Play it by Ear: A Case for Serendipitous Discovery of Places with Musicons

Anupriya Ankolekar
HP Labs
1501 Page Mill Rd, Palo Alto,
CA 94304, USA
anupriya.ankolekar@hp.com

Thomas Sandholm
HP Labs
1501 Page Mill Rd, Palo Alto,
CA 94304, USA
thomas.e.sandholm@hp.com

Louis Yu
Pomona College
185 E. Sixth Street,
Claremont, CA 91711, USA
louis.yu@pomona.edu

ABSTRACT
Current location-based services (LBS) let users find points of interest (POI) in their vicinity but can detract from the user's emotional experience of exploring a new location. In this paper, we examine how cues in the form of popular music (musicons) can emotionally engage users and enhance their experience of discovering nearby POIs serendipitously in unfamiliar places. The primary contribution of this paper is a field study, in which we evaluate the performance and emotional engagement of different types of audio-based cues for directing users' attention to specific POIs. Musicons and mixed-modality cues performed close to visual and speech cues, and significantly better than auditory icons, for POI identification while creating a much more pleasant and engaging user experience. We conclude that cues for POI discovery need not always be as explicit as the baseline visual cues. Indeed, the most challenging cues, auditory icons, led to a heightened sense of autonomy.

Author Keywords
Audio interfaces, mobility, affective computing, emotion

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Sound and Music Computing

General Terms
Design, Human Factors

INTRODUCTION
Location-based applications and services (LBS) [16], which tailor the information they provide to the user's location, are among the most popular and useful of mobile applications[1]. From local news, to the closest gas station, to public tweets posted locally, these applications let users discover people, points of interest (POI) and activities around their current location. The overwhelming majority of LBS, such as Google Maps, Where, Yelp and Foursquare, are visual, informational and focus on the physical characteristics of a place. They typically assume that users either have a specific destination, or have a well-defined need for information about a specific place or for a specific type of place. They are effective at providing comprehensive information and focus on filtering it to optimally present the most relevant fragments on the limited visual real-estate of the user's mobile device.

The visual cues used by current LBS are, however, not optimal in cases where the user does not have a well-defined information need. Sometimes, the goal of a mobile journey is to discover POIs in a pleasing, memorable and serendipitous manner, such as when the user is on vacation, at leisure, or wants to get a feeling or vibe about the people and culture of an unfamiliar place [13]. By requiring the user to constantly consult their mobile device and presenting comprehensive textual information about POIs, visual cues can actually detract from the user's experience. Our goal is to support a rich and emotionally engaging manner of POI discovery that preserves some of the "wonder and enchantment of an individual's exploration" [17].

We define the POI discovery process to be one in which the user wants to discover interesting POIs in the vicinity, but does not have a specific or urgent need. The LBS presents a cue about a specific POI that is nearby, which the user can choose to follow, if it is relevant, or should be able to ignore easily or even enjoy, if it is not. Thus, enjoyable cues that can be correctly mapped to POIs are critical to the success of an ongoing POI discovery process. Note that the process we describe is a constrained form of POI discovery that is particularly suited to systems design. It is applicable to any LBS with a location-aware discovery component and it can also be evaluated in a quantified, controlled manner.

Prior work has examined how non-visual cues, such as speech cues, auditory icons [24], adaption of continuous music [8, 5] and even haptic feedback [17] can be used to alert users to POIs nearby and used for navigation. Music cues or musicons[2] [12] have not yet been systematically explored for POI discovery, although artists have long experimented with the expressive effects of placing music in landscapes (e.g. [19]). Musicons have the potential to carry higher semantic content than some of the other non-visual cues. The rhythms,

[1]http://pewinternet.org/Reports/2012/Location-based-services.aspx

[2]We define musicons as snippets of popular music that convey information about a specific POI.

CHI'13, April 27–May 2, 2013, Paris, France.

instruments, lyrics and style of music can convey subtle emotional, cultural or historical information about a POI. Due to their ambiguity of reference, musicons might, however, be less precise in guiding users' attention towards specific POIs than visual cues and speech. For the same reason, though, musicons can form more subtle cues to POIs than typical visual or speech modes. Users can therefore either discover places implicitly or (given the prevalence of mobile music consumption [4]) tune out the music, and thereby transform the interaction with LBS into something more subtle and entertaining. To our knowledge, there has been no study that compares the effectiveness and emotional appeal of different kinds of cues for POI discovery.

In this paper, we present a field study comparing different kinds of cues in terms of their effectiveness (i.e. the speed, accuracy and confidence with which people identified the POIs) and the emotional response they elicit. The rest of the paper is structured as follows: we first summarize the relevant literature and present some innovative services and apps that are currently available. We then describe the design of the experiment and subsequently our findings. Our discussion section summarizes the findings and supplements them with some insights from the post-experiment survey responses. Finally, we conclude with an overview of the contributions of this paper, their implications and potential future work.

RELATED WORK

In the spirit of [14], we want to use audio to support serendipitous, enjoyable discovery of POIs rather than helping users direct themselves as quickly as possible to a given destination. Prior work has used auditory icons and earcons to convey information about POIs and has shown that navigation with audio is faster and more enjoyable than using a visual map [14, 22]. In contrast, Kurczak et al. [11] show that a continuous stream of audio representing a physical entity reduces player performance in location-based games, but that it increases their sense of immersion in the virtual world.

Music has been less commonly used for navigation, with the exception of systems that adapt the music volume and spatial balance to indicate the direction in which the user should go [23, 10]. Beyond navigation, the use of music for system feedback has also been examined. Sikora et al. [21] found that music sounds were consistently preferred over real-world sounds or auditory icons as system feedback signals.

Music has been shown to convey meaningful information faster than speech in certain situations [2]. In addition, music has a well-recognized effect on engaging the emotions and influencing mood [18]. It can convey rich emotional experiences and itself be a powerful stimulus for creating and evoking emotionally charged memories [20] and personal associations. Johnson [9] has demonstrated that different kinds of music can influence people's perception of their environment.

