A music player user interface based on head-gestures and 3D audio feedback

Jonas Hinge

Supervisor: Thomas Olof Pederson

M.Sc. Thesis

IT University of Copenhagen

March 2014

Abstract

Music applications on smartphones makes listening to your favourite music accessible and mobile. Although the possibility of listening to music at any time and place immediately seems like a positive development this could introduce other challenges. E.g. biking and controlling a music application will conflict in the sense that biking demands hands on the handlebars and eyes on the road, and a smartphone application demands hands (or at least one hand) and eyes for navigating resulting in an increase of the users cognitive load.

At the same time emerging accessories with built in sensor hardware e.g. Google Glass or Intelligent Headset (http://intelligentheadset.com/developer/) offer alternate ways of using gestures in form of GPS location, rotation, acceleration, speech etc.

Encouraged by the biking scenario challenge and todays emerging and accessible mobile technology - alternative ways of controlling a music application should be explored. In this project an alternative way of navigating using head gestures and audio feedback is explored.

Results?...

Conclusion?..... .

Acknowledgements

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Contents

List of Figures

List of Tables

Abbreviations

HCI **H**uman **C**omputer **I**nteraction

Chapter 1

Introduction

1.1 HCI in mobile environments

Mobile and wearable devices has been a growing area in computing in recent years. Compaired to desktop computers these devices have introduced new standards for when and how people interact with especially mobile applications. Suddenly people are able to check the news, navigate via interactive maps, post social messages, listen to music, etc., while they are on the move. At the same time emerging hardware in mobile devices and wearable computing expands application complexity and interaction possibilities.

This mobility factor introduces challenges when interacting with these devices. Although screen resolutions and physical sizes of mobile devices are increasing, the visual work space is limited i.e. screens easily becomes cluttered with information and the input keyboard can be an interaction challenge when moving. More importantly, when moving around e.g. in the traffic, interacting with a mobile device at the same time can create cluttering [TODO: rephrase cluttering] in form of distractions e.g. "eyes off the road" or "hands occupied" and in the worst case cause accidents. Motivated by this problem fines are introduced (in Denmark) for people interacting with their mobile device while biking [7].

Solutions for these cluttering [TODO: rephrase cluttering] challenges could lie in the interaction between users and mobile devices. The emerging hardware (e.g. sensor technology) and software opens up for alternative input modalities e.g. head gestures, gaze tracking, speech recognition making hands-free interaction possible. At the same time output modalities such as audio and haptic feedback could liberate the eyes from the screen.

1.2   Problem statement

Considering mobile interaction cluttering [TODO: rephrase cluttering] challenges, this project will be based on the concrete scenario where people are biking while listening to and controlling their music library. As biking requires eyes on the road and hands for steering the input/output modalities should preferrably not include eyes and hands. Instead head gestures for input and 3d audio for output will be evaluated. [TODO: make a numbered list]

More specifically the following questions should be answered:
Can a user interface based on head gestures and 3d audio compete with existing user interfaces for music players (e.g. touch and vision-based) with respect to for instance a) navigation and control efficiency b) learnability, c) general usability (cognitive/perceptive load), c) suitability to real-world hands-occupied situations.

With the chosen combination of input and output modalities, there is a high risk for the system to misinterpret normal everyday actions performed by the user as commands for controlling the system ("behavioural cluttering" (Janlert et al., in press)). How can features in the user interface prevent unwanted manipulation of the system?

1.3   Goal

The goal of this project is to examine if head gesture based input and audio output modalities in combination can compete with a traditional touch and vision based input/output interface and show which advantages, disadvantages and challenges that arise when using and designing such interaction techniques. More precisely a mobile system that recognises these alternative interaction methods should be designed and implemented in a music application. [TODO: Make goals quantifiable if possible]

1.4   Method

Use triangle framework for HCI design [15]

Test users will compaire this new way of controlling a music application with a traditional music application in form of comfort, efficiency and learnability in a closed lab test and a "real world" scenario.

Project structure description (short)...

Chapter 2

Related work

In this chapter specific research areas and systems related to this thesis are presented. First the current situation on mobile music player user interfaces is discussed. Then a wider description of two research areas; mobile HCI and multimodal interaction are presented before going into details about how previous developed interaction techniques within these two areas can be combined with a mobile music player UI. A graphical overview is shown in fig. 2.1.
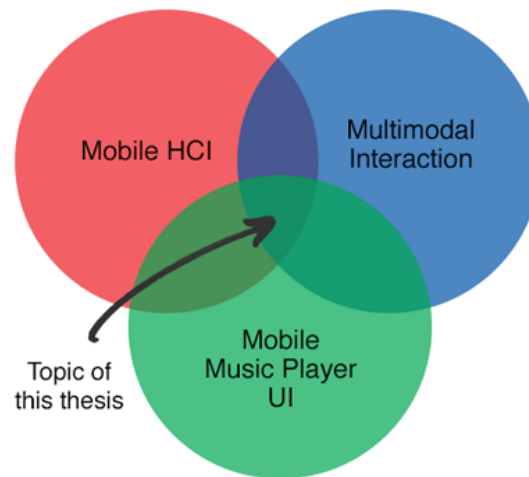

