

AN EXPLORATION OF A HAPTIC AFFECT LOOP THROUGH USE CASES

by

Matthew Keith Xi-Jie Pan

B.A.Sc., The University of Waterloo, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate Studies

(Mechanical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

October 2012

© Matthew Keith Xi-Jie Pan 2012

Abstract

This work investigates a novel interaction paradigm of implicit, low-attention user control, accomplished by monitoring a user's physiological state. Additionally, we provide feedback to the user regarding system changes in response to this implicit control through touch. We explored the implicit interaction concept, termed the 'Haptic-Affect Loop' (HALO), in the context of two use cases.

In the first, we developed a HALO interaction to bookmark then resume listening to an audio stream when interrupted. A user's galvanic skin response is monitored for orienting responses (ORs) to external interruptions; our prototype automatically bookmarks the media, allowing the user to resume listening from the point he/she is interrupted. In a controlled environment, we found an OR detection accuracy of 84%. We further investigated the usefulness of two forms of haptic feedback for bookmarking: notification of bookmark placement and display of bookmarks during navigation. Results show that haptic notification is able to provide a significant performance benefit in terms of navigation speed when paired with visual-spatial indications of where bookmarks have been placed over conditions with no notification; also, performance was no worse with the haptic display of bookmarks than the visual display. Participants tended to prefer haptic notification at interruption time, and both haptic and visual display of bookmarks at resumption.

We used a second use case, music-listening, as a framework for another HALO implementation to estimate user preference for music. In a pilot experiment, we collected physiological and music rating data while participants listened to music. Combining this with structural feature data extracted from the music, we obtained state-space models which were used by a Kalman filter to estimate user rating of music selections. Results currently show poor performance in terms of ability to estimate rating of music both known and unknown to users possibly due to a false assumption of system linearity.

The outcome of this effort was a first pass implementation of HALO which provided insight into its strengths and weaknesses. This work enables further exploration in how HALO can benefit interactions in other contexts by providing validation of the technological feasibility, utility and behavior of the Haptic-Affect Loop.

Preface

Studies and experiments described in this thesis were performed with the approval of the Behavioural Research Ethics Board at the University of British Columbia, under ethics application #H01-80470 “Low-Attention and Affective Communication Using Haptic Interfaces”.

Work presented in Chapter 2 arose from a collaboration with Jih-Shiang Chang, Gokhan Himmetoglu, AJung Moon and Thomas Hazelton. I was responsible for the planning and directing of the experiments conducted, in addition to performing software development used for the experiment setup. This work is available in two works (a full paper and a demonstration) presented at the ACM Conference of Computer-Human Interaction in 2011.

M. K. X. J. Pan, J.-S. Chang, G. H. Himmetoglu, A. Moon, T. W. Hazelton, K. E. MacLean, and E. A. Croft, “Now Where was I? Physiologically-Triggered Bookmarking,” in *Proc. ACM Conf. CHI*, 2011, pp. 363–372.

M. K. X. J. Pan, G. J.-S. Chang, G. H. Himmetoglu, A. Moon, T. W. Hazelton, K. E. MacLean, and E. A. Croft, “Galvanic skin response-derived bookmarking of an audio stream,” in *Proc. ACM Conf. Computer Human Interaction - Extended Abstracts*, 2011, p. 1135.

Chapter 3 is based on work conducted in collaboration with my co-supervisors and with Dr. Joanne McGrenere. I was responsible for planning, carrying out and analysis of the experiment. The construction of equipment used in experiments described in Chapters 3 was performed by Andrew Strang, a then-undergraduate student working in the Collaborative Advanced Robotics and Intelligent Systems Laboratory as a summer research assistant under my direct supervision. A version of Chapter 3 has been submitted to IEEE Transactions on Haptics and is currently pending review.

M. K. X. J. Pan, J. McGrenere, E. A. Croft, and K. E. Maclean, “Exploring the Role of Haptic Feedback in an Implicit HCI-Based Bookmarking Application,” Submitted, pp. 1-12, 2012.

Aside from the works listed above, the author of this thesis was also involved in the following related publications arisen directly due to work presented in the thesis:

K. E. MacLean, S. Yohanan, Y. Sefidgar, M. K. X. J. Pan, E. A. Croft, J. McGrenere. “Emotional Communication and Implicit Control through Touch,” in *Proc. IEEE Haptics Symposium – Workshop on Affective Haptics*, Vancouver, Canada, March 2012.

T. W. Hazelton, I. Karuei, K. E. MacLean, M. A. Baumann, and M. K. X. J. Pan, “Presenting a Biometrically Driven Haptic Interaction Loop,” in *SIGCHI Workshop - Whole Body Interaction*, 2010.

Contents

Abstract	ii
Preface.....	iii
Contents	v
List of Tables	x
List of Figures.....	xi
List of Abbreviations and Symbols.....	xiv
Acknowledgements.....	xv
Dedication.....	xvi
1 Introduction.....	1
1.1 Context.....	1
1.2 Theory and Motivation	2
1.3 Introducing the Haptic Affect Loop (HALO).....	3
1.4 Related Work	5
1.4.1 A Framework for Affect	5
1.4.2 Implicit Human-Computer Interaction Systems	6
1.4.3 Physiologically-Based/Affective Interaction	7
1.5 Objectives	8
1.6 Outline of the Thesis.....	9
1.6.1 Use Case 1: Audio Stream Bookmarking.....	9
1.6.2 Use Case 2: Music Listening	11
1.6.3 Conclusions.....	11

1.7	Contributions.....	12
2	Physiologically-Triggered Bookmarking.....	13
2.1	Research Questions.....	14
2.2	Structure.....	14
2.3	Background and Related Work.....	14
2.4	Experimental Apparatus.....	16
2.4.1	GSR Apparatus and System Architecture.....	16
2.4.2	Audiobook Player and Bookmark Manager	17
2.4.3	Interruption Detection Software	18
2.5	Experiments	20
2.5.1	Experiment 1: Evaluation of GSR Utility.....	21
2.5.2	Experiment 2: Observation of Desired System Behavior	27
2.5.3	Experiment 3: Interaction Utility Assessment	29
2.5.4	Experiment 4: GSR Field Assessment (Pilot).....	31
2.6	Discussion.....	32
3	Exploring the Role of Haptic Feedback in Physiologically-Triggered Bookmarking.....	35
3.1	Purpose.....	37
3.2	Structure.....	37
3.3	Background and Related Work.....	37
3.3.1	Low-Attention Haptic Systems.....	37
3.3.2	Wearable Haptics	38
3.4	Research Questions.....	39

3.5	Experimental Apparatus.....	40
3.5.1	Overall System.....	40
3.5.2	Hardware.....	41
3.5.3	Software	42
3.6	Experimental Methodology	42
3.6.1	Participants.....	42
3.6.2	Conditions.....	43
3.6.3	Apparatus	44
3.6.4	Task.....	45
3.6.5	Procedure	48
3.6.6	Quantitative and Qualitative Measures.....	49
3.6.7	Hypotheses.....	49
3.7	Results.....	50
3.7.1	Effect of Display on Resumptive Navigation.....	50
3.7.2	Effect of Haptic Notification on Resumptive Navigation.....	53
3.7.3	Secondary Analyses	54
3.7.4	Summary.....	56
3.8	Discussion.....	56
3.8.1	Limitations and Constraints of the Experiment	60
3.9	Summary.....	61
4	Music Preference Recognition through Kalman Filtering.....	62
4.1.1	Use Case.....	62

4.1.2	Research Objectives	63
4.2	Background and Literature Review	65
4.2.1	Affect Estimation during Music Listening	65
4.2.2	Kalman Filtering	66
4.2.3	Summary	67
4.3	Methods.....	68
4.3.1	Participants.....	68
4.3.2	Location of Study and Consent.....	68
4.3.3	Experimental Setup and Sensor Equipment.....	68
4.3.4	Trial Task	69
4.3.5	Procedure	70
4.3.6	Quantitative and Qualitative Measures	71
4.4	Modelling and Filtering	72
4.4.1	Modelling.....	72
4.4.2	Unscented Kalman Filter	75
4.5	Results.....	79
4.5.1	Performance Evaluation Criteria.....	79
4.5.2	Self-Validation Testing.....	80
4.5.3	Familiar Music Selection Testing	81
4.5.4	Unfamiliar Music Selection Testing.....	82
4.5.5	Observed Behaviours	84
4.6	Discussion and Lessons Learned	84

4.6.1	Limitations, Improvements and Future Considerations.....	85
5	Conclusions and Future Work	87
5.1	Audio-Stream Bookmarking.....	87
5.2	Music Listening	88
5.3	General Implications.....	88
5.4	Future Work	89
	References.....	90
	Appendix A – Physiologically-Triggered Bookmarking Experiment Materials	101
A1	Participant Consent Form.....	101
A2	Pre-Experiment Questionnaire.....	103
A3	Experiment 2 and 3 Experimenter Instructions.....	104
A4	Post-Experiment Questionnaire and Semi-structured Interview Questions.....	106
	Appendix B - Haptic Feedback during Bookmarking Experiment Materials.....	109
B1	Participant Consent Form.....	109
B2	Participant Instructions.....	111
B3	Post-Experiment Questionnaire.....	114
	Appendix C – Affect Estimation during Music Listening Experiment Materials	118
C1	Participant’s Consent Form.....	118
C2	Post-Trial Questionnaire	120

List of Tables

Table 1. Definitions of test outcomes, conditions and measures.....	25
Table 2. Mean and std. dev. of root-mean-square deviation for each participant.....	81
Table 3. Mean and std. dev. of root mean square deviations per participant, for music selections which are familiar to participants.	82
Table 4. Visual inspection results of estimated rating trajectories for each participant for music selections which were familiar to participants.....	82
Table 5. Mean and std. dev. of root mean square deviations per participant, for music selections which are unfamiliar to participants.	83
Table 6. Visual inspection results of estimated rating trajectories for each participant for music selections which were familiar to participants.....	83

List of Figures

Figure 1. Proposed HALO (haptic-affect) implicit interaction loop structure.....	4
Figure 2. Circumplex model of affect as illustrated by Posner et al. [35].	6
Figure 3. Diagram of bookmarking system architecture.	16
Figure 4. Screen capture of the a) audiobook player; b) and bookmark manager.	17
Figure 5. a) Flowchart of OR detection process. b) Typical smoothed GSR waveform (3200 samples). c) Smoothed first derivative GSR waveform (3200 samples).	18
Figure 6. Approximate timeline of the auto-bookmarking system. Horizontal bars indicate timeline variability, black vertical lines show an example timeline for a participant introduced to vocal stimulus.	20
Figure 7. Experiment 1 Setup.	22
Figure 8. GSR signals overlaid with interruption and detection lines for one representative trial (threshold set at 0.07 μ S).	25
Figure 9. Average sensitivity and precision versus detection threshold. The red line refers to the maximum sensitivity when the threshold is set to 0.07 μ S.	26
Figure 10. Sample GSR trace showing a missed detection of an interruption.	27
Figure 11. Average algorithm sensitivity (proportion of interruptions detected of total) for interruption types. Error bars represent standard deviation of sensitivity across all subjects.	27
Figure 12. Experiment 2 set-up. A similar set-up was also used for E3.	29
Figure 13. Raw GSR data collected from a participant in a) a controlled test environment and b) a bus terminal at rush hour. Thick, dotted lines represents interruptions; thin, solid lines represent bookmarking system OR detections.	32
Figure 14. Flowchart showing progression of events in the audio stream bookmarking use case.	36

Figure 15. System flowchart.	40
Figure 16. Exploded view and photo of wrist-worn navigational device and haptic display.	41
Figure 17. A screen capture of the media player displayed to participants. The scroll bar represents the current location in the audio stream with respect to the current track. Automatically placed bookmarks are shown as green lines above diamonds. Green diamonds signify that the user has navigated to that bookmark.	42
Figure 18. Diagram showing factors within the experiment conducted.	43
Figure 19. Screen capture of the experiment settings window (not seen by participants). This window provides the experimenter with displays such as currently playing track, track time, the user’s GSR activity, a list of bookmarks that have been placed, options regarding how bookmarks are placed, and how they are presented to participants.	45
Figure 20. Representative timeline for data collection portion of experiment, for one participant. Factor levels shown are for the within-subject navigation display conditions only; haptic notification display is between-subjects.	47
Figure 21. Interaction effect between visual and haptic bookmark display type during resumptive navigation.	51
Figure 22. Mean navigation time for each bookmark display condition presented during resumptive navigation. Error bars show 95% confidence intervals. * indicates that the condition is significantly faster than the control condition (D_{None}).	51
Figure 23. Participant preference for bookmark display.	52
Figure 24. Interaction effect between D_V and N_H for navigation time. Error bars show standard error of the mean.	53
Figure 25. a) Mean resumptive navigation time for trials with and without false positive bookmarks. b) Mean resumptive navigation time for each condition. Error bars represent 95% confidence intervals. The star represents a significant difference between navigation times for trials with and without false positive bookmarks within a condition (this distinction is meaningless for the D_{None} condition).	55

Figure 26. The number of occurrences where information was skipped over due to improper navigation across experiment conditions.....	56
Figure 27. Photograph of the music rating device.....	69
Figure 28. Diagram showing input (u – features drawn from music), states (x – user preference for the music) and outputs (y – physiological measures) of the system model.	73
Figure 29. A diagram showing the choosing of sigma points (blue dots) around a data point (white dot).....	77
Figure 30. Examples of UKF results showing a) good correlation between estimated and reported music ratings resulting in a low RMSD value, and b) poor correlation resulting in a high RMSD value.....	79
Figure 31. Examples of UKF results showing a) good estimation performance since both estimated and reported ratings at 40s fall in the same bin, and b) poor estimation performance since both estimated and reported ratings fall in different bins.....	80

List of Abbreviations and Symbols

Abbreviation	Full Name
BVP	Blood Volume Pulse
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
GSR	Galvanic Skin Response
HALO	Haptic-Affect Loop
HCI	Human Computer Interaction
ICICS	Institute for Computing, Information and Cognitive Systems
iHCI	Implicit Human Computer Interaction
KF	Kalman Filter
OR	Orienting Response
RMSD	Root-Mean-Square Deviation
SVT	Self-Validation Testing
SD	Standard Deviation
ST	Skin Temperature
UKF	Unscented Kalman Filter
μS	Micro-Siemens

Acknowledgements

I offer my enduring gratitude to several people who have made significant contributions not only to this thesis, but to my knowledge of the scientific world:

- Supervisors Dr. Elizabeth Croft and Dr. Karon Maclean, who have so graciously accepted me as their graduate student and have provided me with unwavering support throughout my time as a masters student;
- Dr. Joanne McGrenere for her patience in providing coherent answers to my constant barrage of questions on human factors experiment design; and,
- Dr. Ryozo Nagamune, whose knowledge of control theory has been invaluable.

I would like to express additional thanks to Gordon Jih-Shiang Chang, Gökhan Himmetoğlu, AJung Moon, Tom Hazelton, Susana Zoghbi, Gordon Andrew Strang and Jessica Dawson for their support on various parts of the work presented in this thesis.

This thesis is dedicated to my loving parents, for their perpetual patience and support throughout my life, and to Martin, Valerie, Daniel and Thomas, for their encouragement during the most difficult of times.

1 Introduction

1.1 Context

The development of modern user interfaces has largely depended on explicit channels of command and control driven by unambiguous requests from the user [1]. In this type of interaction, a computer or device receives specific, explicit inputs provided by a user who is (ideally) consciously aware of his/her actions within the context of the interaction. Information is passed from users to devices through explicit gestures like key-strokes, verbal commands and finger swipes, and is received by users from devices through visual, auditory and, more recently haptic (touch) channels. The advantages of this ‘explicit interaction’ are readily apparent: virtually every consumer electronic device utilizes explicit communication channels between the user and device.

When well designed, this clear, controlled form of interaction has the important benefit of leaving little room for misinterpretation. However, there are circumstances where the demands and costs of an explicit interface encumbers its utility and efficiency, most noticeably in multitasking scenarios [2]. In one time- and safety-critical example, steering and handling a vehicle often requires most of the driver’s attention; yet a variety of secondary vehicle controls (e.g., adjusting audio system settings) compete for the same perceptual and cognitive resources. The resulting diversion from the primary driving task can impair safety. In other cases, the cost of explicit control demands may be in convenience, privacy or attentional fragmentation [3], [4]: a user may wish to alter a device’s behavior, but cannot or prefers not to use explicit commands (e.g., adjusting an mp3 player when doing chores or exercising); a badly timed device notification may temporarily dip in priority and then be quickly forgotten (e.g., a pop-up meeting notification).

Trends in technological growth and ubiquitous computing indicate that the problems of interruption and fragmentation will become compounded as explicit-based technologies becomes increasingly pervasive in our daily lives [5–10]. Users will need to interact with more explicit interfaces, each demanding its share of the user’s attention and cognitive resources. Human-technology interactions which can reduce the requirement for explicit interaction can help to mitigate the cognitive burden on users, freeing them to focus on the activities that are most important to them within their current context.

To address this and similar issues in human-computer interaction (HCI), in this thesis we consider the design of an implicit interaction loop that aims to move an interaction from the foreground to the background, understanding and adapting to our needs without demanding our full attention. We propose an interaction model employing implicit user-device communication in [11], [12]. This paradigm utilizes physiological signals – indicative of user affect – incorporated into a device’s control loop, thereby allowing the device to respond to changes implied by the user’s affective state. To further our minimal-attention goals, we have suggested using haptic feedback to provide immediate, private and unintrusive (in terms of attention required from the user) indication of the device’s response to the user’s state: a reassuring confirmation which closes the interaction loop.

1.2 Theory and Motivation

Foreseeing the rise of ubiquitous computing, Weiser and Brown in 1996 argued that rather than abruptly demanding a user’s focus, interactions should slide transparently between the attentional periphery and center, “calmly” providing context and orientation [10]. Since then, Schmidt proposed that a system could recognize a user’s actions not primarily aimed to facilitate interaction with a computerized system, and use it as a system input – he termed this concept *implicit human-computer interaction* (iHCI) [13].

Progress in areas such as computer vision and machine learning has allowed for the development of iHCIs where explicit user control and the inherent related cognitive load are significantly reduced. Using such technologies, it can be anticipated that devices will be able to adapt their behavior automatically from sensor-based observations of the user and their current context. For example, some work has been done at having computers recognize affective facial expressions [14], [15] which help the computer in choosing how to act in a socially acceptable and responsible way. There have been several other proposed implicitly-derived input channels beyond facial expression recognition utilizing physiological measurements instead. Physiological sensors such as heart rate, respiration, skin temperature, and electroencephalographic sensors, have been used in the medical and psychological fields for quite some time; however, their use in the domain of human-computer interaction is relatively recent. Only recently have physiological-sensing devices developed for HCI been released to the consumer market – e.g., Emotiv [16] and NeuroSky [17] EEG-based headsets developed for video gaming and educational applications. With training, these devices allow users to implicitly

control computers, electronics and even robots. Several researchers within the HCI field have recognized this potential and have developed interactions based around user physiological sensing [18], [19].

In developing any interaction method, explicit or non-explicit, we recognize that the system must act transparently such that the user understands how interaction is performed and what behaviors are to be expected. Thus, the system should always keep users informed about what is going on, through appropriate feedback within reasonable time – a feedback mechanism is required to notify the user that the behavior, mode or operation of the device has changed in response to alterations in the observed state of the user [20]. In keeping with the goal of low cognitive load for implicit interaction design, it would be suitable to implement such a user-feedback system through a low-attention channel as well: bombarding the user with visual or auditory feedback undermines the goal of reducing cognitive resource requirements of the user in the interaction. Alternatively, there is a wide body of literature which suggests that haptic displays can be used for this purpose and convey information to the user [21–25]. In addition to being underutilized and a natural communicative method, research has shown that humans are able to reliably discriminate and classify haptic signals, without expending significant attentional resources [26], [27].

1.3 Introducing the Haptic Affect Loop (HALO)

In this thesis, we explore and substantiate a method of a human-device interaction that synthesizes the ideas presented above, namely:

- using implicit physiological monitoring to drive the interaction rather than explicit commands; and,
- providing low-attention haptic signaling to provide feedback to the user with regards to system state and/or behaviors.

We directly incorporate these two principles into an iHCI control loop in anticipation that such a system can effectively support the user without requiring considerable cognitive resources, especially in task-laden environments. Building on the knowledge that some human affective (emotional) states can be estimated from physiological signals captured with off-the-shelf biometric sensors [14], [15], [28], our implicit channel is a user’s voluntary and involuntary physiological responses to a situation. We incorporate this biometric response directly into the

application’s interaction loop, such that it responds to changes implied by the user’s affective state as shown in Figure 1. The loop is closed with an immediate but unintrusive indication of system recognition and response to the user’s estimated state using haptic feedback, communicating to a user that the system has responded and performed some action as a result of their estimated affective state. We call this novel approach to device interaction the Haptic-Affect Loop (HALO).

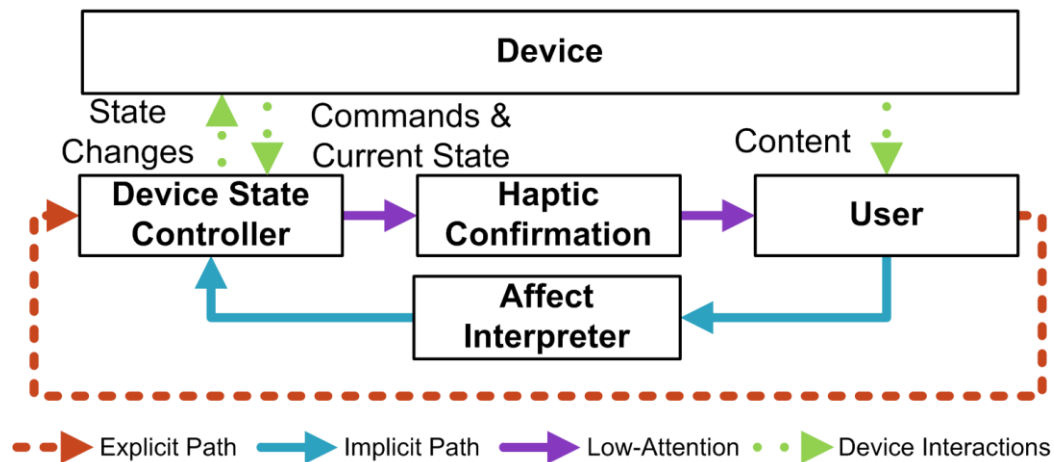


Figure 1. Proposed HALO (haptic-affect) implicit interaction loop structure.

The premise behind the HALO concept can be best illustrated by an example:

While on the job as a programmer, Lily likes to explore and listen to new songs through an internet music service while working. Sometimes while listening, however, she finds that some of the songs do not suit her changing tastes, or are quite distracting from her work. Normally, she would have had to suspend whatever she was working on, and switch windows to the internet browser and tell the music service to jump ahead to the next song. Recently, however, Lily (the user) began using a wrist-mounted device able to save her the hassle of constantly switching between her programming task and the internet music service. The device continuously measures key physiological indicators and, using signal processing and machine learning, is able to estimate Lily’s emotional state (implicit path affect interpreter). Thus, when the device detects that she dislikes a currently playing track or is stressed, it automatically executes a script which tells the music service to play the next song in queue (device state controller enacting state changes). Additionally, the device notifies Lily of this action through a very gentle vibration of the watch-like device (haptic confirmation). If she finds that the action taken by the device is incorrect, she can press a button on the watch to automatically reverse the device’s decision

(explicit path). Lily finds that this new way of exploring music allows her to be more focused on work instead of spending time fiddling with the music service.

In the example, we see that the watch-like device instantiates the HALO concept by mediating the transition between a primary task (programming) and secondary task (music listening). With this proposed technology, we hope to improve users' ability to multitask between various activities. That is, the HALO model discussed herein provides a mechanism for supporting low priority tasks at the periphery of the user's attention, freeing up focused cognitive resources to be used for more urgent/primary tasks.

1.4 Related Work

Although novel, the concept for the HALO paradigm is founded on prior work conducted in the areas of affective classification and computing, iHCI, and physiologically-based interaction. Here, we provide mention of this work to provide a backdrop for the research conducted in this thesis. The topics presented here represent common themes across all chapters within the thesis. Other literature and background information more specific to each chapter will be presented separately in their own sections within the chapter (Sections 2.3, 3.3, and 4.2).

1.4.1 A Framework for Affect

As would be expected, performing research in affect-based systems requires some understanding of what affect is, how it is generated and how it can be quantified or measured. However, herein we find a classic problem in affect research: literature reveals that there is no standard agreement as to how affect or emotions are defined [28–31]. Following from this, we find that the criteria for which we can distinguish an emotion from other emotions or bodily processes, and what constitutes a *base set* of emotions are also ill-defined. However, many researchers agree with a standard *view* of affect, which suggests that:

- affective responses represent an information processing strategy that is potentially valuable for organisms in terms of dealing with certain situations [15]; and,
- emotions are both cognitive and physical - not only can emotions be generated by the brain, but also by changes in physiology and bodily chemistry [28].

Rather than trying to use a definitive classification scheme to assign categorical labels for affective states, we use Russell's *Circumplex Model of Affect* [32]. His model, shown in Figure 2, suggests that all emotions are linear combinations of valence (positive or negative) and

arousal (high and low) dimensions. Strong, positive emotions (e.g., elation) appear in the upper right of the valence-arousal, whereas weak, negative emotions appear in the lower left (e.g., depression). Although other popular models of emotion and affect exist (see [33] for a comprehensive review of emotion models), Russell’s model has been widely used in the social sciences due to its simple structure and uniform representation of affective states on a two-dimensional plane [34]. For this reason, we have chosen to use the circumplex model as a framework for affect understanding in this thesis.

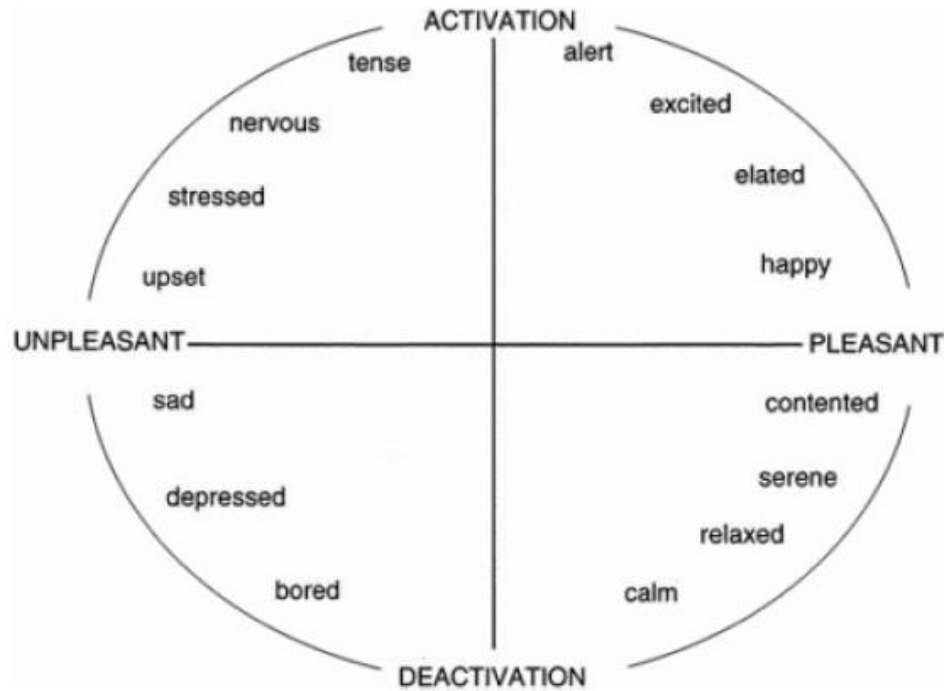


Figure 2. Circumplex model of affect as illustrated by Posner et al. [35].

