Juho Kostiainen

Mobile Auditory Guidance for Public Transportation

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 22.8.2011

Thesis supervisor:

Prof. Vesa Välimäki

Thesis instructor:

Dr. Cumhur Erkut, D.Sc. (Tech.)

Aalto University
School of Electrical Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE MASTER'S THESIS

Author: Juho Kostiainen

Title: Mobile Auditory Guidance for Public Transportation

Date: 22.8.2011 Language: English Number of pages: 6+55

Department of Signal Processing and Acoustics

Professorship: Audio Signal Processing Code: S-89

Supervisor: Prof. Vesa Välimäki

Instructor: Dr. Cumhur Erkut, D.Sc. (Tech.)

Comprehensive information for using public transportation in the Helsinki region is freely available and there are different applications that provide this. These applications, however, typically only provide information in the visual modality, completely neglecting the potential of using audio for conveying (additional) messages. In the context of mobile devices and requiring information while on the move, this can be cumbersome and requires the user to pay a lot of attention to the application in order to get the information.

A completely functional mobile application for journey planning and guidance for travelling along a found route was designed and developed in this thesis. The aim for the application is to reduce the attention requirements to a minimum by using sound to deliver useful information to the user based on schedules, time, and the user's location. The objective is to provide the information by meaningful and easily understandable non-speech auditory cues to free the user's eyes and hands while, at the same time, removing the need to pay attention to the time.

The recognizability of the used sounds, and estimation of whether they would be useful or annoying, was tested informally by a small number of people. After a brief explanation of the purpose and metaphors of the sounds in the application, most auditory icons were recognized well. Using an application with any sort of sound in public without headphones raised doubts in some. Overall the application was found to have potential for being a helpful aid and most of the test subjects saw elements that would be useful for themselves.

Keywords: Audio design, audio systems, mobile communication

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Juho Kostiainen

Työn nimi: Julkisen liikenteen käytön ääniopastus mobiililaitteessa

Päivämäärä: 22.8.2011 Kieli: Englanti Sivumäärä: 6+55

Signaalinkäsittelyn ja akustiikan laitos

Professuuri: Äänenkäsittelytekniikka Koodi: S-89

Valvoja: Prof. Vesa Välimäki

Ohjaaja: TkT Cumhur Erkut

Helsingin seudulla on vapaasti saatavilla kattavaa tietoa julkisesta liikenteestä. Tietoa välitetään käyttäjille monien eri sovellusten avulla. Yleensä nämä sovellukset toimittavat tiedon pelkästään visuaalisessa muodossa, jättäen äänen potentiaalin (lisä)tiedon välityksessä täysin vaille huomiota. Graafinen ratkaisu voi olla epäkäytännöllinen ja tiedon saanti vaatia paljon keskittymistä, kun kyseessä on mobiililaite ja informaatio tarvitaan liikenteessä.

Tässä työssä suunniteltiin ja kehitettiin täysin toiminnallinen mobiilisovellus reittisuunnittelua ja -opastusta varten. Sovelluksen tavoitteena on minimoida visuaalinen huomiointi ja tarkkailu, toimittaen äänen avulla tietoa perustuen aikatauluihin, kellonaikaan sekä käyttäjän sijaintiin. Päämääränä on tarjota informaatio merkityksellisessä ja helposti ymmärrettävässä muodossa, hyödyntäen muuta ääntä kuin puhetta, vapauttaen käyttäjän silmät ja kädet sekä poistaen tarpeen seurata kellonaikaa.

Käytettyjen merkkiäänten tunnistettavuutta ja arvioita niiden ärsyttävyydestä testattiin epämuodollisesti pienellä määrällä koehenkilöitä. Sovelluksen tarkoitus ja äänten metaforat selitettiin lyhyesti, jonka jälkeen suurin osa käytetyistä auditiivisista symboleista tunnistettiin hyvin. Ylipäänsä mitään ääniä hyödyntävän sovelluksen käyttö julkisissa paikoissa ilman kuulokkeita herätti joissakin testihenkilöissä epäilyksiä. Kaiken kaikkiaan sovelluksessa todettiin olevan selviä mahdollisuuksia hyödylliseksi apuvälineeksi ja useimmat koekuuntelijoista havaitsivat ominaisuuksia, joiden he kokivat olevan itselleen käytännöllisiä.

Avainsanat: Audiojärjestelmät, mobiili tiedonvälitys, äänisuunnittelu

Contents

Abbreviations

API	Application Programming Interface
ARA	Augmented Reality Audio
CAF	Core Audio Format
GPS	Global Positioning System
HCI	Human-Computer Interaction
HFES	Human Factors and Ergonomics Society
HSL	Helsinki Regional Transport Authority (Helsingin seudun liikenne)
HTTP	Hypertext Transfer Protocol
ICAD	International Community for Auditory Display
iOS	iPhone OS, a mobile operating system by Apple Inc.
JORE	Register of Public Transport (Joukkoliikennerekisteri)
JSON	JavaScript Object Notation
KKJ	Finnish coordinate system (Karttakoordinaatistojärjestelmä)
OpenAL	Open Audio Library
PCM	Pulse-Code Modulation
SDK	Software Development Kit
WAV	Waveform Audio File Format
WGS84	World Geodetic System
XML	eXtensible Markup Language

1 Introduction

Hearing is an effective way to receive information. We perceive sounds from all directions immediately, whereas to obtain visual information one needs to look at the source of the information, thus losing visual awareness of the surroundings. Visual aids have been designed to assist in the use of public transportation, such as simplified maps for subways where the routes are described more as lines rather than their actual shape to improve clarity, but the use of audio is mostly limited to spoken announcements at, for example, railway stations.

Using sound to provide (additional) information is natural. On the street you can hear a car approaching without seeing it. When you hear rain drops you know to take an umbrella without checking the weather. Backround sounds in a phone conversation give you an idea of the situation the person you are talking to is in. Movies and games use sounds extensively to set the mood and provide the context of scenes. A completely dark scene in a movie can elaborately describe the situation and surroundings by the sound of squeeking floorboards, foot steps, sound of breathing, wind howling, etc. Of course a lot of the information different devices can provide are not specifically related to certain recognizable sounds. This, however, does not mean that using sound should be neglected in these cases. After all, before movies became 'talkies', many movie theaters had a piano player providing the soundscape for the scenes to help set the mood. Even abstract sounds can clearly be descriptive.

As the amount of information around us keeps growing, the more difficult it can become for a user to extrapolate the most relevant pieces of information from a visual display. These days, more and more information is accessed by using small hand-held devices with small visual displays, which means the user will likely have to navigate through different views to access all the information that cannot fit on the screen at the same time. Delivering some of the information in auditory, instead of visual, form can therefore improve the user experience by making displays less cluttered and ensuring that the user gets the information when it is presented. Still, the most convenient advantage with using sound may be that the user does not need to take the device out of the pocket and look at it while, for example, riding a bicycle or carrying grocery bags in both hands. That is why it makes sense to have different ring tones for phone calls and text messages – the latter one tends to be less urgent. Similarly, you can use more complex sounds to provide further details about an event. A sound played to inform about a received e-mail could have some variation to describe its content. The same melody could be used with different timbre to distinguish an e-mail with an attachment or an e-mail marked urgent from a regular one with just text in it.

There are different choices for different situations when using audio as a source of information. The sound can be, e.g., speech (recorded or synthesized), intuitively recognizable sound items (auditory icons) or abstract sounds for which the meaning must be learned (musical and hierarchical earcons). While speech can deliver the information most accurately and unmistakably, it usually takes longer to play than a non-verbal auditory cue and can feel more embarrassing in public.

1.1 Goal of the Thesis

The goal of this thesis is to design and implement a mobile application that uses non-speech sound for conveying useful information to assist the user with using public transportation. The application aims to provide the user with information based on time, location, and the timetables of public transportation to allow the user to pay less attention to the journey and what time it is. Information, such as when to leave, distance to go, and whether there is a need to hurry, can not only help in catching a bus but also relieve stress by giving the reassurance of being on time without the user having to focus on checking the situation. The sound design of the application aims to provide meaningful notifications instead of the commonly used, ambiguous alarm sounds of many electronic devices that often leave the user wondering what the sound was for. When a planned departure time is getting closer, it is natural to check the time more and more often, which can be very distractive if the user loses visual focus when, for example, reading something. Simply shaking a mobile device to get auditory feedback about the remaining time, or if the user is in a meeting, by vibration, can provide the same information without becoming distracted. In route guidance it often is not enough to check the plan for the journey before leaving. Because the information needed is related to time and place, more checking is required, which can be troublesome or annoying when on the move. Mobile devices, these days, often have built-in digital music players, internet access, and even global positioning. This, combined with the common habit of people listening to music on their devices with headphones in buses and trains, provides a great opportunity for the use of audio in delivering a significant amount of information without distracting or requiring much effort by the user.

The created application uses an application programming interface (API) provided by the Helsinki Regional Transport Authority (HSL) and, therefore, includes timetable and route information for all public transportation services in the Helsinki region.

As the application is for a mobile device, and particularly for use on the move, it raises a few particular design concerns which will need to be addressed. The user should be able to use the application while moving, so it should require **minimal attention** to allow the user to keep his eyes available and to not demand a lot of precise touches of buttons. The application should, therefore, be easy to use and let the user know when attention is required rather than having the user focused on it. A key concern for mobile devices, and in particular mobile phones, is **battery consumption**. Limited battery functionality should not be wasted on secondary applications if they do not benefit the user. The application should, then, be designed to eliminate substantial power consumption that does not bring value to the user.

Another important aspect for mobile applications is how to visualize the information on a small screen. Because of the limited space available, providing all the information is often not possible so the designer has to make decisions about which pieces of the information should be shown and how to present it. This thesis, however, is about providing the information in the auditory modality and the graphical design is not focused on. Auditory solutions may, however, affect the graphical part as it can help reduce the amount of information that needs to be fit on the screen.

1.2 Structure of the Thesis

In Section 2, the use of audio to convey information is familiarized along with the different types of auditory cues that can be used for it. The fundamentals of route planning when using public transportation is also presented in this section, mostly from the viewpoint of a mobile user. Section 3 explores two relevant studies in detail to provide a better understanding of what kind of auditory cues are found useful and effective in practice and how well they are learned and remembered. Section 4 discusses how the created application is structured and designed. In Section 5, the evaluation of the application is covered and the results discussed. Section 6 summarizes the work done and discusses possibilities for future research on the subject.

2 Background

This chapter provides the background for the sound design part of the work that was completed. First the concept of auditory display is introduced and its purpose and benefits are explained. After this, the various types of auditory cues are presented along with their appropriate usage in conjunction with the relevant auditory displays and interfaces as different situations and requirements often define what kind of sounds can and should be used. Generally, the auditory alarms and notifications used in machinery and electronic devices have mostly consisted of abstract beeps and other ambiguous sounds. Too often, a user's first response to a device making a sound is to wonder "what was that for?" Designing more meaningful and descriptive audio notifications to, for example, mobile services has been addressed only recently [1]. Finally, some basic information about journey planning and routing is covered, particularly from a mobile user's perspective.

2.1 Auditory Display

Hermann traces back the birth of auditory display as a research field to the first International Community for Auditory Display (ICAD) conference in 1992 [2]. The research field is, then, still relatively new. The vast improvements in technology over the past couple of decades have made it easy to use practically any sounds imaginable in interfaces, rather than the simple tones used earlier. The same applies for audio notifications in factories and warehouses that used to rely on bells and whistles for warning and informing workers. An auditory display uses sound to convey information about the state of the device to the user. McGookin and Brewster distinguish auditory displays from auditory interfaces by the fact that auditory displays only use sound as output, while auditory interfaces may also receive information via sound (mainly by speech recognition) [3]. Hermann suggests that, since for visualization a display is a necessity for the user to see the information, for auditory display the aspect of converting sound signals into audible sound is similarly required. This means that the auditory display itself must have the technical means of creating an audible perception (such as loudspeakers or headphones) [2].

Even simple beeps and warning sounds allow us to act when something happens instead of checking if something has happened or not. When you hear the sound of a microwave, you know your food is ready without having to pay attention to it. As Brewster put it, "our ears tell our eyes where to look" [4]. Using auditory information to accompany visual information can be a very efficient and natural way of using an interface. Directing some of the information to the user's auditory sense reduces the visual workload and the need to divert eyes from the point of focus to receive secondary information which may or may not be relevant. Just as hearing and vision work simultaneously in regular life, displays can provide more information by using both graphics and sound, without impairing one sense to enable the other. Not only can auditory information provide additional information, but experiments have shown that, in some cases, auditory cues can be more effective than visual cues [5]. Experiments have shown that reaction times can be significantly faster in a categorization task when congruent information is provided in both visual and auditory modalities, rather than just visual (e.g., seeing a picture of a dog and hearing the sound of a dog barking as opposed to just seeing the picture) [6]. Rath found auditory feedback to improve performance in balancing a virtual ball on a wooden control track the user tilted in his hands [7]. In his experiment the virtual ball and the track were graphically displayed on a computer screen and different sound models were used to provide a sound of a ball rolling on the track.

Displays that use ambient music aim at expanding the user's attention capacity by presenting information in the periphery in a non-stressful way [8]. Ambient sound, and even simple sounds played for a prolonged period, start to fade out of awareness gradually [9]. It is this habituation that allows you to become unaware and undisturbed by, for example, the sound of ventilation. Ambient sound only grabs attention if it changes or if it is unexpectedly missing. A constant background sound can therefore be unobtrusive, or even mask unwanted sounds, while also giving the reassurance that a system is working. In some situations a familiar sound is expected and can indicate that something is working. For example, if the microwave does not make any sound, it may seem broken. Sometimes the lack of sound can even be a safety issue. Adding sound to quiet hybrid vehicles has been studied, to improve the safety of pedestrians who expect to hear a car approaching [10]. Sometimes an ambient sound with different features describing specific information can subconsciously give the reassurance that things are as they should, because you hear the expected sound [11].

The sound itself does not necessarily provide all the information from an auditory display. The direction of the sound can be one part of the conveyed message as well. Walker et a. designed a mobile calendar application that used the symbolism of a horizontal clock face for providing the time information of a calendar event by the direction from which the auditory notification came from (front would be twelve o'clock, right would be three o'clock and so on) [12]. Some participants in their experiment said that they received the time information "for free" by being able to memorize the event associated with a spatial landmark, as opposed to having to remember the event and the time as two details.

When designing auditory displays, it is particularly important to consider the purpose and context carefully. If the sounds emitted by a device seem annoying or distracting, the user is likely to restrict or completely disable playback. Annoyance has different aspects to it. Noise is, by definition, unwanted sound. If information received from a device by sound is relevant to the user it is least likely to be evaluated as annoying. The same sounds can, however, be meaningless and only act as a distraction to surrounding people. A user can also choose to disable the audio signals on a device to protect personal privacy if the content of the auditory messages would give too much information to other people present. This applies especially to mobile devices, which are often used in public places. The type of the sound and how often it is played also play a part in how annoying it can seem. Hearing speech can seem more distracting to outsiders than more abstract sounds because the content is evident and is, therefore, more likely to be picked up from the peripheral soundscape [13].

2.2 Different Types of Auditory Cues

2.2.1 Speech

Speech, whether it is synthesized or recordings of words spoken by actual people, is the most reliable way to deliver precise information by using sound. Speaking is, after all, a natural way for people to communicate detailed information to each other. If, for example, a value describing some magnitude changes to another, abstract auditory cues can provide a broad sense of direction and magnitude of the change. In order to know the precise values, speech is much more accurate and reliable. Speech, however, has some serious drawbacks. An obvious requirement with speech is that it requires the user to know the language – otherwise it will just be some unintelligible utterance. Speech is, therefore, a suitable choice for situations where all the users speak the same language and understanding the content of the information accurately is critical. With speech, delivering all the information should not need to be instantaneous (i.e., a situation where getting the information would prompt an immediate action) because of some inevitable delay before comprehending the entire message. Another issue is evident when it comes to continually changing or rapid interaction. Using speech to describe continuous changes will inevitably mean that the change has already happened before the full information has been acknowledged because of the time it takes to articulate even short sentences. Depending on the temporal importance of getting the message, the information may become useless, confusing, or misleading. Also, if there is a significant level of background noise, speech has a tendency to get more or less masked and harder to understand. Using non-speech warning sounds has been studied because of the urgency and resistance to masking in cases such as military aircraft cockpits, where spoken messages may not be suitable [14]. Speech is also generally not fast enough in situations such as games with real-time interactive environments that need to communicate events in high tempo [15]. Conveying more than one message simultaneously using speech is not suitable either and queuing the messages of simultaneous events would further decrease the sense of interaction.