Many location-based music applications are starting to appear in the consumer space, e.g. Blicko, Herd.fm, Soundtracking, Spotisquare, SuperGlued, and Roqbot[3]. These services let users tag songs or playlists with locations or venues, thus allowing users to see what friends listen to where. The systems are typically implemented as mashups between social apps (e.g. Facebook), location apps (e.g. Foursquare) and music apps (e.g, Spotify). None of these systems use the actual content of music, i.e. the lyrics, style and instruments of the music to convey semantic information about location.

Despite substantial prior research on the use of audio in navigation and on the psychological effects of music, it is an open question as to how musicons perform compared to other kinds of cues for POI discovery.

METHOD

We conducted a field experiment to investigate the effectiveness and emotional experience of various kinds of audio cues. A field study was necessary because a system to assist in POI identification is highly dependent on its surrounding environment and cannot be tested realistically in a laboratory setting. An experimental setup was used to obtain a quantitative comparison of the effectiveness and experience of different audio cues. Since the focus of this study was on comparing how different kinds of audio cues can aid POI identification, which is an easily quantifiable task, we chose to conduct an experiment rather than an ethnographic field study.

The advantages of field experiments for mobile guides have been extensively discussed in Goodman et al. [6]. While the levels of potential confounding variables, such as light, noise, traffic and weather conditions could not be kept consistent, their variation manifests itself randomly across conditions. Like [7], we consider this to be acceptable because variation in such variables is an integral part of real-world usage. Removing all such essential variation would likely produce unrealistic results. Using actual locations and authentic environmental conditions provides vital data on the usability of such cues in practice.

Hypotheses

Hypothesis A: Musicons and auditory icons are associated with greater errors in identification, slower identification speed and lower confidence than visual cues

Due to their ambiguity, we expect musicons and auditory icons to take longer to interpret and to be associated with greater identification errors. Participants would also be less confident about having found the correct POI compared to conditions where participants were told the name of the POI sought.

Hypothesis B: Musicons and auditory icons are associated with greater feelings of pleasure, arousal and autonomy than visual cues

Due to their entertaining content, we expect musicons and auditory icons to be associated with greater pleasure. Since participants will also need to put in greater effort to interpret

[3] `http://blicko.com`, `herd.fm`, `soundtracking.com`, `spotisquare.com`, `superglued.com`, `roqbot.com`

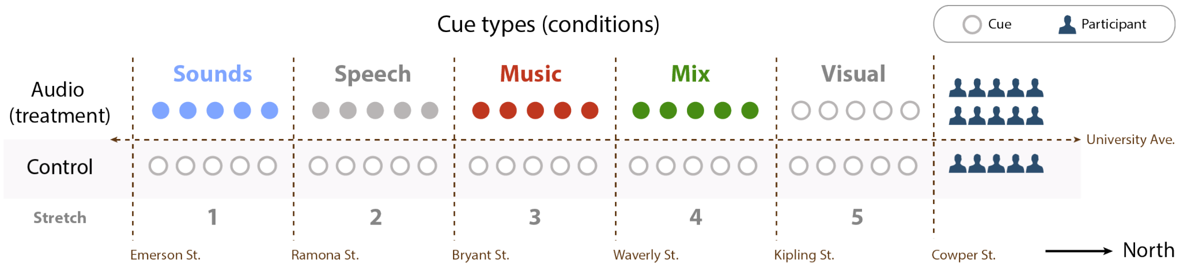
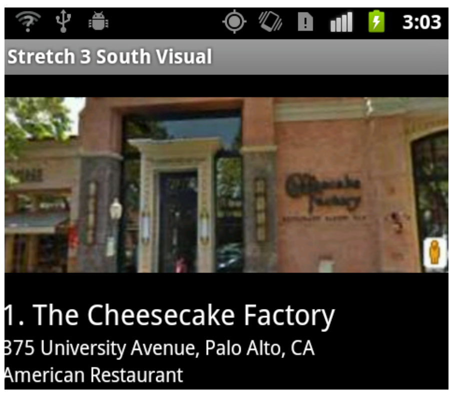
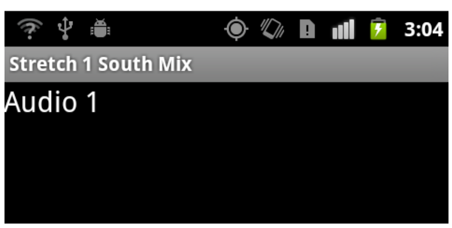


Figure 1. Design of the experiment.

these cues, we expect them to be associated with greater feelings of autonomy and arousal[4].



(a) Visual Condition



(b) Audio Conditions

Figure 2. Visual and audio cue views in the custom Android experiment app.

Participants

The experiment was conducted with 15 able-bodied participants recruited from the Palo Alto area for a 45-minute experiment. The participants were randomly assigned to either the audio treatment or to be controls, such that two-thirds of the participants (10/15) experienced the audio treatment and a third (5/15) were controls. The participants were mostly young (73% of them were between 25 and 34 years) and male (5/16 participants were female), which we consider fairly representative of LBS users on smartphones today[5].

The participants were not overly familiar with the experiment location; 73% of them visited the area once a month or less. During the experiment, participants reported on whether they knew the POI and the results show that in 81% of the trials (i.e. the POI, subject pairs matched in the experiment), the subject was unfamiliar with the POI (74% of the pairs in the control conditions, and 84% of the pairs in the audio treatment were with unfamiliar POIs). Only a single POI (out of 25) was matched with more subjects who were familiar with it than unfamiliar with it. Given that the street experiences high turnover in shops and restaurants, and given that we were requiring participants to interpret unfamiliar references to POIs, we deemed this to be an acceptable level of familiarity with the location.

All the participants had used smartphones before, with iOS being the most commonly used platform (8/15), followed by Android (6/15). The majority of participants were also very familiar with new technologies, using GPS-based systems (12/15), and LBS (11/15) at least once a week; and accessing the mobile Internet daily (11/15). They were therefore very comfortable with receiving cues about POIs in their vicinity via a mobile phone.