1.4.2 Implicit Human-Computer Interaction Systems

Implicit human computer interaction (iHCI) is defined as ‘*an action, performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input*’ [13]. The concept of iHCI is built on the underlying notion that the computer system participating in the interaction has obtained, or been provided with, an internalized representation or information which could be used to construct a model of certain behaviors of a human participant (the user model). These behaviors are usually inferred through sensing and recognition of the user’s physical gestures, facial expression, physiology or a combination thereof and are fed as inputs into the computer system. Behaviors that are recognized through comparison to the user model can be used to enable the system to respond and automatically enact alterations to its state, creating the impression that the system is ‘smart’.

Early work in this area yielded several interaction schemas utilizing iHCI principles [36–41]. For example, Schmidt et al. developed an iHCI scheme which could facilitate ‘real-world bookmarking’ [36]. In this application, passive RFID tags are integrated into physical objects linking them to an object-specific URL or application. A user sporting a wearable RFID tag reader could interact with these RFID-enabled objects, triggering the link/application to be opened on a mobile device. For example, picking up a spoon could bring up a webpage suggesting a recipe and opening a wallet could bring up a user’s stock portfolio. However as reported by Poslad, many researchers cite difficulties in determining user context mainly due to the non-deterministic nature of the human user or the environment which may not correlate well to the system’s internalized model [9]. Additionally, users may find it difficult to interact with a system which purposefully hides some of the interaction from them.

1.4.3 Physiologically-Based/Affective Interaction

In review of work conducted in implicit interaction, we find that there are close ties between the domains of iHCI and affective computing; there is a large body of work dedicated to estimating human affect to be used as a potential iHCI input [14], [15], [19], [42–52]. Prior work indicates that emotions arise from cognitive interpretations of core physiological experiences, which have spawned the development of dimensional models of emotion such as Russell’s Circumplex described above [35], [53]. Naturally, this in turn suggests that examining core physiological features may provide insight into affective experiences. As such, several groups of researchers have focused on developing a mapping between user affect estimates and physiological signals. However, even with the mapping of affective states, researchers have found that correlating physiological signals to affect is not straightforward. Human physiological measurements vary not only between users, but also due to other factors such as the time of the day, environment conditions, or even whether medication was consumed prior to the measurement being taken.

To overcome these challenges, several methods of processing bio-signal sensor measurements have been used to develop user affect models include filtering [19], [52], neural networks [50], [54], machine-learning classifiers [47], [49–51] and statistical analysis [47], [49]. For example, in their early work in affective computing, Picard et al. classified emotions using physiological sensors [28], [47], [48]. With algorithms such as Fisher Projection and Sequential Floating Forward Search, they obtained emotional state recognition accuracy of up to 81% after 2-4 min of algorithm training per subject, for eight categories. Kim et al. aimed to reduce signal

monitoring times and system training requirements through feature extraction and pattern classification of data (electrocardiography, skin temperature variation and electrodermal activity) pooled from multiple subjects [19]. A recognition rate of 78% for three emotion categories was obtained after 50s of monitoring.

There have been attempts to use physiological signal processing in implicit man-machine interface applications. For example, Conati et al. has explored the use of physiological monitoring to measure engagement and emotional states in children while playing educational games [42]. In this work, pedagogical agents act on a probabilistic model of the user's affect to generate interventions which are able to achieve a compromise between a user's learning and engagement. Another example includes a wearable video camera system which monitors a user's galvanic skin response (GSR) to detect physiologically arousing events (e.g., orienting responses) in users [52]. When such an event is detected, the camera records a series of digital images of the user's environment, capturing the event which produced the startle response for the creation of an image-based diary for offline examination. More recently, Liu, Rani and Sarkar presented a framework for closed-loop human-robot interaction through a robot-based basketball game [55]. The objective of this game is for the player to throw a ball into a basket mounted on a robotic arm manipulator that could move in different directions and speeds according to the game's difficulty setting. The group used various physiological measures including skin temperature, GSR, electrocardiographic and electromyographic activity to model player affective state and anxiety using a regression tree. Classification accuracy of player affective states using this model was reported to be 88%. Based on this model, the robot was able to respond to higher anxiety by decreasing difficulty and vice versa. The group reported that the robot could influence lower player anxiety for 79% of the participants during the game.

1.5 Objectives

The research presented in this thesis is grounded in work conducted by Hazelton, who first investigated the potential for the HALO paradigm [56]. Using focus groups, technical validation studies, and participatory design, he was able to define system requirements, preliminary guidelines and design principles for the HALO interaction. However, the scope of his work did not extend beyond rigorously gathering requirements, leaving implementation of the interaction loop as future work. Here, we continue to build upon Hazelton's preliminary work, and expand development through realization of technical elements within the HALO paradigm.

The objectives of this thesis are twofold: to establish the technological viability and efficacy of the HALO paradigm we propose, and to explore its design space.

We examined the HALO in terms of the two main loop components separately: physiological sensing of affect and haptic feedback. For each component, we wanted to determine:

- if the technology for the component can be built/modeled in the context of a use case;
- how well the technology works within the use case;
- what problems arise in terms of technological limitations and/or user experience;
- if the technology developed can be extended to a more generalized HALO model or theme, applicable to a wider range of applications, especially those in multitasking and ubiquitous technology environments as determined from experimental results; and,
- what technologies need to be further improved or developed to improve applicability and robustness of HALO.

We approach these objectives by examining HALO in the context of two use cases where the HALO framework has been applied, descriptions of which appear in the following section.

1.6 Outline of the Thesis

The main body of this thesis presents two primary use cases through which the HALO concept is explored.

1.6.1 Use Case 1: Audio Stream Bookmarking

Chapters 2 and 3 of this thesis describe work where the feasibility of the HALO interaction method is explored in the context of a specific interrupted audio stream listening use-case, where a user, engaged in listening to an audiobook or podcast, is suddenly interrupted by a disruptive element in the environment, such as a friend coming by to say hello, an incoming call on the user's cell phone, or a knock at the door. Often times, users forget to pause or stop the audio stream 'in the heat of the moment,' forcing the user to blindly navigate and rewind to the point where the user had left off prior to the interruption. Of particular interest to us was to explore what improvement HALO could offer to the user following the interruption, particularly in the context of audio stream resumption through the generation of physiologically-triggered bookmarks.

In this stream of work, we have developed an ‘auto-bookmarking’ system which detects and marks a user’s orienting response (OR) to an interruption. For this work, we looked at using a single physiological measure: electrical conductance of the skin – more commonly known as Galvanic Skin Conductance (GSR) – as an indicator of ORs. Bookmarks placed near the estimated point of interruption can be accessed later via a graphical or haptic tool.

In Chapter 2, a series of experiments are described which are primarily designed to determine if the feasibility of HALO paradigm in terms of the audio-stream use case. To this end, we wanted to determine: *Does GSR exhibit a consistent and detectable change in response to interruptions of a level that distract users?* If so, we could then proceed to design and set up elements of the interaction itself – this leads to having to answer research questions specific to the audio-stream interaction use-case itself. For example, *where should a mark be placed for greatest utility;* and, *is there a correlation between interruption duration and how far back users tend to rewind the audio stream?*

In relation to the study of the HALO concept, implicit interaction methods have traditionally not been well defined or understood. Thus, successfully designing and implementing the affect recognition portion of the HALO concept poses a significant challenge. The work described in Chapter 2 is meant to provide an introductory glimpse into the HALO concept’s overall feasibility. Content in this chapter is published in the Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI’11) and can be found in [11]. This work was completed in collaboration with Gordon Chang, Gokhan Himmetoglu, AJung Moon, Karon MacLean and Elizabeth Croft.

Chapter 3 describes an effort to continue and complete the work described in Chapter 2 in terms of ‘closing of the loop’. Here, there was a focus on exploring and developing appropriate haptic feedback practices to notify a user of bookmark placement after an interruption has been detected, and to assist the user in navigating to the bookmark of interest. Additionally, this work serves to complete implementation of the HALO within the media stream bookmarking scenario and investigates its efficacy and viability through user testing.

The work presented in Chapter 3 has been submitted to the IEEE Transactions on Haptics and is, at the time of the writing of this thesis, pending review and approval. Co-authors of this paper are: Joanna McGrenere, Elizabeth Croft and Karon MacLean.

1.6.2 Use Case 2: Music Listening

Music-listening is often a secondary or tertiary task – usually users are attending to more cognitively-demanding tasks while listening to music such as driving home from work or diligently toiling on a take-home assignment from school. However, there are times where users need to interact with the media device to change how or what audio content is being delivered due to changing tastes or contexts. This interaction is generally based on well-established interface design principles where information about system status is explicitly displayed and explicitly visible to the user; device behavior is changed through a series of button presses or touchscreen swipes. As both the act of explicitly transmitting and receiving data to and from the device requires significant attentional resources to process, the music-listening task is no longer a secondary task – it has instantly moved from the periphery to the center of attention. Depending on the importance of the primary task at hand, the severity of this change in task priority may have dire consequences (e.g., driving). Thus, this scenario provides great opportunity for study and application of the HALO interaction paradigm.

Much like the audio stream bookmarking use case discussed in Chapters 2 and 3, the notable feature in this particular use case is the presence of several tasks simultaneously competing for the user's attention. The potential of the HALO concept to provide devices with emotion- and context- aware behavior allows for interactions with the media player to be kept within the periphery.

Chapter 4 of this thesis describes preliminary work on this music-listening use case; the scope of which was limited to developing user affect recognition in response to music. Here, we developed a unique approach in applying a variation of the Kalman filter to recognize user affective states from music features and user physiological signals (e.g., heart rate, skin temperature etc.). This system was used in an exploratory pilot experiment to address questions of technological feasibility.

1.6.3 Conclusions

Over-arching conclusions drawn from these linked explorations are presented in Chapter 5. They involve recommendations for implementing a HALO-style interaction loop in a portable audio system, a discussion of generalizability to other use cases, and suggestions of relevant areas for future work.

1.7 Contributions

To achieve the objectives of the research described in this thesis, the following contributions were made:

- Development of an orienting-response detection algorithm for automatic placement of bookmarks in audio streams based on skin conductance;
- A haptic signaling system for system feedback to the user indicating when and where such bookmarks have been placed, completing implementation of the HALO concept for an audio stream bookmarking application;
- A novel affect recognition platform using an unscented Kalman filter to predict user music preference based on physiological measurements and music features; and,
- Experimental qualitative and quantitative results based on the use cases which demonstrate which situations can be adequately serviced by the HALO interaction concept and where it fails.

The research described here defines new models for a unique user-device interface driven by implicit, low-attention control. To test this paradigm, we have developed experimental setups testing the HALO for two media-based use cases. These use cases have provided us specific understanding as to the viability of the holistic HALO interaction model and how best to refine it such that it may be appropriate for a wide range of applications.

2 Physiologically-Triggered Bookmarking¹

Motivated by the need to substantiate the benefits of the previously unproven HALO paradigm described by Hazelton [56], in this chapter we consider a representative use case to serve as a context for a first pass implementation of the HALO. Here, we are interested in the utility and effectiveness of the HALO implicit control paradigm, illustrated by this scenario:

A woman listens to an audiobook on her portable media player in a dentist's waiting room. Her name is called; her focus shifts from the audiobook to the hygienist. She fumbles in her bag to pull out the device, and eventually finds the "pause" button. The hygienist waits.

Or,

... her focus shifts. Reluctant to make the hygienist wait, she pulls out her earbuds and gets up, leaving her player running until the book ends. Later, she iteratively scrolls and listens backwards through the audio stream, trying to find the last familiar point and estimating forwarding intervals when she overshoots.

Both trajectories highlight how an explicit interface can 'demand' a user's attention, with negative outcomes if attention is not given. But what if instead:

... her focus shifts. A sensor embedded in her bracelet² captures this, signaling the system to mark the audio stream which continues to play. She feels a small pulse on her wrist that signals the placement of a bookmark. She follows the hygienist, pulling out the earbuds when it is convenient. Later when alone again, she scrolls through the audio stream, feeling for discrete vibrations on her wrist indicating bookmarks. She jumps back through two or three automatically placed marks, locates the one placed just before the interruption where she stopped listening, and continues playback from there.

This scenario illustrates the HALO implicit interaction method: an 'auto-bookmarking' system detects and marks a user's orienting response (OR) to an interruption by monitoring galvanic skin response (GSR) in real time. Bookmarks placed near the estimated point of interruption

¹ A version of Chapter 2 has been published. [M. K. X. J. Pan] et al., "Now Where was I? Physiologically-Triggered Bookmarking," in *Proc. ACM Conf. CHI*, 2011, pp. 363-372.

² Pasquero, Stobble and Stonehouse describe a wristwatch capable of haptic signaling in this reference: J. Pasquero, S. J. Stobbe, and N. Stonehouse, "A Haptic Wristwatch for Eyes-Free Interactions," in *Proc. ACM Conf. CHI*, 2011, p. 3257.

can be accessed later via a graphical or haptic tool. This use case highlights the example of a mobile application used in a predominantly hands-free mode. At this stage, we do not seek to change the underlying nature of existing explicit control or replace its channels; but rather to bypass points of dysfunction with a new, lower-effort channel when appropriate.

2.1 Research Questions

This research example of attentional bookmarking, defines a model for a unique user-device interface driven by implicit, low-attention control. With this platform, we have assessed the viability of a holistic interaction model that may be appropriate for a wide range of applications. We have focused on two questions pertaining to our implicit interaction model, with respect to the audiobook-listening use case:

- 1. Does GSR exhibit a consistent and detectable response to interruptions in realistic contexts?*
- 2. How should the system respond to this information?*

In this chapter we demonstrate empirically that GSR can be used to detect orienting responses to interruptions under controlled conditions, and have data suggesting that this ability can be extended to noisier, chaotic environments. We built an OR detection system based on this result, and to explore Question 2, made quantitative and qualitative observations of its use in a setup simulating audiobook listening scenarios. The latter effort gave us specific insights towards how best to refine our implicit interaction model in both this and more general cases. We explored the HALO interaction paradigm without the haptic feedback component shown in Figure 1, as the feedback dynamics warrant a more in-depth examination than can be offered here. This feedback component is explored in detail in Chapter 3.

2.2 Structure

In the remainder of this chapter, we summarize related work regarding the GSR and ORs (Section 2.3), describe our hardware/software setup for detecting and displaying ORs (Section 2.4), and present the experimental methods and results by which we quantified GSR-based ORs to interruptions and applied it to our interaction paradigm use-case (Section 2.5). We conclude this chapter with discussion of results (Section 2.6).

2.3 Background and Related Work

It is well known that the OR is an immediate reaction to the perception of a novel element or stimulus that is not sudden or intrusive enough to elicit a startle reflex. ORs are often examined

to gather insights on human attention shifts and information processing [57]. An OR can be detected in many ways, including heart rate and electroencephalography (EEG), but the simplicity of measuring GSR and its strong relationship to OR is attractive.

The GSR (i.e., electrodermal activity) has been studied since the late 1800s. It is currently believed that the GSR is caused by the electrical activity of the sweat glands, and is connected to the sympathetic nervous system; it has been linked with the physiological instantiations of emotion, arousal and attention [53], [58–60]. Firth and Allen showed that short-term changes in GSR reflect ORs [57]. Previous research has shown that within normal ranges of ambient room temperature and controlled subject state and motion, there is a high correlation between OR and GSR [59]. Other literature suggests that GSR measurements can be more easily discriminated than other physiological measures such as heart rate and EEG, since they can be detected quickly and without complex analysis [60].

GSR measurements are most sensitive on volar surfaces, suggesting a future possibility for sensors that can be worn unintrusively as a ring or shoe insert during daily activities. GSR sensors are inexpensive, wearable and pose no risks for the user. Disadvantages include latency in signal detection (i.e., 1-4s lag periods are common), significant response variance between subject groups (e.g., gender, age), possible habituation over time, and non-specificity [53].

There have also been initial attempts to use GSR-based classification to augment traditional interaction techniques. For example, Healey and Picard's StartleCam - a wearable video camera - monitors a user's GSR to detect when a user is startled [52]. When GSR indicates a user's heightened arousal, a time series of digital images are saved to mimic the user's "flashbulb" memory of the event causing the startle, autonomously generating an image-based diary for offline examination and memory assistance.

We chose GSR as, on balance, the most promising physiological input to our control loop - keeping its limitations in mind. Since we require only OR information, with GSR we can bypass the complexity of a full affective model. Our interaction model builds on the results of the prior work presented in this section with the aim of providing transition support between primary and secondary tasks. Beyond offline review, we seek a fluid continuous interaction where marked moments are used to propel the user towards a goal with minimal disruption.

2.4 Experimental Apparatus

In our use case, marks are placed automatically during interrupted listening to an informative audio stream. Bookmarks are placed near GSR-detected ORs, without explicit user direction. The marking system used here is intended as a platform to demonstrate and study the broader implicit interaction paradigm. In this section, we describe our physical implementation. The ‘loop’ is illustrated in Figure 3: the user’s GSR was sensed, processed and sent by network to a control computer which analyzed the GSR stream, placed bookmarks, and made the bookmarks accessible to the user via a graphical list. A user can control the audio stream directly and/or via bookmark selection. Bookmarks can also be placed manually by a user (“Bookmark Functions” in Figure 3). Some parts of the system were developed specifically for this use case (e.g., the audiobook player and bookmark manager). Others, in particular the interruption detection algorithm and software, are of more general applicability. A general overview of the system is provided here.

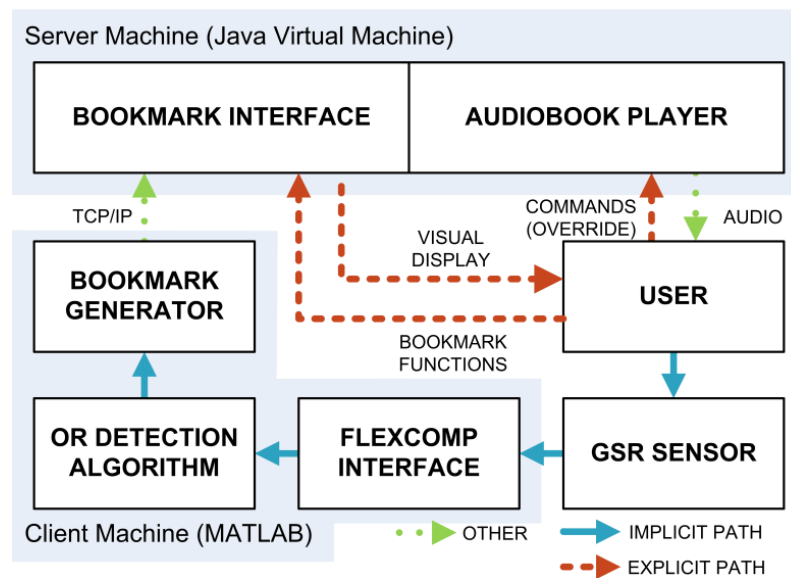


Figure 3. Diagram of bookmarking system architecture.

2.4.1 GSR Apparatus and System Architecture

Our GSR measurements were obtained with Thought Technology’s ProComp Infiniti® physiology-measurement hardware system [61]. The ProComp encoder reads data from a GSR sensor which uses dry electrodes attached to the index and middle fingers of the non-dominant hand. The encoder transmits the filtered, digitized signal to a notebook computer via USB. Skin conductance was measured in microsiemens (μS) and recorded at 256 Hz (the equipment default recording rate).

To test various feature-based OR detection approaches, we developed a distributed system based on a TCP/IP client-server architecture (see Figure 3). The client CPU received GSR data via a custom MATLAB program which performed OR detection and bookmark generation. Bookmarks are sent via TCP/IP to a custom MP3 audiobook player which we developed running on the server notebook. This architecture was designed for flexibility and future use in implicitly-controlled media players in mobile or distributed environments, with the wired USB connection replaced with a Bluetooth link from a wearable sensor.

2.4.2 Audiobook Player and Bookmark Manager

Users interacted with a custom Java audiobook player (Figure 4a), on which the user could place bookmarks explicitly using a custom bookmark-manager graphical interface (Figure 4b). Participants used the bookmark GUI to navigate through existing marks labeled by type (GSR-derived and explicitly user-created) and time. In the future, we plan to render device-to-user communication and bookmark placing through other channels (e.g., haptics) to further reduce demand on a user's visual faculties.

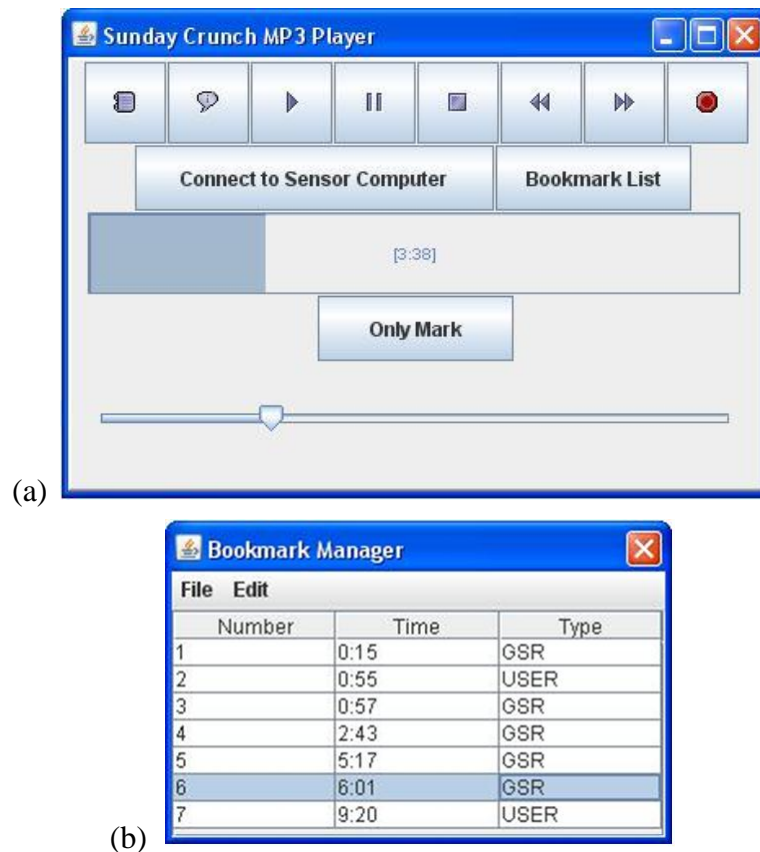


Figure 4. Screen capture of the a) audiobook player; b) and bookmark manager.

2.4.3 Interruption Detection Software

Figure 5a shows a process flow diagram outlining the detection process and subsequent auto-bookmarking behavior, which was implemented in MATLAB. The GSR was measured in real-time, then down-sampled from 256 to 32 Hz to permit online processing (literature and our own results have shown that GSR rise times for ORs are in the range of 1-3s [60]; the Nyquist-Shannon theorem translates this to sampling requirement of >5 Hz). The resampled signal is smoothed by convolution with a 32-point Bartlett window (Figure 5b) to reduce noise in the signal, and then differentiated (Figure 5c) similar to the procedure used by Kim et al. [19].

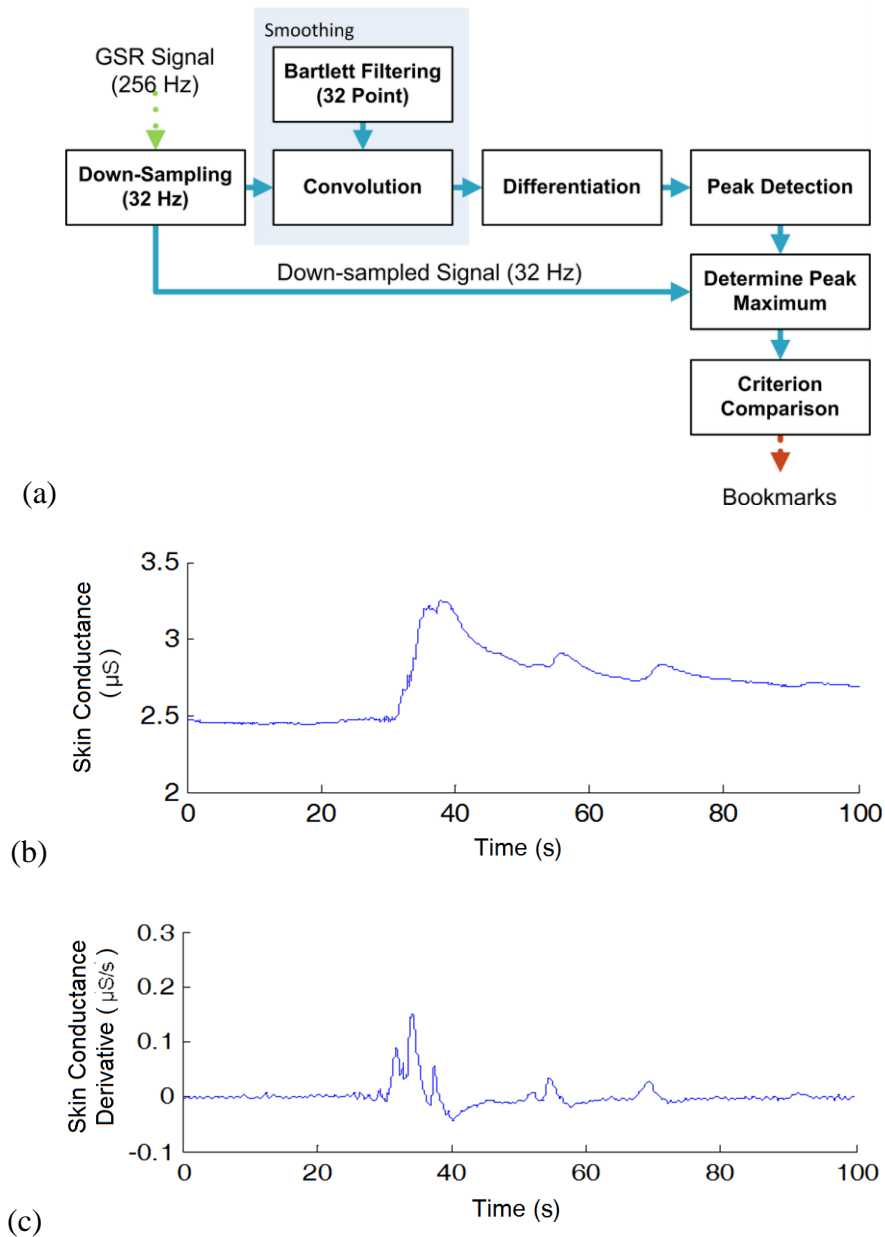


Figure 5. a) Flowchart of OR detection process. b) Typical smoothed GSR waveform (3200 samples). c) Smoothed first derivative GSR waveform (3200 samples).

To detect an OR, we first identified zero-crossings in the first derivative of the smoothed GSR, noting the signal’s direction at these points (‘-’ to ‘+’ or ‘+’ to ‘-’). A positive to negative zero-crossing, signifies a peak in the raw GSR signal. Following this, a bookmark is placed if the following inequality holds true:

$$GSR_{peak} - SMA_{640}(GSR) \geq Threshold \quad \text{Equation 1.}$$

Where *Threshold* is a pre-determined value established in Section 2.5.1, GSR_{peak} is the maximum value of a peak, and $SMA_{640}(GSR)$ represents the moving average of the last 640 samples (20s). Thus, this criteria is met if the peak is sufficiently above the average GSR signal over a 20s window period. To avoid the placement of extraneous bookmarks, new bookmarks were suppressed for a 20s “blackout” window following the most recent placed bookmark. This procedure incorporates contextual information (GSR amplitude) for the signal under investigation, which can vary greatly across individuals and time.