Lengthy commands in speech recognition can be shortened when combined with the recognition of physical gestures. Bolt introduced a system where the user could point at objects on a screen and give the system commands by speaking [16]. It allowed the user to perform simple tasks by pointing and speaking out commands as simple as “put that there” to perform tasks that would, without gestures, require the user to say “move the blue triangle to the right of the green square.”

2.2.2 Sonification

Kramer et al. define sonification briefly as “the use of nonspeech audio to convey information” and more specifically, “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation” [17]. This separates sonification from not only speech but also from music, as the definition emphasizes the purpose of the usage of the sound. The purpose of the sonification is to deliver precise perception in an auditory form of the interaction or data which caused the sound. Sonification uses an organized and well-structured method to render sound to express information. Visualization is a good analogy for sonification. As visualization uses space to depict data by using techniques such as graphs and charts, sonification can deliver the information of data changes on a graph in time [2].

The term sonification does not only mean the mapping of data features to certain acoustic features of sounds. It covers the use of different ways to represent data by using sound, such as auditory icons or earcons, which are explained in detail in following sections. Describing a larger set of data with a single defined sound can be thought of as simply using a lower resolution for mapping the information than having several, more detailed points of the data described with multiple sounds with varying features.

2.2.3 Soundmarks

Soundmark is a term derived from the word *landmark*. Whereas the term landmark refers to locations that are visually easily recognizable, such as geographic features or buildings, soundmarks do this in an auditory form. Baus et al. found that auditory landmarks can be as usable as visual ones in route descriptions and that learning routes can be done using auditory landmarks. In their user experiment 12 out of 13 subjects found that orientation during a navigation task was made easier by using sound sources as auditory landmarks [18]. Soundmarks can then be used to position oneself in an area. Kainulainen et al. used the sound of a fountain as a soundmark in their mobile route guidance application [19]. Their scenario for using this soundmark was to instruct the user to proceed along a specific route the sound of a fountain was heard, at which point the user should then turn. The landmark equivalent for this would be having a picture of the fountain on a map.

2.2.4 Auditory Icons

As a "picture says more than a thousand words", a short sound, too, can provide a clearer portrayal than a longer spoken message. The idea of an auditory icon is the same as that of its visual counterpart, a graphical icon [13]. They both work because of the intuitive connection to the meaning they are conveying. The purpose of an auditory icon is to remind the user of a familiar object or concept in the world we live in. The symbolism for a graphical icon depicting a dog is the same and as identifiable as an auditory icon which sounds like a dog barking. In everyday listening, people associate heard sounds with the objects and events that are understood to be the direct source of it as opposed to the musical or psycho-acoustical parameters of it [20]. A familiar sound is presumed to be a result of a certain action and context. When you hear a ball dropping on the floor, the sound can be enough for you to evaluate if the floor is made of concrete or wood, or if the surface is carpeted, whether the ball was a golf ball, a tennis ball, or a basket ball, and perhaps even how high it was dropped from. If a real-life sound is identifiable enough and the meaning it is associated with is clear, it is able to convey complex information in a single short sound. When using speech, the user has to hear the entire message to understand its full meaning.

Auditory icons need not be natural, recorded sounds. In fact, cartoon-like caricatures or stereotypes of the sounds that naturally occur may often be more easily recognized precisely because they are exaggerated and emphasize the most relevant and unique parts of the sounds that are being imitated [21]. Auditory icons can be parameterized even if they are actual recordings of real events. It is possible to filter out unwanted details or emphasize them. Different elements can be added to the sounds as well to make the sound more accurate or specific for the action and object it is associated with. Modifying the sound of an object hitting another object can, for example, be altered to change the context. The size and material can sound completely different if new artifacts are added to the sound or some characteristics are attenuated. Adding an echo to the sound could give an idea of the space around the objects.

Instead of simply playing a single auditory icon, they can be concatenated to provide further information and context. The sound of a car horn alone can be easily recognized by a listener, but it does not provide a clear idea of why the driver would honk the horn. The situation could be a slow traffic jam where the driver is frustrated or it could mean an impending accident. Adding the sound of tires screeching would give a great deal of additional information implying an accident being the more likely reason for the horn. Even further, the sound of rain and thunder could provide the context so that the sounds are enough for the user to create a hypothesis that a car is veering off the road dangerously because of bad weather conditions.

Auditory icons are not a constant set of existing sounds depicting objects and events. New technology and events create new iconic meanings to sounds while some other sounds that were familiar in the past may become rare and forgotten. A non-meaningful sound can become very strongly associated with an event when it becomes more and more ubiquitous. A small hammer hitting a bell has signified a telephone ringing for about a hundred years. It may well be one of the most recognizable sounds that do not naturally occur in the nature, but has become a stereotype by being distinctly used for the one purpose. Even this iconic sound may, however, be losing its familiarity as new generations are accustomed to mobile phones with practically limitless choice of custom ring tones. Designers should also be conscious of the fact that the familiarity of sounds differ between cultures and subcultures. An auditory icon that is considered obvious in one country may be unheard of in another.

2.2.5 Earcons

A lot of the events and objects present in computing devices are not related to any specific sound in everyday life. New sounds for different meanings need to be created when an intuitive auditory icon is not available or is deemed unsuitable; perhaps because it would be too annoying or too ambiguous.

Blattner et al. proposed methods for creating audio messages that are easy to understand, use, and modify [22]. In these methods, single or rhythmicized sequences of pitches that form a short and distinctive pattern are called *motives*. These motives have rhythm and pitch as fixed parameters and timbre, register, and dynamics as variable parameters. Larger groupings called *families* can then be assembled from related motives. Different families can be recognized when earcons with similar sounds have similar meanings.

Whereas auditory icons are based on metaphorical, intuitive mappings between sound-producing everyday events and the things they represent, earcons are based on learning a new meaning for a sound to create an association with the desired event or object [13]. Auditory icons are parameterized by their natural attributes. The sound of scratching, for example, sounds like scratching, but also naturally provides an idea of the intensity of the motion and pressure used as well as the material being scratched. Earcons, on the other hand, can hold the same information of the same attributes, but they are based on an invented hierarchy that the user needs to learn in order to understand it.

Brewster suggests starting with timbre, register, and rhythm when designing a family of earcons [23]. He proposes creating the basic structure by having a different timbre and default register for each family of earcons and therefore differentiating them from other families. The different families can also be given individual spatial locations either by stereo position or full three-dimensions if the hardware it requires is available. Major sub-groups within the families can be created by using different rhythmic patterns. Further differentiation can be done with pitch, intensity, chords, and various effects. The designer should carefully consider how the users will detect the differences. If no reference is used, pitch, for example, is difficult to evaluate. How well different kinds of effects and chords are recognized depends on practice as well.

Earcons can be categorized into four types [3, 22]. The most simple of these types are *one-element earcons* which can only communicate one bit of information by being only one pitch or having rhythmic qualities. Unlike the other types, one-element earcons cannot be further decomposed to provide more detailed information. If several one-element earcons are to be used with the pitch as the only changing attribute, the interval between the pitches should be at least a couple of octaves because detecting a pitch accurately without any reference is very difficult [23]. The second type is *compound earcons*. These are formed by concatenation of shorter earcons that may be one-element earcons, or any other type, to create more meaningful messages. Compound earcons can be considered analogous to sentences created by putting words together. Melodies as simple as two notes can be assigned different meanings. The first tone acts as a reference point and the detection of the earcon can be done by comparing the relative pitch, which most people are able to do to at least some degree of accuracy. The third type, *hierarchical earcons*, are constructed around a grammar by considering each earcon as a node in a tree. A following node then inherits the properties of the nodes above it. For example a rhythm without a distinct pitch can represent an error and adding a pitch to it could describe what kind of error it is. This example is summarized by Blattner et al. in Figure 1 [22]. The last category for earcon types is *transformational earcons*. Transformational earcons are based around a grammar, like the hierarchical ones. The different acoustical parameters of an earcon, such as timbre, pitch, and rhythm, can be modified or transformed to change their meaning. The same note pattern with the same rhythm can then describe a different scenario based on the timbre. For example, McGookin and Brewster studied how the number of simultaneously played earcons affected the identification of individual earcons [3]. Their research tried using earcons to describe different parameters of rides in an amusement park. The earcons included three different types of cues for the type, intensity, and cost of the ride. The encoding of these parameters was done by timbre, melody, and the register in which the earcon was played, respectively. As an example, they tried using a low-pitched piano rhythm to represent an inexpensive roller-coaster theme-park ride and by changing the timbre to a violin, the meaning was changed to an inexpensive water ride.

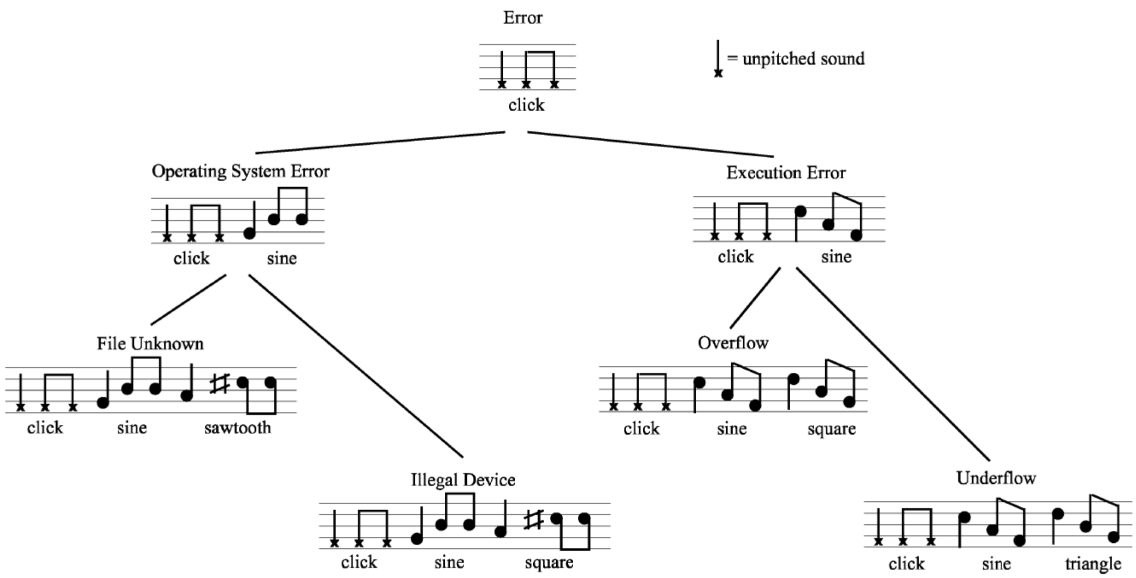

Figure 1: Example of the structure of the hierarchical earcon grammar by Blattner et al. [22].

2.2.6 Musical Cues

Katie Melua sings: "the piano keys are black and white, but they sound like a million colors in your mind." It is a good comparison for how diverse emotions and information can be expressed through music by subtle differences. Music is a complex entity comprised of a multitude of elements, such as rhythm, melodies, bass lines, vocals, several instruments, and sections, such as intros and choruses. Still, with all its complexity, it does not sound confusing and people are able to pay attention to many of its different elements at the same time. Even non-musicians learn to hear more details in music by practicing. You are more likely to detect different nuances in your favorite music than in a song you are hearing for the first time, because familiar sequences of sounds are more likely to be separated from other sounds [3].

Music can be a pleasant way to convey information, as most people can, and many even prefer to, listen to music while performing other tasks such as working, as long as the sound level is not too high. Not only does ambient music have a low level of interference with unrelated activities, but it can even enhance the participation. Just as with earcons, different parameters of music can be assigned different meanings. Ambient music can be composed with the information in mind, so that, for example, a new instrument can be added to signify a chosen meaning, or selected effects can be added to arbitrary music. A person can then recognize the specified cue (e.g., a certain instrument) within the music, while other people that are not aware of the meaning will not notice anything changing. Ambient music can, therefore, be a very non-intrusive way to convey messages. Ambient music can have the added benefit of masking some unwanted background noises. If an office space is too quiet, co-workers' actions and chatter can be heard too clearly, thus creating a distraction. [8, 24, 25]

2.2.7 Augmented Reality Audio

Whereas in virtual reality the user's environment is generated by completely replacing the real enviroment with a virtual one, augmented reality adds to the real environment without removing or replacing the existing objects [26]. Augmented reality, then, exists between reality and a completely synthetic fabrication by having elements of both.

In Augmented Reality Audio (ARA) sound elements are added without preventing the listener from perceiving the natural surrounding sound environment. The added sounds can be noticeable cues that stand out from the expected environmental sounds or they can be completely realistic so that the listener cannot distinguish them from the real ones. Using realistic sounds that are natural to the actual environment may prompt issues depending on the context and the purpose of the used sounds. On one hand, if the user is hearing something that is not present, it may be confusing. On the other hand, if the sound is meant to be a cue for a particular event, it might be mistaken as coming from the environment and the information could be lost. Listening to earphones that have microphones that pass the sounds of the environment to the ear differs from the naturally perceived sound field. Rendering virtual sound elements to the sound field has been studied by Härmä et al. [27]. It is found that detecting virtual sound from the real ones is affected by the direct leakage from the source to the ear canal and the natural habit of turning the head to better localize a sound source.

ARA could, for example, provide the soundscape of a location or event that is still, in reality, too far away from the user to be heard, through headphones. It could, then, be possible for a user to use his current location information and a compass, combined with a database of locations tagged with soundmarks, to listen to different directions around him over various distances – as, essentially, the capability of zooming in on a target in the auditory modality.

2.3 Mobile Journey Planning

Public transportation consists of different vehicle types, planned routes, and predefined schedules. In the Helsinki region the different modes of public transportation include walking, buses (and service line buses), trams, a metro, commuter trains and the ferry to Suomenlinna.

In order to efficiently use the public transportation network, the user needs to have some knowledge about the relevant routes and schedules. Awareness of the routes and schedules as well as being on time is not, however, a guarantee of a successful journey. Being accurately on time is a goal for public transportation but facing delays or being ahead of schedule is sometimes inevitable in real traffic situations. At the time of writing, there is real-time information available about the location of some of the public transportation in the Helsinki region, most notably for tram services. Figure 2 shows an example of a real-time vehicle tracking system by Mattersoft Ltd.

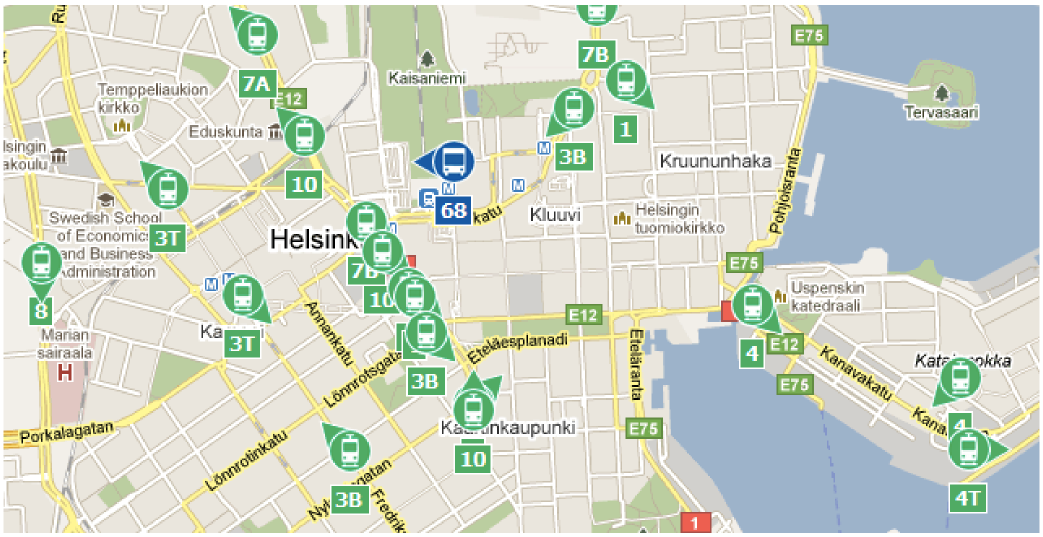

Figure 2: Real-time tracking of some of the public transportation in the Helsinki region by Mattersoft Ltd. [28].

