fingers touched the screen affected the sound playback. For example, a two-finger pinch gesture would sound different if the participant touched down with the leftmost finger first, as opposed to the rightmost finger.

5.2.4. Summary. We extended our two-dimensional sonification mapping (pitch+ stereo) to support both multistroke and multitouch gestures. Most participants were easily able to adapt to multistroke gestures without trouble. Multitouch gestures were especially difficult to understand when multiple fingers were sonified simultaneously. Playing each finger sound independently seemed to improve participants' ability to understand the gesture being played, although some participants had difficulty combining the serial sounds into a single multitouch gesture, and some gestures such as rotation were frequently described by participants as being very difficult to perform.

6. STUDY 3: COMPARING GESTURE SONIFICATION AND VERBAL FEEDBACK

Our third study evaluated two gesture feedback techniques for blind users: (1) *gesture sonification* using the pitch+stereo combination that was best in Study 1, and (2) *corrective verbal feedback* using text-to-speech and generated by automatically analyzing the drawn gesture.

In Study 1 and Study 2, we focused primarily on the effectiveness of sonification for conveying information about a reference gesture. However, participants had requested more concrete feedback about the gestures that they performed. Thus, this study introduced a new form of verbal gesture feedback, which we compared to the sonification methods we previously developed. While we recorded performance data as in Study 1 and Study 2, the primary focus here was to collect subjective feedback regarding the usefulness of sonification and verbal feedback for teaching gestures.

6.1. Method

6.1.1. Participants. Six visually impaired volunteers (three female, three male) participated in this study. The average age was 36.1 (*SD* = 16.2; range 24–62). All participants were totally blind, except for one who had low vision (20/200). On average, participants had 15.8 years of experience with computers (*SD* = 6.2; range 9–25). All participants owned a touchscreen device and four participants reported daily touchscreen use. One participant reported a mild auditory disorder but was able to complete the study tasks. Five participants played at least one musical instrument, one of whom reported having perfect pitch. One participant in this study also participated in Study 2; however, over

12 months separated their participation in the two studies, and we did not believe that they had any significant advantage because of their experience participating in the other study.

6.1.2. Apparatus. We used the same device and physical overlay as in Study 1, except that we did not use a box to shield the device from view, since the participants in this study were all blind or visually impaired.

6.1.3. Feedback Techniques. In this study, we compared two forms of gesture feedback: *gesture sonification* and *corrective verbal feedback*.

- **Gesture sonification.** The gesture sonification was based on the pitch+stereo combination that was preferred by participants in Study 1. Study 3 incorporated gesture sonification in the following ways: (1) a *sonified preview* of the reference gesture, presented before the first gesture trial and accompanied by a text-to-speech description of the reference gesture; (2) *sonification feedback* produced when the user touched the screen; (3) upon an error, a *replay* of the reference gesture sound for comparison to what had been generated by the user.

  For sonification, the screen was divided into a 10×10 grid, where each row mapped to a different pitch and each column mapped to a different stereo position. As with Study 1, pitch was set to D5 (587.33Hz) in the topmost row and dropped by one musical note per row. Gain for stereo panning (x axis) was also manipulated similarly to Study 1, with the exception that the middle column in the grid was set to a gain level of 0.1 in both the left and right channels to create a perceptually smooth horizontal transition.
- **Corrective verbal feedback**. The corrective verbal feedback technique consisted of a text-to-speech description of the gesture, presented before the gesture trial, and text-to-speech corrective feedback after errors. To generate the corrective feedback, the software compared the drawn gesture to the reference one as follows:
  - *Speed*. For gestures that required a specific speed, the software told the user whether the gesture needed to be "faster" or "slower." This feedback was used for swipes and double taps (short vs. fast double taps).
  - *Size and Aspect Ratio*. For swipes of different lengths, feedback was provided to make the gesture "longer" or "shorter." For more complex shapes, feedback was provided on the aspect ratio. When the reference gesture had an aspect ratio of 1:1, feedback consisted of "try wider" or "try taller," as appropriate. For other aspect ratios, both width and height feedback was provided. For example, if the expected aspect ratio was 1:3 but the drawn gesture was 1:2, the feedback would be "try taller and narrower."
  - *Direction*. Directional feedback was based on the angle of rotation. For swipes in this study, we tested only the horizontal direction (right/left), so the feedback was "opposite direction."
  - *Location*. When location was important (for the *Tap Location* task), the system provided feedback using the four cardinal directions (e.g., "more to the right," "higher," or "more to the right and higher").
  - *Repetition*. To correct single or multiple taps, the system asked the user to try more or fewer taps.