Latency and Window Length

GSR measurements experience a natural latency of 1-4s [60]. The auto-bookmarking system introduces a delay due to differencing, smoothing and other operations, measured at 100ms. Computational latency associated with OR detection was experimentally assessed to range from 0.5-2s due to differences in GSR signal rise time after different stimuli. When summed, there is a 1.5-6s lag between the actual occurrence of a stimulus and the GSR-based recognition of the OR.

We have hypothesized that users could more easily re-orient after jumping to a few seconds prior to a perceived interruption, than to the exact point of interruption. Experimentally, we found that users tended to rewind audiobooks to approximately 10s prior to the start of an interruption³. We therefore placed bookmarks 15s prior to the detection of an OR, corresponding to 9-13s before the actual interruption based on our estimate of recognition latency. Figure 6 shows a timeline of these events.

³ Refer to Section 2.5.2.

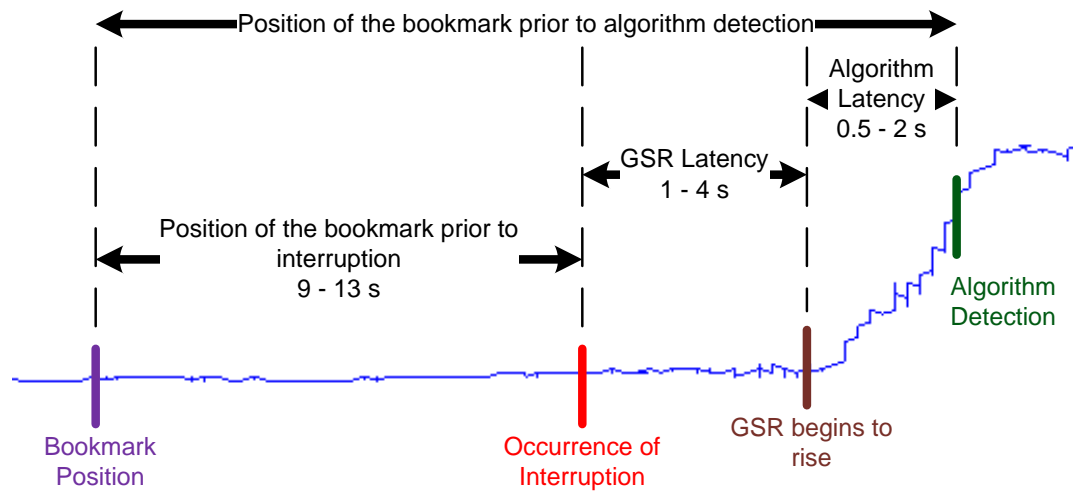


Figure 6. Approximate timeline of the auto-bookmarking system. Horizontal bars indicate timeline variability, black vertical lines show an example timeline for a participant introduced to vocal stimulus.

The choice of averaging window length was the result of a compromise. If the averaging window is too long, false detections arise from large-amplitude, low-frequency GSR changes that are not characteristic of ORs. If the window is too short, the analysis will be too localized and the algorithm will fail to identify ORs. As we were more concerned with failures in detecting ORs rather than producing false positives within this preliminary work (i.e., we preferred sensitivity over precision of our algorithm), longer window lengths were favored over shorter ones⁴. In pilot studies, we tested various window lengths and observed effects on bookmark placement under different interruption conditions. We found that a 20s window provided the most stable algorithm sensitivity; hence, this value was used in Experiment 3 (Section 2.5.3).

2.5 Experiments

Our main objective was to support development of and then validate the general implicit interaction paradigm, within the context of the initial audiobook marking use case. We structured our research questions to be broad, but answerable and relevant from this use case perspective. We conducted three experiments to answer the research questions posted in Section 2.1 above. Experiment 1 (E1) examined the usability of GSR as an indicator of true interruption. We used Experiment 2 (E2) to refine parameters of the bookmark algorithm; e.g., the distance by which

⁴ We realize that by preferring sensitivity over precision, the window length chosen may be overly sensitive to fluctuations in GSR, thus causing a large amount of false positive bookmark placements. However, as we're testing the utility of placed bookmarks in these experiments rather than the effect of false negatives, we felt that this decision was justified in the context of this experiment.

bookmarks should be advanced from the point of measured interruption. In Experiment 3 (E3) we showed our bookmarking implementation to users and invited their qualitative feedback. Finally, in a preliminary pilot experiment (E4), we logged and examined additional GSR data in an uncontrolled, noisy environment to assess the algorithm's viability in more realistic situations.

2.5.1 Experiment 1: Evaluation of GSR Utility

Our first experimental question addressed physiological responses to interruptions - specifically:

***Q1:** Does GSR exhibit a consistent and detectable change in response to interruptions of a level that distract users?*

Bookmarks should be placed at or near the point where the user is interrupted. To do this, we need to know what GSR looks like when the user has really been interrupted. Only some events at some times are interruptive enough to distract users who are focused on a primary task (here, audiobook listening). For example, one may be able to mentally block a nearby conversation in a café, but will be distracted by the sound of a door opening in a quiet room. We would like to differentiate between these two situations using only features extracted from GSR measurements. E1 was conducted to answer Q1 by specifically addressing the following hypotheses:

***H1:** The GSR signal amplitude exhibits a range that corresponds to ranges of interruption levels; and*

***H2:** Disruptive interruptions exhibit a characteristic GSR profile and amplitude range that can be detected with usably high (>80%) true-positive rates.*

This would provide the empirical grounding necessary for development of an OR detection system. An 80% accuracy rate was chosen as baseline as this was comparable to results obtained by Kim et al. [9] for the recognition of three emotion types; thus, we considered this value to be adequate for our purposes practical applications. Usability of detection rates is examined in the Discussion (Section 2.6).

Design

Four female and seven male subjects ($n = 11$) aged 23-30 participated in E1 under consent (Refer to Appendix A for consent form). Subjects were pre-screened for potential confounds including diagnosed attention disorders (obsessive-compulsive or attention-deficit) and

familiarity with the audiobook used through a pre-experiment questionnaire (see Appendix A). Data from one male subject was discarded due to sensor malfunction.

At the beginning of the experiment, half of the subjects were asked to put in their pocket a cell phone provided by the experimenter that was programmed to emit an auditory ringtone triggered by an experimenter. The other half were asked for their own cell phone numbers, then asked to set their cell phones to vibrate mode and place them in a pocket. This was done to determine if there is a net noticeable difference in GSR when a familiar stimulus from the participant's own cell phone is presented versus an unfamiliar one from cell phone provided by experimenters. We observed no discrepancy between the two cases in terms of average peak values of the GSR response.

During the experiment, subjects were asked to sit in a silent experiment room facing a wall and to don a GSR sensor and pair of headphones. To reduce noise in the GSR, subjects were asked to avoid making large physical motions. The entire experiment was video-recorded for post-hoc comparison with GSR data. The experimenters sat on the other side of a visual divider from the subject, as shown in Figure 7, to avoid unintentional distraction or anticipation thereof. To encourage interruption responses typical of focused listening, subjects were instructed to listen to the audiobook carefully as they might be tested on concepts described in the audiobook; no test was actually administered.



Figure 7. Experiment 1 Setup.

After collecting two minutes of baseline GSR data where the participants were asked to silently relax, experimenters started playback of the first chapter of the audiobook “Free” by Chris Anderson [62]. The chapter, which describes the invention and marketing of Jell-O and Gillette

disposable razors, was verified to be of neutral arousal level in a pilot study; that is, listening to the audiobook did not impact the user's GSR signal except at the start or end of play. The content was considered interesting but boring (not inducing any strong, emotional reactions) by most subjects and was reused in E1, E2 and E3.

After approximately a minute of playback, the experimenter began to cause interruptions without pausing or stopping the player. Four different interruptions were used: knocking on the experimenter's desk three times {K}, tapping the subject on the shoulder twice from behind {T}, calling the cell phone in the subject's pocket {C}, and verbally activating the subject {V}. These interruptions were chosen as they were perceived to be common in daily-life settings within the context of an audio-listening task. Each interruption was used twice. Intervals of at least 1 min between interruptions allowed the subject's GSR to settle.

Eight interruptions were presented to all subjects in the following order: {K}-{T}-{C}-{V}-{V}-{C}-{T}-{K}. During piloting, we did not discover any significant difference in any measured quantities as a result of stimulus order in terms of GSR response; therefore, to standardize the experience of the interruption order across users we did not randomize this sequence. Subjects were not informed of the order of interruptions. The same experimenter introduced all interruptions throughout all sessions with consistent volume and tone. The duration of phone rings was constant, and the two verbal instructions in the experiment were scripted as:

V1: "[Subject Name], I just want to let you know that you are doing great."

V2: "[Subject Name], we are going to continue to take more measurements, and let you know when the experiment is over, OK?"

The experiment was followed by a post-hoc questionnaire and a semi-structured interview (refer to Appendix A for questionnaire and interview questions). The questionnaire aimed to discover which interruptions were most disruptive. For each type of interruption, the subjects were asked whether they: a) were distracted, b) found it hard to refocus on the book after the interruption, and c) would have liked to rewind the book to a time just before the occurrence of interruption. In the interview, we collected general feedback on the experiment with a focus on each interruption. Qualitative data from the questionnaires and interview was statistically analyzed to test H1. Quantitative data from the recorded GSR was analyzed with pattern detection algorithms

to test H2. The length of E1 excluding preparation and administration of the questionnaire/interview was 20 min.

Results

Questionnaires & Interviews: Post-experiment questionnaires revealed that 81% of participants agreed or strongly agreed that they paid close attention to content in the audiobook. Of our participant pool, 73% agreed that they were interrupted during the experiment. From a set of yes/no questions for each type of interruptions, 91% of participants agreed that they were interrupted by verbal interaction {V}; 55% of subjects stated that {K}, {T} and {C} were interruptive. Qualitative analysis of the post-experiment interview showed that verbal interruption was found to be the most disruptive; it was also found that {K} did not register as either annoying or disruptive to most participants. Three subjects reported that they did not notice a knock at all. Users found the level of disruption by tapping or cell phone to be in between that of knocking and verbal interruption. Hereafter, we refer to {K} as non-disruptive and {V}, {T}, and {C} as disruptive interruptions.

GSR Signal Analysis: The algorithm's performance was evaluated by sensitivity and precision metrics as defined in Table 1. We considered an interruption "detected" if an OR occurred within -4 to +11s from the presentation of an interruptive stimulus; this range was derived from the sum of the 1-6s detection latency (see Section 2.4.3) and $\pm 5s$ which was allocated to account for variation of when ORs were detected. We tested both the sensitivity and precision of the algorithm's interruption detection. Sensitivity reflects the frequency of true positives, i.e. detection of when an interruption occurred. For example, if 4 out of 6 disruptive interruptions were detected in a single trial, sensitivity = 67% ($N_{TP} = 4$, $N_{FN} = 2$). Precision indicates robustness to false positives. If 4 out of 10 detections correspond to disruptive interruptions, precision = 40% ($N_{TP} = 4$, $N_{FP} = 6$). A false positive proved difficult to define because nonspecific responses – those that occur in the absence of an identifiable stimulus – could be caused by valid internal or otherwise unobservable stimuli.

Figure 8 shows the results of our OR detection algorithm for a typical trial, illustrating disruptive interruptions and detected ORs. Supporting our previous approximation of 1.5 to 6s, the average interval between an interruption's actual occurrence and "true" OR capture was 3.52s ($SD = 3.25$).

Table 1. Definitions of test outcomes, conditions and measures.

Name	Definition
True positive (TP)	A detection occurred within a 15s time window from the start of a labeled interruption (-4 to +11s).
False negative (FN)	No detection occurred within a 15s time window from the start of a labeled interruption.
False positive (FP)	A detection occurred without a corresponding a labeled interruption within a 15s period.
True negative (TN)	No detection occurred when there was no labeled interruption within a 15s period.
SENSITIVITY	Proportion of interruptions that are detected by the algorithm - $N_{TP}/(N_{TP}+N_{FN})$.
PRECISION	Proportion of detections that are interruptions - $N_{TP}/(N_{TP}+N_{FP})$.

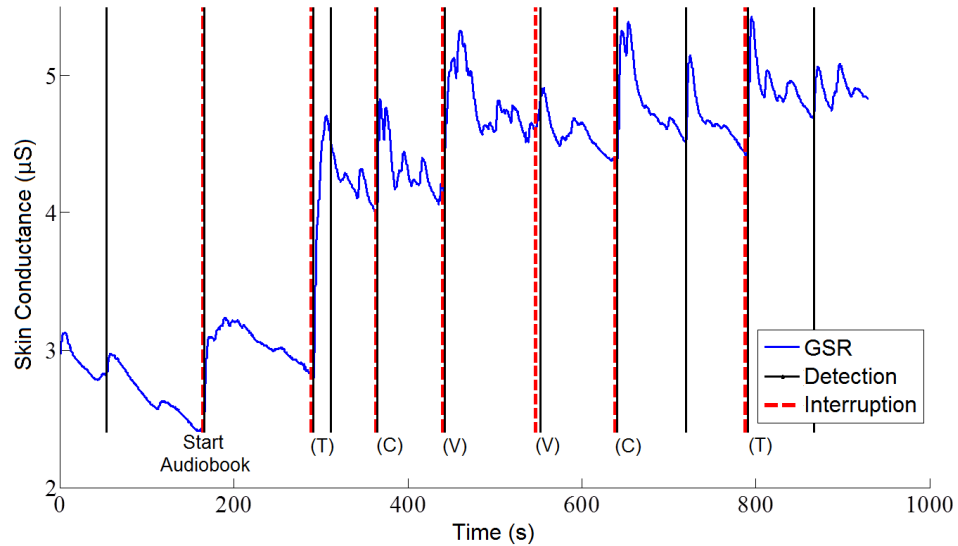


Figure 8. GSR signals overlaid with interruption and detection lines for one representative trial (threshold set at 0.07 μS).

We examined the algorithm's sensitivity to detection threshold by computing the average sensitivity and precision for threshold values ranging from 0.01 to 0.70 μS as shown in Figure 9. Sensitivity and precision are inversely and directly proportional, respectively, to threshold in a roughly linear fashion. A threshold of 0.07 μS (shown by the red line in Figure 9) was found to be optimal in terms of providing a maximum sensitivity of 84%, though at a precision of 32%. For these experiments, we used a threshold value of 0.07 μS since we were more concerned with

detecting the highest number of interruptions over an increased number of false positives to evaluate utility and effect of bookmarks. However, from Figure 9, we see that using a threshold of $0.07 \mu S$ will cause the precision of our algorithm to be at 35%. For application of this bookmarking algorithm in realistic scenarios, we realize that this threshold will be overly sensitive causing a large amount of false positives. Thus, we recommend that the threshold should be increased to provide a more balanced trade-off between sensitivity and precision beyond these experiments. For example, from Figure 9, we see that at a sensitivity of 70%, our precision rises to a more usable 55%.

Still, given this overly sensitive threshold, interruptions were missed by the bookmarking algorithm using a smaller threshold as shown in Figure 10. Upon further analysis, we found that most, if not all, of these interruptions were not detected due to slower rise times of the GSR signal and smaller peaks.

E1 subjects received a total of 80 interruptions, 20 of each type. The average sensitivity of each type of interruption is plotted in Figure 11. Additionally, we calculated the sensitivity of detected interruptions deemed to be disruptive, by excluding the apparently non-disruptive {K} types. With the detection threshold set to $0.07 \mu S$, the average trial sensitivity of disruptive interruptions was 84% (SD = 15%). A two-sample right-tail T-test verified that the sensitivity of disruptive interruptions - {V}, {T} and {C} - is significantly greater than the sensitivity of non-disruptive {K} interruptions ($p=0.02$). However, inspection of GSR data from E1 does not show notable qualitative differences between interruption types.

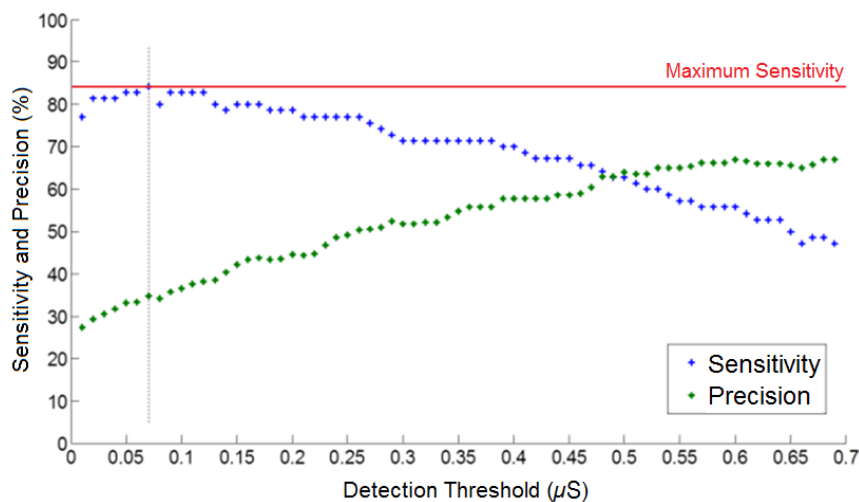


Figure 9. Average sensitivity and precision versus detection threshold. The red line refers to the maximum sensitivity when the threshold is set to $0.07 \mu S$.

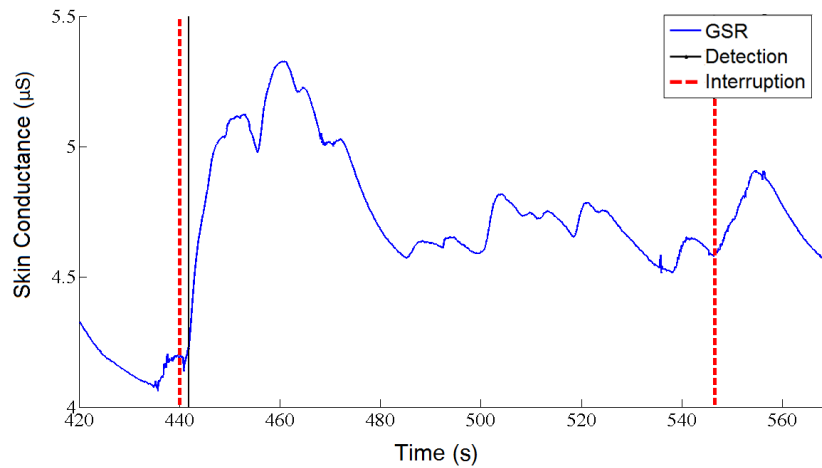


Figure 10. Sample GSR trace showing a missed detection of an interruption.

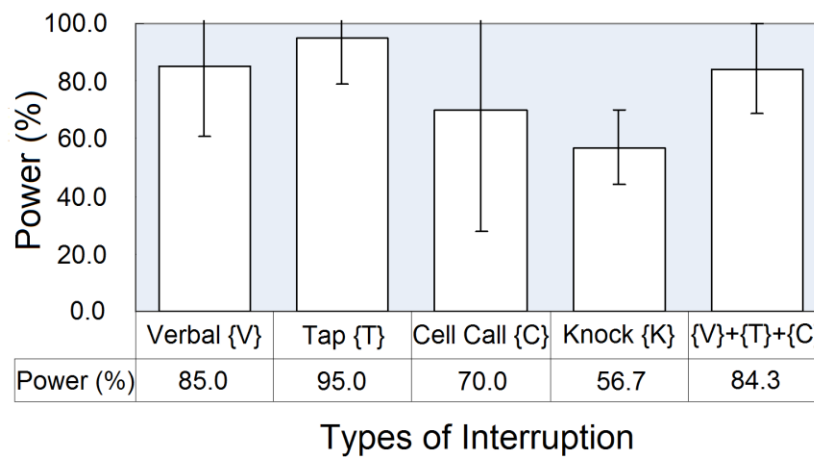


Figure 11. Average algorithm sensitivity (proportion of interruptions detected of total) for interruption types. Error bars represent standard deviation of sensitivity across all subjects.

2.5.2 Experiment 2: Observation of Desired System Behavior

In E2, we sought data that would help guide the design of an auto-bookmarking application for an average user. Specifically,

Q2a: Where should a bookmark be placed for greatest utility?

Q2b: Is there a correlation between interrupt duration and how far back participants rewind the audio stream?

According to E1 results, verbal interruptions (conversation, verbal instructions and questions) were the most disruptive of those tested; verbal interruptions are also controllable in duration. Thus, to maximize the number of rewind events, we used only verbal interruptions in E2.

Design

Four females and seven males (n=11) aged 23 to 30 participated in E2 under consent (see Appendix A for consent form). The same audiobook used in E1 was reused in E2. We required that subjects chosen for E2 had not previously listened to nor read the book prior to the experiment. Additionally, we ensured that no participants from E1 participated in E2 to prevent participants from having prior knowledge and expectation of interruptive events occurring during the experiment.

The setup for E2 is shown in Figure 12. The visual divider used in E1 was omitted as we felt verbal interruptions lacking eye contact was unnatural. A second experimenter recorded interrupt durations. Both experimenters were out of the participant's immediate sight to prevent anticipation of stimuli. Subjects used the MP3 player shown in Figure 4a; the pause function and 'track time' display were disabled to simulate our use case. After a brief training session on the customized player, subjects were instructed to pay close attention to the content of the audiobook, to rewind whenever desired, and that there would be content-related questions following the listening session. The player automatically logged timestamps of user rewind actions. Both visual (i.e., trackbar display) and audio cues (i.e., audio scrubbing) were used to assist subjects in determining how far they had navigated backwards in the audiobook.

Subjects were asked to relax for two minutes at the outset of the experiment to collect baseline GSR data. Experimenter 1 initiated verbal interruptions periodically after the subjects' GSR had settled, with short (2-13s), medium (14-28s), and long (29+s) conversations. A total of six verbal interruptions (two short, two medium and two long) were used for each subject. One medium and one long question were related to the content of the audiobook. Finally, the video recording of the experiment was played back to the subject in a semi-structured interview, to understand the rationale for various rewinding patterns. As in E1, the length of E2 excluding preparation and administration of the questionnaire/interview (and thus the length of the audio stream segment) was 20 min.

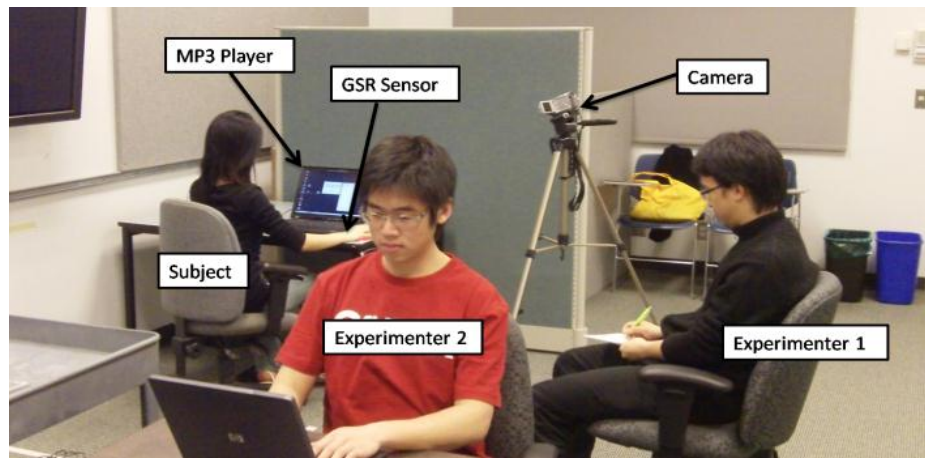


Figure 12. Experiment 2 set-up. A similar set-up was also used for E3.

Results

We expected all 66 stimuli to be interruptive - six per subject across the 11 subjects. 19 of these generated user-driven rewind events that corresponded to an interruption. 33 of the remaining stimuli did not cause any rewind event, and the other 14 were discarded either because subjects reported that they caused unrelated rewinding (i.e., due to mind wandering etc.), or subjects were not able to remember their reasons for rewinding the audiobook at that particular incident during the interview. It was found that users tended to rewind multiple times for a short duration after an interruption for reasons such as seeking for the ‘best’ location. This was still counted as a single valid rewind event.

We measured how many seconds the subject rewound prior to the onset of the interruption. For example, we recorded a value of -2 if a subject rewound the audiobook to a position 2s prior to the start of an interruption. The average rewind period during the 19 valid interruptions was found to be -9.37s ($SD = 7.65$), i.e. on average, a subject rewound the audiobook to 9.37s prior to the occurrence of the disruptive interruption event. We also recorded the duration of each interruption. The duration of the 19 valid interruptions ranged from 7 to 56s. The correlation coefficient between interruption length and extent of the rewind event was $r=0.03$, which suggests that the rewind distance was unrelated to length of interruption.

2.5.3 Experiment 3: Interaction Utility Assessment

In E3, we tested the utility of our tuned GSR-based auto-bookmarking algorithm. E3 re-used E2’s procedure and set-up (Figure 12) but employed the bookmarking algorithm in real-time. Data from E3 provided insight into user reactions to the bookmarking prototype.

Design

One female and four males (n=5) aged 23-29 participated in E3 under consent (refer to Appendix A for consent form). The system used in E3 is described with its parameterizations in Section 2.4. While the previous experiments provided us with information on how to detect disruptive interruptions and how to use that information, E3 specifically addresses the question:

Q3: Do our system-generated bookmarks consistently provide utility after interruptions?

We used the same pre-screening procedure and setup as in E2. In addition, E3 subjects were asked to use the custom MP3 player and associated custom bookmark manager interface (Figure 4b). Subjects were instructed to use this manager as their first method of rewinding when rewinding was desired – reverting to manually navigating through the audiobook only if the bookmarks were unsatisfactory. Subjects could rewind from any bookmark, but were advised to try the most recent one first. Interactions with the MP3 player and bookmark manager were logged, and video and audio recordings of the experiment were saved.

To better understand their strategies for interrupt recovery, we asked subjects to view the recorded video stream while commenting on each interruption in a post-experiment questionnaire. Finally, we conducted an interview to understand user satisfaction and usefulness of each bookmark (refer to Appendix A for questionnaire and interview questions).

Results

Each subject in E3 was exposed to six interruptive stimuli, for a total of 30 events. 26 of these stimuli were reported as disruptive, valid interruptions. 21 of these caused subjects to rewind immediately following the interruption. 18 of the 21 rewinds utilized only system-generated bookmarks; in 13 of these, subjects used only the latest system-generated bookmark, whereas subjects in the other five instances continued to try older bookmarks. In the remaining three rewind events, subjects manually navigated through the audiobook. Subjects rated 76% of the automatically-placed bookmarks that they used as ‘appropriately positioned’.

The semi-structured interviews provided constructive feedback: users were surprised to see such a system, and were interested in seeing further application areas of such interaction. In general, subjects provided positive feedback on the system’s usefulness and utility. Few suggestions on bookmark follow-up were raised. Two subjects suggested the system should pause when the user is interrupted, and to play again by user's manual control. One subject reported that the system’s visualization and/or interaction quality could be improved.