Because the real-time information is provided only for a very limited amount of vehicles, relying on the planned schedules is typically necessary. When the user is mobile and needs to figure out how to get from place A to place B, one option is to carry along the timetables that show the full schedules for different transportations. Another option would be to just have the scheduled times for all the vehicles passing a certain stop the user often uses. These options, however, require the user to carry all the necessary information on his person and to browse through a lot of information to plan his travelling if the route is not familiar beforehand. The modern solution for these problems is using a mobile device to connect to a service that provides the routing information according to a few input details. The required details for a routing request are:

- Origin
- Time of travel
- Destination

In a mobile context the *origin* can often be the user's location at the time. If the mobile device has GPS (Global Positioning System) capabilities or other means of locating the user, the *origin* becomes trivial – the users do not even need to know their origin. When the user is mobile, it is natural to initiate route planning immediately before leaving, which means that the answer to the *time of travel* parameter is often the current time. The one thing that is not automatic is providing the destination. In order to establish the *destination*, the user should either know the name or the address of a target location or be able to pick the location from a map. When using personal devices, such as mobile phones, the *destination* could be provided by the application as a guesstimate based on a learning algorithm that compares the current location and time with earlier routes the user has used [29]. Besides these details, to actually use the route information in the form of travelling, the user must also know when to leave (actual time), where to board (which direction to proceed), which transportation mode and line to take (which vehicle), and when has he reached the target (when to exit the vehicle).

A route planner takes the required information from the user as input and provides as an output the best route options that are available for the specified time and locations based on what is considered the most optimal according to the computational logic used in the software. Because it is a machine, the given options are not always the choices a live person would consider the best. The software chooses different options according to specified weightings on factors such as whether it is worth saving x minutes in travel time if it means having to walk y meters more. The software computes the connections along the routes based on the schedules, taking into account possible optimization parameters from the user, such as minimum transfer margin to account for either the user or the vehicle being a bit late. The useful information can be delivered accurately in a visual form in a variety of different stylistic approaches [29] but utilizing the auditory modality is often completely neglected.

3 Prior Studies

In this chapter, a few relevant studies are examined thoroughly to achieve a better understanding of the requirements and characteristics significant for the design of auditory output for a mobile application.

First, a study on a mobile multimodal route guidance application for pedestrian and public transportation by Kainulainen et al. [19] is analyzed. This study presents an application that has a lot in common with the one created in this thesis. The application uses auditory icons and soundmarks to describe used routes and to improve the user's locational awareness. Secondly, a research by Garzonis et al. on the intuitiveness, learnability, memorability, and preference of auditory icons and earcons as mobile service notifications [1] is looked into.

3.1 Soundmarks in Spoken Route Guidance

Kainulainen et al. presented an already existing multimodal mobile route guidance application for pedestrian and public transportation called TravelMan [19]. They also created and performed user testing on a route description design, based on TravelMan, which utilized soundmarks and auditory icons. Similar to this thesis, their intention was to use the powerful and versatile features of recent mobile devices, such as GPS, to address the interaction limitations of the mobile use context, such as the small screen size and the user's eyes and hands being often occupied. Their goal was to use auditory icons for describing the different methods of transportation, as well as the distances travelled using them, and to identify points of interest. The TravelMan itself used synthesized speech as the auditory form of information. Using the auditory icons was an attempt to complement the synthesized speech and graphical information to provide the users with awareness supporting information in a less intrusive manner.

They listed keynote sounds, signals, and soundmarks as the three basic elements that soundscapes consist of. Soundscape, or acoustical environment, is a perceptual part of one's location awareness and spatial orientation. Keynotes are the constant background features that can quite often be unconsciously heard; e.g., traffic noises. Signals are attention grabbing sounds, often prompted by an action; e.g., a car horn. Soundmarks are sounds, or combinations of sounds, unique to an area, which can, therefore, be used for locational awareness.

The TravelMan displays the route as a reel, showing one part of it on the screen at a time with the previous and following sections easily scrollable to focus. The different parts include information about time, location, and event or action; e.g., transportation type, address, and time, or instruction to get off the transport, address, and time. All the information is synthesized to speech. Having GPS device support, the application has the option of providing real-time guidance about the progress of the journey, or as a reference description guide when used without the locating services. Kainulainen et al. believe that route descriptions and soundmarks would provide enough information to be used for route guidance even without position awareness but this, however, is something they have not yet tested.

In general, pedestrians with normal sight consider landmarks the most useful navigation aid. Kainulainen et al. used soundmarks as additional information to assist users in maintaining awareness of the route. The aim was to provide location awareness in a more subtle way. They believed this could be particularly important for the visually impaired. Figure 3 shows an example of how they added the soundmark information to the spoken message. This would give the user the confirmation and reassurance of being on the right track when hearing the sound of the fountain, or it could act as a caution if not heard even though it was expected.

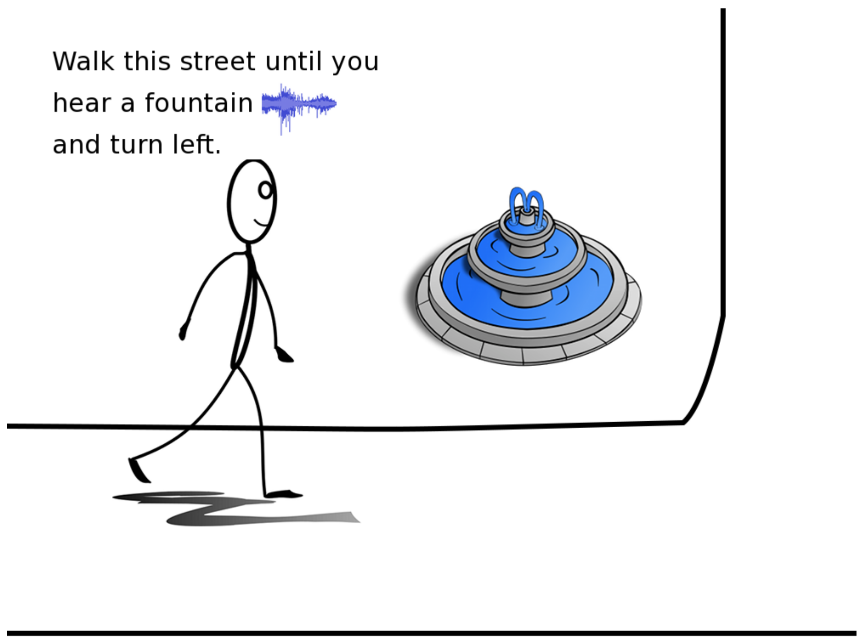

Figure 3: Example of the use of a soundmark in route guidance by Kainulainen et al. [19].

Kainulainen et al. used four vehicle types to describe the mode of public transportation on each leg of a route; walking, metro, tram and bus. Their idea was to replace the vehicle name and the time of travel, that was told by using synthesized speech, by using descriptive auditory icons instead. Their solution was to record the actual sounds of the vehicles from within them as they travel around the city. From the recordings, they picked representative sound elements that would be recognizable as auditory icons. They decided to modify the sounds by doubling the tempo and, as an effect to that, pitch, to make them more iconic and to avoid confusing them with actual traffic sounds. Only mono sounds were used, as the design considered a mobile phone and an optional hands-free set as the only hardware requirements for using the application. They believed that spatial audio might improve guidance precision but also that it might be a potential risk to use headphones that would attenuate the sound of traffic around the user.

Each auditory icon describing a mode of transportation would consist of three parts. The first and last parts were half second long recordings of the vehicle accelerating and decelerating, respectively. The middle part was the sound of the vehicle travelling at a constant speed. The duration of the auditory icon was to indicate the temporal length of that leg of the route. One second of the auditory icon would represent one minute of real travel time. The shortest, one minute long, travel time would only consist of the sound of the vehicle accelerating and decelerating, while longer travel times would include the middle part of the auditory icon played in a loop for the additional duration required to represent the correct temporal length of the leg.

Three different presentation styles were tried. One consisting of only the complete message by speech with no auditory icons. The second used only auditory icons, one after another to describe the entire route. The third presentation was a combination of using both speech and auditory icons. Whereas in the speech-only presentation the information included the mode of transportation, the possible vehicle line number, the address to go to, and the time of travel, in the combination with auditory icons, the spoken part would only include the possible vehicle line number and the target address. The auditory icon would then give the information of what transportation to use and how long the trip should take.

A total of 57 people participated in an initial test for comparing the three different route description styles. Each of the three styles were presented by first completing a practice task, followed by three recognition tasks with increasing difficulty. In each task, after hearing the auditory route description, four graphical presentations of the route were presented as a multiple-choice question. Opinions were gathered by an open questionnaire after the tasks.

Their experiment results showed that respondents negatively evaluated the presentation when auditory icons were used together with speech. This, however, was preferred over the presentation where only auditory icons were used despite the fact that recognition rates were lower in the former. The best recognition rate was associated with the presentation using only speech and the worst for the combined design. Kainulainen et al. considered whether there is a particular difficulty in using two different kinds of sounds, such as the recorded natural sounds (though modified in tempo and pitch) and synthesized speech, together. They note that previous studies have shown similar difficulties even in combining natural speech with synthesized speech [30]. They estimated that combining speech and non-speech is not trivial because they could be seen as different modalities. The auditory icons contain background noise and echo, whereas the synthesized speech is clear. In the combined presentation the speech and auditory icons are played alternately, which means that the soundscape, or the background noises, change abruptly several times, possibly making it more difficult to concentrate on the relevant information.

It was noted that the treatments to the tempo and frequency of the auditory icons made them less natural. As Kainulainen et al. concluded, recognizing the vehicles, especially similar ones like the tram and the metro, may be even more difficult because of the changed frequency range, even though the relative frequency component differences remain the same. A tram may start to sound like a vacuum cleaner and a higher frequency range and faster tempo can turn the sound of a metro into the sound of a jet engine. If the listener is confused by the sound or has to put effort into establishing meaning, it could well lead to missing or forgetting some of the other information. It seems questionable how relevant and functional the temporal information is after a certain period of time. Will the user be able to focus on the constant sound for longer than, say, 10 seconds and be able to tell even roughly how long the time was, while at the same time keeping in mind all the other information, such as the target address.

Kainulainen et al. admit that, even though initially they thought that it would be logical to sonify the actual travel time by the duration of the auditory icon depicting the mode of transportation, their design may have been the worst choice. They surmise that using more finite sounds would be better than the sound of the constant speed, which is the majority of the auditory icon for all legs longer than two minutes. They suspect that suppressing the noise of the auditory icons, to make them sound less like general traffic noises, could make it more difficult to intuitively recognize them, but more distinct and effective after having learned them.

3.2 Auditory Icon and Earcon Mobile Service Audio Notifications

Garzonis et al. designed auditory icons and earcons to be used for mobile service notifications and performed an evaluation of them in terms of intuitiveness, learnability, memorability, and preference in both laboratory and field conditions [1]. The idea was to create meaningful auditory notifications to let a user know when a certain mobile service is available. The need for service-specific auditory cues comes from the rising number of mobile services. If the user received the same notification for all services, eighet time might be lost checking the source of the notification to see if it is of interest or not, or it might be simply ignored altogether. The large number of different notifications, however, leads to the challenge of learning and remembering what each sound is associated to. Their research attempted to get a better understanding of what kind of auditory cues should be used to improve service awareness. As they mention, speech can provide the most precise information, but, due to privacy concerns, it may be the least preferred option for notifications and reminders used in public. In this research, they focused on auditory icons and earcons.

In an earlier, similar study Garzonis et al. started by gathering 52 available services provided by three major network operators in the UK [31]. First they divided the services into categories. They noted that it was not easy to develop a simple categorization, as the services could easily be considered to belong in several very different categories. They decided to categorize services according to purpose with three super-categories based on the origin of the service; from the ‘world’ (e.g., the internet), from another person (e.g., a phone call), and the user himself (e.g., a calendar reminder). Each super-category had three sub-categories based on the purpose the service would serve, leading to a total of nine categories that were all assigned different auditory cues. In the new study, Garzonis et al. used the same categorization as the basis for a new version that had been improved by further evaluations and cluster analysis. Another super-category was added for ‘other services’. The revised categorization is shown in Table 1.

Garzonis et al. selected and experimentally tested auditory icons to describe the

Table 1: The categorization used by Garzonis et al. [1].

	Super-categories		Categories
A	Information Services From the World	1	News Information
		2	Sports Information
		3	Here & Now Information
	Entertainment Services From the World	4	Entertainment Downloads
		5	Entertainment Live
B	Services From Other Users	6	Incoming Calls
		7	Incoming Messages
C	Services From 'myself'	8	Self Reminders
		9	Backup Reminders
D	Other Services	10	Other Services

different services so that they would be as identifiable as possible. For the earcons they used a different timbre for each of the four super-categories with various rhythms and pitches used for the sub-categories in them. The earcons were experimentally tested and improved by a professional musician for better clarity and improved pleasantness. Descriptions of the sounds used for each category are listed in Table 2.

The evaluation consisted of four stages. The accuracy of the responses and the response times were tested in all four stages to measure how well the sounds were recognized. The changing variables were time and training. The participants preference of the sound types was inquired at all four stages also. The first stage of the evaluation tested the intuitiveness of the selected auditory icons and earcons in a laboratory setting before any training, thus also providing baseline measurements for the following stages. Garzonis et al. hypothesized that, since auditory icons, by definition, are descriptive of the events and objects they represent, "auditory icons will be significantly more intuitive notifications than earcons" for the untrained listeners who can only guess the intended associations between earcons and their respective services.

The second stage was a week-long field test where the users received audio notifications on their mobile phone randomly during the day. Roughly once an hour two sounds were played; one auditory icon for one service and one earcon for another service. The participant would then use the interface on the mobile phone to guess which service each sound represented and received feedback confirming the answer if right or telling what the correct answer was. Garzonis et al. figured that playing auditory icons and earcons at the same times would remove the effect that social context may have on comparing which types of cues are more preferable, as both sounds would be played at equally convenient times and situations. No training was done before or during the field testing because the study was trying to conform with real user situations in which everyday devices, such as mobile phones, are usually used without intentional practicing. The second stage was intended to measure the learnability of the sounds. Their hypothesis was that "auditory icons will be significantly

Table 2: Descriptions of the sounds for each category used by Garzonis et al. [1].

	Category	**Auditory Icon**	**Earcon**
1	Information News	BBC News Ident	**Piano** Monophonic Going up
2	Information Sports	Stadium Crowd	**Piano** Monophonic Going down
3	Information "Here and Now"	Public Announcement at an Airport	**Piano** Monophonic Jumps up & down
4	Entertainment Downloads	20th Century Fox	**Piano** Monophonic Going down
5	Entertainment Live	Audience Applauding (e.g., in a theater)	**Piano** Polyphonic Going up
6	Incoming Calls	Old-Fashioned Phone Ringing	**Flute** Monophonic Short notes repeating
7	Incoming SMS Messages	Message Transmitted in Morse Code	**Flute** Monophonic Long notes going down
8	Self Reminders	Windows Mobile Reminder	**Vibraphone** Monophonic Going down
9	Backup Reminders	Truck Reversing	**Vibraphone** Polyphonic Going up
10	Other Services	Wind Chimes	**Violin** Varying pitch, chords, and single notes

more learnable notifications during the field study than earcons."

