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

by

Jonas Hinge

IT University of Copenhagen

February 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

LAH **L**ist **A**bbreviations **H**ere

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 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 [1].

Solutions for these 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 challenges, this project will be based on the concrete scenario where people are biking while listening to and controlling their music libray. 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.

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 Method

Use triangle framework for HCI design [2]

1.4 Project structure

...

1.5 Limitations

...

Chapter 2

Background

This chapter presents first different research areas and topics related to this thesis. Then in the second part related works are presented and compaired.

2.1 Background overview

From the problem statement of this thesis follows two main interaction goals: Hands-free interaction and eyes-free interaction. Common properties of these two areas are the topics and research areas related to this project. This overview is presented in fig. 2.1.

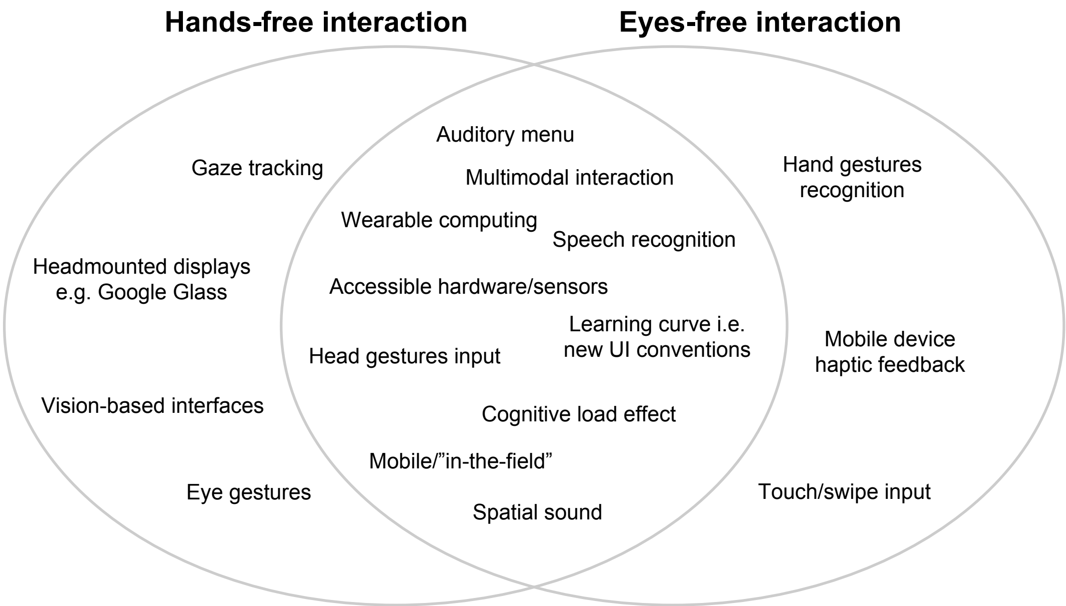

Figure 2.1: A comparison of thesis topics

2.1.1 Hands-free interaction

This term refers to controlling a system without using the hands including no hand gestures although sometimes used this way e.g. refers to not holding a device in the hand [3]. Achieving this hands-freeness interaction often relates to speech recognition or gaze/head tracking techniques but also other body parts are used for simple interactions e.g. leg shifting music track while running [4]. Speech recognition is becoming a more common interaction modality but there exists accuracy and stability challenges especially in mobile noisy environments. In contrast head tracking could possibly provide a more stable interaction detection in mobile environments.

2.1.2 Eyes-free interaction

Several work on both audio [5–9] and haptic [10, 11] 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 [12, 13] and even improve efficiency compaired to traditional visual displays [8].

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 [14]. Eyes-free interfaces can keep the users visual attention on the road while driving [15] or walking around in the city [9].

Much of the interfaces work in wearable computing tends to focus on visual headmounted displays [16] 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 [17]. 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.

2.2 Spatial sound

(Spatial audio, Head Related Transfer Function)...
Good reference for 3d sound [18]

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 [19]. Similarly Graham explores the advantages in reaction time when using auditory icons [20]. In [21] Gaver presents the use of spatial sound icons. In doing so, he draws forward the unutilized potential of creating natural interaction through spatial audio.

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 [22].