Figure 8 shows examples of oval and rectangle gestures drawn by our participants, and illustrates the effects of sonification and verbal feedback on their gestures.

6.1.4. Procedure. Study sessions were designed to last 90 minutes. The order of presentation for the feedback techniques (*Sonification* and *Verbal*) was fully counterbalanced, and participants were randomly assigned to an order. For each feedback condition, the

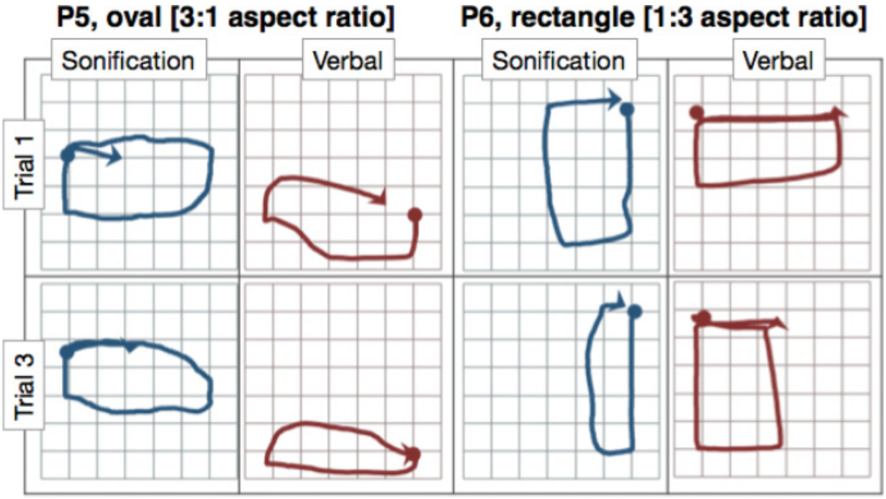

Fig. 8. Example shape gestures by two blind participants in Study 3, showing the first and third of three trials. Note details such as lack of closure in one of P5's verbal feedback trials, and changes between the first and third trial for P6.

procedure mimicked a gesture tutorial scenario. The following tasks were tested in random order, with gestures presented randomly within each task. Participants were asked to replicate each gesture.

Because the study protocol could be quite time consuming, we simplified the gesture set in order to reduce participant fatigue. In Study 1, we found that sonification was effective at enabling users to perform swipe gestures in multiple directions, and to convey multiple aspects of the gestures (angle, shape, and speed). Thus, we combined the swipe task into a single task with varying direction, length, and speed. Because we had previously established that sonification could convey information about swipe gestures in multiple directions, we focused here on horizontal swipe gestures, as they are most common in current systems.

—**Swipe.** Twelve swipe gestures: 2 directions (left, right)×3 lengths (short: 157.5px, medium: 315px, long: 630px)×2 speeds (fast: 1/2 px/ms, slow: 1/6 px/ms).
—**Tap location.** Nine tap locations, one in each cell of a 3×3 grid covering the screen. Locations were described to participants as "top-center of the screen," "bottom-right," etc.
—**Tap type.** Three tap types: single tap, fast double tap, and slow double tap. Each tap lasted 100ms, with 200 and 500ms gaps between taps for slow and fast double tap, respectively.
—**Shape.** Six shapes: 2 base shapes (circle, rectangle)×3 aspect ratios (large: [1:1], short and wide: [3:1], and tall and narrow: [1:3]). As an example, a circle that has a 1:3 ratio was described as "a tall, narrow oval."

Participants performed three trials of each gesture. For the first trial, the software played a verbal description via text-to-speech (e.g., "a tap in the top-left of the screen"). This description was immediately followed by a sonified preview of the reference gesture in the *Sonification* condition (like Study 1). After a correct gesture trial, a chime sound played. After an incorrect trial, the system played a "thunk" sound, followed by either corrective verbal feedback (in the *Verbal* condition) or a replay of the audio prompt (in the *Sonification* condition). In both conditions, gesture sonification feedback was produced while the participant touched the screen. Participants were asked about their subjective experience after each task and at the very end of the study. Because we had previously established that sonification could be used to convey swipe direction, we did not require participants to guess the gesture being sonified.