The field testing was followed by another test in a laboratory setting in the third stage. The users would, again, try to recognize the services related to the selected auditory icons and earcons, first without training and then after having been explained the logic behind the sounds and going through rigorous training. As with stage two, stage three assessed the learnability of the sound–service associations and had the same hypothesis that auditory icons would be easier to learn than earcons. They also expected the performance of both sound types to be significantly better after training.

For the fourth and final stage the participants completed a web-based questionnaire and tried connecting the right sounds to the right services again. As the goal of the fourth stage was to assess the memorability of the associations, it took place on two occasions; first one week after the third stage, and the second, three weeks later. Based on earlier research on learning and retention of auditory warnings [14], they hypothesized auditory icons to be easier to remember than the abstract earcon sounds.

The results for the intuitiveness tested in the first stage of the evaluation showed a significant difference between the auditory icons and earcons. As Garzonis et al.

hypothesized, the auditory icons were quicker and easier to associate with the correct service than the earcons, due to their natural quality of being relatable to familiar events and objects.

Garzonis et al. found that the recognition of both auditory icons and earcons improved during the field testing. The rate of improvement for recognizing auditory icons was significantly higher than for earcons. The training done after the initial test at the beginning of stage three improved the results for both auditory icons and earcons significantly. Recognition of the earcons, even after a very significant improvement from before training, fell short of the near perfect recognition level of the auditory icons after training. The hypothesis favouring the learnability of the auditory icons over earcons was, therefore, confirmed by the empirical evidence.

The results from the fourth stage showed that the correct association of both sound types suffered over time, as they expected. The meaning of the earcons was forgotten significantly more than for the auditory icons, confirming the final hypothesis that retaining the association of auditory icons is easier than of earcons.

In the earlier study [31] Garzonis et al. received inconclusive results about the preference of the type of auditory cues. During their testing they found participants preferring auditory icons over earcons, but in a following questionnaire the results contradicted this with earcons being the more preferred type. The current study, however, found that auditory icons were consistently preferred over earcons. Garzonis et al. also found a strong positive correlation between the correct associations of the notification and the sound preference, indicating that higher intuitiveness leads to higher preference and more successful learning. Because of the memory limitations of the mobile phones used, the sampling rate of the sounds was lowered from 44 kHz used in laboratory testing to a mere 8 kHz for the field testing. This may have affected how pleasant some of the sounds were considered.

Overall, they found auditory icons to perform better than earcons in all aspects. A few of the auditory icons performed badly, which Garzonis et al. concluded to be due to a poor choice of metaphors. Understanding, or being able to learn and easily remember, the metaphor is an important consideration along with how to use it. In developing a computer interface that uses auditory icons, Gaver, for example, used the metaphor of pouring water into a container to represent copying files [32]. In this metaphor the continuously raising frequency of the sound indicated the progress of the copying process, as if completing the copying was like filling a container with water. In the real world, sound does change in frequency depending on the volume of water poured into a container with which a listener is able to estimate capacity. Whether the mapping of the metaphor to the intended event is understood can be problematic, especially for events that do not exist in the real world – such as copying files. Combined with a visual progress bar filling, Gaver found it to be a plausible conceptual mapping.

The least recognized auditory icon Garzonis et al. used performed almost at a chance rate (9 %). The same auditory icon also had the lowest identification score (64 %) of the used auditory icons in the initial surveys that were used for selecting the auditory icons for the experiment. The second least recognized auditory icon performed significantly better (21 %), though still very badly, in the experiment as well as in the initial identification (77 %). Garzonis et al. thought that this result suggested that, among a convenient population sample, the identification rate of sounds to be used for mobile audio notification design should be higher than 70 %.

Other conclusions they made were that basic training can be enough to achieve recognition accuracies of about 70 % for ten abstract earcons but the performance of auditory icons will not be reached even after substantial training. Also, the more seldom an auditory cue is played, the greater the need for it to be based on a familiar metaphor.

Other studies, as well, have found auditory icons to be easier to learn than abstract sounds. Leung et al., for example, found that the most iconic sounds are, as can be expected, the easiest to learn and retain [14]. Edworthy and Hards, however, found that using real, environmental sounds does not necessarily make learning the association easier if the sound does not naturally relate to the event or object in question [33]. In other words, using a naturally recognizable sound as a cue for a randomly chosen event will not make learning the association more likely than with an abstract sound. This suggests that when there is no obvious auditory icon to convey some information, focus should be on considering what kind of sound would be most appropriate in terms of urgency, discreetness, and pleasantness as the meaning has to be learned whatever sound is used, and abstract sounds can be as easy to learn. Using certain characteristics of the sound should be considered, however, even if the sound itself is abstract. Even if the sound is meaningless, different features can impart subtle notions – even if only on a subconscious level. For example, the difference between a minor chord and a major chord can indicate a mood, or a more tense or dissonant sound can give the impression that a sound was intended as a warning or an error notice instead of positive feedback, such as a task having finished successfully.

4 Implementation

This section describes the components, functionality, and design decisions of the mobile application created in this work. Commonly when sound is used to deliver information in user interaction, it is a direct response to a user action, such as pressing a button. In this application, much of the auditory information is prompted by the combination of the user location and time. Even many of the more direct responses to a user pressing a button will also have a significant delay in the response time due to information, such as current weather and bus routes, being retrieved from servers on the internet. Therefore, they often act more as notifications rather than direct responses. The user can start a route search and put the device in his pocket while the search is being performed. The relevant and selected information will then be delivered by sound after it is retrieved and the time for the cue is appropriate.

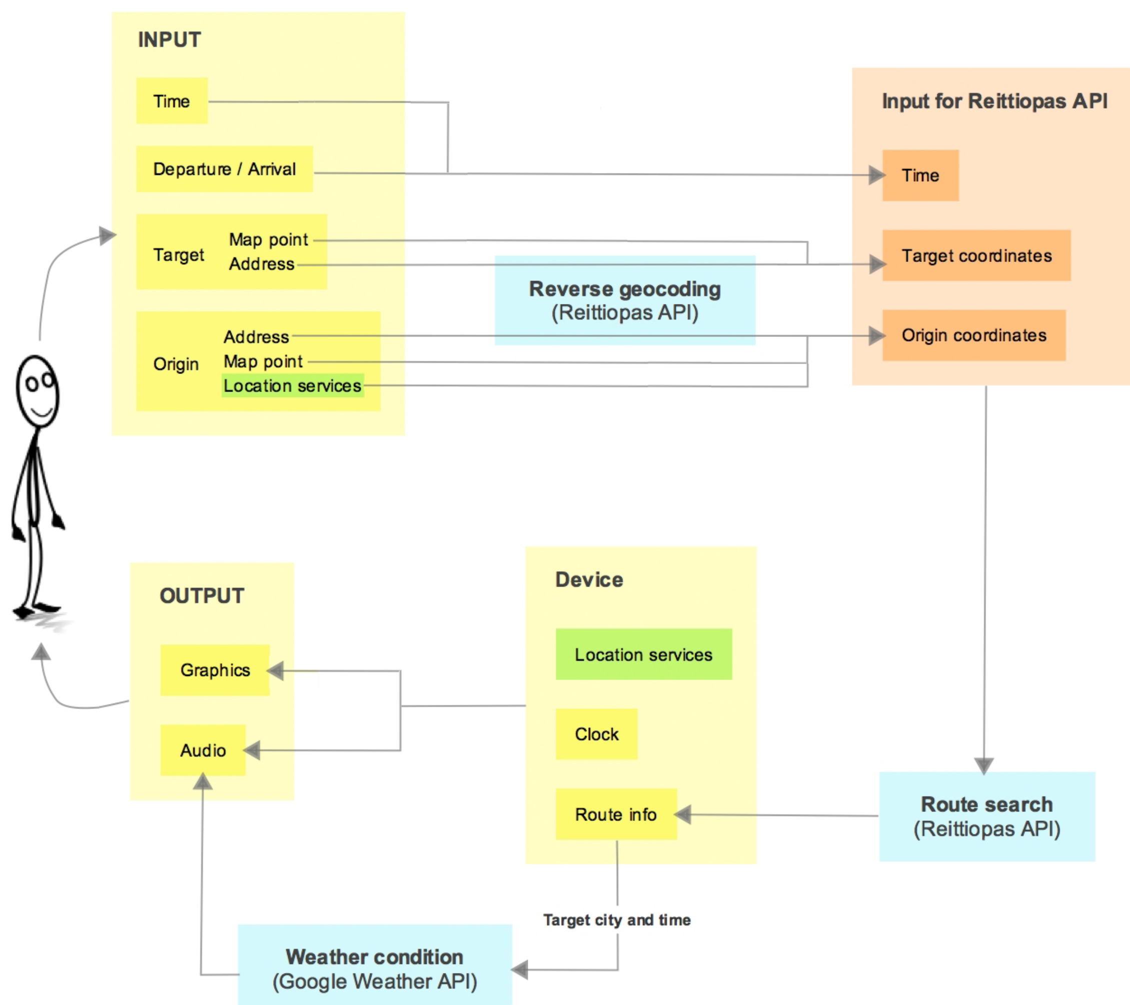

Figure 4: A flowchart illustrating the usage of the application.

The flowchart in Figure 4 represents the basic operation of the application from providing input to receiving output. The structure is described in more detail in this section by first presenting the external application programming interfaces that are used for retrieving the data on which the functionality of the application heavily relies. Next, the relevant frameworks used in the programming are briefly described. After that the functional design of the application is described, explaining the logic and the decisions behind the existing features and the way that the application works. And finally, the auditory cues used in the application are described along with the various situations in which they are played.

4.1 Application Programming Interfaces

This section describes the external application programming interfaces (API) on the internet that are used in the application. First, the API that is used for obtaining all location, route, and timetable information for the Helsinki Region Transport Authority (HSL) is described. Then, a weather API, which is used for obtaining weather conditions in the target city of a searched route, is introduced. Both APIs are only queried when the user searches for locations or routes. An internet connection, therefore, is always required to do a search but once the data has been loaded in the application, the connection is no longer needed until performing another query. This makes using the application possible with devices that do not have the capabilities for a mobile internet connection or if the data plan for the mobile internet connection would be too expensive, if the user can connect to a Wi-Fi network to search and load the necessary information before commencing the journey.

4.1.1 Reittiopas API 1.1.2

The Helsinki Region Transport Authority (HSL) provides free access rights to their Journey Planner interface for development use when the application or service being developed contributes to using public transportation and information availability related to it [34]. At the time of writing this, the latest available version of the API is 1.1.2. The API provides timetable, location, and route data by a HTTP GET interface. The response from the API is in either eXtensible Markup Language (XML) or JavaScript Object Notation (JSON) format, chosen with an input parameter by the user. Both XML and JSON are textual, defined formatting systems used for describing data structures in a machine-readable format that need to be parsed according to the predefined format to make it more human-readable. In this application, the data from the Reittiopas API is requested in JSON format. An XML database dump file that includes all the timetable and location data in one file is also available. However, since the dump file is hundreds of megabytes in size, providing it within a mobile application would not necessarily be appropriate since the data would still need to be updated when there, e.g., are any schedule changes. For that reason, for this case, only the HTTP GET interface is used and all the data is loaded from a server when the user initiates a search for locations or routes. The application, therefore, always requires an internet connection at the time of searching for locations or routes.

Locations can be searched for by address, name of a point of interest, or coordinates. The user can decide the coordinate system used for the input and output of the queries. The available coordinate systems to choose from include the first three sectors (KKJ1, KKJ2, and KKJ3) in the Finnish KKJ coordinate system, the World Geodetic System (WGS84), and the Mercator system. WGS84 is used for this application because the location services of the iOS use this coordinate system as well. Location search returns details, such as, the location address, name (if it is recognized as a point of interest), the name of the city the location is in, and the coordinates of the location.

Route searches require the coordinates of the point of origin, the destination, and an optional via location in between, as input. Time of departure or arrival, different parameters for optimization (e.g., least transfers, least walking), and other details, such as minimum number of minutes between changes, can be used to customize the search further. The API can be queried to display one, three, or five route options at a time. The response will then include the selected amount of consecutive routes with the same input details so that the user can easily compare which departure or arrival time suits best. Because the route search uses only coordinates, a reverse geocoding, by doing a location search, is required in order to get the coordinates when the user selects the location by typing in an address or the name of a point of interest. The user also has the option of selecting locations from a map by opening a map view. A desired location is simply selected by cursor pointing. After that, the marked location can be assigned as the start, via, or destination point.

Figure 5: Different parts of the route are drawn with a color representing the type of transportation.

The response for the route search contains the different legs of the found routes with details including a list of the coordinates that define the path for the leg. By connecting all the coordinates of all the legs, the entire route can be drawn. As the coordinates are provided separately for each leg, it is easy to separate them clearly in a visual representation, as can be seen in Figure 5, as well as for future processing. The other details include the arrival and departure times, duration, and length for each part separately, as well as, for the entire route. For each leg of the route, unless the leg is to be walked, a unique line code is provided in the form they are listed in the Register of Public Transport (JORE). From the JORE-code the relevant information can be parsed. This information includes the line code that is familiar to the passengers, type of transportation, the area the transport is used in, and in which direction it is heading.

4.1.2 Google Weather API

Generally, the APIs that provide weather forecasts for cities in Finland with time frames as accurate as a few hours are not provided free of charge. The Google Weather API is an unofficial interface that provides an XML feed as a response to a HTTP GET request [35]. It was chosen to be used in this application because it is free, very simple to use, and adequate for this application. Calling the API can be done as easily as providing only the name of the desired city as the input parameter. It is used by Google applications but it is undocumented. Still, it is very intuitive to use and accessible for anyone. Being unofficial, there is, however, no guarantee that it will stay as is. The API provides the latest weather condition observation for a requested city and a forecast for the following three days. Since a forecast is not available for different times of the day, the weather data should be taken "with a grain of salt" when it is for a journey that will commence several hours following a database query. For the purposes of this application, however, it can be considered fairly sufficient. Most often when you are searching for a route and care about the weather, you are likely to be leaving shortly, in which case the latest observation that is updated once every 30 minutes is likely to be a good enough indication of the weather for the time you will be travelling.

4.2 Structure of the Application

The most relevant of the different frameworks that were used in the application are presented in this section. Native iOS applications are programmed using Apple Inc.'s iOS Software Development Kit (SDK). The different technologies in iOS can be considered as a set of four layers of the system. The lower layers entail the fundamental services and technologies that all applications rely on. More sophisticated services are available at the higher levels. The highest layer of iOS is Cocoa Touch. The Cocoa Touch is an API, based on the Mac OS X Cocoa API, that provides frameworks for building software programs for Apple Inc.'s mobile devices that utilize touch screens. The frameworks include powerful functions for performing various tasks in only a few lines of code as well as foundational APIs in the programming language C to provide the programmer direct access to the lower level systems when needed. The programming for the higher level frameworks is primarily done in the Objective-C language.

As the application's main functionality is based around information downloaded from APIs on the internet, the data needs to be read and parsed into a usable format. Reading the weather data from the Google Weather API is done with the XML parser class in the foundation framework of the Cocoa Touch. For the more complex and broad location and routing data received from the Reittiopas API, an open source JSON framework [36] is used. The framework provides a strict JSON parser in Objective-C with a very simple interface. The framework also includes a JSON generator which, however, is not used in this application as the data only needs to be read, not written.

Core Location framework provides the means for finding the user's geographical location. The locating uses different technologies together; nearby Wi-Fi networks, cellular towers, and GPS satellites. The technology used affects the accuracy and power consumption. In this application the user can pick from several choices the accuracy used when tracking a selected route. If the user considers an accuracy of one kilometer, for example, sufficient, the GPS might be turned off to conserve the battery while the locating is done by only utilizing the nearby cellular towers and the Wi-Fi networks within reach. The most accurate tracking option consumes so much power that it is recommended to only be used when the device is plugged into a power source.