Work has shown that non-speech audio is effective in improving the interaction with mobile devices [23, 24].

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 [25].

2.3 Head tracking

There exists different kinds of areas when it comes to controlling a system with head gestures. Using cameras it is possible to effectively track head movements via facial recognition [26] and gaze tracking makes it possible to control an object by fixating the eyes on that object while moving the head [27]. 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. This will constrain the use especially in mobile "on-the-move" situations.

2.3.1 Motion gesture recognition

Dynamic Time Warping [28]

2.3.2 Intelligent Headset

...

2.4 Related work

Based on the research areas and topics mentioned this section presents the specific related works in these areas and finally a sum up and comparison of the properties of the related works and this thesis.

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 [7].

Park et al. also experimented, using head gesture input and aural output, with 1D and 2D menu interfaces [29].

2.5 Sum up of related work and background

Table: Summing up references that handles specific research areas...

TABLE 2.1: Related works properties comparison

Related work	Head gesture	Spatial sound	Music application	Accessible hardware
Multimodal eyes-free interaction techniques for wearable devices [7]	+	+	-	-

Chapter 3

Design

3.1 Interaction model

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

3.2 Sound design

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)

4.2 Motion gesture recognition

DTW - Dynamic Time Warping

4.3 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] Cyklistforbundet. Bdetakster for cyklister. URL `http://www.cyklistforbundet.dk/Alt-om-cykling/Love-og-regler/boedetakster`.

[2] Wendy E. Mackay and Anne-Laure Fayard. *HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines*. 1997.

[3] H. Witt, T. Nicolai, and Holger Kenn. Designing a wearable user interface for hands-free interaction in maintenance applications. In *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006*, pages 4 pp.–655, March 2006. doi: 10.1109/PERCOMW.2006.39.

[4] Boris Smus and Vassilis Kostakos. 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, page 433434, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0283-8. doi: 10.1145/1864431.1864473. URL `http://doi.acm.org/10.1145/1864431.1864473`.

[5] Raine Kajastila and Tapio Lokki. Eyes-free interaction with free-hand gestures and auditory menus. *International Journal of Human-Computer Studies*, 71(5): 627–640, May 2013. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2012.11.003. URL `http://www.sciencedirect.com/science/article/pii/S1071581912002042`.

[6] Matthew N. Bonner, Jeremy T. Brudvik, Gregory D. Abowd, and W. Keith Edwards. No-look notes: Accessible eyes-free multi-touch text entry. In Patrik Floren, Antonio Krger, and Mirjana Spasojevic, editors, *Pervasive Computing*, number 6030 in Lecture Notes in Computer Science, pages 409–426. Springer Berlin Heidelberg, January 2010. ISBN 978-3-642-12653-6, 978-3-642-12654-3. URL `http://link.springer.com/chapter/10.1007/978-3-642-12654-3_24`.

[7] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of*

the SIGCHI conference on Human factors in computing systems, page 473480, 2003. URL `http://dl.acm.org/citation.cfm?id=642694`.

[8] Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. 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, page 13951404, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240836. URL `http://doi.acm.org/10.1145/1240624.1240836`.

[9] Yolanda Vazquez-Alvarez and Stephen A. Brewster. 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*, page 21732176, 2011. URL `http://dl.acm.org/citation.cfm?id=1979258`.

[10] Jerome Pasquero, Scott J. Stobbe, and Noel Stonehouse. A haptic wristwatch for eyes-free interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 32573266, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979425. URL `http://doi.acm.org/10.1145/1978942.1979425`.

[11] Martin Pielot, Benjamin Poppinga, Wilko Heuten, and Susanne Boll. A tactile compass for eyes-free pedestrian navigation. In Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction INTERACT 2011*, number 6947 in Lecture Notes in Computer Science, pages 640–656. Springer Berlin Heidelberg, January 2011. ISBN 978-3-642-23770-6, 978-3-642-23771-3. URL `http://link.springer.com/chapter/10.1007/978-3-642-23771-3_47`.

[12] Ian Oakley and Jun-Seok Park. Designing eyes-free interaction. In Ian Oakley and Stephen Brewster, editors, *Haptic and Audio Interaction Design*, number 4813 in Lecture Notes in Computer Science, pages 121–132. Springer Berlin Heidelberg, January 2007. ISBN 978-3-540-76701-5, 978-3-540-76702-2. URL `http://link.springer.com/chapter/10.1007/978-3-540-76702-2_13`.