As in Study 1, swipe gestures and tap type were deemed to be correct if they were closer on all characteristics (e.g., length, speed, direction) to the reference gesture than to any other gesture in the set. For tap locations, the drawn gesture was correct if it was within 4.8mm (60px) of the reference gesture's location, equivalent to falling within the bounds of a reasonably sized touch target centered at that location using a 9.6mm (119px) target size [Parhi et al. 2006]). Finally, for shapes, an aspect ratio between 4:3 and 2:3 was considered a correct square (1:1), narrower than 2:3 was considered tall and narrow, and wider than 4:3 was considered short and wide.

6.2. Results

Due to the sample size, we focus primarily on subjective findings, descriptive statistics, and individual user differences. However, we also report on statistically significant findings where applicable—these should be considered preliminary, but will be useful for informing the design of a future gesture tutorial system.

6.2.1. Performance on Swipe Task. For the *Swipe* task, *Verbal* was particularly effective for correcting swipe length. All six participants improved in length accuracy after receiving corrective verbal feedback, from being on average 102.0px (SD = 13.8) off from the reference gesture length in the first trial to only 73.0px (SD = 15.5) off in the third trial. With *Sonification*, only half the participants improved on this measure from the first to third trials (first trial: M = 89.9px, SD = 37.3; third trial: M = 100.0px, SD = 27.8). There was no evidence that either feedback type had an impact on swipe speed (*Sonification* trial 1: M = 0.17px/ms, SD = 0.09; and trial 3: M = 0.20px/ms, SD = 0.13; *Verbal* trial 1: M = 0.23px/ms, SD = 0.10; and trial 3: M = 0.21, SD = 0.15). Finally, participants always correctly replicated the direction (right/left) of the reference gesture.

Examining where drawn swipes were located, we found that participants exhibited a tendency to begin gestures close to the edge of the device. Participants had been told that swipe gestures were centered on the screen (at an x-axis location of 350px). However, the midpoints of the drawn gestures per participant were offset: for left-to-right swipes the average midpoint was left of center at 275.1px (SD = 34.6), and for right-to-left swipes it was right of center, at 398.3px (SD = 59.2). A two-way repeated measures ANOVA (feedback type×direction) showed that direction (left-to-right or right-to-left) had a significant impact on the midpoint location ($F_{1,5} = 24.79$, $p = 0.004$, $\eta^2 = 0.83$). No other main or interaction effects were significant.

6.2.2. Performance on Tap Location Task. Calculating the Cartesian distance between drawn tap locations and the reference gesture location, participants were off by on average of 43.3px (SD = 15.0) with *Sonification* and 42.5px (SD = 21.7) with *Verbal*, across all three trials. From the first to the third trial, all participants improved in accuracy with *Sonification* (improvement in pixels: M = 38.7, SD = 46.8) and 5/6 improved in accuracy in *Verbal* (improvements in pixels: M = 24.8, SD = 18.6).

6.2.3. Performance on Tap Type Task. For *Tap Type*, participants made no errors in recognizing and performing single tap gestures compared to double taps. For double taps, the difference between the gap lengths of the drawn gestures and those of the reference gestures was similar for both feedback conditions (*Sonification*: M = 118.7ms, SD = 51.8; *Verbal*: M = 129.9ms, SD = 46.8). These differences are less than the 300ms difference between the slow and fast double tap reference gestures (500ms vs. 200ms).

6.2.4. Performance on Shape Task. Both *Verbal* and *Sonification* feedback appeared to help with the accuracy of drawing tall and wide shapes (Table III). The effect was particularly strong with the tall shapes (aspect ratio of [1:3]), where all participants improved with *Sonification* and all but one participant improved with *Verbal*. For perfect squares and circles, that is, an aspect ratio of [1:1], participants did not have much trouble completing the gestures accurately on the first trial.

We also assessed how the feedback impacted form closure for the shapes. To do so, we calculated the Cartesian distance between the start and end points of a gesture; a distance of 0px represents perfect closure, while anything greater is either an open shape or has overlapping start/end points (Figure 8). The gap between start and end points was lower with *Sonification* (M = 209.4px, SD = 83.5) than for *Verbal* (M = 285.6px, SD = 150.9). Although not a statistically significant difference, this data suggests that it would be useful to further explore if *Sonification* is particularly effective at conveying complex shape features such as closure.