Besides using sound, the application also includes a regular, visual guidance system displaying the route and, when tracking is activated, the user's location on a map. The built-in Map Kit framework is used for displaying the map and drawing the paths and annotations, such as icons for the start and destination points, on it. The Map Kit framework uses Google Maps and is able to display a traditional road map, a satellite image, and a hybrid of the two. The last viewed area of the map can be stored in the devices cache memory but mainly the map is loaded from the internet as it needs to be displayed on the screen. The Map Kit framework allows performing reverse-geocoding easily to get the coordinates of a location on the map. This function is used in the application for being able to mark the locations used for route searches simply by touching the screen at the desired map location.

4.3 Sound Programming for the Application

The application is designed so that the user can listen to music on the device without being interrupted by the auditory cues or other events of the application. In order to allow mixing the application sounds with the music, an AVAudioSession is initialized and configured appropriately as the application is launched. The defined properties include setting the session category to *playback*, which allows mixing the external sounds (music played by the music player application) with the application's own sounds as well as allowing sounds to be played while the screen of the device is locked, which is essential because the main purpose for the application is to play sounds without the user actively handling the device. Another property that is set for the AVAudioSession is an audio route change listener to be able to only play certain sounds when headphones are plugged in. The AVAudioSession also registers a callback for detecting and gracefully handling interruptions, such as incoming phone calls, by pausing the music for the duration of the interruption and preventing any application sounds from being played until the interruption has ended. The AVAudioSession can later be called to use properties like “audio ducking”, which drops the sound level of other playing sounds for the duration of a new sound (cf. Figure 6). This allows lowering the music volume for a moment to make sure an auditory cue can be heard without stopping the music entirely. The AVAudioSession is defined in the AV Foundation framework.

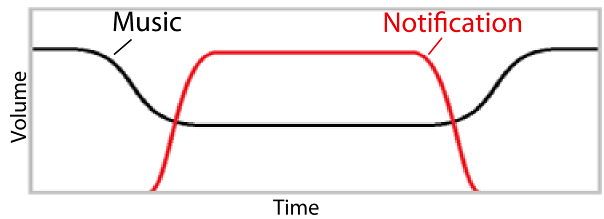

Figure 6: Music level can be lowered for the duration of an auditory cue being played.

The AV Foundation framework is an advanced Objective-C interface, common for Mac OS X and iOS, for playing and managing audio–visual media in applications. It is one of several frameworks in the iOS for creating and playing media files (other frameworks include, e.g., Media Player and Core Media).

Several different systems are used for playing different sounds in the application. The music the user can listen to while using the application is played by the external music player application. A music player controller, from the Media Player framework, is used for adopting the control of the external music player application, however, to allow the user to pause and resume the music from within the application. As it controls the external player, if the music is started from within the application, it will continue playing after quitting the application. Because only the control of the music player, but not the player itself, is adopted, the external player takes care of handling interruptions appropriately.

The iOS has a feature that automatically puts the device in a sleep mode when the screen is locked, unless sounds are being played. The sleep mode stops keeping track of the timers that are used in this application, e,g,, for playing departure cues and turning the tracking on at the right time. In order to keep the timers running properly even when the user is not listening to music, an AVAudioPlayer instance, from the AV Foundation framework, is created and set to play a short silent sound every five seconds. When the music stops playing the AVAudioPlayer is automatically created and it starts playing the silent sound file to keep the application awake. The AVAudioPlayer instance is removed when music starts playing. The AVAudioPlayer is a lightweight system and playing a short sound file at sound level zero does not consume a notable amount of the battery.

The sounds used in the application are preselected audio files and some default sound effects related to the objects used, such as typing on the virtual keyboard and scrolling a time picker. No audio synthesis is performed. Most of the sounds are used as simple auditory cues and played as is without any modification. These sounds are played by using System Sound Services, which is a very simple and compact C interface, in the Audio Toolbox framework, designed specifically for playing short sounds when there is no need for control over features such as sound level, stereo positioning, or looping. System Sound Services are also used for invoking vibration (if supported by the device).

OpenAL (Open Audio Library) is a cross-platform API for three-dimensional audio by Creative Technology Ltd. [37]. The primary purpose of the API is to allow the positioning of multiple audio sources in a three-dimensional space around a listener. This creates a reasonable spatialization for different sound directions when appropriate audio system is used (e.g., headphones or a 5.1 speaker setup). The API is very advanced for controlling different aspects of audio besides just the spatialization.

There are three fundamental objects in using OpenAL; buffers, sources, and a listener. The listener is a position with an orientation in a three-dimensional space. Source positions are then measured as vectors to the listener. The buffers are then loaded in their appropriate sources, resulting in a correct spatial experience as the sounds are played from the intended directions and distances relative to the listener position. Updating the positions and orientations of the sources and listener then enable creating a dynamic and convincing audio environment. The audio data is loaded into the buffers as PCM data. If compressed audio files are used, decompression is first required. The buffers are then connected to the appropriate sources to be played at command. The listener, sources, and buffers have different properties that can by dynamically updated. Properties include, for example, pitch, gain, position, orientation, looping, etc.

The OpenAL framework is used in this application when volume, pitch, sound direction, or looping need to be altered. A sound engine with low latency and a simple interface that is designed for the iOS, called Finch [38], was slightly modified and utilized in this application. Finch includes a decoder for getting the raw PCM data from uncompressed audio files (WAV, CAF) and a simple interface for playback as well as changing the properties for pitch, gain, and looping. In its simplicity, the sound engine initializes and prepares the desired, uncompressed audio file to be played when called. Positioning of the sound source was added to the engine to enable panning the sound for it to be heard from a specific direction.

4.4 Functional Design

This section describes the relevant concerns in designing a mobile application along with the found solutions. The decisions made regarding the functionality and features of the created application are also explained here.

4.4.1 Minimal Attention

The first important aspect about an application for a mobile device is to understand the usual circumstances in which the application will be used and how it is operated. Because of its small size and lack of common, fixed input equipment, such as a full-sized keyboard and a mouse on a solid surface, providing accurate input (e.g., typing) becomes difficult while moving. More focus and attention is required to ensure pressing the right keys when on the move. Unfortunately, this further increases another negative aspect related to the use of a mobile device – paying attention to the surroundings. Because the device is used when mobile, the user often has to be aware of what is going on around him. For example, you need to pay attention to the traffic while walking down the street instead of just looking at the device in the palm of your hand.

The first, and somewhat obvious, solution addressing the problem of giving input to a device while moving is to minimize the amount of actions (like pressing buttons) required from the user and to minimize the amount of time the screen needs to be looked at. The application is designed to work in an automated way so that as little as one touch of a button is often enough. The user can define two default targets to be used when the application is launched; one for the morning (between 05:00 AM and 10:00 AM) and another one for other times. This allows the user to often use the application by only touching one button after opening the application. In the morning the application can, then, find a route from the current location to work and in the afternoon or evening the single touch can find a route from the current location to home. When the user needs to type in an address, an incomplete address with three or more letters is sufficient to perform the search. If only one matching location is found, the application will automatically start searching for routes without requiring further actions or even glances from the user. If, however, more than one location match the given address, the user is notified by a rapid clicking sound indicating the amount of found locations so that he knows a search has finished and requires further attention. The amount of clicks heard are equal to the locations found, although limited to a maximum of 40. The clicks are played very rapidly to simply give the user a rough "ballpark" figure of how many locations were found. The user is also notified about a finished route search by informative auditory cues. The provided information and the sounds used in delivering it are described in detail in the section *Sound Design*. Because the user can choose to have the first found route automatically selected, there is often no need to even wait for the route search to finish before putting the device in the pocket and focusing on something else. When a route is selected, notifications for occasions such as the time of departure and prior ones to be able to prepare to depart, are automatically created and activated based on settings chosen by the user. These are explained further in the *Sound Design* section as well.

4.4.2 Retrieving Information

The address, route, and weather information are loaded from servers on the internet. An internet connection is, therefore, required whenever locations of new addresses or new routes need to be found. The weather information is retrieved only immediately after a new route search has been performed. The time it takes to perform searches depends on the internet connection used, the traffic at the server of the API, and also the type of search. The Reittiopas API caches recent searches so that they need not be calculated again, which means that performing the same search twice has a chance of being faster the second time. The amount of data depends on the search type as well. The response to a route search includes a full list of coordinates for the route, whereas a location search will use much less data. The details for a route with a length of ten kilometers can often have more than 10,000 characters in it. A bicycle route can have up to five locations marked between the starting point and the destination, and the data for the route can become more than 100,000 characters for long routes. The user has the option of retrieving either one, three, or five routes per search. Searching for five long routes can, therefore, mean loading over 500,000 characters of routing information. The JSON parsing that is used for getting the Reittiopas API will cancel the process if it gets no response within 30 seconds of starting the search, and inform the user with a sound indicating an error and a message displayed on the screen.

Once a route search has finished and the weather information has been retrieved, an internet connection is no longer required for using the application until a new location or route needs to be loaded or the user wishes to look at a map. This means that the application can be used on devices without mobile internet connections as long as the device can be connected when doing the search. For example, a route can be searched via Wi-Fi connection before beginning the journey. The user can save locations to the application's memory but no route information is stored so all the searches will require the connection to the internet. If the user quits the application, the locations, routes, and tracking status at that time is saved and automatically loaded when the application is launched again if the arrival time for the routes has not yet passed. If no routes were loaded at the time the application was quit, the status and information at the time of quitting is saved and loaded if the application is launched again within 30 minutes to allow the user to continue from where he left off if, for example, interrupted by a phone call, finding an address from contacts, or checking a time from the calendar.

4.4.3 Threading

Some elements and methods of the application are called in different threads that are separated from the main thread in order to keep the user interface responsive at all times. The main thread is created as the application is launched and, generally, the tasks in the application will be done in this thread unless the program is instructed to create new ones. Creating another thread means separating the processing so that the primary tasks, such as responding to the user using the interface, are handled in the main thread which has priority, while other tasks can be performed in the new thread when possible. If the main thread is performing heavy processing, the separated secondary threads may take longer to finish their tasks to let the main thread work without delays. If the main thread would begin performing a lengthy task that has to finish before moving onto the next one, or a task would tell the thread to "sleep" before performing it, it could seem to the user as if the application had crashed. Any unresponsiveness in an interface is easily annoying. The internet queries can take a significant amount of time. All the processing for these queries are done in separate threads. The application sounds are also played in different threads. Playing the sounds itself would not make the interface unresponsive but initializing the sound systems can take a moment – especially on older devices. In some situations the application will play queued sound items with specified delays between them. These delays are done in a way that would make the user interface unresponsive if it was done in the main thread. Using separate threads for all the sounds allows playing different sounds with the system sound services simultaneously, as well. This means that, in appropriate situations, different auditory cues can be played at the same time instead of queuing them.

Playing the auditory cues in different threads does pose some new challenges to take into account, however. When an interruption happens (e.g., an incoming phone call), it is essential to prevent the application from playing any sounds until the interruption is over. Because the music is played by the external music player application, it is automatically handled gracefully, but the application's own sounds need to be dealt with as well. Normally, the System Sound Services would be automatically stopped when an interruption is detected and it would, in fact, be impossible to keep them playing. As the System Sound Services are being played in a different thread in this application, the normal situation does not apply here. When an interruption begins, a variable is set that prevents new sounds from being initialized until the interruption has ended. Already initialized sounds check the status of the variable whenever they are about to play a sound, to prevent queued sounds from being played during interruptions. The OpenAL framework requires the user to set a callback method to handle interruptions. Dealing with interruptions gracefully is done in the same way for the OpenAL sounds as it is for the System Sound Services, except that instead of checking just before every sound, the OpenAL also checks the status once every second after starting to play. The difference is because the sounds played with System Sound Services are very short while the OpenAL may be playing a sound in a loop up to ten seconds at a time.

4.4.4 Information Related to Time and Place

In order to provide information related to both time and place, locating services are required. The application can be fully used as a regular visual journey planner with auditory cues for situations like departure times even without any locating services available or enabled. Information such as when to get off a bus, however, requires knowing the location of the user. Using GPS along with additional triangulation and tracking of cellular towers and Wi-Fi networks can determine the location accurately and quickly. Having GPS or mobile network capabilities is not an absolute necessity, however. Using just Wi-Fi networks alone is often sufficient enough for establishing present location as the starting point for route searches. In urban areas, even movement tracking is possible if the Wi-Fi networks are sufficiently ubiquitous to provide a practically constant coverage of the path.

The locating accuracy and frequency vary between different device models and even between the same models. The application is designed with a possibly very low update rate and moderate accuracy in mind so that, for example, subsequently found locations are not used to try to determine the direction the user is going. Also, locations further than 500 meters from the path are ignored as incorrect results when tracking a selected route using public transportation. It does, however, mean that the user should try searching for a new route that begins from the current location if more than 500 meters away from the public transportation route that is being tracked, in order to get location-based cues. This means that the functionality of the application is rather robust even with lesser equipment. When tracking the user along the selected route, any location based information is played when the location update happens. Depending on the type of event, there are, however, limits for how frequently certain sounds can be played to avoid annoyance.

The application sources the route information from the Reittiopas API as a set of coordinates. The coordinates are then connected with a line for the visual representation of the path on the map. To use the location information for more than just a distance from the user's location to one particular point "as the crow flies", the coordinates are also considered as points on a path from a mathematical perspective. When the location services find new coordinates for the user, these coordinates are compared with the coordinates of the route to first find the nearest location to the path. A line is formed between each consecutive coordinates along the path (or, if the consecutive coordinates are exactly the same, one is skipped over) and if the user's location is perpendicular to the line, the nearest point on the line and its distance from the user is measured. If the user is not perpendicular to the line, the distance to each coordinates at the end of the line are measured. After going through each coordinate pair the point on the path nearest to the user's location will be known as well as how far from the path the user is.

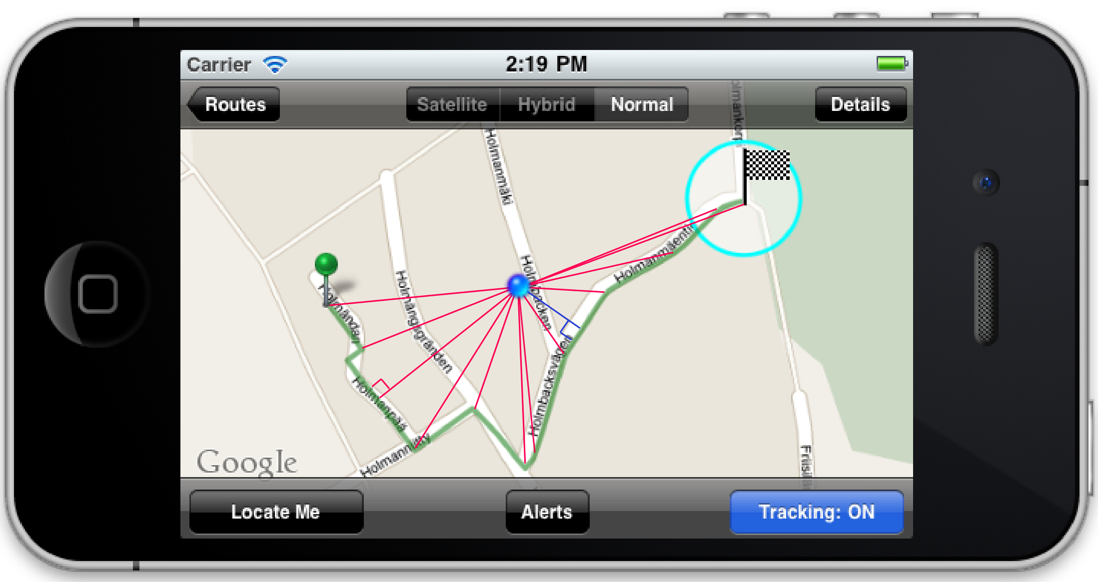

Figure 7: Simplified illustration of the distances measured between user location and path. Blue line selected as the closest position.

Finding the nearest point on the path relative to the user's location is illustrated in Figure 7. By comparing the slope of the vector drawn between two coordinates of the path and the user's location coordinates, it can be determined which side of the path the user is on. The distance from the user's location to the nearest point on the path is added to all the remaining distances from that point to the end of that leg as they are gone through in order to get the actual distance the user has to travel along the path.

