Non-Visual Navigation Using Combined Audio Music and Haptic Cues

Emily Fujimoto
University of California, Santa Barbara
efujimoto@cs.ucsb.edu

Matthew Turk
University of California, Santa Barbara
mturk@cs.ucsb.edu

ABSTRACT

While a great deal of work has been done exploring non-visual navigation interfaces using audio and haptic cues, little is known about the combination of the two. We investigate combining different state-of-the-art interfaces for communicating direction and distance information using vibrotactile and audio music cues, limiting ourselves to interfaces that are possible with current off-the-shelf smartphones. We use experimental logs, subjective task load questionnaires, and user comments to see how users' perceived performance, objective performance, and acceptance of the system varied for different combinations. Users' perceived performance did not differ much between the unimodal and multimodal interfaces, but a few users commented that the multimodal interfaces added some cognitive load. Objective performance showed that some multimodal combinations resulted in significantly less direction or distance error over some of the unimodal ones, especially the purely haptic interface. Based on these findings we propose a few design considerations for multimodal haptic/audio navigation interfaces.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*auditory feedback, evaluation, haptic I/O*

Keywords

Non-visual navigation; spatial audio; vibrotactile feedback

1. INTRODUCTION

Many people today rely on navigation devices, especially their smartphones, to guide them from one place to another. While users can receive spoken instructions, most applications rely on visual information, such as maps or directions lists. However, walking while looking at one's phone, while common, is not safe (e.g. [9]). It also requires users to hold and actively interact with their phone, interrupting whatever else they are doing. Imagine instead an interface that could guide users without needing to be removed from their pocket, one that could integrate itself into another common activity, like listening to music. This would improve safety and require little additional effort from the user.

Audio and tactile cues are the two most commonly used visual replacements. In addition to verbal instructions, spatial audio has also been used for navigation. Usually this is done by making it appear as though there is a sound source on the target location. Vibration cues is another method for giving directions with haptics. However, in many instances this involves custom hardware with multiple vibrators that the user must acquire independently. Because of this, recent research has explored using off-the-shelf smartphones due to their ubiquitous nature. Unfortunately this significantly limits what can be manipulated as current phones only have one vibrating motor that can only be turned on or off; the vibration strength or roughness cannot be modified.

Although many audio and haptic alternatives have been explored, almost no alternatives using a combination of the two have been investigated without custom hardware. The benefit of using audio is that people already listen to music while walking, providing a behavior that could easily be adapted to suit navigational needs. It also requires no extra training to locate a sound source, unlike current methods of single vibrator navigation. However, modifying the user's music might decrease their enjoyment of it. There could also be confusion between between natural musical changes and manipulated ones meant to convey information. A vibrotactile interface does not have this problem and would not interfere with musical enjoyment, but it is not as natural as audio cues. A combination of these two modalities could create a navigational interface that is clearly understandable without detracting from the user's original activity. This paper explores how some combinations affect a user's performance, perceptions, and music listening experience.

2. RELATED WORK

As mentioned, one method of giving directions through audio cues is with spatial audio. For example, instead of saying "Go left," a virtual sound source could be placed to the user's left to indicate their target direction. Since it is not feasible to place speakers everywhere, this is accomplished through audio processing and headphones. One advantage of using spatial audio over spatial language is that spatial audio requires less mental effort to decode [7].

One of the early navigation studies by Holland et al. [5] used spatial audio to pan a tone right or left to indicate the target location. To help users find their goal, they added an extra "chase" tone that matched the pitch of the original tone until the user strayed off course, at which point its pitch would rise or fall. Distance was conveyed by how quickly the tone was repeated, using a geiger counter metaphor to increase the rate of the tone as the user got closer.

One advantage to using tones, like Holland et al. did, is that factors such as pitch and timbre can be used to convey information. However, it far more likely that people would prefer to listen to music rather than tones, so many later studies focus on manipulating music instead. While music could be modified like the tones were, Jones et al. [6] found that users disliked even a slight pitch alteration. This constrains the ways that information can be passed.

Since it would be annoying to pulse music like a geiger counter, distance is usually conveyed through volume, louder signifying closer. While this matches the idea of placing a musical source on the target, there are a couple of drawbacks. One is that, as seen in the work of Jones et al. [6], users can have trouble distinguishing gradual volume changes, resulting in heading in the wrong direction for some time before realizing that the sound is getting softer. In addition, the natural crescendos and decrescendos in the music can be confused with a change in distance information. Some have chosen to address this with further audio augmentation, such as adding a low-pass filter to muffle the music when the user is heading away from their target [3]. Liljedahl et al. [8], however, chose only to give distance information when approaching a turning point. Once at that point, they would pan the sound to one side, mimicking a turning gesture, to indicate where to head next. Furthermore, their use of notification tones rather than music avoided any confusion between natural changes and those they were manipulating.

