Foxtrot: A Soundtrack for Where You Are

Anupriya Ankolekar
HP Labs
Palo Alto, CA, USA
anupriya.ankolekar@hp.com

Thomas Sandholm
HP Labs
Palo Alto, CA, USA
thomas.e.sandholm@hp.com

ABSTRACT

In this paper, we present a mobile location-aware and crowd-sourced audio application, Foxtrot, that allows people to share the sounds and music they enjoy and associate with a particular location. Foxtrot plays an automatically created radio-like stream of geo-tagged music, ambient sounds and comments left by friends and other people. We discuss some of the design considerations for Foxtrot and our approach to selecting and scheduling audio content for playback. In addition, we present a pilot study of Foxtrot, which indicates that a location-aware music system might indeed provide an engaging mobile experience for users.

Categories and Subject Descriptors

H5.5 [**Information interfaces and presentation**]: Sound and Music Computing

General Terms

Design, Human Factors

1. INTRODUCTION

Location-based applications, 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, things and activities around their current location. While this kind of information is extremely useful, most people also foster emotional connections to places they live and work in, and visit. People grow to love the places where they have spent time, and even dreary, impersonal places come alive when imbued with the sounds of people's lives, their music and stories.

Given the powerful effect of music and sounds on our emotions and feelings about places, it is perhaps surprising that the overwhelming majority of location-based applications are visual and informational, and do not capture this emotional aspect of location. A notable exception is RjDj[2], an iPhone app that plays an engaging, dynamically-generated soundtrack that reflects sounds ambient to the user.

In this paper, we propose a mobile audio application, Foxtrot, that allows people to share the sounds and music they associate with and enjoy in a particular location. To provide a compelling, yet unobtrusive experience, Foxtrot mixes the crowd-sourced geo-tagged audio into a radio-like channel that people can tune into to listen to audio for a given location. While the primary thrust of Foxtrot is towards stirring emotions, the sounds in Foxtrot could potentially also carry information content. E.g. people might be alerted to a popular café around the corner by the sounds of conversation and coffee cups before they even see the café. In addition to music and sounds, people can also record and post new snippets, their own personal stories or sound-bites.

Anderson et al. [1] have described how the proper use of music and ambient sounds can evoke 'an atmosphere or sense of place, thereby heightening the overall feeling of immersion in the virtual environment.' While Foxtrot is designed to operate in the real environment, many of the characteristics we want to highlight are virtual, e.g. the popularity of an establishment in a location, or stories recorded by people in a location. These may not always be present in the real context, but contribute to the mood of the place. The media and entertainment industry has long made use of particular, studio-recorded sounds to enhance the reality depicted in movies and radio shows. We would like to achieve the same effect in a more democratic, crowd-sourced fashion.

A system like Foxtrot is ideal for tourists seeking a rich experience of a new place. Tourists have time to spare and tend to be more receptive to entertaining and emotional content. By listening to music, ambient sounds and recorded sound bites of people, both historical and contemporary, the tourist can experience the place as many others before, locals and other tourists, have experienced it. They can discover the interesting stories around a place, get an emotional feel for the place through the music associated with it, and find good restaurants and other commercial enterprises by ear. A local user might of course use Foxtrot very differently. She is likely to care less about historical stories and more about listening to music or curious about new events. Foxtrot in this

[1]http://blog.compete.com/2009/06/02/location-based-services-applications-carriers-advertisers/

[2]http://rjdj.me

IWS '11, August 30, 2011, Stockholm, Sweden

scenario can function as a filter, amplifying ambient sounds to provide an extra sense to guide her quickly to new information about her surroundings.

Automobiles provide an additional interesting use case for Foxtrot. Since a car cuts off much of the natural ambient sounds, when listened to in an automobile, Foxtrot recreates the key parts of the experience of being in a place, playing only the most popular music and sounds, providing hints as to what the place is like that the automobile is driving through. Furthermore, the hands-free mode of Foxtrot does not assume any touch or visual contact with the device and is well suited for the requirements of an in-car experience, e.g. safety concerns of operating a device while driving.