Experiment design

The experiment was a mixed factorial design depicted in Figure 1. All participants were asked to walk along the same route and identify POIs based on the cues they received. Control participants received visual cues for the entire route. The treatment for audio participants was a repeated measures design, where each participant received 5 different kinds of cues on equal-sized, consecutive stretches of the route. The cue types were: (a) visual (*Visual*), (b) pure speech (*Speech*), (c) auditory icons (*Sounds*), (d) musicons (*Music*) and (e) a mixed-modality cue (*Mix*) combining speech, auditory icons and musicons.

Participants who were controls were only exposed to visual cues without audio, with the name, address, category and picture of the POI on the phone screen (see Fig. 2(a)). This information was taken from the Google Places information for the POI. Visual information, and in particular pictures, are known to lead to superior identification and navigation [7], so using

[4]Note that this hypothesis complements Hypothesis A in the sense that it is easy to conceive of cues that have few errors, fast identification and high confidence but suffer in terms of the emotional experience, and vice versa.

[5]Smartphone Adoption and Usage: `pewinternet.org/Reports/2011/Smartphones.aspx`

visual controls allows us to compare user performance on ambiguous cues, such as musicons and auditory icons, with an unambiguous and familiar type of cue. We also intended the identification of POIs to be fast and with high confidence in the control condition, to allow us to have a satisfactory upper bound for performance in the audio treatment.

The audio cues were presented without any identifying information about the POI (see Fig. 2(b)). The speech cues were a Google TTS (text-to-speech) rendition of the name and address of the POI. We introduced this condition because it is an audio modality and yet is unambiguous. Auditory icons are representative sounds of a place, e.g. a hair salon was represented by sounds of scissors clipping and a hair dryer blowing. The musicons were fragments of music that could be representative for that place, e.g. an upscale Italian restaurant was represented by traditional, romantic Italian singing. The *Mix* condition consisted of the auditory icon for that POI, followed by the speech cue and then the musicon. This condition was introduced to investigate whether combining different forms of information can mitigate the potential disadvantages of any single modality[6]. We used 5 different cues in each condition[7]. The actual design of the auditory icons and musicons is discussed in the next section.

Due to constraints on sample size, experiment scope and the fact that we wanted to make the experiment resemble a regular stroll, we counterbalanced the direction in which participants followed the route (northbound or southbound), but fixed condition and location pairings. This kind of design can introduce sequence effects and location bias, since for any given POI, experiment participants receive only either one kind of cue in the audio treatment or the visual cue. We therefore introduced the control condition described in the previous section. With 5 control participants who only received visual cues for the entire route (with direction counterbalanced as before), we can perform a between-subjects comparison of the performance with each type of audio cue to the baseline of visual cues. Thus, the experiment design enables both between-subject and within-subject comparisons of all types of cues to the controls. We also note that neither the subjects exposed to the audio conditions nor the control group subjects had any knowledge of which condition they would be in or when they would hear or see certain types of cues.

The experiment was conducted in a busy shopping street in the primary business area of a medium-sized town, Palo Alto, CA, USA (see Figure 3). This location was chosen because it was densely packed with many POIs, such as shops, banks and restaurants, allowing us to test a large number of cues within a short route. The high density of POIs also ensured that participants had to put in some effort to choose a single, correct POI for each cue and that the experiment was not too long and tiring. The different stretches of the repeated measures design contained a very similar density of POIs to reduce the chance of location bias. The POIs to be identified were themselves chosen to constitute a balanced mix of large and small places, as well as prominent and obscure places. They were generally equally spaced within each stretch, but could be on either side of the street[8].

Figure 3. Experiment location was a densely packed shopping street.

Since the experiment was measuring the emotional experience of cues, we took care to make the experiment setting as natural and unobtrusive as possible. Therefore, the participants were given in-ear headphones that provided good sound quality, but did allow some ambient noise to filter in for safety reasons. Participants were also asked to follow the route at a natural, but leisurely pace.

The route consisted of about 5 street blocks, with an equal density of POIs per block. Within each block, participants received 5 cues associated with 1 POI each for a duration of 1 minute each. Without stopping, the route took about 10 minutes to walk. Altogether, each participant completed 25 trials (POIs) and spent 25 minutes to walk the route. With 15 participants, our study comprised 375 trials all together. Since the blocks were relatively short compared to the experiment time available for each block, participants could potentially walk back and forth in case they walked past a relevant POI. This again was by design to mimic a natural stroll on the street. The experiment POIs in each block were equally spaced and were rarely visible from the traffic lights between blocks.

Cues

While designing the musicons and auditory icons, the goal was to convey enough semantic information about a given POI to let the user uniquely identify it. We decided to focus on conveying the category of the POI using various features of the musicons and sounds. We also wanted to allow the user to identify the correct POI from among a set of POIs located nearby that might potentially be in the same category.

[6] To maintain reasonable experiment scope, we chose to include only 1 combination condition, choosing it to be "the best possible without visual cues", i.e. including all audio cues.

[7] All the audio used in the experiment can be heard at **`www.crowdee.com/foxpoi`**.

[8] A map of the POIs used, including a heatmap of GPS traces from subjects, is available at **`www.crowdee.com/foxpoi/poi.png`**

E.g. one of the POIs to be identified was an Italian restaurant among several other Italian restaurants within the same block. A good cue would allow the user to identify the correct Italian restaurant in that block.

The musicons created for the experiment are snippets of modern and popular music, where the melody, tempo, rhythm and lyrics are used to convey information about the POI. Musicons can differ in their emotional content and can generate affect independent of the experiment, especially if the snippet is familiar to users. To minimize this potential confound, we chose music that was likely to be unfamiliar to users, e.g. mostly using indie music or more obscure tracks of familiar artists. This music was chosen by an indie music DJ. Only the first minute of the track was played and we removed the first few seconds to reduce the variation in the starting time of the primary musical content of the track. Other than this, the choice of music was relatively unrestricted, e.g., we allowed lyrics and different styles of music. Designing the best possible musicons for the chosen POIs was an arduous task, which would have only been more difficult had we constrained ourselves to a particular style or type of music.