[13] Bo Yi, Xiang Cao, Morten Fjeld, and Shengdong Zhao. Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 27892792, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208678. URL `http://doi.acm.org/10.1145/2207676.2208678`.

[14] Jason Pascoe, Nick Ryan, and David Morse. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(3):417437, 2000. URL `http://dl.acm.org/citation.cfm?id=355329`.

[15] Jaka Sodnik, Christina Dicke, Sao Tomai, and Mark Billinghurst. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, 66(5):318–332, May 2008. ISSN 10715819. doi: 10.1016/j.ijhcs.2007.11.001. URL `http://linkinghub.elsevier.com/retrieve/pii/S1071581907001553`.

[16] Woodrow Barfield and Thomas Caudell. *Fundamentals of Wearable Computers and Augumented Reality.* L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2000. ISBN 0805829024.

[17] Erik Geelhoed, Marie Falahee, and Kezzy Latham. Safety and comfort of eyeglass displays. In *Proceedings of the 2Nd International Symposium on Handheld and Ubiquitous Computing*, HUC '00, page 236247, London, UK, UK, 2000. Springer-Verlag. ISBN 3-540-41093-7. URL `http://dl.acm.org/citation.cfm?id=647986.743887`.

[18] Durand R. Begault. *3DD Sound for Virtual Reality and Multimedia.* Academic Press Professional, Inc., San Diego, CA, USA, 1994. ISBN 0-12-084735-3.

[19] William W. Gaver. The SonicFinder: an interface that uses auditory icons. *Hum.-Comput. Interact.*, 4(1):6794, March 1989. ISSN 0737-0024. doi: 10.1207/s15327051hci0401_3. URL `http://dx.doi.org/10.1207/s15327051hci0401_3`.

[20] R Graham. Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics*, 42(9):1233–1248, September 1999. ISSN 0014-0139. doi: 10.1080/001401399185108. PMID: 10503056.

[21] William W. Gaver. Auditory icons: Using sound in computer interfaces. *Hum.-Comput. Interact.*, 2(2):167177, June 1986. ISSN 0737-0024. doi: 10.1207/s15327051hci0202_3. URL `http://dx.doi.org/10.1207/s15327051hci0202_3`.

[22] Raine Kajastila. Interaction with eyes-free and gestural interfaces. 2013. URL `https://aaltodoc.aalto.fi/handle/123456789/7720`.

[23] Antti Pirhonen, Stephen Brewster, and Christopher Holguin. 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, page 291298, New York, NY, USA, 2002. ACM. ISBN 1-58113-453-3. doi: 10.1145/503376.503428. URL `http://doi.acm.org/10.1145/503376.503428`.

[24] Nitin Sawhney and Chris Schmandt. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM transactions on Computer-Human interaction (TOCHI)*, 7(3):353383, 2000. URL `http://dl.acm.org/citation.cfm?id=355327`.

[25] Stephen Brewster. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3):188205, 2002. URL `http://dl.acm.org/citation.cfm?id=594356`.

[26] Carlos Morimoto, Yaser Yacoob, and Larry Davis. Recognition of head gestures using hidden markov models. In *In Proceeding of ICPR*, page 461465, 1996.

[27] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. Eye-based head gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, page 139146, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1221-9. doi: 10.1145/2168556.2168578. URL `http://doi.acm.org/10.1145/2168556.2168578`.

[28] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11(5):561580, October 2007. ISSN 1088-467X. URL `http://dl.acm.org/citation.cfm?id=1367985.1367993`.

[29] Gie-seo Park, Jong-gil Ahn, and Gerard J. Kim. Gaze-directed hands-free interface for mobile interaction. In Julie A. Jacko, editor, *Human-Computer Interaction. Interaction Techniques and Environments*, number 6762 in Lecture Notes in Computer Science, pages 304–313. Springer Berlin Heidelberg, January 2011. ISBN 978-3-642-21604-6, 978-3-642-21605-3. URL `http://link.springer.com/chapter/10.1007/978-3-642-21605-3_34`.