Instead of using audio, some have experimented with using haptic cues for navigation. One advantage with haptics is that tactile senses, unlike auditory ones, are not generally used while navigating. One approach to conveying information has been to place multiple vibrating motors on the user's body, lending a spatial quality to the information. Information can then be conveyed by vibrating them in different patterns [2, 13]. However, this requires custom hardware, so subsequent work has focused on using only a single vibrating device such as cellphones. Since there are no longer spatial cues, this unfortunately means the user requires more training to learn the different signals.

To find the most intuitive interface Pielot et al. [10, 11] have tried conveying different types of information, including approaching or departing and a target direction expressed on either a discreet (left, right, straight ahead) or continuous scale. The final interface, Pocket Navigator, used a geiger counter metaphor for distance and a series of pulses for directions on a continuous scale. If the phone gave two short vibrations, the target was in front of the user. The longer the first vibration was in comparison to the second, the further to the left the target was, and vice versa for targets to the right. If the target was immediately behind the user, the phone would vibrate three times (Figure 1).

Rukzio, Hardy, and Rumelin [12] instead combined various vibration strengths, durations, patterns, and roughness to communicate direction and distance. Their interface, aptly named NaviRadar, would signal the current heading, wait as the imaginary radar "swept" the screen, and then signal again when it had reached the desired direction (Figure 2). Their final design used the vibration intensity for distance and different rhythms (one vibration versus two) to distinguish the current direction from desired.

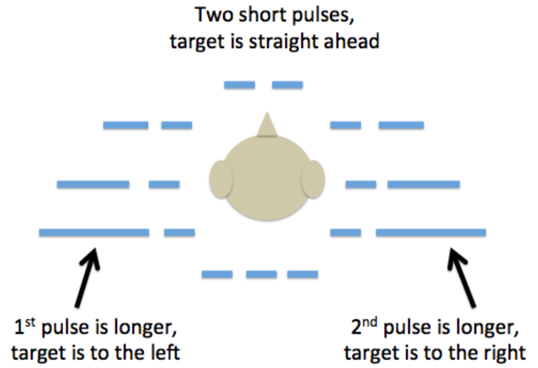


Figure 1: The vibration encodings to communicate different directions used by Pocket Navigator[11].

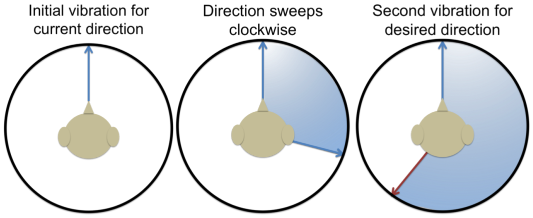


Figure 2: The direction interface used by NaviRadar[12]. The first vibration represents the direction the user is facing. Then an imaginary line sweeps around the user like radar to vibrate again, in a different pattern, in the desired direction.

A great deal of work has been done with both audio and tactile navigation methods individually. However, the only interface that has combined both audio and single-motor vibrations is in the work of Hara et al. [4]. Since their focus was on the testing environment rather than the navigation technique, the interface was relatively simplistic; when the user began to stray off course, the device vibrated and played a chime at the same time, using different patterns to indicate how far off course the user was. To our knowledge, no one has attempted to combine more complicated unimodal interfaces. By joining these two modalities, we can start to find ways to leverage the benefits of each one.

3. EXPERIMENTAL METHODS

Users tested six interfaces in a simulated navigational task. For this task, users walked around a room while responding to periodic navigational cues. The room had two large projector screens taking up the majority of two adjacent walls. The area in front of the projectors was left open with a square outlined on the floor to circle the majority of the space as a rough guideline of where to walk. At the back of the room was a table and lamp for filling out forms and some chairs for resting (see Figure 3). To encourage awareness of their surroundings during the task, as is needed to safely navigate, users were also asked to note when a given target image appeared on one of the projector screens. Both screens displayed the same image at all times, and to make it easier to see the projected images, the overhead lights were turned off. However, the lamp at the back of the room was kept on to provide enough light to see their surroundings.

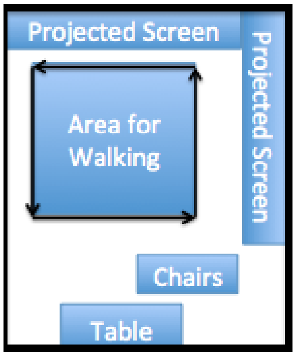


Figure 3: Experimental setup.

Table 1: Summary of direction and distance encodings to be tested.

	Haptic	Audio	
Direction	*DirHapVib* Vibration patterns 1	*DirAudSpatial* Stationary sound source 2	*DirAudSweeps* Moving sound source 3
Distance	*DistHapGeiger* Time between vibrations A	*DistAudVol* Music volume B	

3.1 Interfaces

To test the effectiveness of combined audio/haptic interfaces, we chose to communicate very basic navigational information: direction and distance. Based on previous work we chose to communicate *direction* using three methods and *distance* with two, for six total interfaces. We limited ourselves to methods that could be implemented on off-the-shelf smartphones. This constrained haptic manipulation to turning the vibrator on and off. We also focused on interfaces that did not require the user actively to manipulate their device, meaning the user could receive directions with the phone still in their pocket. This keeps the user's hands free and able to perform other tasks.