2.5.4 Experiment 4: GSR Field Assessment (Pilot)

As a final early validation step, we assessed the potential effectiveness of this implicit interaction implementation in a less controlled environment. We were keen to preview through a pilot what would be involved as we moved to the field to guide our own future work. Specifically, we wanted an initial indication of:

***Q4:** Are there quantifiable differences between GSR measurements obtained from a controlled vs. uncontrolled, possibly noisy environment which may affect utility or effectiveness of the interaction?*

Design

Two male subjects (n=2) aged 23 participated in the E4 pilot. During rush hour at a busy outdoor bus terminal, each participant was seated (alone) on a bus bench for the duration of the study and asked to attend to an audiobook played through headphones. Subjects wore a GSR sensor and were instructed to use a pushbutton marking device to signal moments where they felt interrupted enough by an event to wish the audiobook to be rewound. An experimenter (out of the participant's view) also recorded major events that could have potentially caused ORs.

Results

An informal analysis of data from two male participants showed little qualitative (wave shape, event responsiveness) difference between raw GSR measurements obtained from our controlled tests (E1 and E2 results) and the bus terminal. The primary distinguishing characteristic was absolute GSR amplitude: subjects in E1 and E2 tended to exhibit baseline and OR GSRs in the range of 3-6 μ S, whereas GSR measurements in E4 were 6-15 μ S; an example comparison is shown in Figure 13. As the software algorithm designed for the auto-bookmarking use case does not utilize absolute signal amplitude, but rather amplitude of peaks relative to the moving average, this preliminary result suggests that system operation should be minimally affected by a chaotic, noisy environment.

Although E4 results are informal and preliminary, this pilot study suggests (although does not prove) that at the cost of more sophisticated signal processing, information regarding ORs will still be present in GSR data collected from less controlled environments.

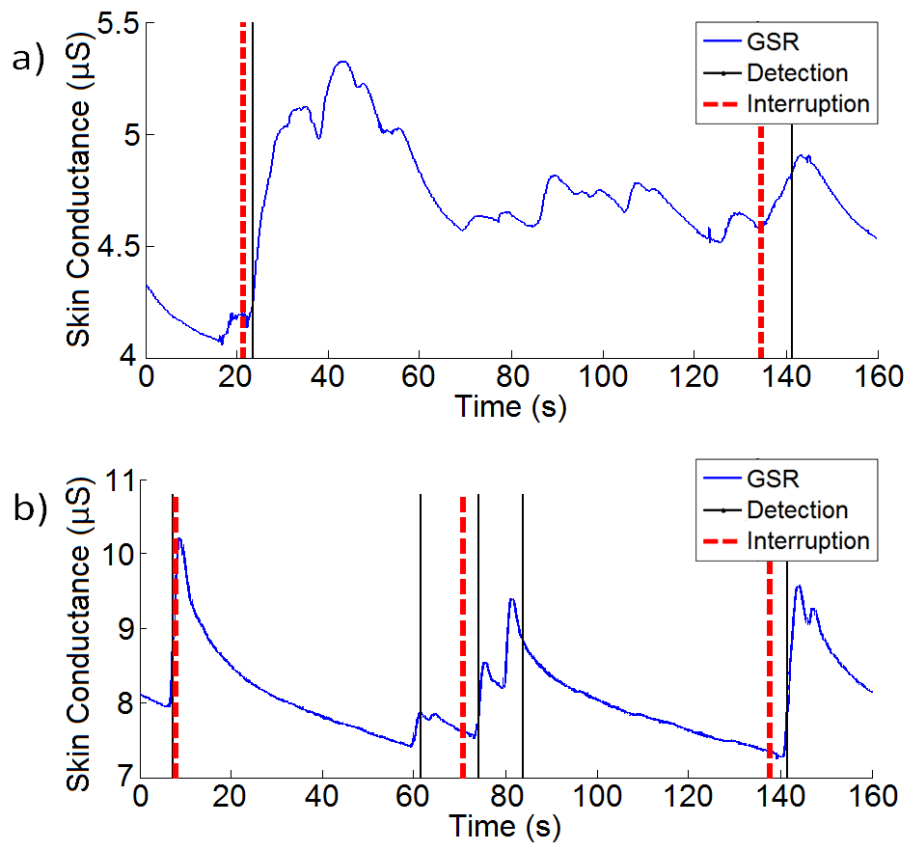


Figure 13. Raw GSR data collected from a participant in a) a controlled test environment and b) a bus terminal at rush hour. Thick, dotted lines represents interruptions; thin, solid lines represent bookmarking system OR detections.

2.6 Discussion

Q1: Does GSR indicate disruptive interruptions?

E1 showed that GSR can be used to detect disruptive interruptions in an audiobook listening context. According to subject interviews and self-reports, our {V}, {T} and {C} interruptions did disrupt subjects. E1 results (Figure 11) confirmed that different types of interruptions correspond to different levels of disruption as indicated by GSR. Thus, H1 – our hypothesis that the GSR signal amplitude exhibits a range that corresponds to ranges of interruption levels – is confirmed for the range of interruptions that we tested. Moreover, for GSR data obtained in E1, we achieved an 84% recognition rate in detecting disruptive interruptions, (i.e., true positives of type {V}, {T} and {C}). This supports H2, where disruptive interruptions produce salient GSR signal features, detectable by our algorithm with an acceptable true-positive rate; false positives were difficult to detect, as explained in Section 2.5.1.

Only four types of interruption stimuli were tested in these experiments; the design space of possible interruption stimuli and contexts is much larger. As such, validity of our results is

coupled to the primary task and interruption stimuli used here. For example, while a gentle knock was not disruptive in this context, it certainly could be in other situations. Nevertheless, our approach – evaluating the GSR signal in context and tuning an algorithm variant to that context – is perhaps general enough to be applied to other primary tasks such as watching television, housecleaning, gardening, working at a computer and perhaps driving.

As seen in Figure 9, there is a tradeoff in sensitivity versus precision as a function of detection threshold, i.e. between false negatives and false positives. Based on subject interviews, we believe that higher sensitivity with slightly lower precision is a reasonable tradeoff. However, the optimal balance will depend on the environment and individual responsiveness to ORs.

Q2: Where should a mark be placed relative to interruption?

E2 provided data for designing the bookmark placement system. Its results indicated that users preferred to rewind the audiobook to a location approximately 10s prior to the onset of an interruption, independently of interruption length. This implication for mark placement advance may be applicable to interruptions in other use contexts. Conveniently, the results also imply that it is not crucial to measure or estimate the length of interruption, a potentially challenging task in the field. However, we believe this finding only applies to interruptions under a minute. Longer interruption may affect user’s behavior and the desire of bookmark location; for instance, a 10-minute interruption may cause the user to replay the entire chapter of the audiobook, because the content has slipped out of his short-term memory. Likewise, our observation of mark-advance independence from interruption length may apply to other use cases, but this will need to be verified.

Q3-Q4: Do our system-generated bookmarks consistently provide utility after interruptions? Is GSR informative in uncontrolled environments?

GSR signals are known to be susceptible to noise from sources such as bodily motion and cognitive workload [60]. We generally found (E1-E3) that our algorithm was quite robust to small hand or arm movements and subtle posture changes, but we had false-positive detections in the absence of an identifiable stimulus. Although it is not feasible to explain each one, we expect that some occur through internal distractions, such as when a phrase or a word in the media being perused initiates mind wandering. E4 results suggest that GSR is robust in a chaotic environment.

Subjects did not always find rewinding the audiobook necessary despite being interrupted – particularly for short interruptions (<11s – as determined by the length of the longest ‘short’ interruptive conversation in E2 without causing users to rewind). We found that no rewinding occurred for 81% of the short interruptions presented to subjects in the experiments. Subjects reported that even if they stopped listening to the audiobook to attend to short interruptions, they could quickly pick up the gist of what was being said afterwards. We believe this was representative of this use case, and it seems that users sometimes use a ‘multi-tasking’ strategy to cope with interruptions.

E3 and E4 served to briefly explore the potential of our implicit control algorithm in controlled and uncontrolled setting; the small sample size limits our ability to provide any concrete evidence pertaining to the usefulness our system in the field and thus caution is required when evaluating the significance of these results. They give some indication that our algorithm provided a usable detection rate; further investigations with larger sample sizes are required. Subjects reported appreciation of, and surprise at, the presence of automatic bookmarking. Perhaps most relevantly, many comments were directed to bookmark follow-up as opposed to the actual bookmark generation, supporting the conclusion that the current bookmarking algorithm is reasonably accurate. We see here that our vision of a loop-closing system is of potential utility and a natural next step is to concentrate on *using* the bookmarks themselves.

Our next step in this work was to provide users with a low-attention notification that the application has acted in response to the user’s physiology (for the use case presented here, that a mark has been placed). The following chapter describes our continued work in developing low-attention display of bookmarks to users.

3 Exploring the Role of Haptic Feedback in Physiologically-Triggered Bookmarking⁵

In Chapter 2, we described two scenarios within an audio stream listening scenario that emphasize how a user's attention could be demanded by an explicit interface and the undesirable result if the user does not fulfill this demand. In these scenarios, the user is forced to interact with the device regardless of environmental context or else is penalized. To further our minimal-attention goals, we have suggested using haptic feedback to provide immediate and unintrusive indication of the device's response to the user's state: a reassuring confirmation which effectively closes the interaction loop as shown in Figure 1. In this work we explore the implementation and effect of this feedback on the implicit interaction experience within the context of this audio stream use case presented in Chapter 2. Given this new interaction paradigm, we present a revised trajectory for our use case:

... her focus shifts. A sensor embedded in her bracelet captures this, signaling the system to mark the audio stream which continues to play. She feels a small pulse on her wrist that signals the placement of a bookmark. She follows the hygienist, pulling out the earbuds when it is convenient. Later when alone again, she scrolls through the audio stream, feeling for discrete vibrations on her wrist indicating bookmarks. She jumps back through two or three automatically placed marks, locates the one placed just before the interruption where she stopped listening, and continues playback from there.

In this scenario, we employ a user's electrodermal activity measurement to drive an auto-bookmarking system for audio streams. Bookmarks are algorithmically placed when an *orienting response* - an organism's immediate response to a change in the environment - to an interruption is detected. We posit that haptic feedback will be usable *at the time of interruption* to inform users when bookmarks will be placed, providing reassurance to users that they can attend to the interruption without worrying about losing their place in the audio stream. We also theorize that haptic signaling will be useful *at resumption of the audio stream listening task* to indicate

⁵ A version of Chapter 3 has been submitted for publication: [M. K. X. J. Pan], J. McGrenere, E. A. Croft, and K. E. Maclean, "Exploring the Role of Haptic Feedback in an Implicit HCI-Based Bookmarking Application," Submitted, p.12, 2012.

positions of automatically-placed bookmarks when navigating through the audio stream, augmenting bookmark display in the visual channel as shown in Figure 14. We chose this use case as our representative scenario as it is able to encompass the main mechanisms and effects of the implicit interaction loop - it provides an example of a mobile application which can benefit from non-explicit communication and haptic feedback. Other examples can be envisioned in the design space of in-vehicle information displays and workplace environments where multi-tasking causing attention fragmentation is common.

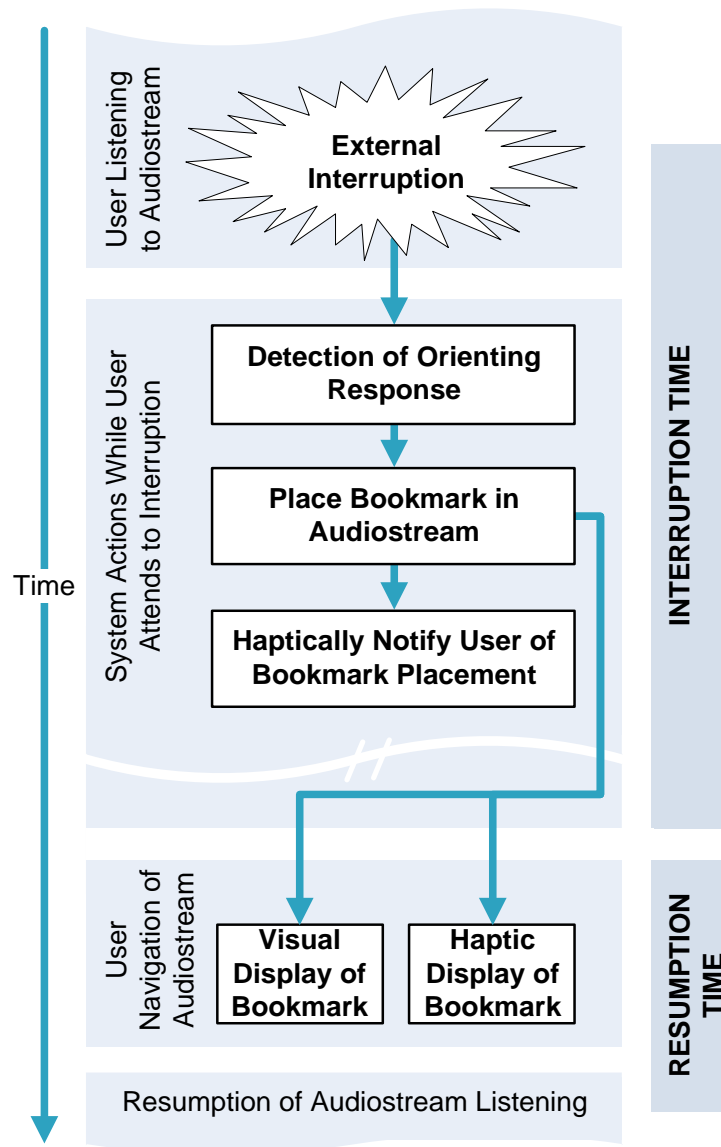


Figure 14. Flowchart showing progression of events in the audio stream bookmarking use case.

3.1 Purpose

The purpose of this work is to present a complete implementation of the HALO paradigm within the audio stream bookmarking scenario, and investigate its efficacy and viability through user testing. We have developed a method of detecting orienting responses to external interruptions that automatically bookmarks streaming media such that the user can attend to the interruption, then resume listening from the point he/she was interrupted (described in Chapter 2). That work substantiated the feasibility of an interaction based on implicit, low-attention user control by demonstrating that the required information is available with reasonable reliability from conventional physiological sensors. However, it left as a next step testing of the closed-loop low-attention interaction paradigm complete with confirmatory haptic feedback. Here, we have run an experiment to test the usability, appropriateness and perceived value of haptic feedback which notifies a user of bookmark placement after an interruption has been detected, and assists the user in navigating to the bookmark of interest.

3.2 Structure

This chapter will provide an overview of prior work in the areas of implicit HCI, low-attention systems, and haptic feedback techniques (Section 3.3). This is followed by a summary of our questions (Section 3.4), a brief description of the hardware/software systems used to explore our interaction paradigm (Section 3.5) and methods used to test our research questions regarding the use of haptics in an implicit HCI system (Section 3.6). We present the results of our experiment (Section 3.7) followed by an examination of how they apply to our original research questions (Section 3.8).

3.3 Background and Related Work

Here, we review several areas of research relevant to, and supportive of, the work presented in this chapter in addition to literature presented in Section 1.4.

3.3.1 Low-Attention Haptic Systems

A review of prior work indicates that much work has been done in the area of low-attention haptic systems in the context of mobile environments; and suggests that haptic feedback augmenting the display of bookmarks in an audio stream could be very effective. MacLean presents a comprehensive review in this area in [27], a few highlights of which are mentioned here.

Using haptic feedback with, or in lieu of, visual displays has the potential to reduce overall mental workload relative to purely visual display of that information, if properly designed [25], [63–66]. For example, Leung et al. found that adding haptic feedback to visual elements such as graphical icons on a touch screen interface made these elements much more usable, especially when users were under cognitive strain [67].

Evidence on the effectiveness of haptic feedback when provided without a visual interface has been mixed; however, Swerdfeger showed that users are able to perceive, learn and interpret ordinal and icon data haptically, even in high workload environments where the user is engaged visually with another task [68]. More recently, Pasquero et al. demonstrated that a haptic wristwatch could successfully be used to convey information in a work environment with simple eyes-free interactions [69]. Brewster et al. and van Erp et al. both established the utility of short, tactile messages or ‘tactons’ to unobtrusively display contextual information to users in attention-demanding environments [22], [23]. Through several experiments where users were performing a primary task (typing or driving), various tactons were signaled to participants indicating the occurrence of ambient events such as the proximity of a friend in a plaza, or upcoming course changes during in-vehicle navigation. Both studies showed that presenting tactile information caused faster reaction times, lowered mental effort and reduced workload.

Taken all together, these various results demonstrate that non-task critical information about ambient events can be presented without disturbing the user’s primary interaction. We are seeking to use informative tactile icons to unintrusively provide temporal information to a user, which is a simpler problem than displaying information requiring an extensive set of tactile icons.

3.3.2 Wearable Haptics

Several research groups have been endeavoring to develop haptic technologies which can be worn unintrusively as documented by MacLean in [70]. For example, as mentioned earlier in this review, Pasquero has developed a wearable wristwatch able to transmit information tactically [69]. Additionally, van Erp et al. has prototyped a tactile belt display for waypoint navigation and explored the ways that distance and direction information can be coded [71]. In examining the design of haptic signals for low-attention notification, Baumann et al. has constructed low-cost, disposable haptic wearables in an effort to explore the design of socially-gradable haptic displays and icons [26]. Peck and collaborators have studied aesthetic touch in the context of

marketing, finding the features that tend to invite touching as well as individual differences in preference for touching [72], [73]. This work has demonstrated users' ability to associate emotional content with haptic sensations and presents a variety of haptic signal profiles that have been associated experimentally with affective descriptions such as 'relaxed' or 'agitated'.

We build on this prior work to demonstrate and test an interaction model that supports online, closed-loop transition support between foreground and background tasks. Our HALO paradigm brings together iHCI, affective computing and low attention haptics to create a new, fluid, low attention iHCI that minimally disrupts the user. This research is novel in that we are examining the role of haptic signaling *specifically for closing an implicit or low-attention interaction loop*.

3.4 Research Questions

The main objective of this study was to understand the benefits and limitations of the general implicit interaction paradigm with haptic feedback, in the context of an audio stream bookmarking use case. To this end, we developed four research questions that when answered, would provide an understanding of how users incorporate haptic information in the closing of an implicit interaction loop. Perhaps more importantly, we wanted to capture whether or not haptics could be utilized and accepted as a method of proactively signaling the user of changes in device behavior in response to changes in user physiological behavior.

- Q1. Do auto-generated bookmarks provide assistance to users in the context of interruption?*
- Q2. Does haptic notification of bookmark placement at interruption time provide value to users?*
- Q3. Does haptically-assisted navigation of bookmarks during post-interruption, resumptive navigation provide value to users over no display or visual display of bookmarks?*
- Q4. Do users feel that the auto-bookmarking/haptic confirmation and navigation feedback system is useful?*

Although these research questions were asked in the context of the audio stream-listening use case, we applied the implications of this study with a more general scope. With this knowledge, we believe it will be possible to construct a fundamentally different approach of device interaction employing closed-loop low attention and implicit communications.

3.5 Experimental Apparatus

3.5.1 Overall System

To test our theories, we developed an audio player system (more fully described in Chapter 2). In this prototyped interaction, a user's galvanic skin response (GSR) is monitored for orienting responses to external interruptions which might redirect the user's attention. Upon detection of an orienting response to an external interruption, our prototype automatically places a bookmark in the audio stream and signals this action to the user unobtrusively through a subtle haptic icon. At this point, the user can direct attention to the interruption without worrying about losing his/her place in the audio stream. Afterwards, the user can resume listening by navigating back to a bookmark signaling the point where he/she was at in the audio stream before being interrupted. The interaction 'loop' architecture that we have implemented for this use case is illustrated in Figure 15. The detection and bookmarking processes are shown concisely in Figure 5 and fully documented in Chapter 2. An overview of the experimental system used in this chapter is presented below (Sections 3.5.2 and 3.5.3).

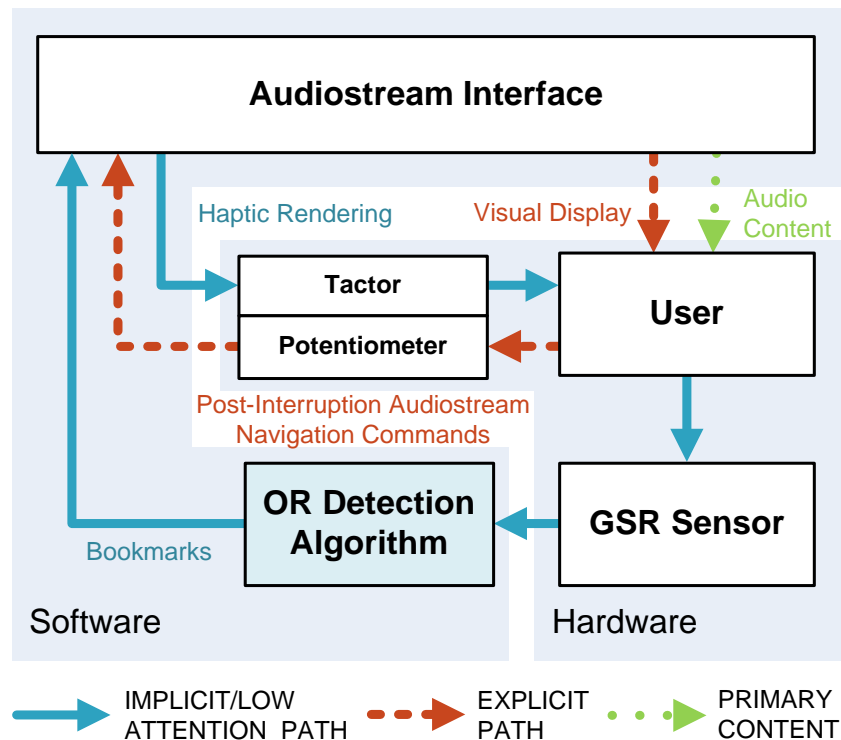


Figure 15. System flowchart.

3.5.2 Hardware

A wrist-worn navigational device was prototyped to provide an interactive interface for which users could navigate through the audio stream intuitively, as shown in Figure 16. The prototype consists of a rotary touch potentiometer placed on the face of the device to supply users with a method of navigating through the audio stream. Users are able to skip through the audio stream by moving their finger in a circular motion; a clockwise finger swipe around the potentiometer causes fast-forwarding while a counter-clockwise motion rewinds the stream. We programmed the navigational interface such that movement speed of the finger directly correlates to the rate of fast-forwarding/rewinding. A button located at the center of the top face allows users to play or pause the audio stream. To provide haptic feedback, an Engineering Acoustics C-2 Tactor [74], a voice coil transducer in which signal frequency and amplitude can be controlled independently, was fitted into the back of the wrist device where it contacts the skin. An Arduino LilyPad drives these components and communicates with our media player software on a nearby experiment laptop via USB (in deployment, the loop could be executed on a mobile handheld device). Users wore the device on the wrist used for a wrist watch – in all cases, the non-dominant hand.



Figure 16. Exploded view and photo of wrist-worn navigational device and haptic display.

GSR measurements were collected using Thought Technology's ProComp Infiniti® physiology-measurement hardware system [61]. Dry electrodes were attached to the index and middle fingers of the non-dominant hand. The ProComp encoder transmits data obtained from the

electrodes to a notebook computer via USB. Electrodermal activity was measured in microsiemens (μS) and recorded at 32 samples per second which was the sampling rate used in work presented in Chapter 2.

3.5.3 Software

We developed a custom audio player based on the freeware cross-platform irrKlang library, with the ability to display GSR-based bookmarks as visual and haptic icons [75] – different from the media player presented in Chapter 2. Its graphical user interface simulated typical mobile smartphone functionality, in choice of audio player viewable area and by using a simplistic design with only a start/pause button, current track position and time, and the name of the track being played (Figure 17).

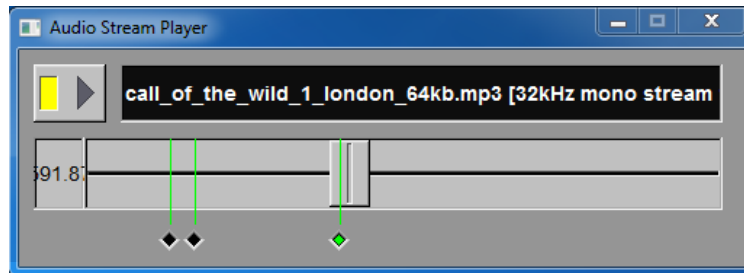


Figure 17. A screen capture of the media player displayed to participants. The scroll bar represents the current location in the audio stream with respect to the current track. Automatically placed bookmarks are shown as green lines above diamonds. Green diamonds signify that the user has navigated to that bookmark.

3.6 Experimental Methodology

To answer our research questions we conducted a controlled lab experiment to clarify the impact of auto-generated bookmarks, haptic notification at interruption time, and haptic assisted bookmark navigation during navigation time.

3.6.1 Participants

16 participants volunteered (6 female), aged 20 - 28 ($\bar{x} = 24.5, \sigma = 2.03$). Participants were recruited through on-campus advertising, and thanked with snacks for their time. All participants were pre-screened for potential confounds including diagnosed attention disorders (e.g., obsessive-compulsive behavior and attention-deficit disorders) and signed a participant's consent form (refer to Appendix B for consent form).

3.6.2 Conditions

The experiment conducted was a 2x2x2 factorial design. As shown in Figure 18, factors consisted of the presence/absence of haptic notification of bookmark placement during listening (tested between subjects to reduce experiment size) and haptic and visual bookmark displays later during resumptive navigation (tested within subjects). A balanced Latin square was used to counterbalance carry-over effects (e.g., interference, learning, fatigue effects).

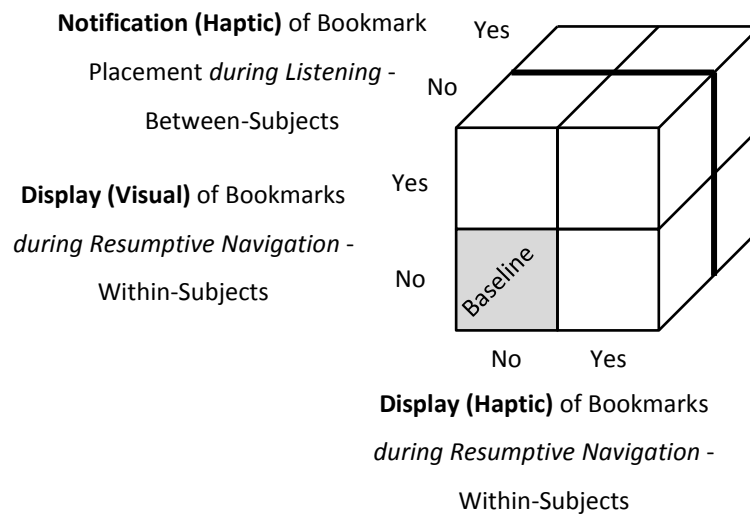


Figure 18. Diagram showing factors within the experiment conducted.

Haptic Notification at Interruption Time (Between Subjects)

We designed a haptic signal to provide notification of bookmark placement to the user at interruption time (as detected by the system in the form of a user’s orienting response). We envisioned that haptic notification would serve two purposes during the interaction. First, the signal could unobtrusively notify and *reassure* the user that a bookmark had been placed and no immediate action was required while he/she attended to the interruption. Second, notification might facilitate search of correctly-placed bookmarks amongst false positives. For example, imagine that a user is peripherally aware of a notification that a bookmark has been placed at the beginning of an actual interruption, and then another (incorrectly) notification a minute or two later. At resumption of audio stream listening, the user can navigate directly to the first (correct) bookmark rather than a second, incorrectly-placed bookmark.