In the next section, we examine related work in mobile music, crowd-sourced audio and augmented reality applications. We then discuss the design of Foxtrot, in particular how audio snippets are selected and scheduled for playback for a given user, based on his/her location and other contextual information, as well as the overall buzz about particular songs and sounds as inferred from explicit ratings and implicit listening user feedback. We then briefly present the implementation of Foxtrot. Finally, we report on a pilot study we performed using Amazon's Mechanical Turk to get user feedback on Foxtrot. Our preliminary results indicate that people do tend to enjoy both sounds and music in this context, but sounds are more closely associated with location. We found that music in particular seemed to engage users more and led to more emotionally rich reactions whereas ambient sounds seemed to provide a better sense of virtually walking down a street. We conclude with some discussion of future work.

2. RELATED WORK

Foxtrot, like other mobile augmented reality applications, such as Layar[3] and Wikitude AR[4], seeks to augment the mobile reality of users with information about things around the user. The key difference is that Foxtrot layers audio content, which is potentially more emotionally engaging, over the user's reality rather than textual, visual content. Nomadic Radio [6] provides an auditory and speech interface to a user's email and other communication services. However, the focus of the work is on letting the user communicate while they are mobile, rather than on providing a location-aware crowd-sourced audio channel.

There are several mobile music applications, such as mobile Internet radio stations (such as Pandora[5] and last.fm[6]) or applications which allow you to play your own music either on the device or in the cloud (e.g. Amazon Cloud Drive[7] and Google Music[8]). Turntable.fm[9] is a crowd-sourced music application, which lets users play DJ or listen in to music being played in various rooms. Unlike these applications, Foxtrot only plays crowd-sourced audio that is associated with the user's location. In addition, Foxtrot also plays ambient sounds and comments left by people in the location, whereas these applications only focus on music.

There are several smaller applications that are closely related in spirit to Foxtrot, in that they encourage recording and playing ambient sounds. The UK SoundMap project of the British Library[10] attempts to archive the soundscape of the UK. Sonic Wonders[11] does a novel spin on the typical travel website, by encouraging people to travel, hear and record unusual sounds in places.

The user comments feature of Foxtrot is similar to that provided by Broadcastr[12], an application that crowd-sources user *stories* associated with geographical coordinates and lets users search for interesting stories and people to follow. Foxtrot, in contrast, develops a dynamically-generated radio channel of music and sounds, in addition to stories.

3. DESIGN

A primary goal of Foxtrot is to enhance the sense of being in a place by creating an emotional aura or atmosphere about the place. Lombard et al. [3] have discussed how sound can enhance the sense of presence in a virtual environment, essentially emphasizing the quality and spatialization of sound. Interestingly, higher sound fidelity does not necessarily lead to a greater sense of being situated in the environment; lower quality sounds can make users feel like 'a part of the action'. Since the user's visual sense is already saturated and since sound can have a powerful effect on our emotions, we use the auditory sense to create the emotional atmosphere around a place. This requires a continuous, emotionally engaging experience where the user is listening to Foxtrot in the background, only dipping in to interact with the user interface when absolutely necessary. Foxtrot, therefore, plays audio like a radio channel, dynamically selecting content to play to the user based on the user's location and preferences, and the popularity of the content.

A secondary goal of Foxtrot is to increase the user's knowledge of the place they are in. Therefore, the (minimal) user interface should allow users to get information about the sounds they heard: their location, artist information, tags, popularity and other relevant metadata.

3.1 Associating audio with locations

A fundamental consideration in designing Foxtrot is how to map audio to locations. We could either require users to specify precise geographic coordinates for each item (as we do currently) or let them specify an arbitrary area. The advantage of using precise geographic coordinates is that it is simple for the user to specify and understand, and for the system to compute the closest tracks. However, audio content may only sometimes have precise and unambiguous locations and consumers may find it difficult to interpret why a particular audio fragment was played at a given location.

We attempt to mitigate this by letting users also specify a visibility range for the track they contribute. This defines a

[3]http://www.layar.com/
[4]http://www.wikitude.com/en/
[5]http://www.pandora.com/
[6]http://www.last.fm/
[7]https://www.amazon.com/clouddrive
[8]http://music.google.com
[9]http://www.turntable.fm
[10]http://sounds.bl.uk/uksoundmap/index.aspx
[11]http://www.sonicwonders.org/
[12]http://beta.broadcastr.com/