4.4.5 Saving the Battery

Power consumption is a very important issue for mobile devices, and especially so if the device is a mobile phone. The locating services, and GPS in particular, use a great amount of power. It is, therefore, very important to consider when the location information is actually needed. For the significant portion of the trip, the user is likely to find location information useless. If you know you need to sit in a bus for 30 minutes, most of that time you, generally, will have no need for tracking where you are. The created application attempts to conserve as much of the battery as possible by automatically activating and deactivating the tracking based on time and place. Although the user can, of course, choose to not have the tracking on at all or to use constant visual tracking on a map.

The times when location information is relevant are usually, apart from finding the current location to use it as a starting point for the route search, only when the user should be walking or establishing where to exit a transport. When a route has been selected and tracking has been initiated (either automatically or by the user), the actual tracking is not activated immediately. Instead, the application establishes at what time the tracking will be useful and sets a timer to begin updating the location accordingly. The appropriate times for starting tracking are the times when the user should start walking and five minutes before a transportation is scheduled to arrive at the destination. This way the user can get real time guidance from the moment he should begin walking as well as real time information about when to get off the bus. The selection to initiate real-time tracking five minutes before arrival at a scheduled destination is arbitrary but is expected to be far enough in advance to avoid a user overshooting the vehicle exit point. If the transportation were to be over five minutes ahead of schedule, the tracking would begin too late. The public transportation schedules of HSL are, however, generally very accurate, as shown by high ratings in customer satisfaction regarding conformance to schedules [39]. Most often the variance between planned and actual arrival times are because of being late (e.g., because of bad weather conditions) rather than ahead of schedule. Further research on the benefit of periodically initiating tracking in order to check whether it is needed could be done if issues arise.

When the user turns the tracking on, the application finds its current location and determines which part of the route, for which the departure time has already passed, is the closest one. The timer for activating the tracking is then set appropriate for this leg of the route. This makes it possible for the user to turn on the tracking after having already began the journey so that the actual activation of the locating services are still done only when it is relevant.

Apart from automatically just turning the tracking on when needed, it is even more important to automatically turn it off as well. When the tracking is activated, the application compares each new updated location with the path of those legs of the selected route for which the departure time has passed. When the location update happens within a 50 meter radius of the end of such leg, the tracking is automatically deactivated to save the battery. Figure 8 shows a part of a route with white circles with radii of 50 meters at the end of its legs, which act as target areas for switching tracking off when reached. When the reached “end zone” is not applicable to the

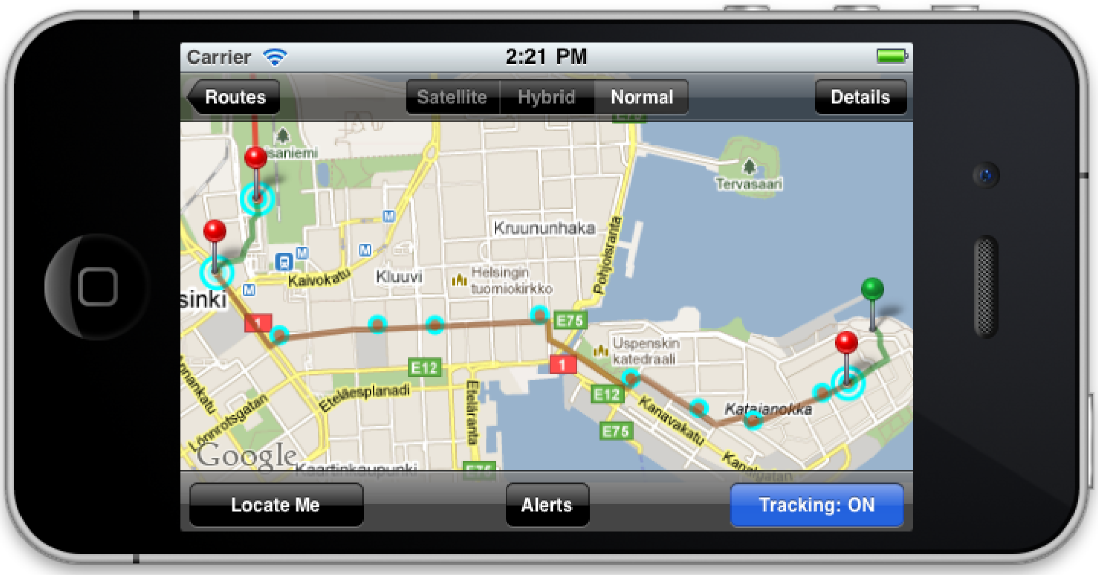

Figure 8: White circles depict the areas at the end of each leg where tracking is deactivated if the departure time for that leg has passed.

end of the entire route, a new timer is created which, again, activates the tracking when appropriate. The paths the Reittiopas API will provide guidance for go along known roads and footpaths very accurately. The stops for different vehicles, however, do not always have precisely accurate coordinates. Most of the time the difference is insignificant, but in some places it can be tens of meters. To comply more with the visual location of the stop on the map and the fact that the stop is, generally, a little off the path instead of in the middle of the road, the stop coordinates are used as another end zone with a similar 50 meter radius.

Once the user is located at an end zone, that leg of the route is ignored, along with all previous ones, when doing further locating. Tracking only the legs for which the departure time has passed and ignoring the ones for which the target area has been reached makes it possible for the application to evaluate which path is actually the one the user is interested in even if different paths are close by or crossing each other. Figure 9 shows such a situation along with a list of the legs the route consists of. If the locating does not update while the user is at the end zone or if, for example, the user decides to go another way than suggested, the application has no way of automatically knowing that the user has reached the end of a leg. The list of the legs of a route can, therefore, be used manually to let the application know where the user is at the time. Alternatively, picking a leg from the list can also be used to select the leg where the tracking is started if the user does not want the tracking to

Figure 9: Left: The paths of different legs of the route crossing. Right: A list of the legs of the route. Check mark indicates the currently selected leg. Tracking status is 'Timed' instead of 'ON' because the departure time for the chosen leg has not yet passed.

be on for the earlier parts of the route. The tracking is automatically deactivated 20 minutes after the scheduled arrival time at the target destination to ensure that the application does not keep on wasting the battery pointlessly even if the user leaves the application running without ever reaching the final end zone or without turning off the tracking manually.

4.5 Sound Design

This section presents the kind of auditory cues that were used for different events and situations as well as the reasoning behind the decisions. The selection and presentation of sounds has been made taking into consideration the balance between the provision of user information and user tolerance to various sounds. Being able to listen to music with headphones without interruptions and still being able to hear the relevant notifications was also considered an important quality for the application. Because some information is more important than other, varying solutions need to be applied. Less urgent information can be blended into the music to a hardly noticeable level, whereas for more essential notifications the music level is lowered for the duration of the cue to ensure full reception. The user is also given the possibility to customize the application by completely disabling certain sound elements that are not considered useful. With all the sounds turned off, the application can still be used as a fully functional graphical journey planner. For some of the navigational cues, the user can also choose to have them only enabled while headphones are plugged in.

In this section, the various sounds that were used are discussed as separate categories. First, the direct auditory feedback for user actions are looked into. Then, the cues used when receiving data from the Reittiopas API are covered. Next, the different scheduled alerts are explained. And finally, the navigational cues for guiding the user along a chosen route are described.

4.5.1 Interface Feedback

The interface feedback sounds are the direct responses the user gets for performing an action. They act as a confirmation and reassurance that the application is responding to the user's input appropriately and alleviates the need for visual inspection to make sure the input was correct, as earlier studies have found effective [6, 7]. Table 3 describes the direct auditory feedback used in the application.

Table 3: Sounds used in category *Interface Feedback*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks.

Interface Feedback								
	Event	**Visual Output**	**Auditory Output**	**AI**	**EC**	**C**	**V**	**D**
I-1	Selecting items on a list	Checkmark added	A very short click.		x			
I-2	Unselecting an item on a list	Checkmark removed	A short brushing or scraping sound.		x			
I-3	Scrolling a list of locations	List of Locations	A short click when displaying a new location. A 16 semitone pitch difference between the clicks used for origin and target lists.		x			
I-4	Reset	Original status	Fading sound of muted guitar strings being hit.		x			

Besides the events listed in the table, the application applies the device settings for using the natural clicking sounds for iOS elements like typing on the virtual keyboard or picking a time by scrolling a wheel. The importance of hearing them is rather insignificant and, as these are used as immediate feedback, the user is likely to be observant when the sounds are played. Therefore, simultaneously played music is not altered in any way.

4.5.2 Searching for Locations and Routes

As the application retrieves information for addresses, schedules, and routes from the internet, the response for them is delayed and, therefore, considered a separate category from the interface feedback. When the user types in an address, or the name of a place, the Reittiopas API is queried for a reverse geocoding to get the corresponding coordinates. When the reply from the API is received, the application can react in a few different ways (cf. Table 4). If the query returned only one hit for the address and a query is still being performed on the other location (origin or target), the application does not give out any sound. When the other search finishes, and also finds a single match for its address, the application automatically starts a route search without playing any sound. If, however, more than one location is found, the application will inform the user by playing a clicking sound in rapid successsion as many times (though limited to 40) as there were locations found. The clicks are the same sound as used for interface feedback when scrolling up and down a list of the respective locations. The sound serves to notify the user that the query resulted several matching locations and further attention is required to pick the desired one. Distinguishing an accurate assessment of the clicks is neither plausible nor intended. It gives the user a broad assessment of whether they are likely to have to browse through a long list of locations (in which case a new search with a more complete address might be appropriate) or just choosing from a couple of options, such as the same address in a couple different cities.

Table 4: Sounds used in category *Location Search Finished*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks.

Location Search Finished								
	Event	**Visual Output**	**Auditory Output**	**AI**	**EC**	**C**	**V**	**D**
L-1	Locations found	Amount of locations	A short click for each found location (limited to 40), played in rapid succession. A 16 semitone pitch difference between the clicks used for origin and target locations.		x			
L-2	Scrolling a list of locations	Addresses displayed	A very short click for each new location appearing on screen.		x			
L-3	No locations found	Text	A tense chord representing an error.		x			

Once the origin and target coordinates are set, the search for the route can begin. If the user looks for a route based on a chosen arrival time, there is a chance that all the departure times for the found routes have already passed. If none of the retrieved routes' departure times are no longer valid, the application will notify the user with a short, tense chord representing an error. Typically the departure time will either be immediately or in the future. If a found route has a departure time within one minute, a distinct chord is played to catch the user's attention to make sure he knows to make a decision to leave right away or to take a later route. When there is more than one minute until the departure, the application will notify the user by first playing a cue representing the time to wait. If the time is longer than ten hours, a sound representing an error is played so that the user takes notice to check whether the found search was as intended. The ten hour limit is somewhat arbitrarily chosen, although it is intended to limit only very unusual route searches off. Close to ten hours until departure may, for some users, be a common time to search for the route if, for example, thinking ahead and, in the morning, finding a route to take after work. The type of the cue depends on how long the wait is. Table 5 describes the alternatives and Figure 10 further illustrates two examples of the varying structures of different time cues. If the route search was performed by arrival time and automatic route selection is selected by the user, the route the cue will be for is the one with the latest departure time. Not only does the cue let the user know how long until departure time but it also acts as a notification that the search has finished and the user can have a look to get more information about the route without having to keep glancing at the device to see if the information is available yet.

Table 5: Sounds used in category *Time Until Event*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks.

Time Until Event								
	Event	Visual Output	Auditory Output	AI	EC	C	V	D
T-1	0 - 9 minutes	-	The sound of a clock ticking, one tick for each minute. As a response to user action: 0.2 seconds between ticks. As a scheduled alert: 0.5 seconds between ticks.	x		x		x
T-2	10 - 59 minutes	-	A set of ten very rapid ticks played for each ten minutes (0.5 seconds between sets).		x	x		x
T-3	1 - 10 hours	-	Sound of a church bell being hit for each hour (0.65 seconds for each).	x		x		x
T-4	> 10 hours	-	A tense chord representing an error.		x			x

Further information about the found route is also available by sound, as listed in Table 6. After the time cue for when to depart, the user gets a broad overview of the weather at the target city, which is automatically fetched (cf. Section 4.1.2) once the route information is received. The read weather data consists merely of a simple text description from a predefined list that tells the general observation or forecast. The weather status can be, e.g., 'Rain', 'Thunderstorm', 'Windy', 'Mist', or 'Sunny'. The application attempts to provide an easily recognizable cue for the usually relevant conditions. As the weather data is dealt with as a limited amount of labels, there is not a sufficient amount of detail to parameterize the audio cues. Instead, constant

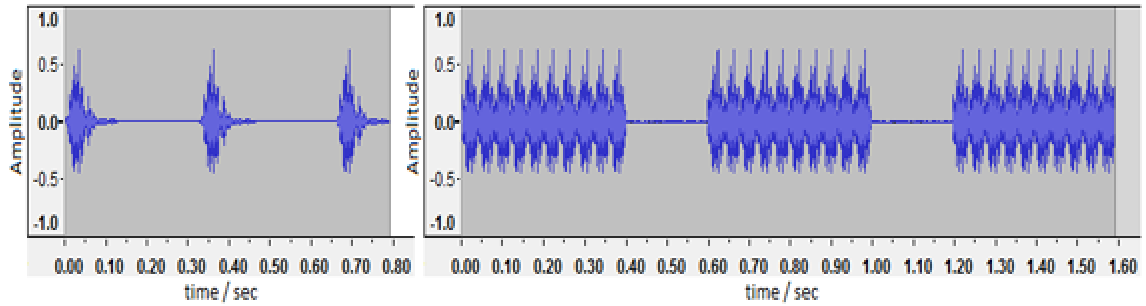

Figure 10: Left: Waveform of a cue for three minutes. Right: Waveform of a cue for 30–39 minutes. Both consist of the same ticks of a clock concatenated.

Table 6: Sounds used in category *Route Search Finished*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks.

Route Search Finished								
	Event	Visual Output	Auditory Output	AI	EC	C	V	D
R-1	Time until departure	Route information	Cue for time until event (cf. table 5). If the departure time 'now', a short chord is played.	x		x		x
R-2	Broad overview of the weather in the target city	-	E.g., birds chirping, rain drops, heavy rain, thunder, or wind.	x				x
R-3	Description of the route (if a route is automatically selected)	Route information	Cues for different vehicles passing by played in succession according to the modes of transportation used on the route.	x		x		x
R-4	No routes found or all departure times have passed	Text	A tense chord representing an error.		x			x

auditory icons are used. The differentiation between details is, therefore, small. For example, the sound of raindrops falling on water are used for weather conditions with labels ‘Chance of Rain’, ‘Light rain’, ‘Drizzle’, ‘Showers’, ‘Chance of Showers’, and ‘Scattered Showers’, while a sound of heavy rain is used for conditions ‘Rain’ and ‘Rain and Snow’. Providing accurate weather data is not the goal here, but merely to give the user a hint about the weather at the target city to help in deciding whether to take a jacket because it is windy, an umbrella because it might rain, or if playing golf is not an option because of thunder. The less common weather conditions, such as ‘Mist’, ‘Dust’, or ‘Fog’, are ignored in the application and a cue consisting of birds chirping, that is used for ‘good weather’ conditions, is played instead. Finding an easily recognizable sound for some conditions is difficult and if the condition is rare, the user would hear it very seldomly, which would make learning and remembering it significantly more difficult, as previous studies have shown [1, 14, 31]. Even for some metaphors that may seem easy to remember, the sound was considered inappropriate because it might be too annoying (e.g., the sound of a foghorn for representing foggy weather might be considered too unpleasant).