6.2.5. Subjective Feedback. Differences between *Sonification* and *Verbal* were clearer in participants' subjective experiences than in the performance data. When asked about overall preference, four of six participants preferred *Verbal*, one wanted both types of feedback, and one preferred *Sonification*. These sentiments were also reflected in ratings of overall satisfaction with the feedback conditions. *Verbal* received more positive ratings than *Sonification* using a 7-point scale (1: "I like it very much" to 7: "I don't like it at all"): *Verbal*'s median was 2 (range 1–4) and *Sonification*'s median was 3 (range 2–5). This difference was statistically significant using a Wilcoxon signed-rank test ($Z = 2.33$, $p = 0.02$, $r = 0.67$). Despite the overall preference for *Verbal*, however, detailed responses highlighted trade-offs between the two techniques.

Overall, participants were positive about the sonified preview that played at the beginning of each trial in the *Sonification* condition. Five reported that it was helpful, particularly for conveying time-related characteristics such as speed and duration between two taps, for example:

"Sound example was helpful for speed" (P2)
"Faster and slower are better with [sonification]" (P3)
"That's good [the sonification]. You can tell how fast, how slow [for taps]" (P5)

There were a smaller number of negative comments; two focused on the utility of *Sonification* for conveying shapes. One of the participants who generally found the sonified preview helpful also felt it provided too much information when used for a shape. The only participant who did not find the sonified preview helpful noted that it was "unnecessary" and required extra time to listen to before completing the shape tasks (P6).

Some participants commented that understanding the sonification required cognitive effort, especially for more complex shapes:

"I need to pay attention" (P2)
"Focusing on two things [pitch and stereo] at the same time was hard" (P3)
"You have to listen to the feedback multiple times to make a correction" (P4)

Participants who preferred *Verbal* feedback overall appreciated its preciseness, because it tells provided clear directions for what to correct. For example:

"[I like] telling exactly what you need to do when you messed up" (P2)
"Easy to correct gestures by hearing feedback only once" (P4)
"It gives more accurate description, more in detail" (P6)

Three participants (P3, P5, and P6), however, commented that a downside of the verbal feedback is that while it offers a specific suggestion for adjusting the gesture (e.g., "draw the gesture faster"), it does not quantify *how much* to adjust. In this respect, *Sonification* offers additional cues. For example:

> *"Audio [sonification] feedback was more helpful for tap location [than verbal feedback]"* (P1)
> *"It [verbal feedback] says it's not narrow enough, but how narrow?"* (P5)

Individual differences may have also influenced participants' experiences. Participants varied in level of vision and in musical training, which could impact subjective experience. For example, there was no visual guidance in the study interface—simply a blank screen—yet a user with limited vision may have different preferences than a user who is completely blind. Only one of the six participants had low vision (P1) and was able to see his fingers on the screen when the device was held at a short distance. This same participant reported having a "slight" auditory processing disorder and, ultimately, preferred the corrective verbal feedback. However, he felt that *Sonification* was more helpful for tapping than for swiping and suggested that combining the two forms of feedback would be useful.

Another participant reported having perfect pitch (P5). She had the most extensive musical experience of all participants, and had earned a college degree in music. She was the only participant who preferred *Sonification* overall, reporting that it was useful for conveying many kinds of information, including width, length, and height. She also felt that the pitch was particularly helpful. For example, she could tell based on the pitch change that her swipes were not perfectly straight even though she was attempting to draw a straight line. In contrast to some other participants who felt the sonification was too complex for conveying shapes, this participant reported that the sonified previews for shapes were useful: *"... it was easy to tell the direction, that's neat."*

6.2.6. Summary. Overall, participants found benefits in both the gesture sonification and the corrective verbal feedback. Perhaps not surprisingly, the corrective verbal feedback was considered to be more straightforward, as it did not require the user to learn a new encoding scheme. However, the gesture sonification feedback did seem to be helpful when drawing gesture shapes. Thus, Study 3 found that both sonification and verbal feedback can be helpful in teaching gestures nonvisually.

7. DISCUSSION

Our three studies demonstrate that audio feedback techniques alone can be used to convey information about a variety of gestures, including taps, swipes, shapes, and multistroke and multitouch gestures. We extend our prior work [Oh et al. 2013] to explore the applicability of gesture sonification techniques to multistroke and multitouch gestures. In this section, we summarize our findings across the three studies and consider the broader impact of this work.