To generate the best ‘feel’ for these purposes, guided by past work suggesting a user assessment of slow, sinusoidal haptic signals as ‘relaxed’ and ‘reassuring’ [26], we chose a sinusoidal pulse to indicate bookmark placement. The pulse was generated by varying the frequency of the factor

vibration within a sinusoid envelope. Based on pilot testing with 4 subjects, we limited the maximum frequency of vibration to 200 Hz and the maximum duration of the pulse to 0.75 seconds. These values were found to provide the best compromise for noticeability and unobtrusiveness.

Display of Bookmarks during Resumptive Navigation (Within Subjects)

We also needed to provide low-attention feedback to users concerning bookmark placement during search-driven movement through the audio stream during navigation time. Currently, MP3 players and smartphone apps which support manual bookmarking of audio streams use some form of graphical display to present bookmarks to the user; while intuitive, this requires visual attention. We speculated that haptic-enabled navigation of bookmarks would allow eyes-free interaction with the system, as claimed by Pasquero et al. [69]. Additionally, we sought to discover if haptic coupled with visual display of bookmarks and/or haptic notification at interruption time could offer a performance advantage over visual or haptic displays individually.

Pilot study results suggested restriction of tactor vibration to a single frequency for the haptic display of bookmarks, based on criteria of salience. Vibration frequency was set at 230 Hz, offset from the frequency of 250 Hz at which the Pacinian corpuscles (nerve endings in the skin which are sensitive to vibration and pressure) are most sensitive, as most participants found 250 Hz too strong. The maximum duration of the pulse was set to 0.5 seconds.

3.6.3 Apparatus

The study was conducted using the bookmarking system and equipment described in Section 3.4. The software was run using a Core i5 laptop with 4 GB of RAM, running Microsoft Windows 7, and connected to a 17" LCD Monitor at 1280 x 1024 resolution.

An experimenter control GUI was implemented to provide a comprehensive view of the system settings including haptic bookmark notification and navigation settings, the user's GSR trace, and GSR bookmarking algorithm sensitivity/thresholding adjustments (Figure 19). In addition to these features, the interface also provided the experimenter with the ability to generate bookmarks and control over bookmark notification/display factors during an experiment trial. When the experimenter selects the visual bookmark display option, the user's audio player window displays bookmarks shown as solid lines directly on the track progress scroll bar as

shown in Figure 17. An icon below the bookmark line lights up when a user navigates to a particular bookmark.

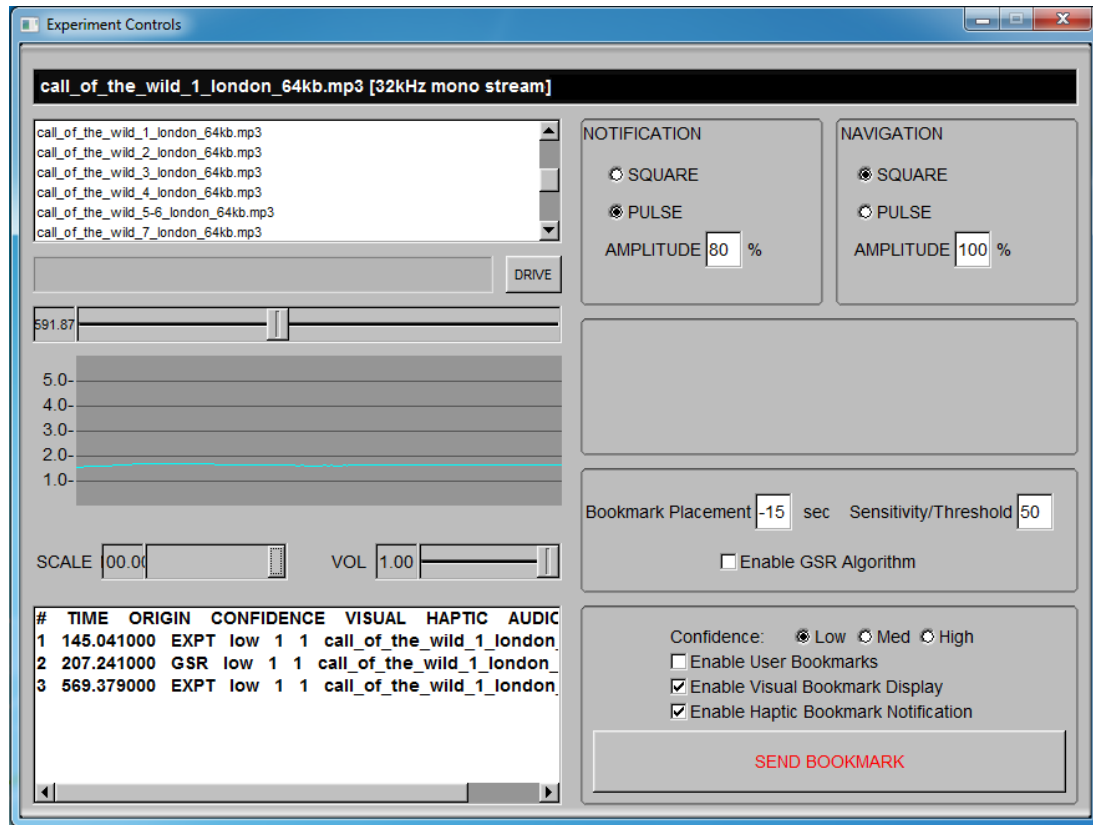


Figure 19. Screen capture of the experiment settings window (not seen by participants). This window provides the experimenter with displays such as currently playing track, track time, the user’s GSR activity, a list of bookmarks that have been placed, options regarding how bookmarks are placed, and how they are presented to participants.

3.6.4 Task

During each session, participants experienced sixteen trials, each 2-3 minutes in duration; in each trial, they carried out a two-part task. The first part was to listen to an audio stream (a podcast segment). They were instructed to attend to interruptions by the experimenter (diverting attention away from the audio stream) – once approximately every 2 minutes (± 1 minute). In the second part following the interruption, participants had to navigate through the audio stream to locate the point they were listening prior to the interruption, given a set of feedback parameters which always included audio stream playback.

Each participant listened to two 20-minute fictional narratives entitled: ‘Dave and the Dentist’ and ‘Sam the Athlete’ in a single session (see Figure 18 for the streams’ allocation to experiment conditions). These stories were obtained from a podcast radio series called The Vinyl Café, a Canadian variety show which features light-hearted stories describing the misadventures of a

fictional family [76]. Unlike the audio content used in Chapter 2, this content was chosen to engage participants in the audio stream content; we were not concerned with GSR artifacts caused by the audio stream itself in this study. The order in which these two narratives were presented to participants was counterbalanced.

Listening and Interruption Phase

A trial comprised of a verbal interruption to the participant listening task resulting in a 50-60s conversation between the experimenter and subject. 16 pre-arranged conversation-starting ‘scripts’ (one for each trial) were used to ensure that the subject matter of the conversations was the same across all subjects; however, ordering of these scripts was randomized across participants. At the beginning of each interruption, a bookmark was placed by the experimenter; bookmark placement was experimenter-simulated to ensure the experiment was properly counterbalanced and to control when bookmarks were placed.

To simulate realistic performance of the automated bookmarking system described in Chapter 2, we provided false-positive bookmarks and their associated notifications/representations (where applicable) during the experiment to test participants’ reactions to them and navigation performance. Two false positive bookmarks were presented to participants at random times in half of the trials under conditions where bookmarks were presented (two trials out of the full four per condition for each condition except for D_{None}). Thus, the precision (the ratio of bookmarks that were correctly placed to the total number of bookmarks placed) for a full participant experiment was 50%. False negatives or ‘miss’ type errors (failure of bookmark placement for a real interruption), were not introduced to participants within the D_H , D_V , and D_V+D_H conditions on the premise that the D_{None} condition essentially emulates false negative errors; hence, we determined that we would not have obtained any useful additional data by introducing these errors into the other conditions. Thus, bookmark placement sensitivity (the ratio of bookmarks that were correctly placed to the total number of interruptions) is 100% for D_H , D_V , and D_V+D_H conditions, and 0% for the D_{None} producing an overall sensitivity of 75%. From Figure 9, we see this combination of precision and sensitivity is consistent with the middle range of precision vs. sensitivity tradeoff (at around 0.35 μ S) obtainable by our GSR bookmarking algorithm described in Chapter 2.

Two bookmark placement notification groups were tested in this experiment during the listening and interruption phase:

- Haptic notification of bookmark placement at interruption (N_H)
- No haptic notification of bookmark placement at interruption (N_{None})

Resumptive Navigation Phase

Following the interruption and conversation with the experimenter, a participant would immediately start scrolling through the audio stream, attempting to navigate to where they believed that they had left off just prior to the interruption; this marked the beginning of the resumptive navigation phase.

Four resumptive navigation conditions were tested in this experiment during the resumptive navigation phase:

- visual display of bookmarks are present during navigation only (D_V);
- haptic display of bookmarks are present during navigation only (D_H);
- both bookmark displays are present during navigation ($D_V + D_H$); and,
- no bookmarks are displayed during navigation (D_{None}).

The order in which each bookmark display type was used was determined by balanced Latin square. Each bookmarking condition was presented to the participant four (4) times.

An example experiment timeline is shown in Figure 20. Half of all participants received haptic notification when bookmarks were placed at interruption time. Each trial session was video recorded for post-hoc analysis to determine navigation times for each display.

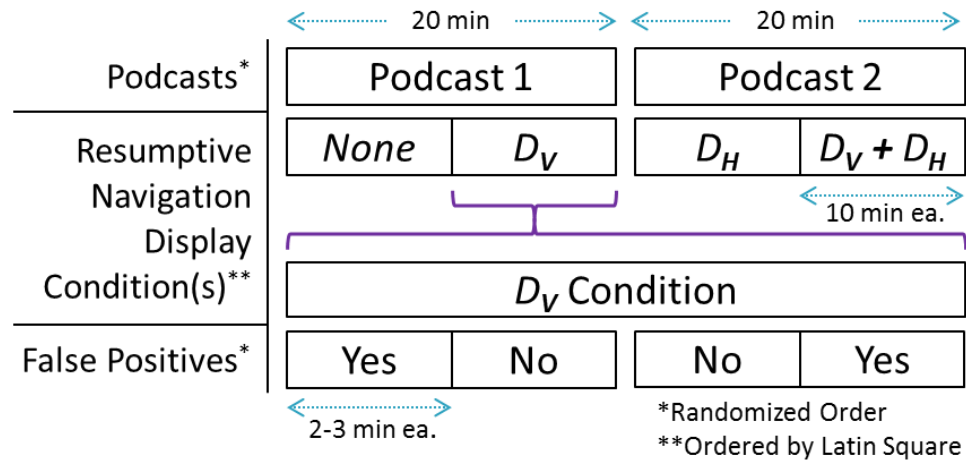


Figure 20. Representative timeline for data collection portion of experiment, for one participant. Factor levels shown are for the within-subject navigation display conditions only; haptic notification display is between-subjects.

3.6.5 Procedure

The experiment was designed to fit within a 1.5 hour session. At the start of the session, participants were introduced to the GSR sensor, the audio player software and the equipment setup. Training was provided on the haptic/visual navigation displays to encourage familiarization with the audio player, the wrist-worn navigational device and the bookmarking system. Participants were then informed that the objective of the experiment was to compare the concept of automatic bookmarking with haptic notification/ navigation to currently available media players (e.g., Apple iPod). They were asked to compare this experience to listening to audiobooks or podcasts over their own device and offer feedback after the trial.

Participants were directed to pay close attention to the information presented in the audio stream as they were to be tested on the content at the conclusion of the experiment. They were informed that during the session, experimenters would be engaging participants in conversation by asking questions. Answering these questions was participants' top priority, requiring them to divert attention away from the audio stream. Participants were also instructed not to pause or stop the audio stream, but instead were told to let playback continue while speaking with the experimenter; a chance was provided for participants to scroll back to listen to any content they might have missed immediately afterwards. Participants were informed that if they did not recognize content at one of the later bookmarks, they should try scrolling to an earlier bookmark.

Following the briefing on the experiment procedure, participants were asked to sit in a silent experiment room facing a computer monitor and to don the GSR sensor, the wrist-worn navigation device and pair of headphones. They were asked to refrain from making large motions during the experiment except while navigating through the audio stream (to avoid triggering excessive GSR noise). The experimenters sat on the other side of a visual divider from the participant to avoid unintentional distraction or anticipation thereof.

After two minutes of rest where the GSR signal was stabilized and baseline established, participants were asked to start listening to the preloaded audio track. Four trials for each of the four bookmark display conditions were presented, with a 5-minute break in between the second and third conditions to prevent fatigue. In each trial, the experimenter interrupted the participant and engaged in conversation, taking their full attention away from the podcast. Following the

interruptions, participants used the wrist interface to scroll back to the location in the audio stream where they believed that they left off.

Following the session, participants were asked to fill out a structured survey regarding the interruptions and device interactions experienced during the study (see Appendix B). They were asked to provide qualitative feedback on the interaction in the form of a structured survey. Specifically, they were requested to:

- Describe the effects of the haptic notification (applicable only to the group having experienced haptic notification of bookmark placement).
- Identify the utility of the haptic/visual navigation display conditions.
- Evaluate how well the overall interaction method works.
- Provide additional comments regarding the utility of haptic notification and comparative feedback of the various display conditions.
- Note the bookmark display condition (i.e., D_{None} , D_V , D_H or D_V+D_H) they preferred.

3.6.6 Quantitative and Qualitative Measures

Two performance metrics were collected for each trial: time required for a participant to acquire a listening location following an interruption, and the acquisition accuracy (distance in the audio stream between the final navigated position and the ‘true’ position corresponding to the start of the interruption). From navigation accuracy, we also determined whether a participant didn’t scroll back enough, signifying that they had missed listening to a portion of the audio stream. Each occurrence of this navigation error was tallied.

The post-experiment survey asked participants to evaluate disruptiveness and usefulness of the haptic notification of bookmark placement, and annoyance and usefulness of the haptic display for bookmark navigation on Likert scales (1-10). Participant preference for each condition (i.e., D_{None} , D_V , D_H , and D_V+D_H) was tallied.

3.6.7 Hypotheses

Users will:

- H1. Perform better when bookmarks are displayed during navigation (D_H , D_V or D_V+D_H).
- H2. Perform at least as well with D_H as with D_V .
- H3. Perform better with D_V+D_H over D_H and D_V alone.

H4. Prefer D_V+D_H over D_H and D_V alone.

H5. Perform better with N_H during navigation so long as bookmarks are displayed (D_H , D_V , or D_V+D_H).

3.7 Results

A 2x2x2x4 (haptic display x visual display x haptic notification x presentation order) repeated measures ANOVA was conducted. No significant effects of presentation order were detected; thus, we simplify our results by examining only the three main factors. Significance is reported at $p < 0.05$. Mauchley's test of sphericity confirmed that the data analyzed did not violate the assumption of sphericity. Pairwise comparisons are protected against Type I error using a Bonferroni adjustment. We also report partial eta-squared (η_p^2), a measure of effect size that can be interpreted using the following rule of thumb: .01 as a small effect, .06 as medium, and .14 as large [77]. Results are organized by main findings.

3.7.1 Effect of Display on Resumptive Navigation

Speed

Any type of bookmark display is better than no display for increasing resumptive navigation speed. Results for navigation speed revealed an interaction between visual and haptic displays of bookmarks during navigation ($F_{1,14} = 10.443, p = 0.006, \eta_p^2 = .427$) as shown in Figure 21. We therefore conducted a 1x4 ANOVA on the four display conditions (i.e., D_{None} , D_V , D_H and D_V+D_H) to test the interaction.

Although it appears that this interaction is disordinal (i.e., conditions co-vary), post hoc testing revealed that the conditions had no additive effect. Tukey's HSD result reveals that D_V ($M = 8.8, SD = 3.41$), D_H ($M = 8.07, SD = 3.68$), and D_V+D_H ($M = 9.62, SD = 3.90$) conditions were all significantly faster than D_{None} ($M = 15.30, SD = 6.94$), but that there were no other significant differences (Figure 22). This result supports our hypothesis that users will perform better when bookmarks are displayed during navigation (H1).

Both D_V and D_H significantly reduced navigation time on their own. Trials with and without D_V and D_H showed a mean difference of -2.4 and -3.2 seconds respectively for resumptive navigation time.

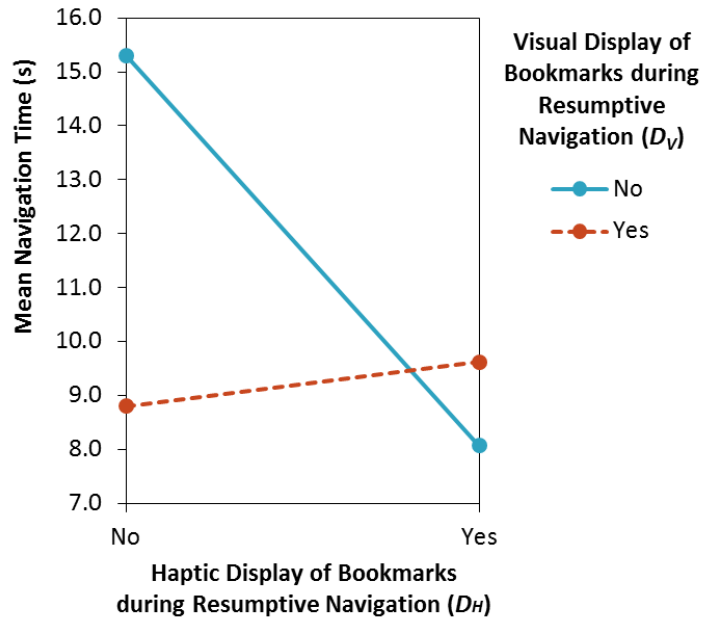


Figure 21. Interaction effect between visual and haptic bookmark display type during resumptive navigation.

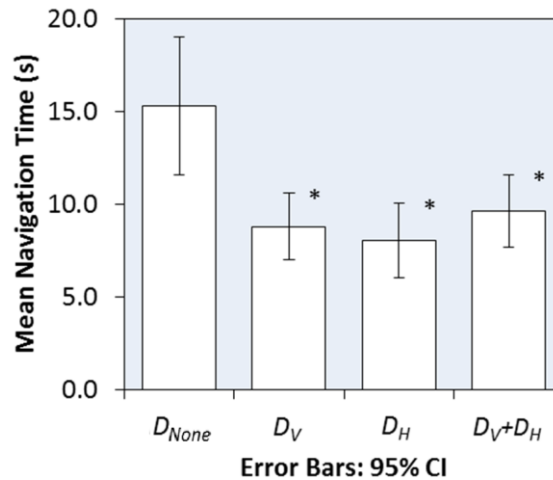


Figure 22. Mean navigation time for each bookmark display condition presented during resumptive navigation. Error bars show 95% confidence intervals. * indicates that the condition is significantly faster than the control condition (D_{None}).

Although D_V , D_H and D_V+D_H bookmark display techniques were all helpful, none offered a clear benefit over the other. No significant differences in resumptive navigation time were detected between D_V and D_H , which supports the hypothesis that users perform at least as well with D_H than with D_V (H2). Surprisingly, navigation time for D_V+D_H was not the fastest condition which contradicts our prediction that users are able to perform better with D_V+D_H over D_H and D_V alone (H3). This, in combination with the main effects described above, signifies that D_V and D_H both perform equally better than D_{None} , but that combining them (D_V+D_H) offers no additional performance benefit in terms of resumptive navigation speed.

Accuracy

D_H , with or without D_V , significantly increased navigation accuracy. A significant main effect of haptic display on navigation accuracy ($F_{1,14} = 5.4, p = 0.024, \eta_p^2 = .080$) revealed that D_H made users more accurate in their navigation with or without D_V . No other effects of bookmark display type impacted accuracy.

Self-Reported Measures and Comments

D_H was perceived as useful and not annoying. On a scale of 1 (not annoying at all) to 10 (very annoying), participants reported a mean annoyance of 2.3 ($SD = 1.9$). On a scale of 1 (not useful at all) to 10 (very useful), participants also reported that they found the haptic display to be useful ($M = 7.6, SD = 2.3$). Data on the annoyance or usefulness of D_V was not collected as it was not considered relevant.

D_V+D_H was preferred over all other resumptive navigation conditions. From 16 responses, we analyzed the frequency with which each resumptive navigation-time bookmark display condition (D_{None} , D_H , D_V and D_V+D_H) was preferred. This was done by calculating the one-dimensional Chi-square statistic to determine if the actual frequency was significantly different from the case where all frequencies were equal. As shown in Figure 23, the majority of participants (75%) reported that they preferred using both haptic and visual display of bookmarks (16 participants, $\chi^2(2,15) = 21.5, p < 0.001$), supporting H4. In the post-experiment survey, a few participants mentioned that they liked the D_V+D_H condition due to its affordance for ballistic yet accurate navigation trajectories. As one participant stated, “visual confirmation with haptics allows me to be much sloppier when getting to a bookmark without overshooting.”

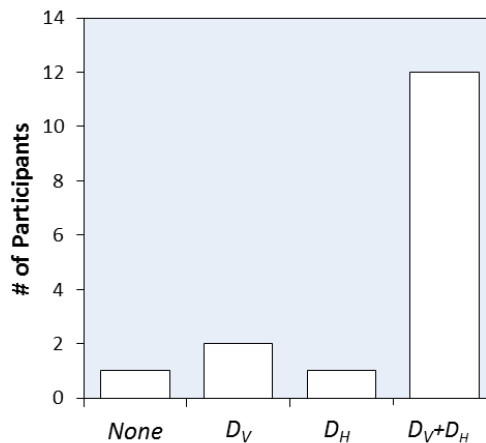


Figure 23. Participant preference for bookmark display.

3.7.2 Effect of Haptic Notification on Resumptive Navigation

Speed

Haptic notification of bookmarks at interruption time (N_H) significantly reduced navigation time if bookmarks were displayed visually (D_V) during resumptive navigation, but did not do so for haptic display during resumptive navigation (D_H). A significant interaction effect between haptic notification of bookmark placement at interruption time (N_H) and visual display of bookmarks during resumptive navigation (D_V) was observed for navigation time ($F_{1,62} = 7.0, p = 0.019, \eta_p^2 = .334$) as shown in Figure 24. A one-way ANOVA was conducted to investigate the interaction which revealed a significant difference of 2.7 seconds or 26% of the total navigation time ($p = 0.045$) between trials with D_V and N_{None} ($M = 10.5, SD = 3.5$), and trials with D_V and N_H ($M = 7.8, SD = 3.3$). In essence, N_H had no significant effect on navigation time when lacking D_V . Because a significant effect was only observed for N_H in the presence of D_V and not D_H and $D_V + D_H$, our hypothesis that users will perform better with N_H during navigation so long as bookmarks are displayed (H5) is only partially supported.

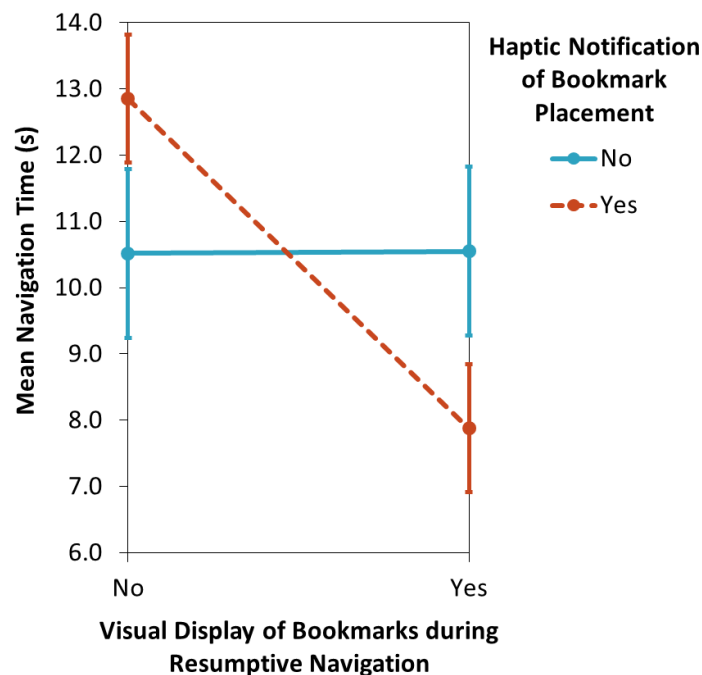


Figure 24. Interaction effect between D_V and N_H for navigation time. Error bars show standard error of the mean.

Self-reported Measures and Comments

Haptic notification of bookmark placement (N_H) was deemed useful and not disruptive. Participants who had experienced N_H reported on the level of disruptiveness and usefulness of N_H using 10-point Likert scales. On a scale of 1 (not disruptive at all) to 10 (very disruptive), participants rated the level of disruptiveness of N_H as low ($M = 2.9, SD = 2.1$). On a scale of 1 (not useful at all) to 10 (very useful), participants also reported that they found haptic notification to be useful ($M = 8.4, SD = 1.1$). One participant commented, ‘Although [the notification was] noticeable enough that I knew that a bookmark was placed, it was not so much that it distracted me from answering [the experimenter’s] questions.’ Another said, ‘It’s good to use vibrations as it helps signal to the reader where they left off without worrying if a bookmark was placed or not.’

3.7.3 Secondary Analyses

False Positives

Introducing false positives into trials reduced navigation speed. A paired-samples T-test was conducted to determine the effect of false positives on navigation speed and accuracy. Results confirmed a significant difference in average resumptive navigation speed between trials *with* ($M = 12.59, SD = 5.03$) and *without* ($M = 10.42, SD = 5.07$) false positives ($t(31) = 2.06, p = 0.047, d = 0.365$) as shown in Figure 25a. Paired T-tests were conducted to determine which conditions exhibited a significant difference in mean navigation time between trials with and without false positives. As shown in Figure 25b, D_V+D_H was the only condition which demonstrated this difference between trials *with* ($M = 11.41, SD = 7.87$) and *without* ($M = 7.84, SD = 5.61$) false positive bookmark placements ($t(31) = 2.05, p = 0.049, d = 0.362$). A trend from Figure 25b shows that false positives slowed navigation speed in conditions where D_H was present; whereas this effect does not appear for D_V . This may suggest that it may be harder for participants to ignore false positive bookmarks when they are presented haptically, rather than visually.

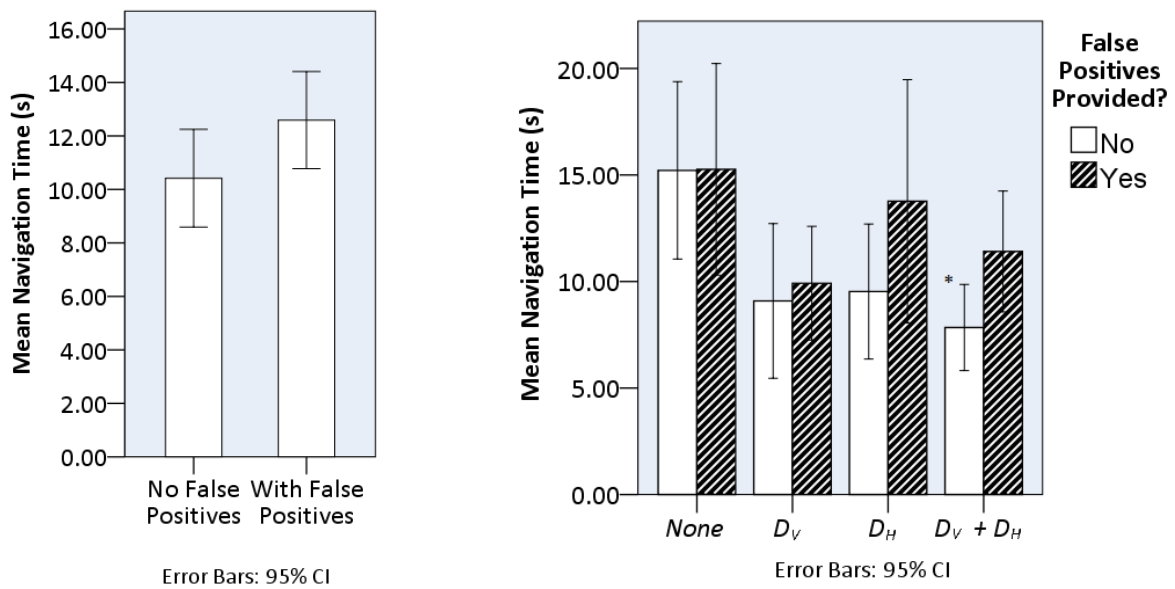


Figure 25. a) Mean resumptive navigation time for trials with and without false positive bookmarks. b) Mean resumptive navigation time for each condition. Error bars represent 95% confidence intervals. The star represents a significant difference between navigation times for trials with and without false positive bookmarks within a condition (this distinction is meaningless for the D_{None} condition).