If the route description sounds are turned on, a set of sounds describing the different vehicles used along the route are played in the corresponding order. The sounds used to represent the different legs of the route are auditory icons of the used vehicles passing by (or a pair of footsteps on a wooden floor for walking). As Kainulainen et al. concluded in their study (cf. Section 3.1), including relevant sound objects, such as doors opening or closing, may help in recognizing between similar transportations, the sounds used to describe different vehicles in this application have been selected to be as distinctive as possible from recordings of the real vehicles travelling. For example, the sound of a train passing by has a rather distinctive sound of the wheels hitting and screeching against the railroad tracks, and the sound for metro has a reverberant character to it, whereas a tram has a more clear sound to it with some clanging noises that can commonly be heard while travelling by it. The cue for the Suomenlinna ferry passing by consists of the sound of a motorboat splashing through waves and faint cries of seagulls. While Kainulainen et al. attempted to separate the sounds in the application from the sounds in the real environment by increasing the tempo and raising the pitch, the sounds used for this application are played at the real tempo and pitch with only the amplitude modified to allow the sound to fade in and out instead of starting and stopping abruptly. Kainulainen et al. were worried that if the sound was played without any modification, it would be confusing as it might be indistinguishable from the environment. Their application, however, used the duration of the sound to describe the length of the leg by looping a second long sound over and over again (between very short sounds of acceleration and deceleration), whereas in this application the cue is only played once. The length of these cues are 1.7–3.9 seconds, except for walking, which is shorter as it is just two footsteps. The longer sounds provide more details along with the more comprehensive character, which should make recognizing them significantly easier. Particularly the rail-based vehicles tend to require longer samples to distinguish them from one another, while the bus, which is very different from the other transportation types, can more easily be represented by a short sound of a car passing by. The chance of mixing the cues for real sounds is still possible, but the surroundings of the user should often provide a basis on whether the sound is realistic for the location and if the corresponding vehicle did, in fact, just pass by.

The route description cues do not include very detailed information and, as it is implemented here, obviously would not be sufficient for a user that is either visually impaired or unfamiliar with the route choices. The sounds of the vehicles passing by provide no information about the line numbers, so a tourist would have little chance of knowing which bus to take, for example. It can, however, be of value for someone who often travels between same places. Consider having three different route options for how to get home from work depending on the time of departure. For example, the three different combinations the available route choices might consist of could be the following:

- Walk – Bus – Walk – Bus – Walk

- Walk – Bus – Bus – Walk
- Walk – Bus – Walk

Hearing the corresponding cues, the user could instantly know which route is in question and be able to make decisions based on his preference and being aware of the usual alternatives. If the user is familiar with the routes and is mainly just interested in knowing the time, the route description can function as a confirmation that the route is similar to the expected.

The weather and route description are played independently after the time information. As the weather data is loaded from a server on the internet, it can have some delay, but typically it is played simultaneously with the route description. Hearing the auditory icons for the weather and the auditory icons for the different vehicles is a fairly natural mix – if you are sitting at a bus stop, you hear both the weather and the traffic, after all. Hearing two familiar sounds simultaneously often does not make a difference in their recognition rates [40]. In some cases problems may occur if the two auditory icons happen to mix together particularly badly, as might be the case with the cue for a windy weather and a metro passing by.

4.5.3 Alerts

When a route search is finished, the application can automatically create various departure notifications (listed in Table 7) according to user preferences. A time between one and ten minutes can be defined to get a notification reminding to prepare for leaving soon, which is presented by concatenated auditory icons providing the informationa about time, as described in Table 5, followed by an auditory icon representing the leaving of the first mode of transportation of the route. When the time of departure is at hand, another auditory icon is played to inform the user that it is time to leave. The cues are designed to provide not only the information that the user should leave but also the mode of transportation. This alleviates confusion if during a journey the user is, for example, going to the shop and hears a cue. The user needs to know whether it meant that he should start walking to the bus stop or if the bus was scheduled to leave. Descriptive and easily recognizable sounds are used where appropriate to conform to suggestions from other studies (cf. Section 3.2) to avoid ambiguity.

The cue for a vehicle leaving after a period of time is different from the cue for immediate departure as the former suggests there is still some time until departure, whereas the latter indicates that the time of departure just passed. The sounds for a vehicle leaving are, for example, an engine starting to represent a bus, the steam engine of a locomotive speeding up, and the rising sound of a metro accelerating. The first two of these examples differ greatly from the recordings of the actual transportations that Kainulainen et al. used (cf. Section 3.1) by being more caricaturistic to be more recognizable and iconic, as they suggested. The sound of a locomotive, for example, is not a realistic sound to be heard with modern trains but the sound is very distinct, just as the sound of an engine starting, for which the association with a bus just has to be learned. The cues for immediate departure are represented by the sound of the vehicle in question passing by (except for walking, which is represented by two footsteps). This acts as a metaphor for missing the departure if the user does not leave right away – hearing the sound of a car passing by indicates that if the user was not at the bus stop, the bus would have been missed if it was on schedule.

If the user has not selected a route (either automatically or manually), the auditory icon for 'depart now' is followed by a cue for how long it is until the departure of the next route (cf. Table 5), so as to help the user in deciding whether to leave right away or take the next scheduled service. When a route is selected, the alerts for other routes are removed and alerts, similar to those of route departure, are added for each leg of the route according to the scheduled departure times. These cues for leg departures help maintain the feel of being in control of the journey. Hearing the notifications that, for example, in three minutes the next connection is leaving, can either give the user the reassurance that he is on schedule or let him know that he should consider other options if it looks unlikely to connect with the service.

Table 7: Sounds used in category *Alerts*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks.

Alerts								
	Event	**Visual Output**	**Auditory Output**	**AI**	**EC**	**C**	**V**	**D**
A-1	Departure: Now	List of alerts	Two footsteps if walking, otherwise the relevant vehicle passing by. If a route isn't selected, cue for time until next departure (cf. table 5) played two seconds later.	x			x	x
A-2	Departure: For route, in x minutes	List of alerts	Cue for time until event (cf. table 5), followed by the relevant vehicle leaving (e.g., engine starting for bus, metro accelerating).	x		x		x
A-3	Departure: For leg, in y minutes	List of alerts	Cue for time until event (cf. table 5), followed by the relevant vehicle leaving (e.g., engine starting for bus, metro accelerating).	x		x		x
A-4	Wake up alert for night time (00:00 - 06:59)	List of alerts	A short, loud, dissonant chord played three, two, and one minutes before transportation is scheduled to arrive at destination.		x		x	x
A-5	Manually added alert	List of alerts	A clean tone.		x		x	x

After having searched for routes, the user can get information about the next departure time by shaking the device (cf. Table 8). When displaying the list of all the found routes or when none of the routes are chosen, the shake gesture will prompt a cue for how long it is until the next scheduled route departure time as described in Table 5. Otherwise, if a route has been selected, the cue will be similar but for the next scheduled departure of a leg and followed by an auditory icon for the mode of transportation for that leg. This allows the user to check how much time he has left without having to look at the device. If the device is on silent and there is less than ten minutes until the departure, the response is given by the device vibrating as many times as there are minutes left. This can work as a discreet and polite way of checking up on your scheduled leaving while, for example, keeping eye contact during video conferencing.

Table 8: Sounds used in category *Feedback for Shake Gesture*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks.

Feedback for Shake Gesture								
	Event	**Visual Output**	**Auditory Output**	**AI**	**EC**	**C**	**V**	**D**
S-1	While a route isn't selected or the list of the routes is being displayed	-	Cue for time until the departure of the next route (cf. table 5).	x		x		x
S-2	While a route is selected and not showing list of found routes	-	Cue for the departure of the next scheduled leg (cf. A-3 in table 7).	x		x		x
S-3	All departure times have passed	-	A tense chord representing an error.		x			x

4.5.4 Navigation

The application will deliver the user information based on time, location, and the schedule of the route (cf. Section 4.4.4) while tracking is active (cf. Section 4.4.5). Table 9 provides descriptions of the navigational cues.

Unless the starting point is a stop or a station, the first part of a route is typically walking. If the nearest part of the route is a walking leg and the user is within 500 meters of that path, information will be provided in the form of footstep sounds. The sound of the footsteps include information about staying on schedule, distance to target, and whether the current location is on the path or astray. As these cues need to be modified to provide the detailed information, OpenAL is used.

The footsteps are played when the location services update to a new location, though with a minimum of 5–60 seconds between starting to play a new set so that the user does not need to listen to them all the time and get annoyed. If the previous cue is still playing as a new location update occurs, the new cue will be ignored. The amount of footsteps represents the distance along the path to the end of the leg (50 meters per step). The used sound file is a pair of footsteps (duration 1.30 seconds), which is then looped for an appropriate time (0.65 seconds for each 50 meters). A pair of footsteps is used instead of a single step as it is significantly easier to associate it with footsteps when the two sounds differ a bit. Repeating one step, as Kainulainen et al. did [19] can be difficult to interpret as a footstep rather than any other tapping

Table 9: Sounds used in category *Navigation*. AI = Auditory Icon, EC = Earcon, C = Concatenated, V = Vibration, D = Music Volume Ducks. *Depends on context.

Navigation								
	Event	**Visual Output**	**Auditory Output**	**AI**	**EC**	**C**	**V**	**D**
N-1	Walking	Location on a map	Footsteps paced to suggested walking speed, panned towards path, amount describing distance.	x		x		x*
N-2	Walking, late	Location on a map	Short, distorted tone. Amount of repeats describing distance.		x	x		x
N-3	Reaching target	Tracking status change	Melody (like a military bugle call), followed by cue for next leg departure (cf. A-3 in table 7).		x			x
N-4	Approaching stop (< 1 km) after which user should press the stop button	Location on a map	Two notes, a semitone apart, played alternately three times each with tempo increasing as distance decreases - similar to the music in the movie *Jaws* as a shark approaches.		x	x		
N-5	User should get off at next stop	-	A tone similar to the those heard in vehicles when pressing the stop button.		x		x	x
N-6	Straying from bicycle path	Location on a map	Increasingly tense sound when further away.		x		x	x

or hitting sound. Again, to avoid being annoying and feeling endless, the sound level of the footsteps are lowered by 0.1 once every second starting from full 1.0. The sound will, therefore, fade to silence within ten seconds. The distance information will, then, only be provided fully for relatively short distances, which quite often is the case in urban settings. Ten seconds will be enough time for playing footsteps representing 750 meters at the regular pace. Longer distances, for which there is not enough time to play all the footsteps, will simply give the user the distance as ‘at least’.

The lack of accurate distance information is not an issue for determining whether the user is on schedule or not because that information, as well, is embedded in the footstep sounds. The application has two possible settings for the user’s stride length, representing the averages for males and females. The stride length, together with the calculated distance along the path and the time until scheduled arrival to the end of the leg, is used to play the footsteps at the actual pace the user should walk in order to arrive at the correct time. As the user is walking, he can easily compare the rhythm of his walk to the footsteps he hears to determine if he is late and should walk faster.

The tempo of the footsteps is changed by multiplying the pitch with a corresponding factor. The time it takes to play the sound is inversely proportionate to the pitch (i.e., doubling the pitch halves the temporal length of the sound). Being ahead of schedule will, then, result in both lower pace and lower pitch which makes the sound seem quieter and easier to blend in to currently playing music, and, as the level is lowered every second, fade-out occurs in less footsteps. Also, the application is set so that if the pitch factor is less than 0.80 (the user is significantly ahead of schedule), currently playing music level is not dropped for playing the footsteps. These three factors all aim at making less important information less prominent so that, again, the user would not be annoyed by the sounds from the application. If the user is behind schedule, the importance of getting the information is greater and having the music level lowered momentarily to play a higher pitched, faster set of footsteps is not as likely to annoy the user as the information can be necessary in order to be on time.

Using the source positioning of OpenAL, the footsteps are played a different direction depending on the user's location relative to the path he should be walking. The application is designed to work on a device without a compass and with locating capabilities with lesser accuracy and update rate than optimal. Because of this, the application does not get information about the user's heading. It assumes the user is aware of the general direction he should be heading and based on this determines which side of the user the path lies. The sound source is then placed between the left and right audio channels depending on the distance and side. When the path is 100 meters or more to the left of the user's location, the sound is panned completely to the left. If the path is 50 meters away on the right side of the user, the sound is positioned to the right of the user at a 45-degree angle, and so on. When the user is on the path, the sound is centered, indicating correct location.

Figure 11: Direction of the sound of footsteps panned towards the path.

Figure 11 illustrates a situation where the user is walking to the correct direction but is informed by the panning of the sound that he should be in the next block. If the user is walking to a direction opposite to where he should be going, the panning, however, will be completely invalid. The pace of the footsteps, however, will be increasing significantly as time passes and distance grows instead of decreasing, which should inform the user that something is wrong.

If the user is still over 150 meters from the target but the scheduled arrival time has passed, a slightly distorted, short tone is used to replace each pair of footsteps (a hundred meters). When the user is further than 500 meters away from the path, no cues are played – it is considered as as erroneous location update or indicative that the user has decided to take another way.

As the user arrives at the end of the leg, a short melody of simple chords is played. If the leg in question was the last of the route, the cue is a slightly prolonged version of the same. The first bar in Figure 12 for the former, both for the latter. If it was not the last travel leg, the cue is followed by information about when the departure from that point is, and by what mode of transportation. If the departure time from that point has already passed, a tense chord representing an error replaces the cue for time and vehicle.

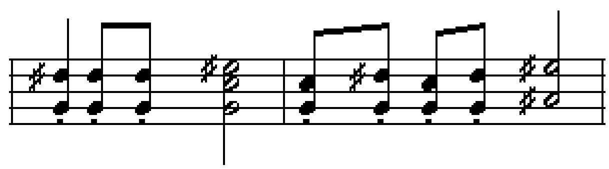

Figure 12: The melody for reaching the end of a leg (first bar) or the end of a route (both bars).

When the user is travelling in a vehicle (other than the Suomenlinna ferry, which only travels between two points), after passing the stop prior to the destination, the application will play a tone similar to those heard in buses when pressing the stop button, to inform the user that he should exit at the next stop. The minimum repeat rate for this cue can be set to 10–60 seconds or to only be played once – so it can be repeated in case the user missed the initial cue but not too often to become annoying.

The application can also be used for cycling and walking routes. The approach for the auditory cues in this case is to avoid the use of sound when everything is going as planned so as to not disturb the user needlessly. In practice this means that while the user is within a hundred meters of the path, no cues are played, except for when reaching target areas the user can define by marking locations on a map. When the user strays further away from the path, a part of a sound file with four tense bending sounds are played consecutively at a rising pitch. The further away from the path the user is, the more of the sounds are played. To allow the user to get back on track without getting annoyed by having to listen to the notification continuously, the repeat rate is limited. As with the footstep cues used for tracking walking when using public transportation, the sound is panned towards the path.

5 Results

The recognition rate and intuitiveness of the used auditory icons were evaluated informally by eight test participants by having them listen to sound examples on a computer. All the participants of the recognition testing were between the age of 25 and 35, none of whom have experience in sonification. All the participants were familiar with using the Reittiopas service and three of them regularly use a mobile application for it. The purpose of the application and the basic concept behind some of the sounds were explained (i.e., vehicle or traffic related sounds refer to the mode of transportation and after a route search is finished, one cue refers to the weather).

The participants were first given an example of a simple route description and the elements explained. The example route consisted of a cue for seven minutes, walking, bus, and more walking. Most of the listeners recognized the sound of footsteps correctly but a few did not understand it. After being reminded that they were to identify the modes of transportation along the route, it was, generally, evaluated correctly. Only one participant was adamant that it sounded like a hammer in use. After explaining it was, in fact, supposed to be the sound of footsteps on a wooden floor to describe a walking part of the route, the participant stated that after becoming aware of this, it was expected there would be no confusion from this point. The example description also included the sound of birds chirping describing the weather. All the listeners understood this indicated good weather either immediately or as soon as they were told it represented the weather. They were then asked to listen to five route description cues listed in Table 10.

Table 10: Route descriptions listened to by test participants.