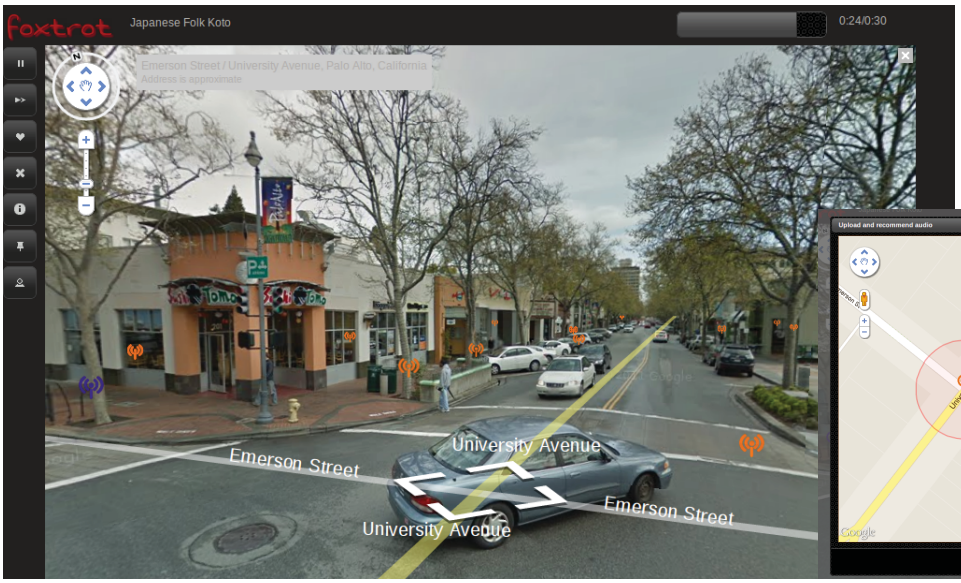

Figure 1: Screenshot of the desktop version of Foxtrot.

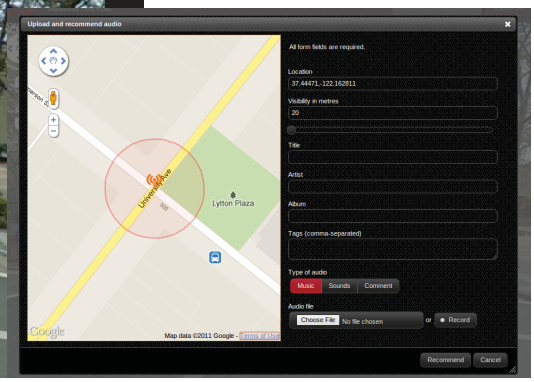

Figure 2: Recommending audio for a location.

circular region centered at the location of the track, representing the area within which the track is relevant. Sounds or stories that are associated with a narrow location can only be interpreted when the user is directly in front of that location, but tracks with a larger visibility are less closely tied to location and are interpretable within a broader area.

3.2 Creating an audio stream

Foxtrot revolves around two primary kinds of objects: users and tracks. Users have a current geographic location (as determined automatically by a web browser or mobile device), a speed, a listening profile (e.g. which tracks they liked), a history (e.g. which tracks they have heard and how long ago) and a list of friends, as determined from Facebook or some other external source. Each audio track is centered at a geographic location, has a visibility range (which defines how far from the center the track is still relevant), a type (music, sound or comment) and user-specified tags.

Our approach to automatically creating a playlist of audio tracks consists of three steps. In the first step, using a location-based content recommendation system, Gloe [5], we retrieve tracks that are located in close geographic proximity to the user's location, sorted in order of increasing distance. If the user is moving, we retrieve items that are closer to the user's trajectory than solely to his or her original position. Next, we score each track with respect to how likely it is that the user will want to listen to this track next at their current location. This step results in a list of tracks, sorted in descending order by score. In the next step, we rearrange the tracks to achieve a pleasant-sounding mix for a good user experience. The last two steps are explained in greater detail in the next sections.

3.2.1 Selecting and scoring tracks

Each track in Foxtrot is assigned a score based on its relevance, i.e. using the likelihood that the user will want to listen to this track next in this location. The score is a weighted combination of three components: (a) a likability score, (b) a geographic relevance score, and (c) a freshness score. The likability score takes into account factors such as whether the user has explicitly 'like'd the track, how many times they have heard the track when it was played to them, how many of their friends have liked the track and heard it and the number of times the track has been liked and heard globally in Foxtrot.

The geographic relevance of the track is again a weighted combination of the distance of the track from the user's current location and the visibility range of the track. The greater the distance, the lower the score; the greater the visibility, the lower its contribution to the score. In the latter case, we want to encourage people to be more specific in the visibility they assign to tracks. Also, the narrower the track's visibility, the more likely it is that the user will miss listening to the track within its visibility range, but a track with broader visibility can be listened to in other locations as well. We could take into account other factors too, such as the user's expected trajectory, based on the recent history of their movements.