No significant difference was observed in the mean navigation accuracy for false positive vs. no false positive bookmark trials.

Information Transmission Loss

Visual + Haptic display of bookmarks during navigation reduced the number of times that audio stream information was lost. During navigation, some participants did not scroll far enough in the audio stream to reach the point where they were interrupted, resulting in their ‘skipping over’ a segment of the audio stream. In most cases, the effect of this error caused participants to be disoriented, lacking contextual awareness of new information being presented in the audio stream as reported by participants in interviews. Figure 26 shows the D_{None} condition produced the most occurrences (13) where information had been skipped following an interruption. A reduction in occurrences of 38% and 53% were observed in the D_V and D_H conditions; the $D_V + D_H$ condition resulted in the least number of occurrences (1) providing some evidence that users perform better with $D_V + D_H$ (H3).

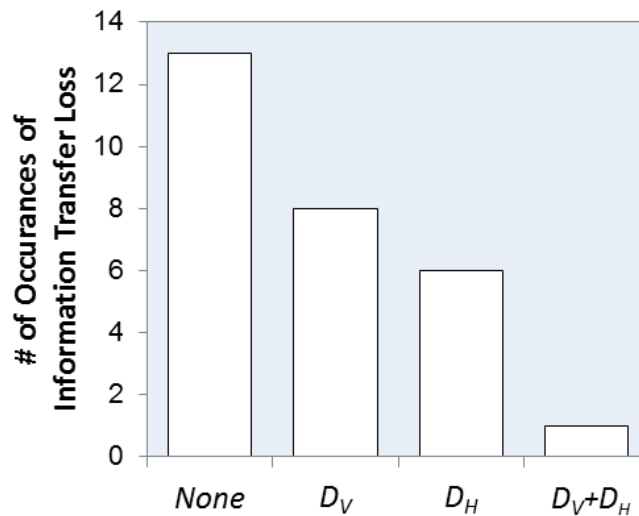


Figure 26. The number of occurrences where information was skipped over due to improper navigation across experiment conditions.

3.7.4 Summary

We summarize our main findings according to our hypotheses.

H1. Users perform better when bookmarks are displayed. SUPPORTED.

H2. Users perform no worse with the haptic display of bookmarks than the visual display. SUPPORTED.

H3. Users perform better with D_V+D_H over D_H and D_V alone. PARTIALLY SUPPORTED.

H4. Users prefer D_V+D_H over D_H and D_V alone. SUPPORTED.

H5. Users perform better with N_H during navigation so long as bookmarks are displayed. PARTIALLY SUPPORTED.

Secondary analysis indicates that false positives may slow navigation speed when D_H is present. Additionally, users perform better with D_V+D_H in terms of information transmission loss than any other form of bookmarking.

3.8 Discussion

Given the results of Section 3.7, we review our research questions and discuss our interpretation of our findings.

Q1. Do auto-generated bookmarks provide assistance to users in the context of interruption?

In response to: ‘Did you find the bookmarks helpful?’ asked in the post-experiment questionnaire, all participants (100%) stated that they felt that the bookmarks in and of

themselves provided utility. Additionally, there is a performance benefit: bookmarks displayed to participants in any form reduced average time required to navigate to the correct position after attending to an interruption (Figure 22). It is not surprising that this was the case for the D_V condition, but the similarity between all of the bookmark display techniques was unexpected. From Figure 26, we see that use of any bookmarking display reduced the number of occurrences of information transmission loss by at least 30%.

These results were obtained under conditions where precision of bookmark placement was 50%. Overall sensitivity of bookmark placement was 75%; ‘correct’ bookmarks placed at the time of interruption were always placed under conditions that displayed bookmarks (i.e., D_V , D_H , and D_V+D_H conditions) and never placed for the D_{None} condition⁶. From graph showing the precision-sensitivity trade-off in Figure 9, we see this level of precision and sensitivity is obtainable by our algorithm described in Chapter 2, and is consistent with the middle range of the precision versus sensitivity tradeoff (at around 0.35 μ S). Based on this, the results presented here indicate that our GSR-based bookmarking algorithm will be sufficient in providing value in bookmarking, offering a substantive performance and subjective improvement to the audio stream navigation experience.

Q2. Does haptic notification of bookmark placement at interruption time provide value to users?

In most cases, the presence of N_H made no perceivable difference in either navigation time or accuracy (i.e., *haptic* notification provided no significant main effect or interaction effect with the *haptic* bookmark display condition in resumptive navigation). However, N_H in resumptive navigation coupled with D_V at bookmark placement time does produce a benefit over the N_H without D_V condition (i.e., an interaction effect was present) as shown in Figure 24.

One possible explanation for this difference may be related to the types of mental models developed for the navigation task. Having bookmarks displayed with D_V during resumptive navigation may allow participants to develop a spatial mental map of where both useful and non-useful bookmarks have been placed, whereas participants lack such spatial information when

⁶ As described in Section 3.6.4, the D_{None} condition essentially emulates false negative bookmark placements; thus, we did not introduce false negatives into any of the other display conditions to reduce since we determined that we would not have obtained any useful additional data. Thus, bookmark placement sensitivity is 100% for D_H , D_V , and D_V+D_H conditions, and 0% for the D_{None} producing an overall sensitivity of 75%.

working with only D_H . As N_H provides information that has direct correlation with a spatial mapping of bookmarks, the benefit of notification in terms of performance may only be apparent with the D_V and D_V+D_H conditions rather than the D_H condition. Thus, it may be that haptic notification provides a performance benefit to participants only when they are also provided with additional means to access or unlock that information; and in the conditions provided here, that extra key may have been best obtained through D_V , or the visual sensory modality more broadly. More examination is required to substantiate this theory.

With regards to subjective results, all participants except one who received haptic notification reported that they found it to be useful. Additionally, general comments from participants provided some evidence that N_H is able to reassure users that a bookmark has been placed such that they could attend to the interruption without worry as stated in our hypothesis H2.

Q3. Does haptically-assisted navigation of bookmarks provide value to users over no display or visual display of bookmark conditions?

D_V , D_H and D_V+D_H produced significant reductions in navigation time over control (D_{None}) but not over each other; only D_H (with or without D_V) significantly increased navigation accuracy. From review of video footage, navigation data and user comments, we believe that haptic-based bookmark display afforded navigation using a ballistic trajectory where scrolling could be faster and less accurate - the haptic feedback provided a ‘safety net’ which would very noticeably inform participants when to stop scrolling. In the visual condition, participants used a different navigation method to predictively fine tune scrolling, producing a smoother trajectory. In other words, a similar level of speedup was achieved in D_H and D_V , but through different mechanisms.

We did not see a significant performance difference (either in acquisition time or accuracy) of the D_V+D_H condition when compared to D_H and D_V alone – there is not an additive performance benefit from combining haptic and visual display of bookmarks. This outcome was unexpected and is contrary to our hypothesis that the D_V+D_H condition would outperform both D_H and D_V conditions. We theorize that this outcome could be explained by the different physical scrolling strategies afforded by the nature of the information provided haptically and visually respectively. However, participants reported both visual and haptic modalities can be used in parallel: they were able to benefit from the ballistic scrolling afforded by the haptic modality, while being able to predict when to start slowing down through visual cues to prevent overshooting. Further

investigation is required to see if greater experience with the interface would allow users to integrate these scrolling strategies more effectively.

On average, participants felt that more could have been done to reduce the annoyance of the haptic signal indicating bookmark placement (e.g., shorter duration of signal, weaker strength). Additionally, some participants suggested implementing a virtual wall in addition to a haptic signal to prevent overshooting of the bookmarks.⁷ In spite of these remarks, participants reported that they felt that haptic display of bookmarks offered much utility. Further, we must keep in mind that their experience was in a quiet controlled environment, and the haptic signal's salience might be less adequate in more active surroundings. In fact, another large topic for further work (and with utility much wider than the present application) is how to adjust haptic signal salience to accommodate ambient noise and activity levels.

Q4. Do users feel that the auto-bookmarking/haptic confirmation and navigation feedback system is useful?

Pending some improvements to the system, the large majority of participants (88%) reported that they would use an auto-bookmarking player like the one presented to them. Many of the concerns were related to the high number of false positive bookmarks (50% precision of bookmark placement); this shows that future work on this bookmarking application should be directed at increasing algorithm precision while maintaining or also increasing sensitivity. Regardless, results showed that while inaccurate bookmarks do slow users down overall, *the impact is not worse than if no bookmarks were provided at all*; and in general, it may be a technically practical strategy to strive for an interaction style that 'lives with' substantial false-positive rates and makes them tolerable, rather than trying to eliminate them. One example of this could be to devise a method of distinguishing between bookmarks with lower statistical probabilities of a being a true positive from ones with higher probabilities as judged from the GSR signal, and to assign a different haptic profile to each bookmark 'strength' category – i.e., lower probability bookmarks could feature lighter, less noticeable vibrations and higher probability bookmarks could use stronger vibrations.

⁷ This idea was considered but intentionally not implemented in the experiment as we wanted to obtain an accurate measure of performance for navigation time and accuracy.

General utility of HALO model:

Reflecting on the larger context of this work, our results suggest that the generalized HALO model with low-attention haptic feedback is viable; reactions from participants and quantitative performance data both suggest that feedback within an implicit interaction framework is beneficial, even perhaps necessary in order to introduce transparency.

In terms of our use case, we see that providing low-attention haptic feedback for *notification* of bookmark placement at interruption time offers a significant and practically useful performance benefit, when combined with visual bookmark display at resumptive navigation time. It is also seen subjectively as a benefit to users as a low-disruption, ‘reassurance’ mechanism, and this may be its greatest value. In terms of resumptive navigation itself, haptic display of bookmarks improves acquisition time and accuracy following an interruption without significantly annoying users in spite of the high number of false positives presented to participants in the current experiment. In terms of the navigational interface design, the D_V+D_H bookmark display condition provides the best support for users: it offers the best performance in terms of navigation time and accuracy. This indicates that it may be worth investigating porting the audio stream system to a mobile platform containing both visual and haptic displays, such as a smartphone. However, our results suggest that the D_H alone could be nearly as helpful, implying that the interaction can be performed almost as effectively without visual guidance.

3.8.1 Limitations and Constraints of the Experiment

A primary constraint in terms of immediately generalizing our results to practice come from the controlled lab setting in which this experiment was conducted. While this setting was an appropriate first step, further research is required to confirm these results in a more distracting environment such as in a busy café or on a bus. In addition, while a pool of 16 participants was a reasonable number for this first study, it would be interesting to consider a larger number of participants in either a highly controlled or certainly for a less controlled environment. The task of returning to a location in the audio stream after attending to an interruption inherently has considerable variation across participants, and a large experiment would be required to examine the full range of user types in a broad population, especially if it includes those without previous experience with circular scrolling devices.

3.9 Summary

In Chapters 2 and 3, we have developed and tested an instance of the HALO paradigm, complete with haptic feedback, within the context of an audio stream bookmarking use case. The results obtained from experiments conducted confirm that consistent features corresponding to user interruption can be derived from physiological measurements. Leveraging this result, we found that an automatic monitoring system can be used to detect when a user is interrupted with a high degree of accuracy. In context of the use case, we used this system for placing bookmarks in an audio stream at any indication of an external interruption to the user. To alert the user to the autonomous behavior of the system, we explored the use of haptic feedback (in combination with traditional visual displays) to present these bookmarks to the user in a non-intrusive way.

From a general scope, the results indicate that the HALO exhibits potential to elegantly manage user transition between primary and secondary tasks; the implementation of this implicit interaction with low-attention haptic feedback represents a first, but large step in confirming that the HALO is able to *calmly* assist the user in an effective manner. However, validation of this new interaction method cannot be done on a single use case; more work needs to be done to verify the faculties and value of the HALO concept. In the following chapter, we attempt to explore the HALO paradigm within the context of another use case: to determine and react to user preference during music listening.

4 Music Preference Recognition through Kalman Filtering

For the automatic audio stream bookmarking use case described in Chapters 2 and 3, we greatly simplified the loop by reducing dimensionality of the affect estimation problem, i.e., we were able to limit the amount of information which requires processing and interpretation by looking only at a user's GSR activity. Although this allowed us to test aspects of the implicit interaction loop without having to tackle the complexities of a fully-featured affect recognition system, there were some disadvantages to using GSR as the only physiological measure in an interaction. Above all, we are only able to obtain information of the user's physiological arousal - no estimate of user valence can be acquired reliably through GSR. Secondly, using only one physiological measure reduces robustness of our ability to estimate user affective state - noise in the GSR measurement can cause erroneous system behavior, whereas the effect of noise can be reduced when using several combined measurements to determine affect.

As a next step in the work to obtain validation of the technological feasibility of real-time affect detection and classification, we investigated the development of a system to model a user's affective state using multiple physiological channel inputs. This chapter describes an exploratory pilot to categorize emotional content to determine time-varying user preference during music listening.

4.1.1 Use Case

Music plays a significant role in the daily lives of many people, yet the exact reason for its importance or appeal is not well understood: it is not a necessary requirement for survival and it has no known tangible, addictive properties. One commonly accepted theory explains that the widespread fascination of music stems from its ability to evoke, project or enhance emotion, which in itself could be rewarding [78–80]. This phenomenon, known as musical emotion induction, has been studied extensively to address aspects of emotional expression and communication in music that are not well understood (i.e., physiological and psychological impact) [81–83]. Because music is consistently cited as medium that is able to evoke strong emotions and moods within people [80], we chose to consider our HALO system development in a music-listening use case context. In particular, we felt that it could adequately serve as a test bed for our affect recognition methods while demonstrating the value of the HALO paradigm.

The movement of music storage away from ‘hard’ physical media formats (e.g., compact discs, cassettes) to ‘soft’ digital files (e.g., mp3, wma) has allowed for music to be easily amassed and transferred to digital music players and online. No longer are users restricted to listen to what they buy on CD, instead, they are able to explore and personalize playlists using internet music services such as Pandora [84], Last.fm [85], Grooveshark [86] and 8tracks [87] which promote/encourage exploration of expansive music collections spanning a vast assortment of genres. Given the vast collection of songs, these services require user self-reports and ratings (i.e., thumbs up, thumbs down) to notify the service which music selections are liked/disliked. This information is then used to algorithmically personalize and tailor playlists for the user. Alternatively, many of the self-reports are amassed across users to create recommender systems where users are associated with other like-minded people. Thus, users are not truly given their own customization.

Often, this ‘music exploration’ experience is quite demanding – if, for example, a user at work is performing data entry in a spreadsheet, he/she would have to switch windows and rate the experience, interrupting workflow every few seconds to notify the system that the currently playing track is liked/disliked. The problem that is faced here is similar to the one described in Chapters 2 and 3: the result of this explicit interaction is increased cognitive demand on the part of the user, exactly in a situation when the opposite is desired. In order for proper functioning, the music service demands that the user focus on the interface, rather than having the system do its job quietly in the background. The solution that we propose is also similar in nature to the one provided for the bookmarking application: to reduce this cognitive demand, we suggest having the system automatically recognize user music preference through user physiological signals and to act appropriately provided this information.

4.1.2 Research Objectives

Much work performed by researchers such as Picard and Sarkar strongly indicate that accurate estimates of affect can be drawn from user physiological measures [19], [45], [47], [48], [51], [54], [88]. Building upon this prior work, we attempt to estimate human affective response with regards to music preference in real-time (i.e., like vs. dislike) through a pilot study. However, unlike much of the research conducted within this field, we have chosen to depart from traditional machine learning techniques which operate offline. Instead, we have opted to use an optimal state estimator, in particular a Kalman filtering technique [89], to estimate user

preference for music. Kalman filtering describes a set of computationally-efficient algorithms which are able to optimally generate a best estimate of a system's current state based on optimal use of imprecise linear (or nearly linear) data describing prior states of the system. The filtering method is unique compared to other state estimation techniques (e.g., machine learning) as it supports estimations of past, present, and even future states, even when the precise nature of the modeled system is unknown. We chose to use a Kalman filtering approach for several reasons – in particular:

- the ability of the filter to robustly forecast time series data [90];
- its ability to be run in real time using measurements as they arrive, allowing the HALO to operate online rather than being forced to react asynchronously [90];
- its capacity to incorporate representations of uncertainty and provides a mechanism for incrementally reducing this uncertainty over time [91]; and,
- its use in other areas of intelligent/implicit HCI research such as gesture and facial expression recognition [92–96].

Our system inputs include music selection features (i.e. centroid, spread etc.). Biometrically-obtained physiological features serve as the outputs of the system which include galvanic skin conductance (GSR), the rate of change or velocity of skin conductance (GSR_{Vel}), heart rate (HR), heart rate acceleration (HR_{Accel}), electromyography (EMG) and skin temperature ($Temp$). Online, continuous participant rating of the music selection serves as our system state.

A controlled pilot study was undertaken to address the practicality of using Kalman Filtering to recognize user preference to music pieces in real-time. An important caveat is that the purpose of the study was not to test the feasibility of affect estimation and recognition as a whole; rather, we hoped that the experiment could indicate whether time-varying music preference can be estimated accurately through Kalman filtering based on physiological and music feature inputs. In particular, can the time-varying level of enjoyment of music be inferred using physiological data and structural characteristics of the music in real time and how reliably? This research will explore whether several physiological signals – such as, heart rate, skin conductance, respiration rate and muscle activity – which evolve while listening to music with different user's levels of enjoyment (e.g., like, neutral, dislike) can be recognized and used by a Kalman filter to predict music preference in real time.

The music listening use case used in this work was first described by Hazelton in his MSc thesis [56] to assist in understanding requirements for the HALO paradigm and developing the feedback-supported interaction. Hazelton conducted focus groups to determine the perceived utility of the HALO paradigm; he mentions that focus group participants indicated that robust identification of continuously changing musical preferences is required for the HALO paradigm to be effective and valuable in a portable music consumption use case. This objective of this work, in some ways, is to build and expand upon the research conducted by Zhogbi and Hazelton in [56].

4.2 Background and Literature Review

The technical readiness of the proposed HALO interaction paradigm depends greatly on the technological feasibility of real-time affect estimation and classification through physiological sensing. Significant work has been performed in the area of affect computing, much of which has been mentioned in Section 1.4. Here, we present a survey of prior work more central to the work presented in this chapter, namely affect estimation during music listening and Kalman filtering.

4.2.1 Affect Estimation during Music Listening

Much of the work done on emotion in musical contexts has been largely restricted to analyzing subjective survey responses from experiment participants [56], [80], [97–99]. However, more recent forays into correlating music with affective response have used physiological measures. Chung and Vercoe developed a novel, real-time music arranging system that selects music on the basis of physical and physiological cues [100]. The goal of the system was to continuously transition the listener to a goal (enjoyable) state based on foot tapping as recorded by a microphone, GSR, and subjective evaluation data through probabilistic state transition modelling. In other work, Salimpoor et al. have used recordings of sympathetic nervous system activity (e.g., GSR, temperature, heart rate, respiration rate) to reveal that there is a strong, positive correlation between pleasure experienced by a music listener and their emotional arousal [101]. Similarly, Kim and André have performed feature analysis on physiological data sets (e.g., EMG, ECG, GSR and respiration) to obtain a set of best features for automatic recognition of emotion [102]. To classify musical emotions, they developed a categorization method which serially divides a set of training patterns into dyadic groups corresponding to arousal and valence classes. Then, multiple binary classifiers are classified corresponding from each dyadic grouping. With this emotion-specific multilevel dichotomous classification scheme, they achieved

recognition accuracies of 95% and 75% for subject dependent and independent classifications of four emotion classes, respectively. Although Kim and André investigate emotion in the context of music-listening, they do not explore user preference to music and its emotional implications.

There have also been several attempts to predict emotional ratings by examining characteristics of the music itself. For example in [103], Yang and Lee present a method of mining and annotating music emotional content through machine learning analysis of music lyrics, achieving a one-dimension emotion intensity classification accuracy of 90%. In a different stream, Eerola and colleagues used structural features extracted from the music itself (e.g., timbre, harmony, register and rhythm) and tested three data reduction techniques – stepwise regression, principal component analysis, and partial least squares regression – to predict emotional representation in music [104].

4.2.2 Kalman Filtering

The Kalman filter (KF) is a time-series recursive, optimal estimator – inferring parameters from indirect, inaccurate or uncertain observations and measurements which can be processed as they arrive. The filter, named after Rudolf E. Kálmán, was originally simply described as a recursive solution to a discrete-data linear filtering problem, but has since had been widely applied to applications such as radar [90], computer vision [105], navigation and guidance [106] and economics [107], [108]. The KF is causal, only using present input measurements and past states to estimate future state, allowing it to be run in real time. The filter relies upon a state-space dynamic model which has known inputs (u), outputs or observed measurements (y) and a hidden state (x) that is to be estimated from the former two. The KF recursively runs through two phases alternately: the *time-step propagation phase* (also known as the predict phase) and the *measurement update phase*. The time-step propagation phase uses the state estimate x from the previous phase/time-step to produce an *a priori* estimate of state for the current time-step. During the measurement update phase, the KF produces a refined *a posteriori* state estimate which combines the a priori estimate with current observed information – i.e., a trajectory correction based on the most current measurements.

Although the KF is a popular estimation tool, it comes with a caveat that it only performs well in cases where state-space models are linear, or approximately linear in nature. To provide better estimation for non-linear models, a series of variations on the KF have been developed. In this work, we focus on a more recent variation developed by Julier and Uhlmann termed the

unscented Kalman filter [109]. The UKF uses a deterministic sampling technique known as the unscented transform (see [110]) to pick a set of sample points, designated sigma points, which encircle the mean data point. These sigma points are then pushed through the non-linear state transition/observation functions, from which the mean and covariance of the estimate are obtained. The result is a filter which more accurately captures the true mean and covariance. Additionally, it can be shown that the UKF performs equivalently to the generic Kalman filter for linear systems, yet generalizes elegantly to non-linear systems [109]. Because of its robustness and generalizability, the UKF has been steadily replacing the generic KF and other non-linear KF variants. Its use can be seen in non-linear applications such as tracking of anatomical features and deformable objects [111], [112], vehicle trajectory estimation and planning [113] and speech processing [114], [115]. We use the UKF in this work due to its ability to robustly handle non-linear state-space equations which will arise from modeling as shown later in this chapter. A comprehensive description of the UKF and its processes can be found in Section 4.4.2.

As mentioned earlier, the KF and its variants has proved to be a very popular estimation method being used in a wide range of fields, though its use in emotion estimation and recognition has been a relatively new development. Much of the work which uses KF to estimate emotion looks at interpreting facial expressions. For example, Maghami and colleagues developed a facial recognition system based on KF, where features are selected and tracked continuously [92]. Similarly, Essa and Pentland have fed a KF with geometric, physical and motion-based models to estimate optical flow of facial motion to analyze facial expressions [116]. Using a KF, Schmidt and Kim have been able to form robust estimates of emotion labels for music in the arousal-valence space which change over time, during progression of the sound track [96]. In their work, a KF attempts to predict emotion labels based on a learned dynamics model and octave-based spectral contrast features of the music.

4.2.3 Summary

The literature presented in this section highlights advances for music-based emotion recognition and estimation. However, we see that relatively little attention has been given to systems that account for the time-varying nature of music and musical emotions – most of the work presented here aimed at recognizing an emotion or rating over an entire music selection or clip. Such generalizations oppose the time-varying nature of music and make emotion-based

recommendation difficult, as it is very common for emotion to vary temporally throughout a song. Additionally, most methods for categorizing music based on emotion and user preference described in the literature above cannot be computed in real-time, especially for non-linear systems. Due to these observations, we take a departure from traditional feature extraction methods and explore an approach of applying an unscented Kalman filter (UKF) for affect estimation.

4.3 Methods

4.3.1 Participants

As this was an exploratory pilot study, only 3 participants (2 male) were recruited for this exploratory study (ages 23-32). All subjects were undergraduate students at the university.

4.3.2 Location of Study and Consent

The study sessions were run in a controlled environment free from noise and interruption in the Usability Lab (X727) of the ICICS building at the University of British Columbia Point Grey campus. Participants were briefed on their ethical rights upon arriving and provided consent at the outset of the session (see Appendix C for consent form).

4.3.3 Experimental Setup and Sensor Equipment

Participants were provided with an online music rating device which consisted of a mounted, 11 cm long linear potentiometer (i.e., fader). This device (shown in Figure 27) was constructed to allow participants to continuously self-report musical enjoyment during music playback on a scale of 0 (dislike) to 1 (like). A high potentiometer contact position (1) indicates participant enjoyment while a low potentiometer (0) position represents dislike. There was only one detent slightly resisting motion of the linear potentiometer knob present at the half-way point (0.5) to represent neutral.

Physiological data was collected through a Thought Technology ProComp physiological measurement system [61]. Four sensors were used for this experiment, which included:

- Galvanic skin response (GSR) sensor;
- Surface Electromyographic (EMG) sensor;
- Blood Volume Pulse (BVP) sensor; and,
- Skin temperature (ST) sensor.

These sensors were selected for use based on their fast response rates ($< 1s$) and use in similar applications in prior work [19], [43]. All signals were recorded at 32 Hz.

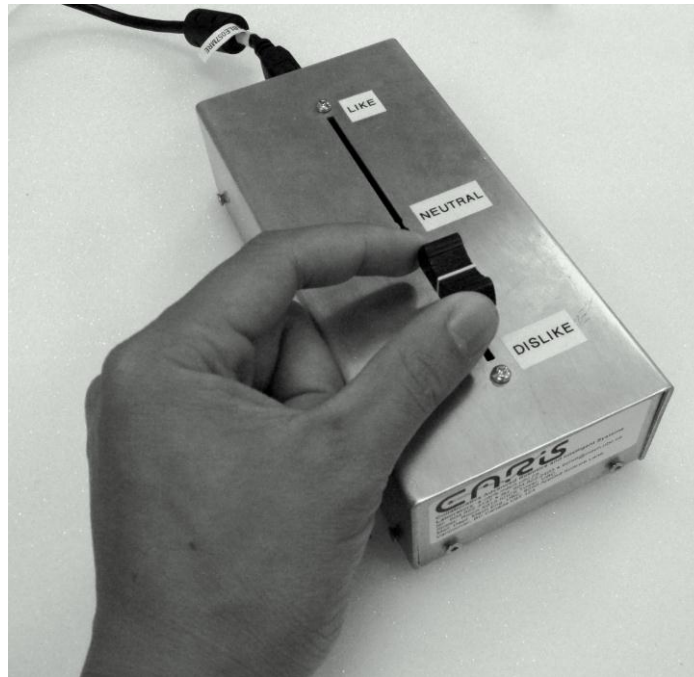


Figure 27. Photograph of the music rating device.