We refer to the three direction methods as DirHapVib (1), DirAudSpatial (2), and DirAudSweeps (3). DirHapVib (1) is from Pocket Navigator, which uses two vibrations to indicate direction as described above. DirAudSpatial (2) is the method most commonly used for spatial audio navigation: make sound come from the target direction. DirAudSweeps (3) is from the work of Liljedahl et al. which also uses spatial audio. Instead of placing a stationary sound source, the interface places the target in front of the user and gradually "sweeps" around until it is at the target location before resetting and repeating. The two distance methods are DistHapGeiger (A) and DistAudVol (B). DistHapGeiger (A) users the geiger counter metaphor used by Pocket Navigator. DistAudioVol (B) relies on the metaphor of placing an audio beacon on a target, so the music is louder the closer the user is to the target (see Table 1 for a summary).

Direction information was limited to the 180 degrees in front of the user. We did this for two reasons. First, since most travel is walking forward with occasional turns, most of the cues are likely to fall in this range. The user rarely needs to turn more than 90 degrees unless they have made a mistake. Second, while front-back errors are common in spatial audio, there are a number of different ways in which one could try to compensate for this, including those previously mentioned. We did not want to further complicate the study, so we controlled for this factor by limiting the possible directions. Distance used a linear scale from 0.0 ("near") to 1.0 ("far"). We chose to communicate relative rather than absolute distance because the degree of precision appropriate for absolute distance varies based on context, and because absolute distance does not map well to our problem formulation of navigation as a background task. Labels such as near and far are more flexible and can easily be changed to percent completion if desired. By using relative distance, we believe that our results will be more generalizable to a variety of situations and contexts.

All music must have a volume level and some distribution of sound between the user's ears. All vibrations have some rhythm and interval. Because of this, for interfaces where a given direction or distance method was not being used, the manipulated variable was set to the middle or "straight ahead" value. For example, DirHapVib/DistAudVol (1B) had the audio balanced between both ears ("straight ahead") and the vibrations came at medium distance intervals.

We investigated which interface users prefer and perform best with. For this, we have the following hypotheses.

H1) Because humans do not need consciously to retrieve signal-meaning mappings, when comparing DirAudSpatial (2) and DirHapVib (1), DirAudSpatial (2) will
  a) require less effort
  b) take less time to identify
  c) be more accurate

H2a-c) Same as H1a-c but with DirAudSweeps (3) instead of DirAudSpatial (2).

H3) Since it provides a moving target, users will rate changes in direction as more easily perceived for DirAudSweeps (3) than DirAudSpatial (2).

H4) DirAudSweeps (3) will interfere more with the music listening experience than DirAudSpatial (2).

H5) Since many people already listen to music, making adoption of DirAudSpatial (2) and DistAudVol (B) require little added effort, when compared with DirHapVib (1) and DistHapGeiger (A), users will
  a) be more open to adopting interfaces using DirAudSpatial (2) and DistAudVol (B)
  b) prefer interfaces with DirAudSpatial (2) and DistAudVol (B)

H6) Since modified cues will not be confused with musical changes, interfaces using DistHapGeiger (A) as compared to DistAudVol (B) for distance will
  a) be easier to use
  b) take less time to identify

3.2 Study Design

To simulate actual navigation, we had users perform two tasks while walking around the test room: a distractor task and a navigation task. This distractor task was to measure situational awareness since one goal is to have the user pay more attention to their surroundings rather than their navigation tool. For this, the user was asked to watch for a specific target image. When they saw their target, users then left-clicked a wireless mouse. All of the target images were cars of various shapes and colors (see Figure 4). To measure the effect of different levels of cognitive load on performance, users were randomly placed into one of two

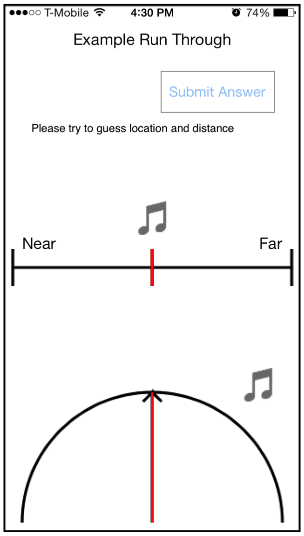
Figure 4: All possible targets for the distractor task. The first row is the "easy" condition where just a color, but not a specific shape, is the target.

conditions. In the "easy" condition, the user's target was a single color, meaning they had to indicate when any car of the given color appeared. In the "hard" condition, the target was one specific car shape of a given color. The projector screen would remain blank for a random amount of time (between 2 and 12 seconds) before a car would appear in a random location for a random period of time (between 3 and 10 seconds) before disappearing. Regardless of which condition the user was in, their target image would appear with a frequency of about 20%. When the user left-clicked to indicate that they had spotted their target image, the screen would flash green briefly to indicate that click had registered. Users also had the option to right-click to pause the task. To further distract the users and add to the realism of their task, we played ambient sounds similar to those that might be heard while walking along a street.