Since the user in Foxtrot may be travelling on foot or by automobile, we do not take the raw distance into account when computing the geographic relevance score of a track. Rather we use the user's speed to calculate the time we expect the user to take to get to the location of the track. This means that if the user is travelling at great speed, e.g. in a car, we automatically cast a wider net and choose popular tracks that are potentially further away. Thus, Foxtrot plays different content when the user is walking versus when the user is travelling in a car. This is all done automatically, but we could potentially identify and tag certain audio to be better played in an automobile environment rather than when walking. As mentioned before, ambient sounds are more interesting when the user is cut off from the sounds of the street than otherwise.

Finally, the freshness score takes into account factors such as how many times the track was heard by the user and the last time the track was heard. The more times and the more recently the track has been heard, the lower the freshness score of the track. This makes sense, since we want to generally play fresh and interesting music. The decay function however is not linear, e.g. if a track has been listened to only a couple of times, we treat it almost as a fresh track, but the score then decays very quickly until it plateaus to a very low score after the track has been heard some fifteen times. Using a freshness score allows us to automatically differentiate the playlist heard by a tourist visiting a place for the first time from a playlist heard by a local, who walks down the street several times a day.

3.2.2 *Scheduling track playback*

Even with highly relevant audio content that is closely associated with the user's current location, jarring transitions between heterogeneous audio items can detract considerably from a smooth user experience. Prior work has investigated dynamically generating playlists for a setting where there is either a single user listening to music [2] and giving feedback or multiple users submitting requests [4]. However, there is no work we know of that aims to minimize jarring transitions in automatically generated playlists.

Our approach is to follow the strategy of a radio station in real life. Just as a radio station airs music, news and advertisements in clustered segments, we partially smoothen the transitions by clustering audio into music, sounds and stories and playing them in segments, what we call radio *programs*. The algorithm we use to rearrange audio tracks is fairly straightforward. We assume that we have a template for radio programs, i.e. the type of program (music, sounds, or comments) and the duration of each program. To create a playlist of radio programs for the user, we simply work down the list of scored tracks, filling each radio program with as many top-scored items of the same type as we can, before moving on to the next program.

4. IMPLEMENTATION

The current desktop version of Foxtrot[13] is shown in Fig. 1. Foxtrot plays a mix of music and ambient sounds for a street in downtown Palo Alto. The desktop version uses Google Maps Streetview[14] to create a simulated mobile experience. Users can rate music up or down for the location they are in. They can pause and resume Foxtrot, as well as skip tracks. In addition, they can recommend audio for their location, as shown in Fig. 2.

Foxtrot runs on most modern mobile web browsers, including Android 2.3+, iOS 4.1+ and webOS 2+ (see Fig. 3), using the HTML5[15] geolocation API and its support for native browser audio (i.e. the audio tag), and JQuery Mobile[16]. On the server side, Foxtrot has been implemented as a Django[17] web service that uses an Apache web server to serve the audio and other media files. Our audio content consists mostly of snippets of music under 30 seconds long, and ambient sounds that have been recorded by a professional sound artist on a street in downtown Palo Alto. Foxtrot uses a general purpose, location-aware rating system, called Gloe [5], as a platform for associating audio with locations and for recording user ratings of audio. Users can log in using Facebook Authentication[18], i.e. using their Facebook credentials, which both authenticates them and identifies their friends for Foxtrot.

[13]http://www.crowdee.com/foxtrot/
[14]http://maps.google.com/
[15]http://dev.w3.org/html5/spec/Overview.html
[16]http://jquerymobile.com/
[17]https://www.djangoproject.com/

5. EXPERIMENT

To understand how people experience a mix of music and ambient sounds associated with location, we conducted a small pilot study, where we solicited 100 Amazon Mechanical Turk[19] users to experience a virtual audio tour online[20]. Within the tour, Foxtrot took users on a 3-minute-long walk through downtown Palo Alto, CA, USA. A screenshot of the website seen by the study participants is shown in Fig. 4. Each user was randomly assigned to a condition to hear either only music, only sounds or a mix of both. In this study, we only performed the first of the three steps discussed in Section 3.2. In other words, the tracks were only ranked by increasing distance and we did not rearrange the track playback in any way.