The process used to design the cues for specific POIs was to first identify key distinguishing features of the POI. The key features to convey about an Indian restaurant e.g., could be 'restaurant' or 'Indian', but if there were many restaurants in the vicinity, we chose to convey the feature 'Indian'. We then chose audio to convey those features. For concrete features, e.g. 'books' (for a book store) or 'cheesecakes' (for 'The Cheesecake Factory'), we relied on lyrics to communicate the feature, e.g. choosing Louis Armstrong's 'Cheesecake' song where the word "cheesecakes' is mentioned prominently and repeatedly in the song. For abstract features that are more difficult to convey explicitly, e.g. 'Indian' or 'Thai', we relied on instrumental music to remind listeners of those places and in doing so, evoke more of an emotional response from the participants. If there were multiple proximate POIs with similar features, e.g. a traditional Italian restaurant and a modern pasta house, we relied on music and sounds that best conveyed the feel of the restaurant.

Auditory icons were non-speech recordings of objects or places, such as stir-frying, chopping food etc., which attempted to communicate the category of the POI with authentic sounds. We followed the same basic process of first identifying key distinguishing features of a POI, and then recording non-speech sounds to convey those features. Equal effort was spent on choosing musicons and auditory icons, which were designed and picked by music and sound engineering professionals by sampling a wide range of audio and with a best-effort commitment to creating a good user experience in a realistic mobile application scenario.

We created a set of about 15-20 musicons and auditory icons to represent a subset of the POIs in the experiment blocks. The set was pilot-tested with several people and filtered to a subset of 10 musicons and auditory icons that were easiest to interpret. These were then used during the experiment for the *Music*, *Sounds* and *Mix* conditions.

Experiment procedure

The experiment was initiated at either the north or the south end of the route. After an initial briefing on experiment procedure, the experiment device was introduced to the participants. The experiment device was an Android smartphone with a custom-built Android experiment app[9] that presented audio and visual cues to the participant depending on the condition they were in. The experiment app simulated the behavior of an LBS without using GPS[10] or mobile Internet connectivity. Since the goal of the experiment was to compare the relative performance of different kinds of cues rather than evaluate the performance of a prototype, this allowed us to eliminate a potential source of confounds that might be caused by technical issues.

The experiment app played each cue for exactly one minute and then switched to the next cue. The smartphone would vibrate between each cue, so that participants in both *Visual* and *Sounds* conditions could safely keep the phone in their pocket and be notified when the next cue began. In order to play each cue for exactly one minute, speech cues and auditory icons were played twice, once at the beginning of the cue and then after 30 seconds. Musicons and visual cues were simply played or displayed for the whole minute. The mixed cue played the auditory icon referring to the POI first, then the speech cue and then the musicon for that POI for the remainder of the minute.

For each cue, participants were asked to identify the POI that the cue referred to. Each block contained a mix of what we considered were easy and difficult cues. Participants were therefore instructed to devote at most the 1-minute of the cue duration to identifying the POI, so as not to miss the next cue. In order to minimize the need to visually and tactually interact with the experiment app, participants used a paper map[11] to mark the POI they identified and reported the speed and confidence of their identification on a 5-point Likert scale (see Figure 4). *A priori* familiarity with the POI was noted as a covariate[12]. Although this does introduce additional artifacts for the participants to handle, we believe that use of paper is more familiar and therefore less distracting for participants than marking the POIs they identified on an unfamiliar smartphone app. Participants were also asked to self-report the speed at which they identified POIs rather than being timed by the experimenter to create a more natural and less stressful environment. The experiment was conducted during weekday mornings and afternoons, but the volume of traffic was such that participants had to be careful not to bump into people. Having the experimenter time the identification for participants would have been quite difficult, without the experimenter closely following the participant and potentially influencing their behavior.

[9]The app requires Android 2.1+ and can be downloaded from **www.crowdee.com/foxpoi/FoxPOI.apk**.

[10]except for logging purposes

[11]with streets and their names marked, but without buildings

[12]The questionnaire we used including maps can be found at **www.crowdee.com/foxpoi/InconditionQuestionnaire.pdf**.

Figure 4. Subject marking POI on map

On the routes, the experimenter walked at a distance away from the participant to avoid disturbing and impacting the experiment, while still being able to observe and detect problems. To minimize experimenter demand effects and erroneous samples, the experimenter was available only at the end of each stretch to answer questions, while the participant filled out a mood questionnaire rating the emotional experience of that stretch. The experimenter would then also start the experiment app for the next stretch.

At the end of the experiment, participants filled out a concluding survey with some open-ended questions about the experiment and their background, as well as preferred stretches and modes of information delivery.

Experience sampling

At the end of each block, the emotional response of participants was gauged by experience sampling, using a modified PAD (Pleasure, Arousal, Dominance) Semantic Differential Scale (PAD scale) [15] [1]. The PAD scale consists of a set of bipolar adjective pairs that are rated along a five-point scale, which corresponds to 3 dimensions of emotional response: pleasure, arousal and autonomy[13]. To make it amenable to frequent experience sampling, we modified and condensed the scale to only include word pairs that were appropriate and easy to interpret for our experiment. This resulted in a modified scale, consisting of six pairs of words in random order.

Analysis

We conducted two kinds of analysis on the experiment results: (1) a between-subjects comparison for each block between the audio treatment and the control and (2) a within-subjects analysis for participants in the audio treatment to compare the outcomes of different conditions for the same users. The performance data and the emotional response data used as dependent variables were factorial and not normally distributed[14], so we used (ordinal) logistic regression for the between-subjects comparisons. Since logistic regression cannot be used on within subjects data, as it violates the assumption about independent samples, we relied on the Aligned Rank Transform method [25] to transform the data and subsequently applied ANOVA to detect main effects of condition on outcomes. Finally, we conducted sanity checks to determine whether there were any treatment or order effects.