This work builds on the now established idea that touchscreen interfaces should be designed to support users with vision impairments. Prior work has shown that touchscreen interfaces may be adapted to support visually impaired users [Azenkot et al. 2012; Bonner et al. 2010; Kane et al. 2008, 2011a]. In developing these interfaces we must consider the entire lifecycle of their use: not just developing accessible gestures, but creating ways for users to learn about, practice, and become experts at performing gestures. These studies came about through our prior research that has explored the design of touchscreen gestures for people with vision impairments. During these studies, we sometimes found that participants had difficulty performing gestures based on our verbal descriptions, and observed that participants who owned touchscreen devices sometimes had difficulty performing gestures on their own devices consistently. Thus, we concluded that there was an opportunity to explore ways to better convey information about how to perform touchscreen gestures. Our work demonstrates that both real-time sonification and post-gesture corrective verbal feedback can influence the performance of gestures.

The outcomes of this research may benefit people with vision impairments in several ways. First, as we have previously mentioned, providing feedback about how to perform gestures more accurately can improve usability, and may also increase the number of gestures that can be used in a user interface. More importantly, perhaps, is the hope that providing improved training for using accessible touchscreen interfaces can improve independence for people with vision impairments when using technology. Providing integrated instructions for performing touchscreen gestures may increase the likelihood that a user with vision impairment may be able to learn and use a device without requiring help from a sighted person.

7.1. Best Parameters for Gesture Sonification

We tested a variety of sound parameters for mapping two-dimensional touchscreen gestures to sound. Based on our collected data, we recommend using pitch to represent movement along the y axis, and stereo panning to represent movement along the x axis. This combination resulted in the best performance on a gesture replication task and was unanimously preferred by the sighted participants in Study 1. In cases where stereo is not usable, such as when the user does not wish to wear headphones, either volume or timbre could be used to represent movement along the x axis; no differences were found between these two combinations. During pilot tests before the full study, we tested and excluded several additional sound parameters that were not appropriate for mapping some dynamic gesture characteristics: vibrato and tempo, for example, are both periodic and thus interfere with communicating gesture speed. Finally, participants had difficulty replicating shapes based purely on a sonification of the shape, regardless of sound parameters used, suggesting that a more realistic training scenario should also include verbal descriptions of the reference gesture (as in Study 3).

While our recommendations are based on a study with sighted participants, Walker and Mauney [2010] have shown that perception of sound mappings is usually consistent between blind and sighted people. Our findings should thus be useful in informing future work with both sighted and visually impaired participants.

7.2. Sonification of Multistroke and Multitouch Gestures

In general, our second study demonstrated that the gesture sonification techniques can be extended to multistroke and multitouch gestures, although understanding the sonification becomes more difficult as gestures become more complex. Gesture sonification seems especially difficult when the gesture involves multiple fingers and the sounds are played simultaneously. Our *Serial* presentation technique, in which each finger's sound is played serially, similar to multistroke gesture sonification, was preferred by the majority of participants. However, it is unclear whether this serialized representation can always convey all parameters of a gesture, as we focused on whether the participant could identify the type of gesture and number of fingers only.

As noted in the previous section, including verbal descriptions of the gestures can greatly help in understanding the sonification. In this study, participants heard the verbal description of the gesture after their first guess, and were better able to perform gestures once they had heard the verbal description. While one goal of this study was to determine how robust the sonification technique might be to different types of gestures, it is clear that a real-world system should include multiple types of information. In Study 3, we present sonification combined with a verbal description.

In Study 2, we found that a few of our participants experienced significant difficulties understanding the sonification. These difficulties persisted throughout the study, suggesting that the minimal amount of training cannot completely account for their lower performance. Although we cannot clearly say why this discrepancy occurs, it is worth noting that the participants with the lowest scores tended to be older: the two participants who had the fewest correct guesses were among the oldest participants (67 and 49). While it is possible that age or some related factor (such as hearing ability or experience with touchscreen computers) affected performance, it is difficult to say from this limited dataset.

7.3. Gesture Sonification versus Corrective Verbal Feedback

As shown in our third study, both corrective verbal feedback and gesture sonification offered performance and subjective advantages, suggesting that a combination of the two may ultimately be useful for a gesture-training system. Either one or both of the feedback techniques improved gesture accuracy for participants from the first to the third trial in Study 2 in terms of swipe length, tap location, and shape aspect ratio. Overall preference was skewed toward the verbal feedback, though almost all participants also appreciated sonification for some tasks. While participants considered the verbal feedback to be precise and easy to understand, they perceived the sonification to be useful for conveying speed (e.g., slow vs. fast taps) and magnitude of change, that is, in communicating not only that a correction needs to be made but by how much.

The verbal feedback reflected relatively simple analysis of shapes, focusing on location, size, speed, and aspect ratio, but did not assess other shape characteristics, such as shape closure or the “roundness” of a circle versus a square. Although not statistically significant, the data in Study 3 suggests that gesture sonification may be superior to verbal feedback at communicating shape closure.