4.3.4 Trial Task

Each trial began with two minutes of baseline data collection where the participant was instructed to relax, allowing physiological measures to stabilize at normal, resting levels. Following this rest period, a selection of music began to play for 90 seconds through the pair of headphones worn by the participant. During this time, participants continuously self-reported their enjoyment/dislike of the music selection in real time using the continuous music rating device. Physiological data recording continued for another 30 seconds. Following each trial, the participant would be given some time to fill out a short questionnaire regarding the music selection just heard (Appendix D). The questionnaire issued allowed them to:

- Self-report their affective state during the trial, both visually and descriptively;
- Indicate whether they recognized the song playing during the trial;
- Indicate their level of enjoyment of the song played (using a continuous Likert scale);
- Report whether they had ‘chills’ when listening to the selection and when;
- Describe any emotional or cognitive associations (if any) with the song if they recognized the selection.

4.3.5 Procedure

Pre-Experiment

Prior to the experiment, participants were asked to provide 10 songs which they enjoyed listening to. Additionally, they were also asked to provide another 10 songs which they disliked. The songs collected were used to personalize the experiment for each participant. Participants were instructed that all selections had to be devoid of vocal lyrics; this was done to eliminate the possibility that participants held preferences to music selection based on appreciation of the selections' lyrics (a feature that would be unrecognizable by the UKF) , rather than on its structure. Additionally, all music selections had to be from the same genre in an attempt to limit the diversity of musical features that the UKF would have to recognize.

The 20 songs selected by the participant were shuffled to create a customized playlist for each participant. In addition to these selections, five additional songs of the same genre (selected by the experimenter) were added randomly into the playlist.

Experiment

At the start of each session, the experimenter explained that the purpose of the experiment was to collect physiological data to determine if we can predict whether or not a machine can recognize music preference. The experimenter introduced the function of each of the physiological sensors as they were donned by the participant. The participant was provided with an explanation of how to operate the online music rating device: a high fader knob position represents that he/she likes the currently playing song, while a low fader knob position represents dislike. They were instructed to continuously self-report their preference for the music selection in each trial (thus to were asked keep their hand on the knob at all times) and reset the fader to half-level (neutral) before and after every trial.

Following the briefing on the experiment procedure, the participant was asked to sit in the silent experiment room facing the online music rating device which had been fixed to a desk. He/she was asked to don a pair of headphones and was asked to refrain from making any more large motions during the experiment (to avoid triggering excessive bio-sensor noise). The experimenters sat behind the participant, away from view or audition of the participant to avoid unintentional distraction.

After two minutes of rest where the GSR signal was stabilized and baseline established, the experimenter then started playing the first song in the prearranged, randomly ordered playlist, marking the start of a trial (described in Section 4.3.4). The music was stopped after 90 seconds, and the participant was asked to complete a post-trial questionnaire. After the questionnaire was completed, participants were asked to rest in silence for one minute while they remain seated before the next trajectory was presented to ensure that the user's physiological data returned to a near-baseline state. Each participant experienced 25 trials, each taking approximately 3 minutes to complete, plus the time required by the participant to complete the post-trial questionnaire; an experiment session was designed to take a contiguous time period of approximately 90 minutes.

4.3.6 Quantitative and Qualitative Measures

Physiological Measurements (y)

Several physiological measurements were recorded continuously during the music-listening portion of each trial for offline processing, including:

- Normalized Galvanic skin response (GSR);
- Galvanic skin response derivative (GSR_{speed});
- Heart rate, measured in beats per minute (HR);
- Heart rate acceleration (HR_{Accel});
- Electromyographic voltage, measured in μV (EMG);
- Skin temperature, measured in $^{\circ}C$ ($Temp$).

Measurements were segmented into 90 second portions corresponding to music selection.

Online Self-Reported Music Preference Data (x)

Continuous self-reported music preference data generated by the participants was captured during the music-listening portion of each trial at 32 Hz using a rating device (shown in Figure 27). This data appeared as a continuous stream of decimal values ranging from 0 (dislike) to 1 (like) representing the participants' preference for the music selection recorded in real-time.

Music Feature Extraction Data (u)

Each music selection played to participants was processed using MIRtoolbox – a Matlab toolbox which offers a set of functions specifically designed to extract structural features from music such as tonality, rhythm, structures, etc. from audio files [117], [118]. Using MIRtoolbox, we

extracted a selection of features chosen based on traditional categories of musical elements (rhythm, timbre, pitch, form, etc.). A representation of these categories was done by selecting only a few basic, continuous-time features which have been described as indicators of emotional content within music selections [104], [119]. These features include:

- root-mean-square of the audio signal energy (*RMS*);
- spectral centroid of the audio signal amplitude (*Centroid*);
- spectral spread or standard deviation (*Spread*);
- relative Shannon entropy of the smoothed and collapsed spectrogram (*Entropy*);
- musical key (*Key*)
- key clarity (*Clarity*);
- key modality, i.e., major vs. minor (*Mode*);
- flux of the tonal centroid as detected by the Harmonic Change Detection Function [120] (*HCDF*).

Post-Trial Survey Data

We use Russell's dimensional circumplex model of affect described in Section 1.4.1 and [32] in post-trial surveys as an intuitive way for users to report their emotional response to the music. In addition to an emotional reflection of the music, users were asked to report on:

- whether or not the music selection was recognized;
- how much they enjoyed the music on a Likert scale of 1-10;
- if they would have changed the song if they had a choice;
- if they had experienced chills; and,
- whether they had any strong emotional or cognitive associations with the song (e.g., the song had been played at high school prom).

4.4 Modelling and Filtering

4.4.1 Modelling

State Space Model Equations

For our music preference estimation, we needed to obtain a state-space model structure which can be used by the UKF. We assumed an explicit discrete time-invariant system; thus, our state space model template consisted of the following equations:

$$x[k + 1] = Ax[k] + Bu[k] \quad \text{Equation 2.}$$

$$y[k] = Cx[k] + Du[k] \quad \text{Equation 3.}$$

Where:

- $x[k]$ is the state vector;
- $x[k + 1]$ is the state vector of the next timestep;
- $y[k]$ is the output vector;
- $u[k]$ is the input vector;
- A is the state matrix;
- B is the input matrix;
- C is the output matrix; and
- D is the feedforward matrix.

Equation 2 represents the state transition equation and Equation 3 represents the output equation, also known as the observation equation.

For this work, our models were developed with respect to a user’s viewpoint, where he/she would listen to music ($u[k]$ – input), develop an internal emotional assessment of the music ($x[k]$ – state), and exhibit this state in the form of measurable physiological indicators ($y[k]$ – output). This model structure is reflected in Figure 28. To obtain our state-space model (namely, the A , B , C , and D matrices) we use a least-squares approximation detailed below.

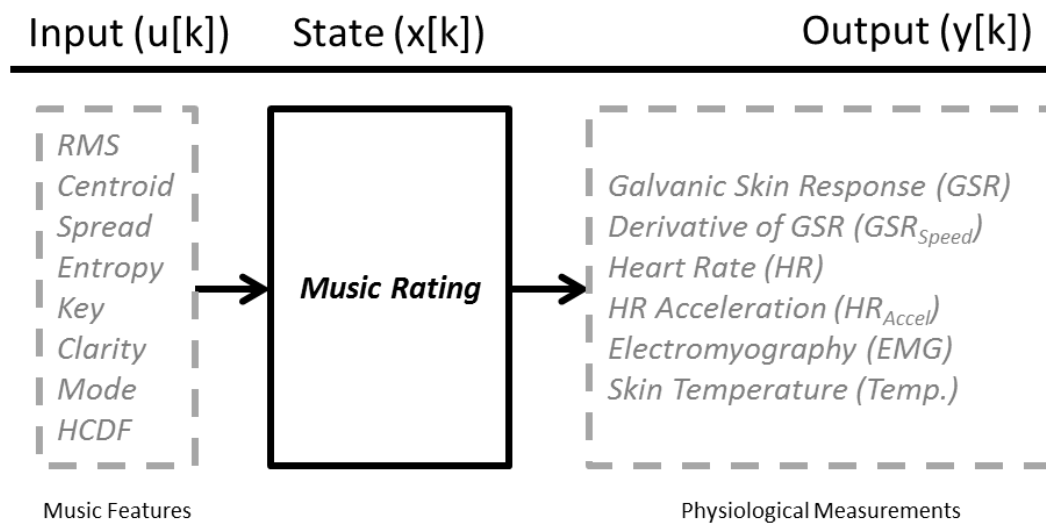


Figure 28. Diagram showing input (u – features drawn from music), states (x – user preference for the music) and outputs (y – physiological measures) of the system model.

Least Squares Approximation

Employing the data collected from the trials, we use a least squares approximation method to generate our state-space models for each participant. From time-series data composed of physiological, song feature and music selection rating measurements, we generated an over-determined system of equations:

$$b = Mv \quad \text{Equation 4.}$$

Where:

$$b = \begin{bmatrix} \textit{Rating}[1] & \textit{Rating}[2] & \cdots & \textit{Rating}[n] \\ \textit{GSR}[0] & \textit{GSR}[1] & \cdots & \textit{GSR}[n-1] \\ \textit{GSR}_{\textit{Speed}}[0] & \textit{GSR}_{\textit{Speed}}[1] & \cdots & \textit{GSR}_{\textit{Speed}}[n-1] \\ \textit{HR}[0] & \textit{HR}[1] & \cdots & \textit{HR}[n-1] \\ \textit{HR}_{\textit{Accel}}[0] & \textit{HR}_{\textit{Accel}}[1] & \cdots & \textit{HR}_{\textit{Accel}}[n-1] \\ \textit{EMG}[0] & \textit{EMG}[1] & \cdots & \textit{EMG}[n-1] \\ \textit{Temp}[0] & \textit{Temp}[1] & \cdots & \textit{Temp}[n-1] \end{bmatrix}$$

$$M = \begin{bmatrix} A_{11} & B_{11} & \cdots & B_{18} \\ C_{11} & D_{11} & \cdots & D_{18} \\ C_{21} & D_{21} & \cdots & D_{28} \\ C_{31} & D_{31} & \cdots & D_{38} \\ C_{41} & D_{41} & \cdots & D_{48} \\ C_{51} & D_{51} & \cdots & D_{58} \\ C_{61} & D_{61} & \cdots & D_{68} \end{bmatrix}$$

$$v = \begin{bmatrix} \textit{Rating}[0] & \textit{Rating}[1] & \cdots & \textit{Rating}[n-1] \\ \textit{RMS}[0] & \textit{RMS}[1] & \cdots & \textit{RMS}[n-1] \\ \textit{Centroid}[0] & \textit{Centroid}[1] & \cdots & \textit{Centroid}[n-1] \\ \textit{Spread}[0] & \textit{Spread}[1] & \cdots & \textit{Spread}[n-1] \\ \textit{Entropy}[0] & \textit{Entropy}[1] & \cdots & \textit{Entropy}[n-1] \\ \textit{Key}[0] & \textit{Key}[1] & \cdots & \textit{Key}[n-1] \\ \textit{Clarity}[0] & \textit{Clarity}[1] & \cdots & \textit{Clarity}[n-1] \\ \textit{Mode}[0] & \textit{Mode}[1] & \cdots & \textit{Mode}[n-1] \\ \textit{HCDF}[0] & \textit{HCDF}[1] & \cdots & \textit{HCDF}[n-1] \end{bmatrix}$$

M is a matrix which amalgamates the state (A), input (B), output (C) and feedforward (D) matrices of the discrete-time linear state-space model – it fully describes the model by which we estimate user music selection preference. v contains input (u) and state (x) data at timestep k , while b contains output (y) and state (x) data at timestep $k + 1$. Thus, Equation 4 simply defines that when M is multiplied by the inputs and state at timestep k , we should obtain outputs and

state at timestep $k + 1$. As the matrix M is initially unknown, it is estimated numerically by obtaining the least squares approximated solution of systems of simultaneous linear equations. Data points where music selection ratings (x) were saturated (rating at 0 or 1) were removed from the self-reported rating data sets during modelling; this was done to remove non-linear elements associated with the data during model generation.

4.4.2 Unscented Kalman Filter

For this work, we use the UKF to predict a user's internal emotional assessment of the music selection being presented based on: the state model developed through least squares approximation on training data, structural features of the music selection and the user's real-time physiological measurements. As mentioned in Section 4.2.2, the UKF uses a deterministic sampling technique known as the unscented transform to pick a minimal set of 'sigma' points around a data point. These sigma points are then pushed through non-linear functions (in our case, our state equations), from which the mean data value and covariance of the estimate are then recovered [110]. We offer a more in depth procedure below.

Initialization

From the least square approximate solution, we obtain the following model:

$$x[k + 1] = \text{sat}(Ax[k] + Bu[k]) + w[k] \quad \text{Equation 5.}$$

$$y[k] = Cx[k] + Du[k] + v[k] \quad \text{Equation 6.}$$

$$w[k] \sim (0, Q[k]) \quad \text{Equation 7.}$$

$$v[k] \sim (0, R[k]) \quad \text{Equation 8.}$$

Where $w[k], v[k]$ are the averages of the least squares error of states and outputs and $Q[k]$ and $R[k]$, are the co-variances of the least squares error of states and outputs respectively [110]. Non-linear saturation of the $Ax[k] + Bu[k]$ term in Equation 5 occurs at 0 and 1. We initialize the UKF with the following parameters:

$$\hat{x}_+[0] = E(x[0]) \quad \text{Equation 9.}$$

$$P_+[0] = E[(x[0] - \hat{x}_+[0])(x[0] - \hat{x}_+[0])^T] \quad \text{Equation 10.}$$

Where,

$\hat{x}_+[0]$ = estimated posteriori state mean at $k = 0$

$E(x[0])$ = expected value of x at $k = 0$

$P_+[0]$ = a posteriori state covariance at $k = 0$

Time-Step Propagation

Next, we choose sigma points $\hat{x}_{(i)}[k]$ for $i = 1, \dots, 2n$ using the following equations:

$$\hat{x}_{(i)}[k] = \hat{x}_+[k] + \tilde{x}_{(i)} \quad i = 1, \dots, 2n \quad \text{Equation 11.}$$

$$\tilde{x}_{(i)} = (\sqrt{nP_+[k]})_i^T \quad i = 1, \dots, n \quad \text{Equation 12.}$$

$$\tilde{x}_{(n+i)} = -(\sqrt{nP_+[k]})_i^T \quad i = 1, \dots, n \quad \text{Equation 13.}$$

Where,

$\hat{x}_+[k]$ = a posteriori state mean

$P_+[k]$ = a posteriori state covariance

n = number of system states

To demonstrate the relation between sigma points and the mean, Figure 29 shows a visualization of these sigma points in the case where we have two states (thus two state axes). Since our state model only uses one state (participant self-reported rating of music), $n = 1$ and the state plane is simplified to one dimension only. These sigma points are then propagated through the state transition equation to transform them into $\hat{x}_{(i)}[k + 1]$ vectors as shown in Equation 14.

$$\hat{x}_{(i)}[k + 1] = A\hat{x}_{(i)}[k] + Bu[k] \quad \text{Equation 14.}$$

Having acquired the $\hat{x}_{(i)}[k + 1]$ vectors, we can combine them to obtain our a priori estimate $\hat{x}_-[k + 1]$:

$$\hat{x}_-[k + 1] = sat \left(\frac{1}{2n} \sum_{i=1}^{2n} \hat{x}_{(i)}[k + 1] \right) \quad \text{Equation 15.}$$

and covariance, taking into account $Q[k]$, the covariance of the least squares error of the output:

$$P_-[k+1] = \frac{1}{2n} \sum_{i=1}^{2n} (\hat{x}_{(i)}[k+1] - \hat{x}_-[k+1]) (\hat{x}_{(i)}[k+1] - \hat{x}_-[k+1])^T + Q[k] \quad \text{Equation 16.}$$

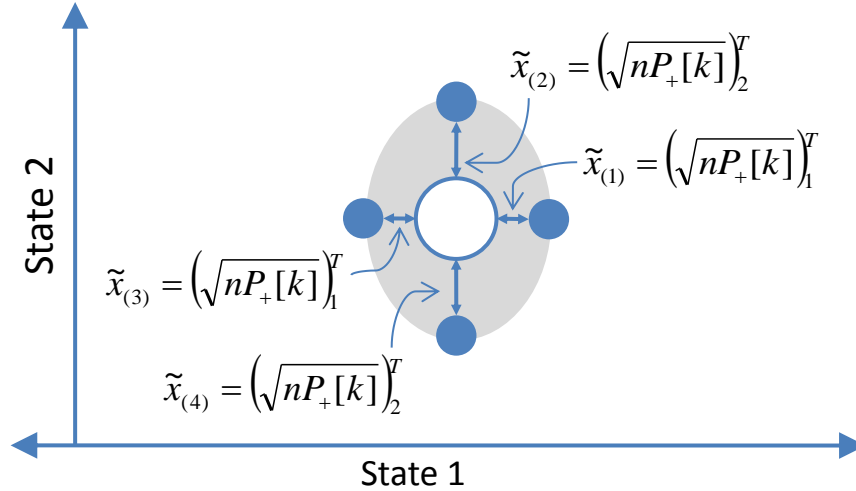


Figure 29. A diagram showing the choosing of sigma points (blue dots) around a data point (white dot).

Measurement Update

Much like in the time step propagation, we choose sigma points, this time using our a priori estimate and covariance acquired in the time step propagation:

$$\hat{x}_{(i)}[k+1] = \hat{x}_+[k+1] + \tilde{x}_{(i)} \quad i = 1, \dots, 2n \quad \text{Equation 17.}$$

$$\tilde{x}_{(i)} = (\sqrt{nP_-[k+1]})_i^T \quad i = 1, \dots, n \quad \text{Equation 18.}$$

$$\tilde{x}_{(n+i)} = -(\sqrt{nP_-[k+1]})_i^T \quad i = 1, \dots, n \quad \text{Equation 19.}$$

These sigma points are then applied to the output equation to obtain $\hat{y}_{(i)}[k+1]$ vectors:

$$\hat{y}_{(i)}[k+1] = C\hat{x}_{(i)}[k+1] + Du[k+1] \quad \text{Equation 20.}$$

The vectors are recombined to obtain our output estimate at time step $k+1$:

$$\hat{y}[k + 1] = \frac{1}{2n} \sum_{i=1}^{2n} \hat{y}_{(i)}[k + 1] \quad \text{Equation 21.}$$

The covariance can then be obtained - taking into account $R[k]$, the covariance of the least squares error of the output:

$$P_y[k + 1] = \frac{1}{2n} \sum_{i=1}^{2n} (\hat{y}_{(i)}[k + 1] - \hat{y}[k + 1]) (\hat{y}_{(i)}[k + 1] - \hat{y}[k + 1])^T + R[k] \quad \text{Equation 22.}$$

We find that the cross-covariance between x and y can then be calculated to be used in determining the Kalman gain:

$$P_{xy}[k + 1] = \frac{1}{2n} \sum_{i=1}^{2n} (\hat{x}_{(i)}[k + 1] - \hat{x}_-[k + 1]) (\hat{y}_{(i)}[k + 1] - \hat{y}[k + 1])^T \quad \text{Equation 23.}$$

Finally, a slightly modified version of the original KF equations can be used to perform the rest of the measurement update:

$$K[k + 1] = P_{xy}P_y^{-1} \quad \text{Equation 24.}$$

$$\hat{x}_+[k + 1] = \hat{x}_-[k + 1] + K[k + 1](y[k + 1] - \hat{y}[k + 1]) \quad \text{Equation 25.}$$

$$P_+[k + 1] = P_-[k + 1] - K[k + 1]P_yK[k + 1]^T \quad \text{Equation 26.}$$

Where,

K = Kalman gain

$y[k + 1]$ = output at time-step $k + 1$

Following this calculation, we move forward one time-step – e.g., $k + 1$ becomes k , and restart the algorithm, performing the time-step propagation and measurement update steps again.

4.5 Results

4.5.1 Performance Evaluation Criteria

To obtain some measure of how well the UKF estimated user preference for music selections, we developed two criteria by which we could test our estimates based on how we could realistically use this data in a human-computer interaction: root-mean-square difference and visual inspection.

Root-Mean-Square Difference (RMSD)

The root-mean square difference (RMSD) is a measure of the average difference between the participant's self-reported and UKF-estimated rating of the music selection, calculated using the following equation:

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (x_{estimated,i} - x_{reported,i})^2}{n}} \quad \text{Equation 27.}$$

As shown in Figure 30, the greater the correlation between the estimated and reported ratings, the smaller the value; the less the correlation, the closer the RMSD value will be to 1. We use 0.20 as a benchmark value which categorizes an estimate as good (<0.20) or poor (>0.20). Any difference greater than this is more likely to indicate that the estimated and reported ratings exist in different categorization 'bands' (0-0.4 = dislike, 0.4-0.6 = neutral, 0.6-1 = like) and thus the estimate is generally invalid.

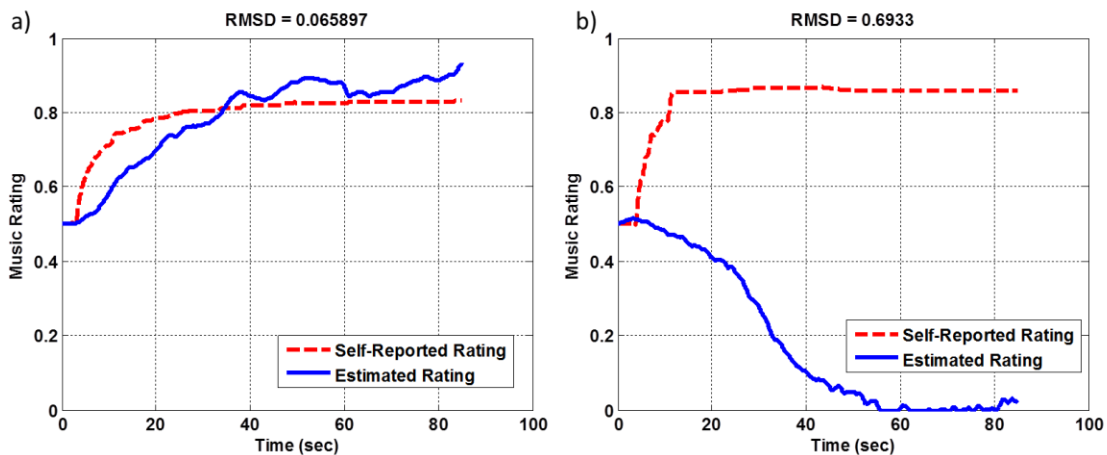


Figure 30. Examples of UKF results showing a) good correlation between estimated and reported music ratings resulting in a low RMSD value, and b) poor correlation resulting in a high RMSD value.

Visual Inspection

We used a quick, visual method of inspecting if the estimate is in the same ‘band’ (0-0.4 = dislike, 0.4-0.6 = neutral, 0.6-1 = like) as the reported measurement. This measurement technique was carried out as a simple indicator of whether the UKF can correctly categorize the music selection into an appropriate ‘bin’ and select the correct course of action - e.g., continue playing the current selection or to skip to the next song. To take this measurement, we use the value of the rating estimate at 40 seconds (work by Hazelton in [56] indicates that users are able to develop an emotional response indicating preference within this time period), and compare the value to the reported rating at that time as shown in Figure 31. If the two measurements are in the same bin, we would categorize the estimate as successful. If the estimated and reported ratings are in different bins, the estimate would be considered to have failed.

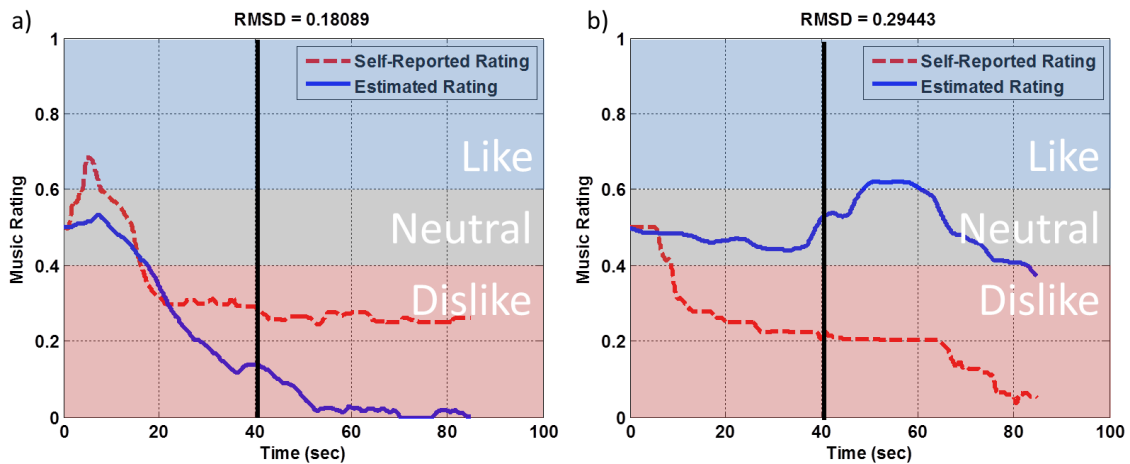


Figure 31. Examples of UKF results showing a) good estimation performance since both estimated and reported ratings at 40s fall in the same bin, and b) poor estimation performance since both estimated and reported ratings fall in different bins.

4.5.2 Self-Validation Testing

Self-Validation Testing (SVT) was used as a simple test to determine if the UKF could indeed produce a reasonable estimate of affect. An SVT consisted of applying a UKF seeded with a system model obtained from a specified data set on the same data set. Although a high degree of correlation between the participant-reported and estimated music selection ratings in the SVT doesn't guarantee that a more general system model will perform well, a poor correlation will definitively indicate that a generalized UKF will also perform poorly.

From analysis of all 75 data sets (3 participants x 25 music selections per participant), we found that the average RMSD across all subjects was quite small ($M=0.040$, $SD = 0.026$); Table 2

shows the RMSD mean and standard deviation for each participant. Only five trials of the total 75 (7%) exhibited an RMSD greater than 0.1. This indicated excellent correlation between participant reported and estimated music selection ratings in this first benchmark test.

Table 2. Mean and std. dev. of root-mean-square deviation for each participant.

	Participant 1	Participant 2	Participant 3
Genre	Soul/R&B	Hip Hop	Classical
Mean	0.037	0.051	0.031
RMSD			
SD	0.022	0.029	0.025

4.5.3 Familiar Music Selection Testing

Encouraged by the fairly low RMSD results from self-validation testing, we felt that proceeding to use the UKF in a more generalized sense was justified. For this test, we generated a general state space model for a participant and applied it to all trials where the music selection was familiar (as indicated in the post-trial questionnaire). The general state space model was obtained by concatenating all of the trial data sets from a participant except for the trial music selection being tested and those trials which had music selections which were unfamiliar to the participant.

RMSD Result

From RMSD results obtained from data where music selections were familiar to the participants, we found that the UKF using generalized models performed poorly: the average RMSD across all participants was 0.313 ($SD = 0.235$), which is greater than our benchmark value of 0.20.

A one sample t-test revealed that the mean RMSD was statistically different between music that was reported as ‘liked’ ($n=30, M=0.288, SD=0.246$) and ‘disliked’ ($n=30, M=0.358, SD=0.217$) by the participants ($p<0.0001$), with disliked music having a greater RMSD. This signifies that the UKF performs more poorly on music that is disliked by participants. A breakdown of mean RMSDs for each participant and ‘liked’ vs. ‘disliked’ music selections is shown in Table 3⁸.