[13] Autonomy is important to measure in order to determine how obtrusive versus serendipitous the experience was. We tried to mimic a natural walk. If participants still felt very guided it would be an indication that the experiment was not a natural serendipitous experience.

[14] thus disallowing the use of t-tests

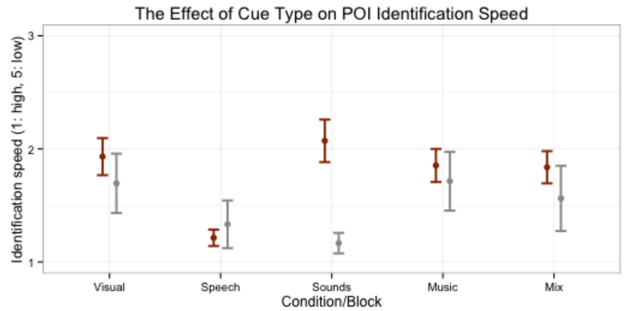


Figure 5. The mean reported identification speed by condition with bars showing the standard error. A speed of 1 denotes instant identification, while 5 corresponds to taking about 1 minute for identification. The maroon dots and bars (left) represent participants in the audio treatment, while the grey dots and lines (right) represent control participants for that same stretch. The Sounds condition refers to auditory icons.

RESULTS

Effectiveness

When studying the percentage of identification errors for the different conditions, musicons and auditory icons were associated with greater errors (7% and 15% resp.) compared to the control participants. However, logistic regression comparing identification errors made with musicons and auditory icons to the errors by control participants did not reveal a main effect of condition on identification errors. Comparing results within subjects using an aligned rank transform followed by an ANOVA, we find a significant main effect of condition on identification (F(4, 203)=3.58, p=0.00759). The contradiction here is likely due to the ANOVA analysis failing to discount some location bias. In other words, overall the controls had fewer errors but in a stretch-by-stretch and condition-by-condition comparison the differences were relatively small, making it hard to make any condition-specific conclusion.

Figure 5 shows the results of reported identification speed by condition. Speech was associated with faster identification and auditory icons with slower identification. For auditory icons, an ordinal logistic regression does indeed result in a significant effect of condition on identification speed compared to controls (χ^2(1,N=60)=10.67, p=0.0011). The effect size is moderate, a difference of 1 point on a 5 point scale. The mix cues also resulted in a small main effect (difference of ~0.3 points) on identification speed compared to controls (χ^2(1,N=53)=3.93, p=0.0473). There is no significant effect for the other factors. Comparing results within subjects using an aligned rank transform followed by an ANOVA, we find that there is again a significant main effect of condition on identification speed, namely (F(4,201)=6.06, p=0.000126), with auditory icons taking much longer to identify and speech cues being much quicker than the others.

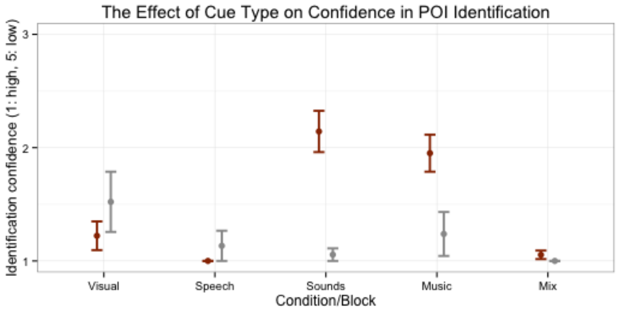


Figure 6. Mean identification confidence by condition. 1 is high confidence, 5 corresponds to low confidence.

The mean identification confidence by condition is shown in Figure 6. As expected, the conditions which identified the POI for the participants, i.e. the mix, visual and speech conditions were consistently associated with very high confidence. The more ambiguous cues, i.e. the musicons and auditory icons were associated with lower confidence. For both auditory icons and musicons, an ordinal logistic regression results in a significant main effect of condition on confidence compared to controls (for auditory icons, (χ^2(1,N=60)=18.50, p<0.0001) and for musicons, (χ^2(1,N=62)=12.32, p=0.0004)). Comparing results within subjects using an aligned rank transform followed by an ANOVA, we find that there is again a significant effect of condition on confidence, namely (F(4,203)=25.97, p<2e-16), where the musicons and auditory icons are associated with much lower confidence than the other conditions.

To sum up, our hypothesis is partially confirmed. Auditory icons are associated with slower and less confident POI identification compared to the visual controls. Musicons are only associated with lower confidence in identification, but not with slower identification. Auditory icons and musicons did not lead to significantly more errors in a stretch-by-stretch comparison to the visual treatment group.

Affect

The mean reported pleasure scores by condition are shown in Figure 7. For musicons, an ordinal logistic regression results in a significant effect of condition on pleasure compared to controls (for the pairing *Annoyed(1) - Pleased(5)*, (χ^2(1,N=15)=6.33, p=0.0118); for the pairing *Bored(1) - Engaged(5)*, (χ^2(1,N=15)=19.09, p<0.0001)). The results for auditory icons are not significant. The difference in pleasure scores for the mix cues is not significant at a 5% level [15] (χ^2(1,N=15)=2.80, p=0.0945). Comparing results within subjects using an aligned rank transform followed by an ANOVA, we find that there is again a significant effect of condition on pleasure, namely (F(4,50)=3.30, p=0.0178) for the pairing *Annoyed(1) - Pleased(5)*, where the music condition is associated with high values and the speech condition with low values. And (F(4,50)=8.65, p<0.0001) for the pairing *Bored(1) - Engaged(5)*, with the music condition associated with higher values.