It is important to note that the verbal feedback and sonification techniques we tested were not informationally equivalent. In designing these feedback methods, we attempted to provide the best feedback for each modality. For sonification, our goal was to provide an understandable mapping between movement on the touchscreen and audio feedback, whereas for verbal feedback we focused on aspects of the gestures that could be automatically detected and communicated clearly. However, it may be possible to provide equivalent information across these two modalities: for example, the corrective verbal feedback could provide specific guidance about the position of a gesture’s start or end points, or the curvature of the gesture shape. Another difference between the two modalities is that currently the sonification provides real-time continuous feedback, while the verbal feedback only occurs at the end of a gesture. It may be possible to design real-time continuous verbal feedback that provides clearer instructions to users, but that also can provide feedback during a gesture. This notion of continuous verbal gesture feedback presents an exciting opportunity for future research.

7.4. Alternative Sonification Mappings

We used an *absolute* mapping in this research to provide location information of a gesture, but *relative* sonification has also been used previously to communicate shape trajectories [Harada et al. 2011]. Location information is important both because some interfaces designed for blind users make use of location-specific gestures and because location can help users position their finger appropriately for the start of a gesture, for example, so the entire gesture fits within the screen bounds. Relative sonification, however, may also be beneficial, especially for gestures that are mostly location insensitive (e.g., gestures used for scrolling or screen transition). Considering these trade-offs, a comprehensive gesture tutorial system may need to include both absolute and relative sonification depending on the type of gesture being taught.

7.5. Gesture Feedback and Long-Term Use

Our findings suggest that sonification and corrective verbal feedback could be useful for helping blind users to replicate gestures more accurately, but we did not evaluate whether this effect remains after the feedback is removed or whether users retain their knowledge over time. A comprehensive gesture tutorial system should extend to these learning contexts. As well, unlike corrective verbal feedback, sonification could be employed during regular use and it would be interesting to explore whether there is a benefit in doing so. Kane et al. [2011b] have shown that blind and sighted users exhibit differences in performing gestures, such as in size and variability, which could cause gesture recognition problems particularly for the blind users. Sonification feedback could potentially address this problem.

7.6. How Accurately Must Users Reproduce or Recognize Gestures?

The studies we have documented in this article present a variety of metrics for success. We have measured participants' ability to identify characteristics of a gesture based on a sonification of that gesture, their ability to reproduce a gesture based on sonification or corrective verbal feedback, and their subjective preferences for receiving feedback about the gestures that they performed. One question surrounding these studies is how accurately a gesture must be reproduced in order for the gesture training to be considered successful.

In practice, this question is difficult to answer in an absolute way, as the required precision needed to correctly perform a gesture is based on the underlying gesture recognizer, which users may not have any control over. We may, however, identify minimum levels of precision required for some types of gestures. For example, for a directional swipe gesture, a correct gesture would need to be within 45° of the ideal angle if the gesture recognizer recognized only four directional gestures, but would need to be within 22.5° of the ideal angle if diagonal gestures were also recognized. Thus, the ability of the target user to accurately and consistently produce gestures is directly related to the number of different gestures that may be recognized, and increasing users' ability to learn and consistently produce gestures could increase the efficiency of accessible gesture-based interfaces by enabling more gestures to be used.

In considering this estimate ($\pm 22.5^\circ$), Study 1 shows that pitch+stereo feedback enabled participants to accurately perform gestures in eight directions. Of course, this represents a naive estimate of how accurately the user must draw a gesture in order for that gesture to be correctly recognized; the actual accuracy required could be even higher due to limitations of the gesture recognizer or the choice of the designer. The requirements for reproducing gestures accurately may be exacerbated when considering mobile devices, as using a device while mobile can reduce the user's ability to tap or gesture accurately [Lin et al. 2007]. One possibility would be to use sighted individuals' performance when performing gestures that they have learned visually as a benchmark for how accurately visually impaired participants must repeat gestures; collecting this type of information between sighted users who learned gestures visually and visually impaired users who learned gestures through sound presents an interesting opportunity for future work.

In examining participants' ability to recognize gestures based on their sonification, once again the target level of accuracy depends upon the desired application. For training scenarios, it should be possible to describe the gesture in text first, and to use sonification to communicate more details about the gestures, as we tested in Studies 1 and 3. However, sonifications that enable users to identify a gesture based on sound alone could be useful in situations such as enabling awareness of a partner's gestures during a collaborative activity. Furthermore, participants' ability to identify gestures also gives us a general sense of how well the sonification is communicating information about the gesture. For example, Study 2 showed that participants often incorrectly identified sonifications of multitouch gestures, which suggests that the current sonification does not clearly convey all of the important aspects of those gestures.