FIGURE 2.1: Thesis topic position

2.1   Mobile Music Player UI

Although mobile audio players have developed since the first portable cassette player by Sony in 1979 towards todays digital audio players with storage capacity e.g. smartphones, there is still one property that remains - the way in which we interact with the device. E.g. like we needed the hands and eyes for rewinding or turning the tape on a cassette player, we need them as well for swiping to the preferred track in our smartphone music application.

- iPod shuffle, no visual interface - leg shifting music track while running [26]

2.2   Mobile Human Computer Interaction

[TODO]

2.2.1   Interaction in motion

- Cognitive load

2.2.2   Mobil device limitations

2.3   Multimodal interaction

[TODO: multimodal interaction in general]

- What kind of modalities, combinations

2.4   Eyes- and handsfree interaction

Several work on both audio [4, 6, 13, 28, 30] and haptic [21, 22] displays use the term eyes-free which refers to controlling the state of a system without visual attention. This kind of interaction has shown to be desirable in some situations [18, 29] and even improve efficiency compaired to traditional visual displays [30].

One of the main motivations behind this eyes-free use is to design interfaces that do not compete with the users visual attention. That is this "visual competition" could introduce risks when people are on the move e.g. travelling in traffic. In these situations

a vital factor is to minimize the amount of distraction for interaction modes [20]. Eyes-free interfaces can keep the users visual attention on the road while driving [27] or walking around in the city [28]. [TODO: rephrase, at this point you should also tell the reader about the complexity of human attention which can to some degree sometimes be decoupled from modalities and that just because information comes in a modality different from the one currently in use doesnt mean that it doesnt distract cognitively]

Much of the interfaces work in wearable computing tends to focus on visual headmounted displays [2] e.g. Google Project Glass. But not only as mentioned does visual displays occupy the users visual attention, they can also be obtrusive and hard to use in bright daylight [10]. Another disadvantage with visual displays is that their power consumption is high i.e. they drain a mobile device battery and they are expensive. By using eyes-free interfaces it is possible to use cheaper and less power consuming hardware.

2.5 Spatial sound

[TODO]

William W. Gaver, a pioneer in audio interfaces, has explored several aspects of using sound in interfaces including the intuitiveness of presenting complex information to users in the form of audio [9]. Similarly Graham explores the advantages in reaction time when using auditory icons [11]. In [8] Gaver presents the use of spatial sound icons. In doing so, he draws forward the unutilized potential of creating natural interaction through spatial audio.

By compairing visual and audio feedback when pushing buttons on the same GUI, Brewster showed that it was difficult for users to devote all their visual attention to an interface while walking, running og driving and that the interaction workload decreased with audio feedback [5].

2.5.1 Non-speech audio

Work has shown that non-speech audio is effective in improving the interaction with mobile devices [23, 25].

2.5.2 Head Related Transfer Function

[TODO]

Good reference for 3d sound [3]

2.6 Head gesture interaction

There exists different kinds of approaches when it comes to controlling a system with head gestures. Using cameras it is possible to effectively track head movements via facial recognition [17] and gaze tracking makes it possible to control an object by fixating the eyes on that object while moving the head [16]. Thus these techniques do not require any hardware sensors e.g. accelerometer and gyroscope but in return a camera placed in front of the user. [TODO: you need to explain this better. our recently graduated phd student diakonhas developed a wearable solution]

2.6.1 Intelligent Headset

[TODO]

2.7 Related systems

[TODO: intro]

Brewster et al. showed that novel interaction techniques based on sound and gesture can significantly improve the usability of a wearable device in particular under "eyes-free" mobile conditions and that head gestures was a successful interaction technique with egocentric sounds the most effective [6].

Park et al. also experimented, using head gesture input and aural output, with 1D and 2D menu interfaces [19].

Kajastila and Lokki has done a user study comparing auditory and visual menus controlled by the same free-hand gestures where the majority of the participants felt that an auditory circular menu was faster than a visual based menu [12].

2.7.1 Systems properties overview

[TODO]

Table compairing properties of related work (and this project) example fig. 2.1 (NOTE: temp, i need some input for this...).