We evaluated user engagement for Foxtrot by monitoring users' click traces, tracking clicks such as 'like'ing, 'unlike'ing, pausing, stopping, and forwarding the audio. After about one minute through the experiment, Foxtrot switched the Google Maps view to Streetview to let users experience Foxtrot in both an offline mode as well as a simulated mobile mode. After completing the tour, users were asked to answer questions on what the Foxtrot user experience was like.

5.1 User Engagement

To assess user engagement, we examined the overall number of user actions (such as pausing, stopping and skipping audio), the number of times audio content was liked, and the number of times audio content was 'unlike'd[21]. Actions that happened by default and not as a result of user activity were deducted from these numbers. Overall, we recorded 2935 actions from 152[22] users (i.e. an averagely active user perfromed an action roughly every 10 seconds). Table 1 lists the average number of actions, 'likes' and 'unlikes' performed per user for each of the three conditions. The numbers are fairly similar across conditions. The only statistically significant result ($p < 0.05$ in pairwise t-tests) is the low number of actions in the sounds-only group.

From the table, it appears that in the music-only condition, users were more engaged, resulting in more actions and more 'like' clicks. Sounds-only, on the other hand, resulted in very

[18]http://developers.facebook.com/docs/authentication/
[19]https://www.mturk.com/mturk/welcome
[20]http://www.crowdee.com/media/mturk
[21]The most popular music and sounds heard during the study as rated by the participants can be viewed at: http://www.crowdee.com/json/login/#lat=37.441883|lon=-122.143019|cha=foxtrot_mturk|rad=50
[22]This higher number includes several additional Mechanical Turk users, who completed the tour even though they did not answer the survey questions.

Figure 3: Screenshot of the mobile version of Foxtrot.

Figure 4: Screenshot of the Foxtrot pilot study.

Table 1: Average number of actions, 'likes' and 'unlikes' performed by users in each of the three experimental conditions.

Condition	Music & Sounds	Music	Sounds
# actions	2.67	5.49	**.65**
# 'likes'	3.95	4.76	4.21
# 'unlikes'	3.27	3.11	2.6

few actions, perhaps because most sounds were short. In this condition, users also had fewer 'unlike' clicks, but that again may be due to lack of engagement rather than because listening to sounds is less annoying.

5.2 User Experience

After the audio tour, the study participants were asked to rate their experience via the survey shown in Table 2. Our survey questions attempted to measure how users felt about the Foxtrot experience as a whole, whether there were any technical issues that hampered the auditory experience, the walking experience, how comprehensible the association of audio to location was and finally, the emotional experience of the audio tracks Foxtrot played. After removing users who did not vote on all items or submit valid results, we were left with 84 users (out of 100). Their responses are summarized in Table 3.

It appears that listening to only music or only sounds resulted in a better user experience than listening to a mix of the two. Users also found it easier to associate locations with only sounds rather than only music, which lowered the quality of the walking experience. Listening to both music and sounds resulted in a surprisingly good walking experience; an encouraging finding that is difficult to explain given the other responses. All of these observations are significant at an approximate 10% level ($p \approx 0.1$).

Listening to only music seems to have led to a better emotional experience of the audio; however this is not statistically significant. Otherwise, there was almost no difference between the conditions. In terms of overall experience, the conditions were also very similar. Within all three conditions, users were on the whole slightly more happy with the system than unhappy with it. (2.5 is the inflection point.)

Table 3: Average rating of the experience by users for each of the three experimental conditions. The higher the value, the better the experience. The mininum value is 1 and the maximum value is 4.

Condition	Music & Sounds	Music	Sounds
User experience	**1.8**	3	2.8
Location association	2.6	2	**3**
Walking experience	**3.2**	2.2	2.8
Emotional experience	2.6	2.8	2.4
Sound experience	3	3.2	3.2
Average experience	2.6	2.6	2.8

6. DISCUSSION

The results of the pilot study seem to indicate that a combined music and sounds condition leads to a significantly poor user experience. However, since this condition also seems to yield a significantly better walking experience, this suggests that the poor user experience is not due to the fact that both music and sounds are being played, but that there is potentially some other confounding factor, either in the way the playlist is created or in the transitions between music and sounds that has led to this result. We plan to tease out this factor in a follow-up field study. It might be that the same audio mix that leads to a good music experience does not lead to a good walking experience and this is something that needs to be traded off in Foxtrot.

Although a location and its ambient sounds are closely associated in memory for most people, it is not clear how individual sounds and music tracks recreate and reinforce a sense of the place, especially when played in a different context, e.g.