7.7. Limitations of the Studies

As with any small set of studies, our work has limitations. Study 1 included only sighted participants, although past work [Walker and Mauney 2010] suggests that the findings should also apply to blind participants. Additionally, we did not control for previous touchscreen experience in either study, which means that participants had varying degrees of experience. That Study 3 showed an improvement in gesture performance even for these more experienced participants is promising. For completely novice participants, we expect to see a similar or larger effect, although more work is required to confirm this prediction. Finally, the findings from Study 2 are useful to inform the design of a comprehensive gesture tutorial system, but it will be important to evaluate such a system with a larger number of users.

Another limitation is the feedback we provided during the study. The corrective verbal feedback was hand-coded for a set of possible mistakes participants could make based on relative difference between a prompted gesture and a performed gesture. We plan to investigate in automatically generated feedback for our future work.

8. FUTURE WORK

The three studies presented here provide a basis for incorporating helpful audio feedback into touchscreen-based applications in order to assist blind and visually impaired users in correctly performing touchscreen gestures. Our approach includes both verbal and nonverbal audio, as well as prospective and corrective feedback. Overall, we found that each of these modes was useful in certain contexts, and that participants often had individual preferences for feedback modes. Thus, it may be valuable to allow users to control the types of feedback that they receive from the system. Beyond changing the feedback mode, users may benefit from even more control over feedback, such as being able to break a complex gesture down into steps, or to replay a gesture at different speeds.

A second opportunity for future work is to explore haptic feedback. Crossan and Brewster [2008], for example, used pen-based haptic feedback and stereo+pitch sonification to teach shape trajectories to people with visual impairments. It may be worthwhile, though not necessarily straightforward, to adapt their approach for touchscreen gestures (including location, size, speed) and to use the simpler vibration motor found on most touchscreen devices.

A third opportunity exists in providing a closer coupling between gesture feedback and the underlying gesture recognizer. In our studies, we chose gestures and characteristics (speed, size, location, etc.) based on common gestures used in current touchscreen-based devices. However, different computing systems may have different gesture sets, or may have different constraints on what constitutes a "correct" gesture. For example, some applications differentiate between the number of fingers used to perform a gesture, while others do not. A future version of this technique might generate feedback based on the system's individual gesture set or gesture recognizer. This system might also be extended to support user-defined gestures.

Finally, the system might be extended to provide instruction or help based on an individual user's performance using the system. Such a system may identify that a user is having difficulty performing a certain gesture, and may offer to provide a tutorial on the problematic set of gestures.

9. CONCLUSION

While accessible interfaces have improved touchscreen-based devices for blind users, challenges to true equal access remain. Improving the ability for blind users to learn to use their devices independently will more fully establish touchscreen devices as an option for users of all abilities. The techniques proposed and evaluated in this article—gesture sonification and corrective verbal feedback—show promise toward this goal. Our studies found that these audio representations of touchscreen gestures can enable blind and visually impaired people to perform a variety of gestures including taps, swipes, shapes, and multistroke and multitouch gestures. Overall, we found that sonification alone can be helpful, but that participants may benefit most from a combination of verbal and nonverbal audio instruction. A fruitful direction for future work will be to integrate both techniques into a full tutorial system for evaluation with novice touchscreen users.

ACKNOWLEDGMENTS

YooJin Kim, Judith Odili Uchidiuno, and Sarah Spencer provided valuable help and feedback on this work. We thank our participants for their help with the user studies. Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES

L. Anthony, Q. Brown, J. Nias, and B. Tate. 2013. Examining the need for visual feedback during gesture interaction on mobile touchscreen devices for kids. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC'13)*. ACM, 157–164.

S. Azenkot, J. O. Wobbrock, S. Prasain, and R. E. Ladner. 2012. Input finger detection for nonvisual touch screen text entry in Perkinput. In *Proceedings of Graphics Interface 2012*, Canadian Information Processing Society, 121–129.

O. Bau and W. 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*. ACM, 37–46.

M. N. Bonner, J. T. Brudvik, G. D. Abowd, and W. K. Edwards. 2010. No-look notes: Accessible eyes-free multi-touch text entry. In *Proceedings of the 8th International Conference on Pervasive Computing*. Springer-Verlag, 409–426.