For the navigation task the user interpreted the music and/or vibration cues of a given interface and inputted the corresponding direction and distance into a smartphone. For a random period of time (between 10 and 20 seconds) the signal given was a neutral signal indicating that the target was directly ahead of the user and at a medium distance away. After that the direction and distance would randomly change, and the user could input the new information into the phone using the interface shown in Figure 5. When they entered an answer, the signal would return to neutral again and the process would repeat. To avoid situations where the user could not detect the change in stimulus, we added a small visual cue to the application. When the stimulus was neutral, the submit button was grayed out and could not be pressed. When the button became active, it was an indication that the user should enter information.

While direction randomly changing is realistic, random distance seems less so. Normally the distance would slowly decrease until it jumped far away when a new target had been acquired. However, most of the time users only need to affirm that they are on the correct path, meaning that it would be most important to have distance information right after reaching a target. This would tell the user how long they had before they needed to focus on navigation again, and this is consistent with randomly changing distances.

Since users did not need to interact with the device constantly for the navigation task, it was recommended that they spend most of their time looking for their visual targets and pause when they wanted to use the phone. We chose to allow users to do this because in practice, the user would not need to look at the device to enter information upon receiving a signal. Instead they would simply change where they were headed, which would require minimal visual attention and therefore not subtract from the user's awareness of their surroundings as much as needing to look at the phone screen. This is also similar to users finding a safe location where they can pause to interact with the device.



Example Run Through

Please try to guess location and distance

Figure 5: The application interface to input the user's responses during the Trial phase. The user drags the red lines to indicate a distance and direction.

While the distractor task difficulty was between-subjects, the interface was varied within subjects, so each user tried all six interfaces. The order of interfaces was counterbalanced using a Latin square design. For each interface, the user went through four phases. The first was a *training phase* where the user could control the direction and distance information to become familiar with the feedback. When they felt comfortable, they moved on to the *quiz phase*. During this part, the user was asked to identify what direction and distance the interface had randomly chosen to communicate. They were then shown the correct answer. If they were within 20 degrees of the actual direction and 0.15 of the actual distance, which was measured on a scale from 0.0 to 1.0, their answer was deemed "correct." They continued to receive random directions and distances until either they had ten correct answers or five minutes had passed. After that, they moved on to the *trial phase*. For this, users were asked to walk in a circle around the test room while performing both the distractor and navigation tasks described above. While there was a box outlined on the floor to give users a rough guideline for where to walk, remaining on that path was not strictly enforced. Users continued to walk while performing their two tasks for the duration of two songs before moving on to the *final phase*, which is where the user filled out the NASA Task Load Index for the trial phase and was given the opportunity to comment on that interface.

3.3 Procedure

After filling out a demographic survey and listening to instructions about the task they would do, the users adjusted the volume of the device so they could still hear music at its softest but did not find the loudest volume painful. They then went through the four phases (detailed above) for each of the six interfaces. We recorded how accurately the user entered both distance and direction information, how long it took them to enter the information, how many of the distractor targets were found, missed, or incorrectly identified, and how long it took them to respond to the distractor task.

Once users had completed all the phases for each of the interfaces, they were finally given a post study questionnaire to fill out. The questionnaire had the following instructions. All ratings were on a 7-point Likert scale.

- Please rate each of the interfaces on how easy you found it to use.
- Please rate each of the interfaces on how much you think you'd use it for navigation if given the chance.
- Please rate each of the interfaces on how annoying you found them.
- Please rate each of the interfaces on how easily you could identify a change in distance or direction.
- Please rank each of the interfaces in order from your favorite (1) to least favorite (6).
- Why did you choose the interface that you did as your favorite?
- Why did you choose the interface that you did as your least favorite?
- For the interfaces using music, to what extent do you think using them would detract from your music listening experience?
- If you have any other comments that you'd like to add, write them here.

4. RESULTS

A total of 41 people (16 male, 25 female) age 18 to 24 ($\mu = 20.3, \sigma = 1.44$) were recruited for the study. The task took about an hour and a half to complete, and each participant was paid 10 dollars for their time. Most of the participants were familiar with smartphones (36/41).

A three-way ANOVA was run with difficulty, direction method, and distance method as the factors. Bonferroni corrected post-hoc pairwise comparisons were then done using a significance level of 0.05. Graphs with error bars show a 95% confidence interval.

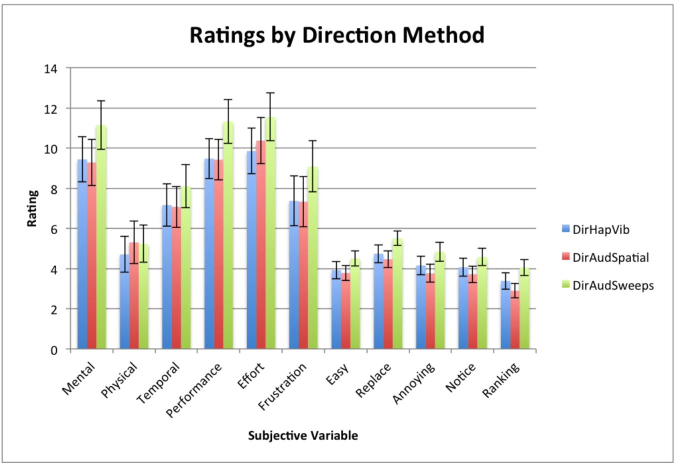
4.1 Subjective

Three of the qualitative variables taken from the NASA TLX showed no significant difference between interfaces. These were physical demand, temporal demand, and effort. Unless otherwise noted, there were no significant interaction between factors or a significant effect on difficulty.