#	**Time**	**Weather**	**Leg 1**	**Leg 2**	**Leg 3**	**Leg 4**	**Leg 5**
1	4 min	Thunder	Walk	Bus	Walk	Train	Walk
2	7 min	Good weather	Walk	Tram	Metro	Walk	Bus
3	20 min	Heavy rain	Metro	Walk	Bus	Ferry	-
4	2 h	Windy	Bus	Metro	Walk	Ferry	Walk
5	3 min	Light rain	Bus	Train	Tram	Tram	Metro

The participants were told what the possible transportation types were, but not the weather conditions. They were asked to give their impression of the time until departure, the weather, and what the parts of the route were after listening to the concatenated auditory icons describing a route once. If the answer was not correct, the participant could listen to it again (as many times as wanted) and then provide a new answer. All six different transportation types were presented in the first three of routes and listeners were told the correct answer before proceeding to the next route. Thus, the users were familiar with all the transportation types but not the weather conditions before the last two route descriptions.

The participants recognized the auditory icons for walking, bus, train, and ferry almost always but mixed tram and metro more than half the time at first listen. After listening to the cue again, they usually did notice the hollow, reverberant sound as a metro. The participants listened to a different auditory icon, used as an alert for an upcoming departure, for the metro. While the sound used in the route description was representative of a metro passing by, the other cue was the sound of an accelerating metro. The accelerating metro was recognized as very distinct and unmistakable for other vehicles.

All the vehicle types (excluding walking) had two auditory icons related to them; one describing the approaching departure time (sound of the vehicle starting or accelerating after a time cue) and one for representing the departure time passing (sound of the vehicle passing by the user). Whether the sounds should have been the other way around was discussed and the answers were inconclusive – both sounds were considered logical in their own way, and many of the participants had trouble deciding which made more sense to them. This illustrates the difficulty of coming up with metaphors that are widely understandable by intuition. Deciding what would seem logical to many people can be very challenging as the interpretation of just one person can vary based on the way he approaches the metaphor. On the one hand, the sound of an engine starting seems logical for indicating immediate departure, but on the other hand, the sound of a vehicle passing by conveys urgency better. Whether there is a need for two separate cues for immediate and upcoming departure is questionable. It would be more efficient, in terms of easier learning and having only better cues, to only have one sound for each vehicle – the one that is most iconic. It might, however, be misleading if the user misses the part of the cue indicating the time until departure. This could, of course, be fixed by having a time cue representing ‘now’ for the immediately departure. Then the user would need to pay more attention to the time information as it would provide the (usually) most important detail.

The example description played for the listeners before the tests had a time cue for departure in seven minutes, described by seven ticks of a clock. This having been explained to them, they had no trouble recognizing the times for the first, second, and fifth routes. They had not been told about the different ways of representing longer periods of time so that the intuitiveness could be tested. The cue for 20 minutes (two sets of ten rapid ticks) and the cue for two hours (two rings of a church bell) were not understood at all. The cue for 20 minutes was, generally, considered fairly logical and evaluated as easy to remember after it was explained. Reactions such as “oh, of course!” followed the explanation of the cue for hours.

The weather cues were all recognized apart from ‘windy’ in route four. None of the participants recognized it correctly. The guesses included, e.g., ‘winter’, ‘snow storm’, and ‘storm’. The weather cue was intentionally played simultaneously with the cue for the metro to test an assumed bad combination. The reverberant sound of the metro was not recognized with the sound of the wind blowing either – though the sound of the metro was mixed with the tram often anyway even without simultaneous weather cues interfering. The participants did consider having the weather cue played during the route description as a surprisingly good and pleasant solution. It was stated, however, that for making sure that all the cues were clearly understood, the cue could be played afterwards, where it might seem more logical as well. As one participant said, “you would first hear when you should go, then how to get to where you are going, and finally what the weather is like over there.”

A couple of the participants were happily surprised by how much information they can easily obtain from a short non-speech set of sounds. The participants all agreed that, even though sometimes it may be difficult to interpret (or remember) all the cues in the route description, it is likely to be very effective as a confirmation that the found route was like the one the user was expecting. A couple of the listeners also believed that if they were first told real, familiar locations where the route begins and ends, the chances of recognizing even the vehicles on rails would increase significantly by having some foreknowledge of what the realistic choices were. The used route description sounds did not directly refer to any real route but completely unrealistic options were avoided (e.g., a metro or a train could not be on both sides of a ferry because the ferry goes to a small island).

As the users were able to recognize the different vehicle cues well (especially if the cue for the metro was replaced by the accelerating one), it should be very easy to recognize them when they are played individually to notify a departure.

Comments were requested for the overall usefulness of the auditory icons and the earcons that are used in the application. The recognition of the earcons was not tested as, by definition, they tend to require practice. Knowing the context in which the sounds would be used, the listeners were asked to focus on whether they felt that the use of the earcons would be helpful in the situations they would be played, as well as the likelihood of them sounding too annoying. One person seemed very skeptic of whether any of the sounds would provide understandable enough information, though the respondent did seem rather stuck in the misconception that the sounds would be used as a complete guidance system for the blind, which was the first impression after the first route description sound was explained. The idea that when one travel leg is completed users are immediately informed about the time to the commencement of the next travel leg was found potentially useful. It can easily help the user evaluate whether there might be enough time to undertake incidental activities, e.g., go to the store in the meanwhile – especially in the winter when accessing the device is more cumbersome.

The common opinion of the test listeners was that the application should come with at least a small tutorial explaining the use of the sounds. It was stated that it may be difficult to learn to use the application because people are not used to audio being used to deliver information. Receiving information by sound is, perhaps, not as unfamiliar as the participants initially thought, however. The lack of other devices and applications that sonify information seems likely to affect the difficulty of placing a meaning to the heard sound where people typically do not expect any further meaning. Learning the meaning of sounds is, after all, a matter of hearing them in the right context. After the association is built, distinguishing, for example, the microwave oven, telephone, and doorbell from one another is easy enough.

The potential for using audio while riding a bicycle was acknowledged but whether any sound can be heard without headphones was questioned. The usefulness of a notification increasing in intensity when straying further from the bicycle path was mostly doubted at first, however, after considering for a while the opinions changed towards positive. At first, the general opinion was that the user would know when he is lost without getting a notification or that getting the cues when deciding to take a slightly different path would be annoying. After considering the fact that the repeat rate of the cue is limited instead of constantly playing and that the case in which someone would use tracking for a bike path would be a situation where he does not know how to get somewhere. In this case the user would not choose a different route by impulse decisions.

One of the participants, unaware of how the application functions, stated that even though the application seems like it could be useful in using public transportation, they would not use it, just as they do not use other similar applications currently. This is because the participant wished to be able to use their mobile phone for speaking and felt that the tracking drains the battery too fast. After being described how this application handles turning the tracking on and off, the participant was pleasantly surprised and said that this might change their stance. Most of the participants considered the auditory cues useful in mobile contexts.

6 Conclusions and Further Research

6.1 Conclusions

Route planning and guidance are typically carried out by providing textual and graphical information. In case of mobile devices, auditory and vibrotactile alarms are also added sometimes. The alarm usually indicates when the user should begin the journey or exit a vehicle. For a desktop solution where the screen real estate is not significantly limited and the user is likely to either be looking at it or able to glance at it with little effort, the graphical solutions work well. When using a mobile device, however, the practicality of accessing the screen and accommodating all relevant information within a limited display space can become an issue, for which audio can provide relief.

In this work, a mobile journey planning application was developed for testing the concept of using auditory cues to assist the user in a natural way. The aim of the application is to reduce the amount of effort required by the user in order to stay aware of the time and the progression of the journey. By the use of easy-to-learn, recognizable auditory icons and more abstract earcons, the user is provided meaningful information based on time, location, and the schedule of the chosen route.

In an informal setting, a set of eight participants took part in a test and listened to some of the used auditory icons to determine the intuitiveness and, after learning the association, the identification of them. Overall, the auditory icons were considered rather distinct and iconic, with some exceptions (recognizing the sound of a tram was particularly difficult, as was expected). Majority considered that at least some of the cues would be either useful or very useful for themselves.

6.2 Further Research

The next major step in the further development of a mobile application using sound to assist the use of public transportation would be to consider the needs of the visually impaired. Hopefully the real time position of all the vehicles used in public transportation will be available in the near future. Using the actual locations of the vehicles, more accurate locating of the user, and other instruments such as a compass, the detail and reliability of the information could be improved greatly.

Providing sufficient information for the visually impaired would require further details, such as line numbers, delivered as well. Another important, challenging part would be a practical and accurate means of giving input to the device. Touch screen devices lack the physical features that distinguish between different buttons so using the application without vision would either require accurate auditory feedback, recognition of learned swiping or shaking gestures, or, without recourse to the use of the screen, speech recognition capability.

References

[1] Garzonis, S., Jones, S., Jay, T. & O'Neill, E. Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. In *Proceedings of the CHI '09,* pp. 1513–1522. Boston, Massachusetts, USA, April 2009.

[2] Hermann, T. Taxonomy and Definitions for Sonification and Auditory Display. In *Proceedings of the ICAD 2008,* Paris, France, June 2008.

[3] McGookin, D.& K., Brewster, S. A. Understanding Concurrent Earcons: Applying Auditory Scene Analysis Principles to Concurrent Earcon Recognition. *ACM Transactions on Applied Perception,* vol. 1, Issue 2, pp. 130–155, October 2004.

[4] Brewster, S. A. Non-Speech Auditory Output. In Jacko, J. A. & Sears, A. *Human–Computer Interaction Handbook,* Ch. 12, pp. 220–239. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, 2002.

[5] Brown, M. L., Newsome, S. L. & Glinert, E. P. An Experiment into the Use of Auditory Cues to Reduce Visual Workload. *ACM SIGCHI Bulletin,* vol. 20, Issue SI, pp. 339–346, 1989.

[6] Bussemakers, M. P. & de Haan, A. When it Sounds Like a Duck and it Looks Like a Dog... Auditory icons vs. Earcons in Multimedia Environments. In *Proceedings of the ICAD 2000,* pp. 184–189, April 2000.

[7] Rath, M. On the Relevance of Auditory Feedback for Subjective and Objective Quality of Control in a Balancing Task. In *2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems,* Berlin, Germany, September 2006.

[8] Barrington, L., Lyons, M. J., Diegmann, D. & Abe, S. Ambient Display using Musical Effects. In *Proceedings of the IUI '06,* pp. 372–374. Sydney, Australia, January–February 2006.

[9] Pfleiderer, B., Ostermann, J., Michael, N. & Heindel, W. Visualization of Auditory Habituation by fMRI. In *NeuroImage,* vol. 17, Issue 4, pp. 1705–1710, December 2002.

[10] Nyeste, P. & Wogalter, M. S. On Adding Sound to Quiet Vehicles. In *Proceedings of the HFES '08,* pp. 1747–1750. New York City, New York, USA, September 2008.

[11] Bakker, S., van den Berg, R., Pijnappel, S. & van den Hoven, E. Sounds Like Home: Sonification and Physical Interaction in the Periphery and Center of the Attention. *Proceedings of ISon 2010, 3rd Interactive Sonification Workshop* KTH, Stockholm, Sweden, April 2010.

[12] Walker, A., Brewster, S., McGookin, D. & Ng, A. Diary in the Sky: A Spatial Audio Display for a Mobile Calendar. In *People and Computers XV: Interactions Without Frontiers,* pp. 531–540, 2001.

[13] Gaver, W. W. Auditory Interfaces. In Helander, M. G., Landauer, T. K. & Prabhu, P. V. *Handbook of Human–Computer Interaction,* 2nd Ed. Amsterdam: Elsevier, 1997.

[14] Leung, Y., Smith, S., Parker, S. & Martin, R. Learning and Retention of Auditory Warnings. In *Proceedings of the ICAD 97,* Palo Alto, California, USA, November 1997.

[15] Gärdenfors, D. Designing Sound-Based Computer Games. In *Digital Creativity,* vol. 14, Issue 2, pp. 111–114, June 2003.

[16] Bolt, R. A. "Put-that-there": Voice and Gesture at the Graphics Interface. In *Proceedings of the SIGGRAPH '80,* pp. 262–270. Seattle, Washington, USA, July 1980.

[17] Kramer, G. et al. Sonification Report: Status of the Field and Research Agenda. Prepared for the National Science Foundation by members of the International Community for Auditory Display. 1999.

[18] Baus, J. et al. Auditory Perceptible Landmarks in Mobile Navigation. In *Proceedings of the IUI '07,* pp. 302–304. Hawaii, USA, January 2007.

[19] Kainulainen, A., Turunen, M., Hakulinen, J. & Melto, A. Soundmarks In Spoken Route Guidance. In *Proceedings of the ICAD 2007,* Montréal, Canada, June 2007.

[20] Mynatt, E. D. Designing with Auditory Icons: How Well Do We Identify Auditory Cues? In *Proceedings of the CHI '94,* pp. 269–270. Boston, Massachusetts, USA, April 1994.

[21] Buxton, W., Gaver, W. & Bly, S. Auditory Interfaces: The use of Non-Speech Audio at the Interface. 1994. http://www.billbuxton.com/Audio.TOC.html. Last accessed 13 June 2011.

[22] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. Earcons and Icons: Their Structure and Common Design Principles. *Human–Computer Interaction,* vol. 4, Issue 1, pp. 11–44, 1989.

[23] Brewster, S. A., Wright, P. C., & Edwards, A. D. N. Experimentally Derived Guidelines for the Creation of Earcons. In *Proceedings of BCS-HCI '95,* vol. 2, pp. 155–159. Springer, Huddersfield, UK, August 1995.

[24] Butz, A. & Jung, R. Seamless User Notification in Ambient Soundscapes. In *Proceedings of the IUI '05,* pp. 320–322. San Diego, California, USA, January 2005.

[25] Jung, R. Ambience for Auditory Displays: Embedded Musical Instruments as Peripheral Audio Cues. In *Proceedings of the ICAD 2008,* Paris, France, June 2008.

[26] Azuma, R. T. A Survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments,* vol. 6, Issue 4, pp. 355–385, 1997.

[27] Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J. & Lorho, G. Augmented Reality Audio for Mobile and Wearable Appliances. In *Journal of the Audio Engineering Society,* vol. 52, Issue 6, pp. 618–639, June 2004.

[28] HSL Live. http://transport.wspgroup.fi/hklkartta/. Last accessed 13 June 2011.

[29] Käki, J. Designing Mobile Human–Information Interfaces for Public Transportation Users. Master's thesis, Aalto University, 2010.

[30] Gong, L. & Lai, J. To Mix or Not to Mix Synthetic Speech and Human Speech? Contrasting Impact on Judge-Rated Task Performance versus Self-Rated Performance and Attitudinal Responses. In *International Journal of Speech Technology,* vol. 6, Issue 2, pp. 123–131, April 2003.

[31] Garzonis, S., Bevan, C. & O'Neill, E. Mobile Service Audio Notifications: Intuitive Semantics and Noises. In *Proceedings of the OZCHI '08,* pp. 156–163. Cairns, Australia, December 2008.

[32] Gaver, W. W. The SonicFinder: An Interface That Uses Auditory Icons. In *Human–Computer Interaction,* vol. 4, Issue 1, pp. 67–94. L. Erlbaum Associates Inc. Hillsdale, New Jersey, USA, March 1989.

[33] Edworthy, J. & Hards, R. Learning Auditory Warnings: The Effects of Sound Type, Verbal Labelling and Imagery on the Identification of Alarm Sounds. In *International Journal of Industrial Ergonomics,* vol. 24, Issue 6, pp. 603–618, October 1999.

[34] Reittiopas API. http://developer.reittiopas.fi/. Last accessed 13 June 2011.

[35] Google Weather, unofficial API. http://www.google.com/ig/api?weather=Helsinki. Last accessed 13 June 2011.

[36] JSON-framework. http://code.google.com/p/json-framework/. Last accessed 13 June 2011.

[37] OpenAL. http://connect.creativelabs.com/openal/default.aspx. Last accessed 13 June 2011.

[38] Finch. https://github.com/zoul/Finch/. Last accessed 13 June 2011.

[39] Nikula, J. & Vuorela, A. Passengers' Satisfaction with Public Transport Services in Helsinki in Spring 2010. HSL Publications 6/2011.

[40] Brewster, S. A., Wright, P. C., & Edwards, A. D. N. Parallel Earcons: Reducing the Length of Audio Messages. *International Journal of Human–Computer Studies,* vol. 43, Issue 22, pp. 153–175, 1995.