L. M. Brown and S. A. Brewster. 2003. Drawing by ear: Interpreting sonified line graphs. In *Proceedings of the International Conference on Auditory Display*.

G. A. Calvert, P. C. Hansen, S. D. Iversen, and M. J. Brammer. 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14, 2 (2001), 427–438.

A. Crossan and S. Brewster. 2008. Multimodal trajectory playback for teaching shape information and trajectories to visually impaired computer users. *ACM Transactions on Accessible Computing* 1, 2 (2008), 12:1–12:34.

B. Frey, C. Southern, and M. Romero. 2011. Brailletouch: Mobile texting for the visually impaired. In *Proceedings of the 6th International Conference on Universal Access in Human-Computer Interaction: Context Diversity*. Springer-Verlag, 19–25.

T. Guerreiro, P. Lagoá, H. Nicolau, D. Gonçalves, and J. A. Jorge. 2008. From tapping to touching: Making touch screens accessible to blind users. *IEEE Multimedia* 15, 4 (2008), 48–50.

S. Harada, H. Takagi, and C. Asakawa. 2011. On the audio representation of radial direction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2779–2788.

E. Hoggan and S. Brewster. 2012. Non-speech auditory and crossmodal output. In *The Human-Computer Interaction Handbook*, Julie A. Jacko and Andrew Sears (Eds.). L. Erlbaum Associates Inc., Hillsdale, NJ, 220–239.

H. M. Kamel, P. Roth, and R. R. Sinha. 2001. Graphics and user's exploration via simple sonics (GUESS): Providing interrelational representation of objects in a non-visual environment. In *Proceedings of the 7th International Conference on Auditory Display*.

S. K. Kane, J. P. Bigham, and J. 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*. ACM, 73–80.

S. K. Kane, M. R. Morris, A. Z. Perkins, D. Wigdor, R. E. Ladner, and J. O. Wobbrock. 2011a. 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*. ACM, 273–282.

S. K. Kane, J. O. Wobbrock, and R. E. Ladner. 2011b. Usable gestures for blind people: Understanding preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 413–422.

P. O. Kristensson and L. C. Denby. 2011. Continuous recognition and visualization of pen strokes and touch-screen gestures. In *Proceedings of the 8th Eurographics Symposium on Sketch-Based Interfaces and Modeling*. ACM, 95–102.

B. Leporini, M. C. Buzzi, and M. Buzzi. 2012. Interacting with mobile devices via voiceover: Usability and accessibility issues. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*. ACM, 339–348.

M. Lin, R. Goldman, K. J. Price, A. Sears, and J. Jacko. 2007. How do people tap when walking? An empirical investigation of nomadic data entry. *International Journal of Human-Computer Studies* 65, 9 (Sept. 2007), 759–769.

N. Noble and B. Martin. 2006. Shape discovering using tactile guidance. In *Proceedings of the 6th International Conference on EuroHaptics*.

D. A. Norman. 2010. Natural user interfaces are not natural. *Interactions* 17, 3 (2010), 6–10.

U. Oh, S. K. Kane, and L. 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*, ACM, Article 13 (2013), 8 pages.

P. Parhi, A. K. Karlson, and B. B. Bederson. 2006. Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 203–210.

B. Plimmer, P. Reid, R. Blagojevic, A. Crossan, and S. Brewster. 2011. Signing on the tactile line: A multimodal system for teaching handwriting to blind children. *ACM Transactions on Computer-Human Interaction* 18, 3 (2011), 17:1–17:29.

J. Su, A. Rosenzweig, A. Goel, E. de Lara, and K. N. Truong. 2010. Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 17–26.

B. N. Walker, J. Lindsay, and S. Lab. 2005. Navigation performance in a virtual environment with bonephones. In *Proceedings of the International Conference on Auditory Display*. 260–263.

B. N. Walker and L. M. Mauney. 2010. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing* 2, 3 (2010), 12:1–12:16.

K. Yatani, N. Banovic, and K. Truong. 2012. SpaceSense: Representing geographical information to visually impaired people using spatial tactile feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 415–424.

K. Yatani and K. N. Truong. 2009. SemFeel: A user interface with semantic tactile feedback for mobile touch-screen devices. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 111–120.

H. Zhao, C. Plaisant, B. Shneiderman, and J. Lazar. 2008. Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Transactions on Computer-Human Interaction* 15, 1 (2008), 4:1–4:28.

Received June 2014; revised December 2014; accepted April 2015