For conveying direction, when the DirAudSpatial (2) method was compared to the DirAudSweeps (3) method, it was rated significantly less annoying and easier to use. Furthermore, both the DirAudSpatial (2) and DirHapVib (1) methods were given a higher preference and rated more likely to be adopted than the DirAudSweeps (3) method (Figure 6).

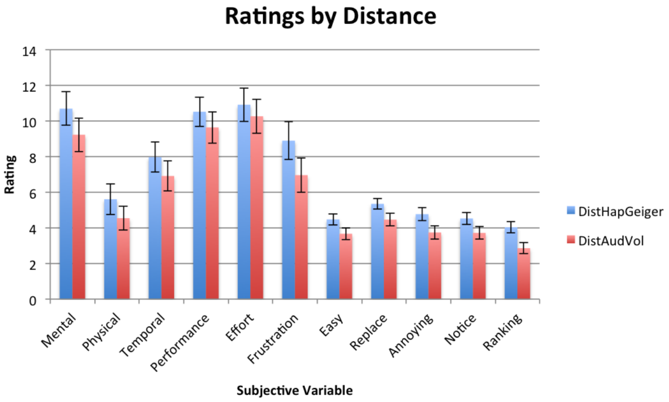
For distance, DistAudVol (B) was rated significantly less annoying, easier to use, less frustrating, less mentally demanding, and more likely to be adopted compared to DistHapGeiger (A). In addition, DistAudVol (B) tended to receive higher preference in order of favorites (Figure 7).

For perceived performance, the distance method did not affect ratings while both difficulty and direction method did. As would be expected, users in the hard condition thought they performed worse than those in the easy condition. Also, like many of the other variables, users thought they did better with both the DirAudSpatial (2) and DirHapVib (1) direction methods rather than DirAudSweeps (3).



..

Figure 6: Summary of ratings for subjection questionnaire by direction method. The first 6 are from the NASA TLX with possible scores from 0.5 to 20.5. The last 4 are from the final questionnaire and range from 1 to 7. For all a lower score is preferable.



..

Figure 7: Summary of ratings for subjection questionnaire by distance method. The first 6 are from the NASA TLX with possible scores from 0.5 to 20.5. The last 4 are from the final questionnaire and range from 1 to 7. For all a lower score is preferable.

The final qualitative variable was how easily users were able to notice a change in navigation information. Unlike previous variables, this one had interaction in all three of the main factors. Across both difficulties, users said they could recognize changes more easily when using DistAudVol (B) rather than DistHapGeiger (A). Direction, however, was only significant in the hard condition where DirAudSpatial (2) performed better than DirAudSweeps (3).

It is clear that the users did not like the DirAudSweeps (3) method; they rated it significantly worse than the other two direction methods on a number of factors. While there was no significant difference between DirAudSpatial (2) and

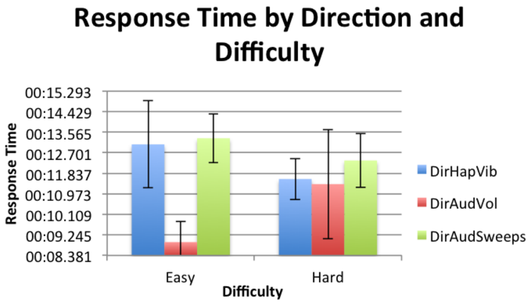


Figure 8: Time taken to decide upon a direction and distance, ignoring the distance method.

DirHapVib (1), DirAudSpatial (2) performed significantly better than DirAudSweeps (3) on more factors than DirHapVib (1) did, suggesting that it might be a slightly better choice. Users also thought that DistAudVol (B) was better than DistHapGeiger (A). This information suggests that users preferred either the DirAudSpatial/DistAudVol (2B) or the DirHapVib/DistAudVol (1B) interface.

4.2 Objective

None of the data related to the distractor task showed any significant difference based on interface or difficulty. However, there were significant differences in the performance for the main task in the amount of time taken to input a user's answer and their direction and distance accuracy.

4.2.1 Response Time

For the amount of time taken to respond to the feedback, there was an interaction between the difficulty level of the distractor task and the method of conveying the direction. In the easy condition, users responded significantly faster with DirAudSpatial (2) than either DirHapVib (1) or DirAudSweeps (3). However this advantage disappeared when looking at the hard condition (Figure 8).