Table 2.1: Related works properties comparison

Related work	Head gesture interaction	Spatial sound feedback	Music application	Accessible hardware
Multimodal eyes-free interaction techniques for wearable devices [6]	+	+	-	-
This project	+	+	+	+

Chapter 3

Design

Plan

Iterative design process:

- An experimental prototype should be developed and tested. 3 types of menus where a test user should perform head gesture interaction to solve small tasks. User is observed and feedback should be given.

- A final prototype should be designed with knowledge from the experimental evaluation. This prototype will go through a "real life" evaluation. It will be evaluated through several days where the user will use the new head gesture based music application and the traditional music application while biking. Again small task could be performed and it could be tested through the users normal use of his/her music application, ending up in a comparison of the traditional vs the new interaction system.

3.1 Interaction model

TODO...

Horizontal 180 degrees head movement, nod/shake...

Human head normally can be rotated about 140 degrees for shaking and 100 degrees for nodding [14].

3.2 Sound design

TODO...

Several studies show that circular auditory menus are the way to go because of horizontally positioned sounds , HRTF, 3D audio...

Chapter 4

Implementation

SDK's, APIs, Processing sensor data...

4.1 Libraries

IHS sdk, (maybe Spotify lib, deezer api,sdk)

4.2 Intelligent Headset

[TODO]

4.3 Motion gesture recognition

DTW - Dynamic Time Warping

Dynamic Time Warping [24]

Accelerometer-based DTW [1]

4.4 iOS application

System design overview...

Chapter 5

Evaluation

...

Chapter 6

Discussion

Other scenarios e.g. visual impaired people, car driving...

Chapter 7

Conclusion

...

Appendix A

An Appendix

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus at pulvinar nisi. Phasellus hendrerit, diam placerat interdum iaculis, mauris justo cursus risus, in viverra purus eros at ligula. Ut metus justo, consequat a tristique posuere, laoreet nec nibh. Etiam et scelerisque mauris. Phasellus vel massa magna. Ut non neque id tortor pharetra bibendum vitae sit amet nisi. Duis nec quam quam, sed euismod justo. Pellentesque eu tellus vitae ante tempus malesuada. Nunc accumsan, quam in congue consequat, lectus lectus dapibus erat, id aliquet urna neque at massa. Nulla facilisi. Morbi ullamcorper eleifend posuere. Donec libero leo, faucibus nec bibendum at, mattis et urna. Proin consectetur, nunc ut imperdiet lobortis, magna neque tincidunt lectus, id iaculis nisi justo id nibh. Pellentesque vel sem in erat vulputate faucibus molestie ut lorem.

Quisque tristique urna in lorem laoreet at laoreet quam congue. Donec dolor turpis, blandit non imperdiet aliquet, blandit et felis. In lorem nisi, pretium sit amet vestibulum sed, tempus et sem. Proin non ante turpis. Nulla imperdiet fringilla convallis. Vivamus vel bibendum nisl. Pellentesque justo lectus, molestie vel luctus sed, lobortis in libero. Nulla facilisi. Aliquam erat volutpat. Suspendisse vitae nunc nunc. Sed aliquet est suscipit sapien rhoncus non adipiscing nibh consequat. Aliquam metus urna, faucibus eu vulputate non, luctus eu justo.

Donec urna leo, vulputate vitae porta eu, vehicula blandit libero. Phasellus eget massa et leo condimentum mollis. Nullam molestie, justo at pellentesque vulputate, sapien velit ornare diam, nec gravida lacus augue non diam. Integer mattis lacus id libero ultrices sit amet mollis neque molestie. Integer ut leo eget mi volutpat congue. Vivamus sodales, turpis id venenatis placerat, tellus purus adipiscing magna, eu aliquam nibh dolor id nibh. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed cursus convallis quam nec vehicula. Sed vulputate neque eget odio fringilla ac sodales urna feugiat.

Phasellus nisi quam, volutpat non ullamcorper eget, congue fringilla leo. Cras et erat et nibh placerat commodo id ornare est. Nulla facilisi. Aenean pulvinar scelerisque eros eget interdum. Nunc pulvinar magna ut felis varius in hendrerit dolor accumsan. Nunc pellentesque magna quis magna bibendum non laoreet erat tincidunt. Nulla facilisi.

Duis eget massa sem, gravida interdum ipsum. Nulla nunc nisl, hendrerit sit amet commodo vel, varius id tellus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ac dolor est. Suspendisse ultrices tincidunt metus eget accumsan. Nullam facilisis, justo vitae convallis sollicitudin, eros augue malesuada metus, nec sagittis diam nibh ut sapien. Duis blandit lectus vitae lorem aliquam nec euismod nisi volutpat. Vestibulum ornare dictum tortor, at faucibus justo tempor non. Nulla facilisi. Cras non massa nunc, eget euismod purus. Nunc metus ipsum, euismod a consectetur vel, hendrerit nec nunc.