[15] but at a 10% level

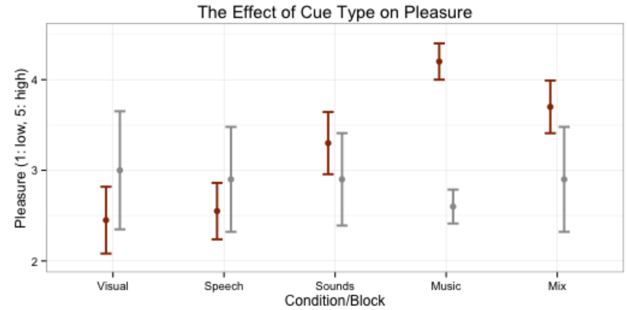


Figure 7. Mean reported pleasure by condition with bars showing the standard error. Pleasure was measured via two pairings: *Annoyed(1) - Pleased(5)* and *Bored(1) -Engaged(5)*. The maroon dots and bars represent participants in the audio treatment, while the grey dots and lines represent control participants for that same stretch.

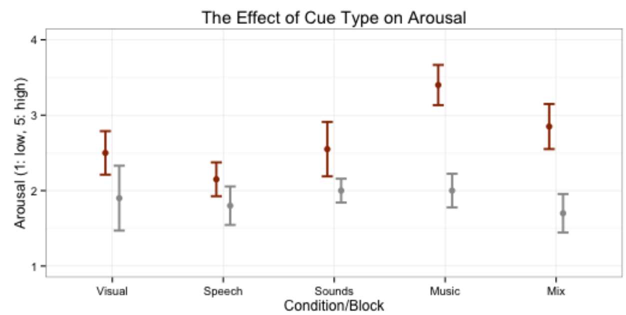


Figure 8. Results of arousal by condition. This graph shows the mean results of the two pairings: *Calm(1) - Excited(5)* and *Relaxed(1) - Stimulated(2)*.

Figure 8 shows the results of mean reported arousal by condition. For musicons, an ordinal logistic regression results in a significant effect of condition on arousal compared to controls for the pairing *Calm(1) - Excited(5)* (χ^2(1,N=15)=9.66, p=0.0019) and a significant effect on a 10% level for the pairing *Relaxed(1) - Stimulated(5)*, (χ^2(1,N=15)=2.99, p=0.083)). The mix cues also resulted in a significant main effect of condition on arousal compared to controls for the pairing *Relaxed(1) - Stimulated(5)*, (χ^2(1,N=15)=8.51, p=0.0035)). The results for auditory icons are again not significant. Comparing results within subjects using an aligned rank transform followed by an ANOVA, we find that there is again a significant effect of condition on arousal for the pairing *Calm(1) - Excited(5)*, namely (F(4,50)=2.94, p=0.0294), where the music condition is associated with high values and the speech condition with low values.

The results of mean reported autonomy, shown in Figure 9, are less clear. The ordinal logistic regression results this time in a significant effect of condition on autonomy compared to controls for the auditory icons, in particular for the pairing *Guided(1) - Autonomous(5)* (χ^2(1,N=15)=4.58, p=0.0324). Comparing results within subjects using an aligned rank transform followed by an ANOVA, we find that there was no significant effect of condition on autonomy. This apparent contradiction can be explained by observing that the visual control in the *Sounds* stretch was lower than the other stretches causing a significant difference in that stretch, which

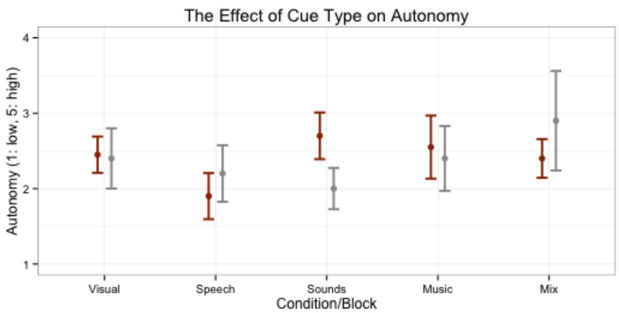


Figure 9. Results of autonomy by condition. This graph shows the mean results of the two pairings: *Influenced(1) - Influential(5)* and *Guided(1) -Autonomous(5)*.

is missed when all stretch/condition differences are lumped together by ANOVA.

Our initial hypothesis is again partially confirmed. Musicons are associated with greater feelings of pleasure and arousal, but no difference in feelings of autonomy compared to the visual control group. Auditory icons are only associated with a greater feeling of autonomy, but no difference in feelings of pleasure or arousal.

Survey

Most participants felt the walking experience was fairly natural and unobtrusive (mean 3.7 on a 5-point scale, sd=0.96). All but one of the participants (14/15) reported discovering new places as a result of participating in the experiment. More than half (6/10) of the participants in the audio treatment (exposed to musicons, mix cues, auditory icons, and visual cues) mentioned the stretch with musicons as the most enjoyable stretch. Control participants (exposed to visual cues) reported enjoying stretch 4 (2/5) and 5 (2/5) most, indicating that location did not account for the greater enjoyment of the music stretch (which was stretch 3).

Participants were also asked whether they had more 'positive associations' with any subset of the blocks after the experiment. This question was kept deliberately vague to get some indication of whether certain conditions had an effect on the way people perceived a place. 60% (6/10) of the participants in the audio treatment reported having more positive associations with the music block after doing the experiment. The same numbers were 30% (3/10) for the block with auditory icons, 20% (2/10) for the speech block and 10% (1/10) each for the other stretches. Note that participants could pick more than one stretch.

When asked how they would prefer to receive information from an LBS, participants were split between the combination of auditory icons and musicons (40%), and the combination of speech and musicons (33%). Overall, 80% (12/15) chose to include music as a preferred way to receive information, 53% (8/15) chose auditory icons and visual information, and 47% (7/15) chose speech.

Post-experiment observations

Block 5 of the experiment showed the same visual cues to both control participants and participants in the audio treatment. We examined this data and found that there was no statistically significant difference between the responses. We did notice, though, that the participants with the audio treatment had greater spread of emotional responses than the controls. This could be due to the larger number of participants in the audio treatment or might indicate a slight carryover effect, i.e. simply having the audio treatment led to heightened emotional responses, whereas control participants seemed to have more stable and consistent responses. Since we did not randomize the order in which the participants were exposed to the conditions, another potential concern is whether the data exhibits any order effects, i.e. participants doing the stretches in opposite orders have different responses. There was no statistically significant effect of order on the responses overall.