It makes sense that people would identify DirAudSpatial (2) cues faster since, unlike DirHapVib (1), the user does not consciously have to "decode" its meaning. However, this would not explain why DirAudSweeps (3) did not also perform well. Another explanation is that DirAudSpatial (2) is the only interface that does not have a minimum response time; for both DirHapVib (1) and DirAudSweeps (3) a certain amount of time must pass as the user waits to see either how long the vibration lasts or at what angle the audio beacon eventually stops. This requirement is not there for the DirAudSpatial (2) condition, but its advantage seems to disappear when the users are in the hard condition, suggesting that the added cognitive load of the distractor task masks any benefit gained by using this method.

4.2.2 Distance Error

For distance error, as might be expected, users in the hard condition had significantly higher error than those in the easy condition. Oddly enough, the direction method also significantly affected distance error. In particular, error went up significantly when DirAudSweeps (3) was used (Figure 9). If either distance or direction were haptic, there was also less error if the other dimension was DirAudSpatial (2) or DistAudVol (B) rather than also haptic (Figure 10).

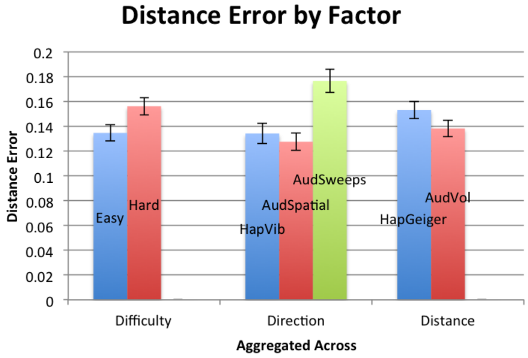


Figure 9: Distance error collapsed across the three main factors.

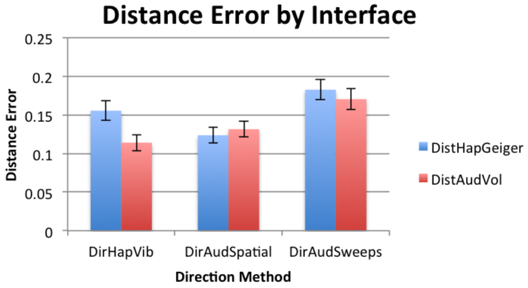


Figure 10: Distance error by interface ignoring difficulty.

One of the more interesting outcomes is the improvement interfaces with one of DirAudSpatial (2) or DistAudVol (B) combined with either DirHapVib (1) or DistHapGeiger (A) had over a unimodal haptic interface. They also performed better than DirAudSpatial/DistAudVol (2B) but not significantly so. However, if it was just the cross modal nature that caused an improvement, one would expect DirAudSweeps/DistHapGeiger (3A) to also perform better since DirAudSweeps (3) also uses audio cues, which is not the case.

One explanation for why DirAudSweeps (3) so drastically decreases distance accuracy might be because humans are naturally inclined to notice changes and movement as a survival mechanism. Since the DirAudSweeps (3) method involves constant movement, it is possible that some amount of the users' attention was unconsciously diverted to monitor the change. Since DirAudSweeps (3) by itself does not convey any distance information, this would mean users' attention was being drawn to the direction over the distance information. This explanation could also help account for DirAudSweeps' (3) performance with directional error.

4.2.3 Direction Error

For directional error, surprisingly, DirAudSweeps (3) performed fairly well. In the hard condition, DirAudSweeps/DistAudVol (3B) performed significantly better than either DirAudSpatial/DistAudVol (2B) or DirHapVib/DistHapGeiger (1A). In the easy condition though, using DirHapVib/DistAudVol (1B) produced the least error, performing significantly better than all the other interfaces except for DirHapVib /DistHapGeiger (2A) (Figure 11).

Since attention to moving objects is often largely unconscious, it would make sense for users to do better with DirAudSweeps (3) in the hard condition, where more of their conscious attention would need to be focused on the dis-

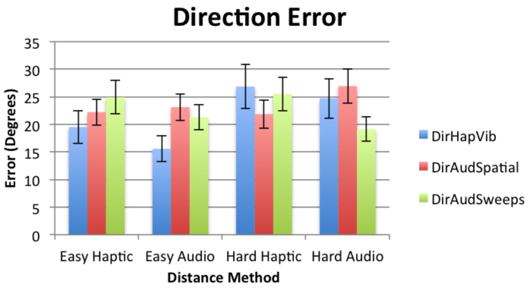


Figure 11: Direction error for all interfaces in both the easy and hard condition.

tractor task. When that extra burden is not there, as in the easy condition, the added benefit of having more unconscious attention allocated to the DirAudSweeps (3) method might not make a noticeable difference as there are more attentional resources to devote to the navigation task.

Another explanation for why DirAudSweeps (3) would do better as compared to all the others is that in a way, DirAudSweeps (3) is a combination of information given in both DirAudSpatial (2) and DirHapVib (1) methods; the user still gets the spatial sound information from DirAudSpatial (2), but they also get timing information based on how long the audio beacon moves for before resetting. Timing information is what the DirHapVib (1) method is based off of since users must determine direction based on the amount of time the phone vibrates. However, if this was the main reason for DirAudSweeps' (3) performance, one would expect it to outperform DirHapVib (1) and DirAudSpatial (2) in both difficulty conditions instead of just the hard one, although it has comparable performance to all the other interfaces except for DirHapVib/DistAudVol (1B).