Table 2: Questions answered by study participants during a required post-tour survey.

Question	1	2	3	4
Overall user experience	poor	fair	good	great
Overall sound quality experience	poor	fair	good	great
I felt like I was walking down the street	never	sometimes	most of the time	all the time
I understood why the sound was placed in this location	never	sometimes	most of the time	all the time
The sounds I heard were	unpleasant	boring	realistic	fun

virtually or in an automobile. Given that different aspects of their environment are salient to different people, who have different sound preferences, it is also not clear to what extent the audio-location associations are generalizable. We expect that crowd-sourcing audio content and letting people rate it will automatically bring the most general associations into prominence, something we will also investigate in the field.

Finally, the same experience that works in a mobile setting, where the audio is merely enhancing the user's current, rich view of the location, is unlikely to work in an automobile, where the user cannot always see his/her surroundings in detail. While we currently do not adapt to the context of the user, this is also something we plan to do in future work.

In the next stage of Foxtrot, we plan to refine our scoring and scheduling algorithms, hand-tuning the weights of individual components as well as updating them using usage data, as far as possible. We plan to also evaluate the effect of our scheduling algorithm against a baseline of unchanged tracks within a larger Mechanical Turk study. Finally, we plan to conduct a field study, where we aim to get detailed qualitative feedback on the perceived quality of the audio recommendations, the user experience and whether using Foxtrot leads to a better emotional connection and knowledge of the place. We plan to evaluate how users react to Foxtrot differently in familiar places versus new places.

As mentioned in Section 3, spatialization is a key component of making the auditory information comprehensible and realistic. In future work, we plan to work on techniques to spatialize the audio tracks, in particular the ambient sounds, to guide users' attention in the correct direction of the sound. Many modern mobile devices contain compasses, which allow to determine where the user is facing and thus correctly spatialize the audio appropiately. Another area of future work is the idea of using tags as radio channels. Foxtrot allows users to tag the audio they upload, in order to make it easier for other users to search for audio. The audio tracks retrieved can then be automatically converted into a radio playlist. Thus, a search for 'country music' would find all the nearby tracks that are tagged 'country music', creating a 'country music' radio channel for users. Using Gloe, Foxtrot can also expose the popular tags in an area to guide users in selecting channels with interesting audio.

7. CONCLUSION

In this paper, we proposed a mobile crowd-sourced location-aware application, Foxtrot, that augments the user's context with an engaging audio stream. As the user moves around in a location, she hears an automatically curated stream of crowd-sourced geo-tagged music, ambient sounds and comments left by friends and other people. Foxtrot employs a radio metaphor to organize the crowd-sourced content presented to the user as a series of radio programs on music, sounds and comments. However, Foxtrot could potentially utilize any meaningful clustering of tracks as radio programs.

We discussed some of the design considerations for Foxtrot and our approach to selecting and scheduling audio content for playback. We conducted a pilot study of Foxtrot, where we found that music in particular seemed to engage users more and led to more emotionally rich reactions whereas ambient sounds seemed to provide a better sense of virtually walking down a street. Overall, the study indicated that a location-aware music system might indeed provide an engaging mobile experience for users. We intend to develop Foxtrot further with better selection and scheduling algorithms, and with more work on the user experience aspects, and conduct field studies.

Acknowledgments

We thank Priscilla Huber for recording and geo-tagging the audio content in Foxtrot, and Zahan Malkani, who developed an early prototype that was a key influence on the current Foxtrot design. We also thank Bernardo Huberman and Alex Vorbau for many fruitful discussions, and the anonymous reviewers for their useful references and suggestions.

8. REFERENCES

[1] D. B. Anderson and M. A. Casey. The sound dimension. *IEEE Spectrum*, 34(3):46–51, 1997.

[2] K. Bosteels, E. Pampalk, and E. E. Kerre. Evaluating and analysing dynamic playlist generation heuristics using radio logs and fuzzy set theory. In *Proceedings of ISMIR*, 2009.

[3] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), September 1997.

[4] F. Pestoni, J. Wolf, M. Habib, and A. Mueller. Karc: radio research. In *Web Delivering of Music, 2001. Proceedings. First International Conference on*, pages 139 – 146, nov. 2001.

[5] T. Sandholm, H. Ung, C. Aperjis, and B. A. Huberman. Global budgets for local recommendations. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 13–20, New York, NY, USA, 2010. ACM.

[6] N. Sawhney and C. Schmandt. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, September 2000.