A secondary concern was that the participants were from diverse cultures and might lack the cultural context required to interpret musicons. However, we found minimal effects of cultural bias (1 auditory icon and 1 musicon were consistently misinterpreted), lending credence to the aphorism of music being a "universal medium."

DISCUSSION

Visual and speech cues are very effective in that users were able to quickly, correctly and confidently identify the relevant POIs. Speech cues were associated with the fastest identification, faster even than visual cues, which included a picture of the POI. This is probably because people could directly look around while receiving the speech cue. People reported that both hearing the name and address of a place and seeing the photo of the POI helped them identify POIs quickly. Visual cues were also quickly identified, but there was a larger spread. Some users did report difficulty in interpreting the picture, feeling that the photo was too small or that the angle of the picture was unexpected.

Ambiguous cues, namely the auditory icons and musicons, did seem to lead to some identification errors, but the difference was not statistically significant. Overall, the performance of auditory icons did suffer a difference of about 1 point (on the 5-point scale) in the speed and confidence of cue identification. Musicons, which are also ambiguous, also resulted in lower confidence scores, but did not take detectably longer to identify. When asked about the difficult portions of the experiment, nearly all participants mentioned the auditory icons as being difficult to identify, but only a third (3/10) of the participants mentioned musicons as being difficult to identify. Interestingly, users also took slightly longer to interpret the mix cues. Since the mix cues incorporated three different kinds of cues (speech, auditory icon and musicon), it might have been too much information for users to interpret quickly, suggesting that more information is not always better. To recap, auditory icons do seem to be associated with worse performance. Users seem to be able to identify POIs relatively quickly with musicons, but are less certain of correct identification.

Musicons were also associated with greater feelings of pleasure and arousal, unlike the auditory icons. Several users mentioned in the concluding survey that the auditory icons were "obscure and vague", "confusing and unpleasant" and one of the least enjoyable things about the walk. Unlike musicons, which can convey information via a wealth of signals, such as lyrics, melody, mood, and choice of instruments, realistic sounds can be too difficult to interpret without supporting context. Some users reported informally that the auditory icons were very enjoyable and immediate when they were easy to interpret. Thus, auditory icons might be more enjoyable when complemented with another unambiguous cue.

Intriguingly, the auditory icons were associated with greater feeling of autonomy. Even though musicons were also ambiguous, they did not lead to greater feelings of autonomy. Given that auditory icons were also most difficult to interpret, one speculation is that the challenge of interpreting auditory icons itself gives participants a greater sense of autonomy and less of a hand-holding feeling than the more explicit cues do. Thus, when LBS present information explicitly, in visual or speech form, they might be unintentionally decreasing the user's sense of autonomy. In contrast, in a setting where users are exploring or discovering new areas, using ambiguous sound cues might actually cause users to feel more in control of their own discovery process.

Although not statistically significant, we observe that visual cues lead to slightly better emotional responses than speech cues, which might be due to greater acceptance of and familiarity with the visual mode of information presentation. We introduced mix cues to examine whether combining different kinds of cues can effectively mitigate some of the drawbacks of any single pure cue. Indeed, mix cues perform almost as well as speech and visual cues in terms of effectiveness as a cue. They also lead to greater pleasure and arousal compared to speech cues. Mix cues did take longer to interpret since we were combining 3 different kinds of cues. Different designs that, e.g., combine one implicit and one explicit cue, or a visual cue and an aural cue, might lead to even better results. More research is needed to determine how to best combine different kinds of cues for the best experience.

IMPLICATIONS AND LIMITATIONS

Despite the limited sample size, we obtained statistically significantly results both in between- and within-subject tests. The difference between auditory icons and musicons is particularly striking. The inclusion of a small control group to remove location bias was crucial in obtaining these results. It was beyond the scope of our experiment to examine the effects of cue types, particularly musicons, on recall and longer-term affect. Similarly, we were limited to a fixed route by our controlled experiment setup. We expect multiple or less constrained experiment routes to only accentuate the differences in emotional experience we observed.

A valid concern with field studies like ours is how specific the results are to the individual cues chosen. Our approach was to provide a best-effort design of both auditory icons and music, consulting both sound engineers and DJs. We therefore consider the difference in task performance and affect of auditory icons and musicons to be stable. Designing effective auditory icons and musicons remains an art form. Even if it were possible to design better auditory icons and worse musicons that reverse the results, our approach demonstrates that it is still *easier* to produce efficient and pleasurable musicons than to produce good auditory icons.

The results of our study could help in LBS design by providing guidance on choosing the right cue for the experience that the LBS wants to evoke. For example, auditory icons give a greater sense of autonomy, musicons greater pleasure and speech cues lead to greater efficiency. For current LBS, we expect musicons to complement rather than substitute the familiar visual cues. The surprising success of musicons for POI discovery suggests that music might also be fruitfully used in non-LBS applications.

Our field study has shown that POI discovery can be done efficiently with more subtle and yet more pleasurable cues than the de-facto visual cues. To our surprise, ambiguous cues only seemed to increase the engagement and sense of fun, in particular for entertaining music tracks. It may be that they create a game-like[16] sense of challenge, which evokes fun and relaxing emotions. Subtle cues may also be easily disregarded by users who are not currently in a discovery mode, without causing annoyance. Explicit cues, by contrast, can be very intrusive, like pop-up ads on the Web, if you are not explicitly looking for information.

There are certain drawbacks to using audio cues, e.g. using it while roaming around in a group, and potentially having it mask other sounds in the environment. It may therefore not be appropriate in all applications. More research is needed to determine the right level of obtrusiveness and distraction that is appropriate for any particular use case. If music or audio can be chosen that users enjoy regardless of whether they want to discover a POI, then this distraction can be elegantly minimized. More work is also needed to determine what role music can play in augmented reality settings, where both video and audio, synthetic and real, may coexist in the same interface.