Bibliography

[1] Akl, A., and Valaee, S. Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, #x00026; compressive sensing. In *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* (Mar. 2010), 2270–2273.

[2] Barfield, W., and Caudell, T. *Fundamentals of Wearable Computers and Augmented Reality*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2000.

[3] Begault, D. R. *3DD Sound for Virtual Reality and Multimedia*. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[4] Bonner, M. N., Brudvik, J. T., Abowd, G. D., and Edwards, W. K. No-look notes: Accessible eyes-free multi-touch text entry. In *Pervasive Computing*, P. Floren, A. Krger, and M. Spasojevic, Eds., no. 6030 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2010, 409–426.

[5] Brewster, S. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing 6*, 3 (2002), 188205.

[6] Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2003), 473480.

[7] Cyklistforbundet. Bdetakster for cyklister.

[8] Gaver, W. W. Auditory icons: Using sound in computer interfaces. *Hum.-Comput. Interact. 2*, 2 (June 1986), 167177.

[9] Gaver, W. W. The SonicFinder: an interface that uses auditory icons. *Hum.-Comput. Interact. 4*, 1 (Mar. 1989), 6794.

[10] Geelhoed, E., Falahee, M., and Latham, K. Safety and comfort of eyeglass displays. In *Proceedings of the 2Nd International Symposium on Handheld and Ubiquitous Computing*, HUC '00, Springer-Verlag (London, UK, UK, 2000), 236247.

[11] Graham, R. Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics 42*, 9 (Sept. 1999), 1233–1248. PMID: 10503056.

[12] Kajastila, R. Interaction with eyes-free and gestural interfaces.

[13] Kajastila, R., and Lokki, T. Eyes-free interaction with free-hand gestures and auditory menus. *International Journal of Human-Computer Studies 71*, 5 (May 2013), 627–640.

[14] LoPresti, E., Brienza, D. M., Angelo, J., Gilbertson, L., and Sakai, J. Neck range of motion and use of computer head controls. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies*, Assets '00, ACM (New York, NY, USA, 2000), 121128.

[15] Mackay, W. E., and Fayard, A.-L. *HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines.* 1997.

[16] Mardanbegi, D., Hansen, D. W., and Pederson, T. Eye-based head gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, ACM (New York, NY, USA, 2012), 139146.

[17] Morimoto, C., Yacoob, Y., and Davis, L. Recognition of head gestures using hidden markov models. In *In Proceeding of ICPR* (1996), 461465.

[18] Oakley, I., and Park, J.-S. Designing eyes-free interaction. In *Haptic and Audio Interaction Design*, I. Oakley and S. Brewster, Eds., no. 4813 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2007, 121–132.

[19] Park, G.-s., Ahn, J.-g., and Kim, G. J. Gaze-directed hands-free interface for mobile interaction. In *Human-Computer Interaction. Interaction Techniques and Environments*, J. A. Jacko, Ed., no. 6762 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2011, 304–313.

[20] Pascoe, J., Ryan, N., and Morse, D. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction (TOCHI) 7*, 3 (2000), 417437.

[21] Pasquero, J., Stobbe, S. J., and Stonehouse, N. A haptic wristwatch for eyes-free interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 32573266.

[22] Pielot, M., Poppinga, B., Heuten, W., and Boll, S. A tactile compass for eyes-free pedestrian navigation. In *Human-Computer Interaction INTERACT 2011*,

P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, Eds., no. 6947 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, Jan. 2011, 640–656.

[23] Pirhonen, A., Brewster, S., and Holguin, C. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, ACM (New York, NY, USA, 2002), 291298.

[24] Salvador, S., and Chan, P. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal. 11*, 5 (Oct. 2007), 561580.

[25] Sawhney, N., and Schmandt, C. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM transactions on Computer-Human interaction (TOCHI) 7*, 3 (2000), 353383.

[26] Smus, B., and Kostakos, V. Running gestures: Hands-free interaction during physical activity. In *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing - Adjunct*, Ubicomp '10 Adjunct, ACM (New York, NY, USA, 2010), 433434.

[27] Sodnik, J., Dicke, C., Tomai, S., and Billinghurst, M. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies 66*, 5 (May 2008), 318–332.

[28] Vazquez-Alvarez, Y., and Brewster, S. A. Eyes-free multitasking: the effect of cognitive load on mobile spatial audio interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), 21732176.

[29] Yi, B., Cao, X., Fjeld, M., and Zhao, S. Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 27892792.

[30] Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P. Earpod: Eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, ACM (New York, NY, USA, 2007), 13951404.