4.3 User Comments

While cross modal DirHapVib/DistAudVol (1B) and DirAudSpatial/DistHapGeiger (2A) interfaces did not perform worse than their unimodal DirHapVib/DistHapGeiger (1A) and DirAudSpatial/DistAudVol (2B) companions, and in fact sometimes had significantly better performance, a few users commented that they thought they were harder to use. User 15 thought that "*combining input from multiple sensory modalities in a task like this can be somewhat overwhelming*" while user 12 mentioned that he "*[h]ad to asses distance and direction in two separate trains of thought.*"

A few users also commented that they did not like waiting for vibration signals, especially for DistHapGeiger (A). This suggests that the haptic interfaces suffered for not being as instantaneous as DirAudSpatial (2) or DistAudVol (B). This means that haptic interfaces might benefit greatly if modern cell phones allowed for more control over the vibrating motor's behavior. This would allow applications to use channels of communication other than timing or patterns.

When asked to what extent they thought the musical interfaces would detract from their enjoyment of the music, 18 participants said they did not think it would affect it that much, 4 thought it would detract some but not a lot, 9 said that it would detract a lot, and 4 did not respond. The remaining 6 participants had mixed reactions. Of those 6, half broke the audio interfaces into their different types. All 3 mentioned that the DirAudSweeps (3) would detract a lot from the experience. The two that mentioned interfaces other than DirAudSweeps (3) both thought the DirAudVol (B) was ok, but had differing responses as to how detracting the DirAudSpatial (2) was. Of the other 3 with mixed reactions, two seemed to think that the interfaces' effect on the music would be ok if the primary goal was navigation but not if it was just for pleasure. The last one only thought it would be a problem if she were listening to new music.

At first, these responses seem discouraging; about 25% of the participants who responded thought audio manipulations would interfere a great deal with their musical enjoyment. However, the question did not separate out the different audio interfaces, meaning that this result may be skewed by DirAudSweeps (3), which likely does detract from the musical experience for most people. Since a great deal of research shows that negative experiences are more likely to be remembered and more likely to influence judgements [1], it is possible that these users based their answers largely on their reaction to DirAudSweeps (3) rather than all three audio interfaces as a whole. In light of this negativity bias, the results look fairly good as nearly 50% of responding participants reported that the audio interfaces would not detract much, if at all, from the overall experience.

5. DISCUSSION

While we did a lab study, we believe that the results can still tell us about performance in an outdoor navigation task. We included ambient noises such as might be heard while walking along a street as well as visual distractors. In addition, users actually walked about the room as they would have to when actually navigating. Furthermore, by using a lab study, we were able to test more interfaces without unduly increasing user fatigue.

H1: Audio Spatial vs Haptic Direction Subjectively users did not think DirAudSpatial (2) took less effort than DirHapVib (1), resulting in no evidence for hypothesis H1a. Users also were not more accurate with DirAudSpatial (2) and actually were worse in some interfaces in direct contradiction to hypothesis H1c. However, they were significantly faster, at least under low mental demand, supporting H1b.

H2: Audio Sweeps vs Haptic Direction When comparing DirAudSweeps (3) with DirHapVib (1), there was again no evidence that users thought it took less effort or less time to use (H2a and H2b). There is conflicting data in regards to the users' accuracy though with DirAudSweeps/ DistAudVol (3B) doing better than DirHapVib /DistHapGeiger (1A) in the hard condition, but being outperformed by DirHapVib /DistAudVol (1B) in the easy condition, giving inconclusive evidence about H2c.

H3 and H4: Spatial vs Sweeps Audio Direction When comparing DirAudSweeps (3) and DirAudSpatial (2), surprisingly DirAudSpatial (2) was rated easier to notice in direct opposition to H3. However, there was support for H4 since users preferred DirAudSpatial (2) and rated it as less annoying, a point that was emphasized in some of the written comments, than DirAudSweeps (3).

H5 and H6: Audio vs Haptic While there was no significant difference in preference or likelihood of adoption for DirAudSpatial (2) compared to DirHapVib (1), users ranked DistAudVol (B) better than DistHapGeiger (A), giving mild support in favor of both H5a and H5b. DistAudVol (B) was also rated easier to use than DistHapGeiger (A) in direct contradiction to H6a, although a few users did specifically comment that they had trouble telling distance information from musical fade-outs despite being familiar with the music. There was no evidence either way as to if DirAudSpatial (2) and DistAudVol (B) or DirHapVib (1) and DistHapGeiger (A) took less time to identify, which was H5b.