CONCLUSION

Our initial hypotheses were as follows: *Musicons and auditory icons are associated with greater errors in identification, slower identification speed and lower confidence than visual cues* (Hypothesis A) and *Musicons and auditory icons are associated with greater feelings of pleasure, arousal and autonomy than visual cues* (Hypothesis B). Hypothesis A was confirmed, and the results were as expected with one exception: the identification speed of musicons was significantly better than that of auditory icons. In fact, in our experiment, designed to mimic a realistic stroll on a busy shopping street, the identification speed of musicons was even comparable to visual cues. Hypothesis B was only partially confirmed. Here the main discrepancies were: auditory icons did not lead to greater feelings of pleasure or arousal, and musicons did not lead to greater autonomy. In summary, our experiment was

[16] without being in a real game setting, which may incur stress

able to highlight some surprising differences between auditory icons and musicons that may help the design of future location-based music applications (e.g. [3]). For instance, for game-like scenarios, autonomy is important and thus auditory icons may be appropriate, but for serendipitous discovery, pleasure and identification accuracy are more important and hence musicons may be a better choice.

As future work, we plan to incorporate lessons learned in this experiment in a real system [3] to produce the right cues at the right time for the right user based on various scoring and scheduling (i.e. the ordering and bundling of tracks) rules. We are also working with musicians and DJs to bring some of their music selection and scheduling knowledge to the location-based services community.

Acknowledgments

We thank Priscilla Huber for her work on the auditory icons and the experiment subjects for their participation.

REFERENCES

1. Agarwal, A., and Meyer, A. Beyond usability: Evaluating emotional response as an integral part of the user experience. In *CHI '09*, ACM (2009), 2919–2930.
2. Alty, J. L., and Rigas, D. I. Communicating graphical information to blind users using music: the role of context. In *CHI '98*, ACM Press/Addison-Wesley Publishing Co. (New York, NY, USA, 1998), 574–581.
3. Ankolekar, A., and Sandholm, T. Foxtrot: a soundtrack for where you are. In *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*, IwS '11, ACM (New York, NY, USA, 2011), 26–31.
4. Bull, M. *Sound moves : iPod culture and urban experience / Michael Bull*. Routledge, Abingdon, Oxon ; New York :, 2008.
5. Garzonis, S., Jones, S., Jay, T., and O'Neill, E. Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. In *CHI '09*, ACM (New York, NY, USA, 2009), 1513–1522.
6. Goodman, J., Brewster, S., and Gray, P. Using field experiments to evaluate mobile guides. In *Proc. of HCI in Mobile Guides, workshop at Mobile HCI 2004* (2004).
7. Goodman, J., Brewster, S. A., and Gray, P. How can we best use landmarks to support older people in navigation? *Journal of Behaviour and Information Technology 24* (2005), 3–20.
8. Holland, S., Morse, D. R., and Gedenryd, H. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal Ubiquitous Comput. 6*, 4 (Jan. 2002), 253–259.
9. Johnson, F. D. The effects of music on temporary disposition (unpublished manuscript), 2001.
10. Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D., and Holmes, G. Ontrack: Dynamically adapting music playback to support navigation. *Personal Ubiquitous Comput. 12* (October 2008), 513–525.
11. Kurczak, J., Graham, T. C. N., Joly, C., and Mandryk, R. L. Hearing is believing: evaluating ambient audio for location-based games. In *ACE '11*, ACM (New York, NY, USA, 2011), 32:1–32:8.
12. McGee-Lennon, M., Wolters, M., McLachlan, R., Brewster, S., and Hall, C. Name that tune: musicons as reminders in the home. In *CHI '11*, ACM (New York, NY, USA, 2011), 2803–2806.
13. McGookin, D., and Brewster, S. Pulse: an auditory display to provide a social vibe. In *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*, IwS '11, ACM (New York, NY, USA, 2011), 12–15.
14. Mcgookin, D., Brewster, S., and Priego, P. Audio bubbles: Employing non-speech audio to support tourist wayfinding. In *HAID '09*, Springer-Verlag (Berlin, Heidelberg, 2009), 41–50.
15. Mehrabian, A., and Russell, J. A. *An approach to environmental psychology*. M.I.T. Press, Cambridge, MA, USA, 1974.
16. Raper, J., Gartner, G., Karimi, H., and Rizos, C. A critical evaluation of location based services and their potential. *Journal of Location Based Services 1*, 1 (2007), 5–45.
17. Robinson, S., Jones, M., Williamson, J., Murray-Smith, R., Eslambolchilar, P., and Lindborg, M. Navigation your way: from spontaneous independent exploration to dynamic social journeys. *Personal and Ubiquitous Computing* (Sept. 2011), 1–13.
18. Roe, K. Swedish Youth and Music: Listening Patterns and Motivations. *Communication Research 12* (July 1985), 353–362.
19. Rueb, T. Trace. Yoho National Park, Canada, 1999.
20. Scherer, K. R., and Zentner, M. K. *Emotional Effects of Music: production rules*. Oxford University Press, 2001, ch. Music and Emotion: Theory and Research, 361–392.
21. Sikora, C. A., Roberts, L., and Murray, L. T. Musical vs. real world feedback signals. In *CHI '95*, ACM (New York, NY, USA, 1995), 220–221.
22. Stahl, C. The roaring navigator: a group guide for the zoo with shared auditory landmark display. In *MobileHCI '07*, ACM (New York, NY, USA, 2007), 383–386.
23. Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., and O'Modhrain, S. Gpstunes: controlling navigation via audio feedback. In *MobileHCI '05*, ACM (New York, NY, USA, 2005), 275–278.
24. Warren, N., Jones, M., Jones, S., and Bainbridge, D. Navigation via continuously adapted music. In *CHI EA '05*, ACM (New York, NY, USA, 2005), 1849–1852.
25. Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *CHI '11*, ACM (New York, NY, USA, 2011), 143–146.