Best Interface One remaining question is which interface is the best? While DirAudSweeps (3) does perform fairly well with directional error, its poor performance in nearly every other parameter outweighs its benefit, making it a bad choice. Furthermore, one user also commented that it made him feel physically ill. Of the remaining four interfaces, DirHapVib/DistHapGeiger (1A) is also not a particularly good choice; subjectively users preferred having DistAudVol (B), and it was either comparable or significantly worse for all objective measures. The last three interfaces are much closer to each other in preference, performing both significantly better and significantly worse on different parameters. However, DirAudSpatial/DistAudVol (2B) and DirHapVib/DistAudVol (1B) came out on top for more parameters; DirAudSpatial/DistAudVol (2B) was better in the subjective measures and response time in the easy condition, DirHapVib/DistAudVol (1B) was better in the subjective measures and direction error in the easy condition, and DirAudSpatial/DistHapGeiger (2A) was only better in the response time in the easy condition. This only leaves DirAudSpatial/DistAudVol (2B) and DirHapVib/DistAudVol (1B). The two of them seem to fall into an accuracy/time tradeoff under low cognitive load; DirAudSpatial (2) is faster to identify but DirHapVib (1) is more accurate. Under high cognitive load, their performance is about the same.

6. CONCLUSIONS

The data suggest a couple of design considerations. The first is that interfaces that convey information though temporal cues should aim to take up as little of the users' time as possible. Users seemed to respond well to DirHapVib(1), which communicated through temporal feedback in the relative lengths of vibration over a short time frame, but they responded poorly to DistHapGeiger(A), for which the pauses between signals were often significantly longer. This is supported by user comments about disliking needing to wait.

Another is that while a multimodal interface might benefit user performance, users might think it requires extra effort compared to a unimodal one. Some struggle with simultaneously interpreting data from two modalities, and must expend effort to mentally switch tracks as they concentrate on one modality and then the other. However, it is possible that this overhead is overshadowed by the overall effort of interpreting the data signals. This would explain the lack of significance in subjective effort, but more work would need to be done to determine to what extent this is true.

There are a number of directions future work could take. One would be to study how user preferences and performance change both over a longer period of time and when navigating a real environment. While we added a number of distractions similar to an outdoor environment, users might still respond differently simply by knowing that there are hazards that are not present in a controlled lab setting. Although these might not appear even with an outdoor experiment, as there is some expectation that the experimenter will be looking out for the subject, providing supervised safety that is not normally present. Another direction would involve testing out different distance scales. While we used a simple, linear scale, an exponential scale, for example, could emphasize when the user is particularly close to their target and might improve performance. This work provides a foundation for exploring these other areas while keeping a focus on creating a usable yet nonintrusive interface.

7. REFERENCES

[1] R. F. Baumeister, E. Bratslavsky, C. Finkenauer, and K. D. Vohs. Bad is stronger than good. *Review of General Psychology*, 5(4):323, 2001.

[2] S. Bosman, B. Groenendaal, J. Findlater, T. Visser, M. De Graaf, and P. Markopoulos. Gentleguide: An exploration of haptic output for indoors pedestrian guidance. *International Conference on Human Computer Interaction with Mobile Devices and Services*, 191(204):358–362, 2003.

[3] R. Etter and M. Specht. Melodious walkabout-implicit navigation with contextualized personal audio contents. *Third International Conference on Pervasive Computing*, 191(204), 2005.

[4] M. Hara, S. Shokur, A. Yamamoto, T. Higuchi, R. Gassert, and H. Bleuler. Virtual environment to evaluate multimodal feedback strategies for augmented navigation of the visually impaired. In *Engineering in Medicine and Biology Society*, pages 975–978, Aug 2010.

[5] S. Holland, D. R. Morse, and H. Gedenryd. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, 6(4):253–259, 2002.

[6] M. Jones, S. Jones, G. Bradley, N. Warren, D. Bainbridge, and G. Holmes. Ontrack: Dynamically adapting music playback to support navigation. *Personal and Ubiquitous Computing*, 12(7):513–525, 2008.

[7] R. L. Klatzky, J. R. Marston, N. A. Giudice, R. G. Golledge, and J. M. Loomis. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology: Applied*, 12(4):223–32, Dec. 2006.

[8] M. Liljedahl, S. Lindberg, K. Delsing, M. Polojärvi, T. Saloranta, and I. Alakärppä. Testing two tools for multimodal navigation. volume 2012, Jan. 2012.

[9] J. L. Nasar and D. Troyer. Pedestrian injuries due to mobile phone use in public places. *Accident Analysis & Prevention*, 57:91–95, 2013.

[10] M. Pielot, B. Poppinga, W. Heuten, and S. Boll. 6th senses for everyone!: The value of multimodal feedback in handheld navigation aids. In *International Conference on Multimodal Interfaces*, pages 65–72, 2011.

[11] M. Pielot, B. Poppinga, W. Heuten, and S. Boll. A tactile compass for eyes-free pedestrian navigation. In *Human-Computer Interaction, INTERACT 2011*, volume 6947, pages 640–656. 2011.

[12] S. Rümelin, E. Rukzio, and R. Hardy. Naviradar: A novel tactile information display for pedestrian navigation. In *Symposium on User Interface Software and Technology*, pages 293–302, 2011.

[13] K. Tsukada and M. Yasumura. Activebelt: Belt-type wearable tactile display for directional navigation. In *Ubiquitous Computing*, volume 3205, pages 384–399. 2004.