⁸ Table 3 does not display ‘Neutral’ rating trajectories as all of the familiar music selections were provided by the participant themselves which consisted only of music that was ‘liked’ or ‘disliked’ (See Section 4.3.5).

Table 3. Mean and std. dev. of root mean square deviations per participant, for music selections which are familiar to participants.

	Participant 1	Participant 2	Participant 3
Genre	Soul/R&B	Hip Hop	Classical
Mean RMSD (Std. Dev.)	0.303 (0.230) N=20	0.355 (0.255) N=20	0.276 (0.223) N=20
Mean RMSD for ‘Liked’ Music (Std. Dev.)	0.290 (0.263) N=10	0.347 (0.275) N=10	0.241 (0.205) N=10
Mean RMSD for ‘Disliked’ Music (Std. Dev.)	0.346 (0.203) N=10	0.401 (0.238) N=10	0.371 (0.249) N=10

Visual Inspection

From visual inspection of the 60 data sets which used music provided by participants themselves, we found that the generalized models were mediocre in performance; 34 (57%) of estimated rating trajectories would have correctly identified the participants’ reported ratings using the criteria described in Section 4.5.1. Table 4 shows the number of rating trajectories correctly tracked for each participant⁹.

Table 4. Visual inspection results of estimated rating trajectories for each participant for music selections which were familiar to participants.

	Participant 1	Participant 2	Participant 3
Genre	Soul/R&B	Hip Hop	Classical
No. of ‘Liked’ Trajectories Tracked Correctly	8/10 (80%)	6/10 (60%)	7/10 (70%)
No. of ‘Disliked’ Trajectories Tracked Correctly	4/10 (40%)	5/10 (50%)	4/10 (40%)
Total No. of Trajectories Tracked Correctly	12/20 (60%)	11/20 (55%)	11/20 (55%)

4.5.4 Unfamiliar Music Selection Testing

In addition to estimating the participant rating of recognized music selections, we wanted to test the UKF within a context that was more representative of our use case - where the system may have to estimate rating for music selections which were not familiar to the user. To perform this test, we generated a system model using data collected from trials with familiar music selections for each participant. This model was then used by the UKF to estimate the rating for selections which were unfamiliar to the participant (as indicated by the participant in the post-trial questionnaire).

⁹ Table 4 does not display ‘Neutral’ rating trajectories as all of the familiar music selections were provided by the participant themselves which consisted only of music that was ‘liked’ or ‘disliked’ (See Section 4.3.5).

RMSD Result

RMSD results obtained from data where music selections were unfamiliar to participants showed that the generalized models derived from familiar music selection data provided only mediocre performance: the average RMSD across all participants was 0.296 ($n=15$, $SD = 0.229$). Again, this did not fall below our specified benchmark value of 0.20. Mean RMSDs for each participant and ‘liked’ vs. ‘disliked’ music selections is shown in Table 5.

Table 5. Mean and std. dev. of root mean square deviations per participant, for music selections which are unfamiliar to participants.

	Participant 1	Participant 2	Participant 3
Mean RMSD (Std. Dev.)	0.211 (0.221) N=5	0.445 (0.272) N=5	0.233 (0.217) N=5
Mean RMSD for ‘Liked’ Music	N=0	0.689 N=1	0.327 N=1
Mean RMSD for ‘Neutral’ Music	0.0278 N=2	0.389 N=2	0.158 N=2
Mean RMSD for ‘Disliked’ Music	0.2043 N=3	0.381 N=2	0.524 N=2

Visual Inspection

Visual inspection of the 15 data sets which used music which was unfamiliar to participants revealed that categorization abilities of the UKF is poor; only 5 (30%) of estimated rating trajectories would have correctly identified the participants’ reported ratings using the criteria described in Section 4.5.1. Table 6 shows the number of rating trajectories correctly tracked for each participant.

Table 6. Visual inspection results of estimated rating trajectories for each participant for music selections which were familiar to participants.

	Participant 1	Participant 2	Participant 3
No. of ‘Liked’ Trajectories Tracked Correctly	0/0	0/1 (0%)	0/1 (0%)
No. of ‘Neutral’ Trajectories Tracked Correctly	2/2 (100%)	0/2 (0%)	1/2 (50%)
No. of ‘Disliked’ Trajectories Tracked Correctly	0/3 (0%)	1/2 (50%)	1/2 (50%)
Total No. of Trajectories Tracked Correctly	2/5 (40%)	1/5 (20%)	2/5 (40%)

4.5.5 Observed Behaviours

During the experiment, we observed a wide range of physical reactions between and within participants during trials, particularly to trials with music selections which were reported as being disliked. In some trials, participants seemed visibly frustrated, manifested through expressions such as frowning, furrowing their brows, squirming etc. Other times, participants displayed no noticeable reaction to the music selection.

4.6 Discussion and Lessons Learned

Although this study was a preliminary one, we feel the results show clear indications that the UKF method implemented here does not perform well; namely, high RMSD values and large differences in trajectory paths between estimated and reported continuous ratings are observed when trying to track both familiar and unfamiliar music selections. There are many possible reasons to explain why our technique failed to estimate participant music preference correctly; such reasons may include:

- the likelihood that there is a non-linear relation between music preference and physiological/musical features which cannot be accounted for adequately using a saturated least squares modeling technique;
- the absence of a strong correlation between similar musical features and music preference – a user listening to a song with similar musical features as another song that is liked does not necessarily mean that the user will enjoy listening to it;
- there being a psycho-physiological difference in the type of experience a user has between listening to music selections that are familiar and well-known to the user and those that are being encountered and explored for the first time (this explain the extremely poor performance of the UKF when trying to estimate participant ratings of unfamiliar songs).

Despite the poor results obtained with using the UKF, we cannot claim that Kalman filtering would not be a suitable approach to estimating user preference of music - we have simply developed user models using one combination of input and output parameters. Using other measures derived from physiology and music structure or using non-linear modeling may offer better performance.

It is interesting to note that performance of the UKF is worse for music selections that are disliked. We believe this may be a direct result of what was observed behaviorally and through self-reports across participants in this pilot: music that is not appealing can cause a wide range of negative reactions varying in type of expression and scale. Negative emotions leading to someone disliking a music selection can arise from a number of psycho-physiological pathways; for example, a listener could detest a song because of its genre (i.e., structural characteristics of the music), or because it reminds him/her of an emotionally distressful event (e.g., a breakup). These two pathways could cause very different physiological reactions which the UKF may not do well in recognizing and handling.

Although the use of the unscented Kalman filter yielded quantitative results which were below expectations, we feel that this pilot study was able to provide valuable insight as to how an affect-recognition system could/should behave, and the problems that will be faced when developing such a system.

4.6.1 Limitations, Improvements and Future Considerations

As this was primarily a pilot study, there is much work that could be done to further investigate and build upon the findings of this work.

In terms of the study conducted, we recognize that a small participant group size was a limitation of this study. A larger pool of participants would have provided more conclusive evidence in determining effectiveness of the UKF method for affect recognition as implemented. Additionally, future studies could be performed which test in conditions outside a controlled, structured lab environment which is unable to reflect a more realistic listening scenario (e.g., busy café, on a bus, during a jog), identify differences in the effect of listening to new and unfamiliar music selections versus music selections which are recognized and familiar, and investigate the effects of a wider variation in music genre presented to participants.

With regards to improving performance of the UKF, we propose that a more careful examination of physiological features should be conducted in determining which physiological measures are most indicative of music preference, and reducing the output matrix to only a few, key physiological signal measurements. In this experiment, we have used a ‘shotgun’ approach in determining which physiological measures to use, hoping to obtain a comprehensive model of the user’s physiological response to musical stimulation. In retrospect, a reduction in the number

of output states which would be monitored would greatly lessen the noise present in the system and could strengthen results in addition to simplifying computation of the UKF. Additionally, we have found it extremely difficult to use structural music features both in terms of the computational challenge of extracting this information from digital music files and using it to reliably predict user preference for the music. In particular, this pilot exercise has led us to discover that the correlation between music features and preference is not intuitive. Participant self-reports and quantitative musical data have indicated that there is a wide variety of genres and preferences to music – the user relationships with music seems to be highly complex and transcends simple, mathematical music features. Also, prior work on developing a correlation between music features and emotion is still largely theoretical rather than practical. Thus, we feel that it would be better to more formally examine the relationship between structural music features, user preference and physiological responses. It may very well be that these features need to be further refined and modulated in order to obtain a more direct path to automatically and non-intrusively determining user preference to music.

As a result of conducting this pilot experiment, we have grasped that current results do not support estimation of user rating of music selection through Kalman filtering of physiological and music feature data as a viable approach. However, we cannot completely discount the notion that Kalman filtering can provide a reliable method to estimate user affect and preference: the list of variables which can be explored has not been exhausted. Alternatively, we propose that future work investigate other predictive modeling methods (i.e. neural networks) which can use physiological measurements to estimate user preferences in music – particularly those methods that allow the system to recognize, accommodate and learn from errors identified by the user. We foresee that modeling techniques which allow the user to provide corrective feedback to the model and have the system ‘learn from past mistakes’ could provide immense benefit in improving system rating accuracy.

5 Conclusions and Future Work

In this work, we have proposed and demonstrated a novel, implicit human-computer interaction paradigm named the ‘Haptic-Affect Loop’ which includes a feedback mechanism utilizing the haptic channel. We attempt to demonstrate the feasibility of this implicitly-controlled interaction model which provides a possible solution for managing the interplay between primary and secondary tasks in various environments. The research documented in this thesis tries to validate the technological feasibility, utility and behavior of the Haptic-Affect Loop through two use cases: audio stream bookmarking and music-listening.

5.1 Audio-Stream Bookmarking

In our audio stream bookmarking use case, we have developed and validated our theory that a computer algorithm can, in real-time, detect features within an electrodermal measurement stream that correspond to ORs caused by external disruptive interruptions. We have exploited these findings with a system that provides OR-based auto-marking for a representative audiobook use case. We have found that with our algorithm, we can achieve a true-positive interruption detection sensitivity of 84%. We anticipate that the false-positive rate found in this use case may be addressable by dynamically tailoring the OR detection algorithm to individuals, through adjusting sensitivity, bookmarking delay and dead-zone period until another bookmark can be placed. Improvements to the present system’s usability in the field (e.g., further reducing noise in the GSR introduced by bodily movements) may be assisted by integrating contextual information such as accelerometer and location data from a worn or carried device.

As a next step in this work, we have also examined the role of low-attention haptic feedback for this use case; this feedback provides users with a low-attention notification that the application has acted in response to changes in the user’s physiology. Two forms of feedback were tested through an experiment in which bookmark placement is simulated rather than automated, for purpose of experiment control: haptic notification of bookmark placement and haptic display of bookmarks during navigation. The work presented demonstrates the feasibility of using haptic feedback for this use case. We found that conditions providing haptic display of bookmarks during resumptive navigation significantly reduced navigation time and accuracy. Visual and haptic bookmark displays in resumptive navigation, alone and in combination, were all found to be faster than no bookmark display, although the combination of haptic and visual bookmark

displays provided the strongest statistical results. These findings confirm previous work regarding haptic augmentation of visual displays, and also show promising results regarding haptic-only interactions for simple eyes-free navigation tasks.

5.2 Music Listening

A more general technological validation test for affect estimation was assessed on a music-listening use case in the form of a pilot study. From physiological and music feature data, we constructed a model intended to estimate a user's rating of a music selections using an unscented Kalman filter. This attempt provided mediocre to poor results when rating estimates generated by the unscented Kalman filter were compared with user reported music rating data. We hypothesize that this poor performance can be attributed to confounding elements such as the wide range of non-linear physiological responses correlating to emotional reactions to music, and an absence of a strong correlation between structural features of music, which are utilized within the model, and music preference, which this modeling approach could not adequately account for. The exploration gave insight into possible explanations, and pointed towards other approaches which might be more successful.

5.3 General Implications

This work presented in this thesis provides a model for an implicit/low-attention audio stream bookmarking system and identifies potential challenges for applying it to other contexts and use cases. Through this research, we have developed a complete, functional instance of a HALO loop (for the audio stream bookmarking use case) which was able to provide valuable insight as to how the HALO behaves, and how users may expect it to behave. Additionally, through an initial attempt to algorithmically estimate affect (in the music-listening use case), we have identified several problems that will be faced when developing other systems which use implicitly-communicative user-computer interactions.

To our knowledge, the system we present here is the first implicit/low-attention interaction loop for human-device interaction using haptic feedback for closing of the loop. The work described in this thesis, although aimed towards use cases, represents a first step in a larger effort in which we aim to develop an implicit interaction loop driven by physiological parameters; we have simply used these use cases as a framework for development of HALO technologies. Thus, we feel that the application of this research can be readily applied to other contexts – e.g., athlete

training, vehicle operation, classroom learning, information retrieval, industrial processes and robotics.

5.4 Future Work

With respect to the audio stream bookmarking application, future work will include the examination of the usefulness of the system in a more realistic setting with the implementation of GSR detection to place bookmarks, implementing a fully closed-loop interaction system and studied with a larger cohort of users. Several hypotheses have arisen in the course of the research which invites further investigation, including the premise that with haptic notification of bookmark placement at interruption time, users are able to develop a spatial mental map of where both useful and non-useful bookmarks have been placed. Thus, when provided with a visual-spatial mapping of these bookmarks during resumptive navigation, they are able to use this mental model to achieve a performance benefit in terms of navigation speed and accuracy.

Within the music-listening scenario, we hope to explore other machine learning approaches aside from Kalman filtering to improve estimation performance (e.g., particle filtering, neural networks, etc.). In relation to this, future work can be directed towards modeling techniques which can learn from mistakes, allowing corrective input from the user to improve the model. This form of user correction may have immense benefit for HALO in general, reducing the occurrence of errors over time.

Finally, we hope to generalize the application of the HALO interaction paradigm, extending it beyond media playback applications. To further test and validate this model, we hope to apply the implicit closed-loop dynamics of HALO to other use cases such as streaming media, classroom learning, video gaming, internet browsing, and vehicle operation.

References

- [1] S. Chatty and P. Dewan, Eds., *Engineering for Human-Computer Interaction*. Springer, 1999, p. 392.
- [2] C. Wickens and J. Hollands, *Engineering Psychology and Human Performance*, 3rd ed. Upper Saddle River: Prentice Hall, 1999.
- [3] E. M. Altmann and G. J. Trafton, “Task Interruption: Resumption Lag and the Role of Cues,” *Proc. Conf. Cognitive Science Society*, Aug. 2004.
- [4] B. Bailey and J. Konstan, “On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state,” *Computers in Human Behavior*, vol. 22, no. 4, pp. 685–708, Jul. 2006.
- [5] M. A. Mazmanian, W. J. Orlikowski, and J. Yates, “Crackberries: The Social Implications of Ubiquitous Wireless E-Mail Devices,” in *Designing Ubiquitous Information Environments: Socio-Technical Issues and Challenges*, vol. 185, C. Sørensen, Y. Yoo, K. Lyytinen, and J. I. DeGross, Eds. New York: Springer-Verlag, 2005, pp. 337–343.
- [6] S. L. Jarvenpaa and K. R. Lang, “Managing the Paradoxes of Mobile Technology,” *Information Systems Management*, vol. 22, no. 4, pp. 7–23, Sep. 2005.
- [7] S. L. Jarvenpaa, L. K. Reiner, and V. K. Tuunainen, “Designing Ubiquitous Information Environments: Socio-Technical Issues and Challenges,” in *Designing Ubiquitous Information Environments: Socio-Technical Issues and Challenges*, vol. 185, C. Sørensen, Y. Yoo, K. Lyytinen, and J. I. DeGross, Eds. New York: Springer-Verlag, 2005, pp. 29–32.
- [8] G. D. Abowd, E. D. Mynatt, and T. Rodden, “The human experience [of ubiquitous computing],” *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 48–57, Jan. 2002.
- [9] S. Poslad, *Ubiquitous Computing*. Chichester, UK: John Wiley & Sons, Ltd, 2009.
- [10] M. Weiser and J. S. Brown, “Designing Calm Technology,” *Powergrid Journal*, vol. 1.01, no. July, pp. 94–110, 1996.

- [11] M. K. X. J. Pan, J.-S. Chang, G. H. Himmetoglu, Aj. Moon, T. W. Hazelton, K. E. MacLean, and E. A. Croft, "Now Where was I? Physiologically-Triggered Bookmarking," in *Proc. ACM Conf. CHI*, 2011, pp. 363–372.
- [12] M. K. X. J. Pan, G. J.-S. Chang, G. H. Himmetoglu, Aj. Moon, T. W. Hazelton, K. E. MacLean, and E. A. Croft, "Galvanic skin response-derived bookmarking of an audio stream," in *Proc. ACM Conf. CHI EA*, 2011, p. 1135.
- [13] A. Schmidt, "Implicit Human Computer Interaction Through Context," *Personal Technologies*, vol. 4, no. 2–3, pp. 191–199, Jun. 2000.
- [14] N. S. Shami, J. T. Hancock, C. Peter, M. Muller, and R. Mandryk, "Measuring Affect in HCI: going beyond the individual," in *Proc. ACM Conf. CHI EA*, 2008, pp. 3901–3904.
- [15] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion Recognition in Human-Computer Interaction," *IEEE Signal Processing Mag.*, vol. 18, no. 1, pp. 32–80, 2001.
- [16] "Emotiv," 2012. [Online]. Available: <http://www.emotiv.com/>. [Accessed: 15-Aug-2012].
- [17] "NeuroSky," 2012. [Online]. Available: www.neurosky.com. [Accessed: 20-Jul-2012].
- [18] S. Yohanan, M. Chan, J. Hopkins, H. Sun, and K. Maclean, "Hapticat: Exploration of Affective Touch," *Computer*, pp. 222–229, 2005.
- [19] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion Recognition System using Short-Term Monitoring of Physiological Signals," *Medical & Biological Engineering & Computing*, vol. 42, no. 3, pp. 419–27, May 2004.
- [20] A. Dix, *Human-Computer Interaction*. Pearson Education, 2004, p. 834.
- [21] S. Brewster and A. Constantin, "Tactile Feedback for Ambient Awareness in Mobile Interactions." .
- [22] J. B. F. Van Erp and H. A. H. C. Van Veen, "Vibrotactile In-vehicle Navigation System," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 7, no. 4–5, pp. 247–256, Jul. 2004.

- [23] S. Brewster and L. M. Brown, "Tactons: structured tactile messages for non-visual information display," in *Proc. AUIC*, 2004, vol. 4, pp. 15–23.
- [24] A. Tang, P. Mclachlan, and K. Lowe, "Perceiving Ordinal Data Haptically Under Workload," in *Proc. Int. Conf. Multimodal Interfaces*, 2005, pp. 317–324.
- [25] B. M. Davis, "Effects of Visual, Auditory, and Tactile Navigation Cues on Navigation Performance, Situation Awareness, and Mental Workload," 2007.
- [26] M. A. Baumann, K. E. Maclean, T. W. Hazelton, and A. Mckay, "Emulating Human Attention-Getting Practices with Wearable Haptics," in *IEEE Haptics Symposium*, 2010.
- [27] K. E. MacLean, "Foundations of Transparency in Tactile Information Design," *IEEE Trans. Haptics*, vol. 1, no. 2, pp. 84–95, Jul. 2008.
- [28] R. W. Picard, *Affective Computing*, 1st ed. Cambridge, MA: MIT Press, 1998, p. 292.
- [29] P. Ekman, "Are there basic emotions?," *Psychological review*, vol. 99, no. 3, pp. 550–553, 1992.
- [30] L. F. Barrett, "Are Emotions Natural Kinds?," *Perspectives on Psychological Science*, vol. 1, no. 1, pp. 28–58, Mar. 2006.
- [31] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and Emotion*, vol. 5, no. 4, pp. 345–379, Dec. 1981.
- [32] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [33] K. R. Scherer, "Psychological Models of Emotion," in *The Neuropsychology of Emotion - Series in Affective Science*, New York, New York, USA: Oxford University Press, 2000, pp. 137–162.
- [34] P. J. Lang, "The Emotion Probe: Studies of Motivation and Attention," *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.

- [35] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 03, pp. 715–734, Sep. 2005.
- [36] A. Schmidt, H.-W. Gellersen, and C. Merz, "Enabling Implicit Human Computer Interaction: A Wearable RFID-Tag Reader," in *Int. Symp. Wearable Computers*, 2000, pp. 193–194.
- [37] K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz, "Sensing techniques for mobile interaction," in *Proc. ACM Symp. UIST*, 2000, pp. 91–100.
- [38] D. McFarlane and K. Latorella, "The Scope and Importance of Human Interruption in Human-Computer Interaction Design," *Human-Computer Interaction*, vol. 17, no. 1, pp. 1–61, Mar. 2002.
- [39] G. Fischer, "User Modeling in Human-Computer Interaction," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1, pp. 65–86, Mar. 2001.
- [40] K. Hinckley, J. Pierce, E. Horvitz, and M. Sinclair, "Foreground and Background Interaction with Sensor-Enhanced Mobile Devices," *ACM Trans. CHI*, vol. 12, no. 1, pp. 31–52, Mar. 2005.
- [41] J. A. Fails and D. R. Olsen, "Interactive machine learning," in *Proc. Int. Conf. IUI*, 2003, p. 39.
- [42] C. Conati, "Probabilistic Assessment of User's Emotions in Educational Games," *Applied Artificial Intelligence*, vol. 16, no. 7–8, pp. 555–575, Aug. 2002.
- [43] D. Kulic and E. Croft, "Estimating Intent for Human-Robot Interaction," in *Proc. IEEE Int. Conf. Advanced Robotics*, 2003, pp. 810–815.
- [44] C. L. Lisetti and F. Nasoz, "MAUI: A Multimodal Affective User Interface," in *ACM Multimedia*, 2002, pp. 161–170.
- [45] M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proc. IEEE*, vol. 91, no. 9, pp. 1370–1390, Sep. 2003.

- [46] N. Park, W. Zhu, Y. Jung, M. Mclaughlin, and S. Jin, "Utility of Haptic Data in Recognition of User State," in *Proc. HCI Int.*, 2005.
- [47] R. W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [48] R. W. Picard, "Toward Agents that Recognize Emotion," *MIT Media Lab Perceptual Computing Section Technical Report*, no. 515, 1998.
- [49] E. L. Broek, V. Lisý, J. H. Janssen, J. H. D. M. Westerink, M. H. Schut, K. Tuinenbreijer, A. Fred, J. Filipe, and H. Gamboa, *Biomedical Engineering Systems and Technologies*, vol. 52. Heidelberg: Springer, 2010, pp. 21–47.
- [50] K. Takahashi and A. Tsukaguchi, "Remarks on Emotion Recognition from Multi-Modal Bio-Potential Signals," in *IEEE Int. Conf. SMC*, 2003, vol. 2, pp. 1654–1659.
- [51] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction," in *HCI*, 2007, pp. 29–38.
- [52] J. Healey and R. W. Picard, "StartleCam: a cybernetic wearable camera," in *Int. Symp. Wearable Computers*, 1998, no. 468, pp. 42–49.
- [53] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, *The Handbook of Psychophysiology*, 3rd Editio., vol. 54. New York, New York, USA: Cambridge University Press, 2007.
- [54] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion Recognition Using Biosensors : First Steps towards an Automatic System," *Affective Dialogue Systems*, vol. 3068, pp. 36–48, 2004.
- [55] C. Liu, P. Rani, and N. Sarkar, "Human-Robot Interaction Using Affective Cues," in *IEEE Int. Symp. Robot and Human Interactive Communication*, 2006, pp. 285–290.
- [56] T. W. Hazelton, "Investigating, designing, and validating a haptic-affect interaction loop using three experimental methods." University of British Columbia, 05-Nov-2010.

- [57] C. D. Frith and H. A. Allen, "The skin conductance orienting response as an index of attention," *Biological Psychology*, vol. 17, no. 1, pp. 27–39, 1983.
- [58] W. Boucsein, *Electrodermal activity*. Plenum Press, 1992, p. 442.
- [59] J. T. Cacioppo and L. G. Tassinari, "Inferring psychological significance from physiological signals.," *The American psychologist*, vol. 45, no. 1, pp. 16–28, Jan. 1990.
- [60] A. S. Bernstein, C. D. Frith, J. H. Gruzelier, T. Patterson, E. Straube, P. H. Venables, and T. P. Zahn, "An analysis of the skin conductance orienting response in samples of American, British, and German schizophrenics.," *Biological psychology*, vol. 14, no. 3–4, pp. 155–211.
- [61] "Thought Technology," 2012. [Online]. Available: www.thoughttechnology.com. [Accessed: 20-Jul-2012].
- [62] C. Anderson, *Free: The Future of a Radical Price*. New York, New York, USA: HarperCollins, 2009.
- [63] J. Vanerp and H. Vanveen, "Vibrotactile in-vehicle navigation system," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 7, no. 4–5, pp. 247–256, 2004.
- [64] A. Tang, P. Mclachlan, and K. Lowe, "Perceiving Ordinal Data Haptically Under Workload," in *Proceedings of the 7th international conference on Multimodal interfaces*, 2005, pp. 317–324.
- [65] K. E. MacLean, "Putting Haptics into the Ambience," *IEEE Transactions on Haptics*, vol. 2, no. 3, pp. 123–135, Jul. 2009.
- [66] K. Kahol, V. Hayward, and S. Brewster, "Ambient Haptic Systems," *IEEE Trans. Haptics*, vol. 2, no. 3, pp. 121–122, 2009.
- [67] R. Leung, K. MacLean, M. B. Bertelsen, and M. Saubhasik, "Evaluation of Haptically Augmented Touchscreen GUI Elements under Cognitive Load," in *Proc. ICMI*, 2007, p. 374.
- [68] B. A. Swerdfeger, "A First and Second Longitudinal Study of Haptic Icon Learnability: The Impact of Rhythm and Melody," University of British Columbia, 2009.

- [69] J. Pasquero, S. J. Stobbe, and N. Stonehouse, "A Haptic Wristwatch for Eyes-Free Interactions," in *Proc. ACM Conf. CHI*, 2011, p. 3257.
- [70] K. E. Maclean, "Putting Haptics into the Ambience," *IEEE Trans. Haptics*, vol. 2, no. 3, pp. 123–135, 2009.
- [71] J. B. F. van Erp, H. A. H. C. van Veen, C. Jansen, and T. Dobbins, "Waypoint navigation with a vibrotactile waist belt," *ACM Trans. Applied Perception*, vol. 2, no. 2, pp. 106–117, Apr. 2005.
- [72] J. Peck and J. Wiggins, "It Just Feels Good: Customers' Affective Response to Touch and Its Influence on Persuasion," *Journal of Marketing*, vol. 70, no. 4, pp. 56–69, 2006.
- [73] R. L. Klatzky and J. Peck, "Please Touch: Object Properties that Invite Touch," *IEEE Trans. Haptics*, vol. 5, no. 2, pp. 139–147, Apr. 2012.
- [74] "Engineering Acoustics," 2012. [Online]. Available: www.eaiinfo.com. [Accessed: 20-Jul-2012].
- [75] "irrKlang," 2012. [Online]. Available: www.ambiera.com/irrklang/. [Accessed: 20-Jul-2012].
- [76] S. McLean, "Vinyl Cafe Family Pack," CBC, 2011.
- [77] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 1988, p. 567.
- [78] J. A. Sloboda and P. N. Juslin, "Psychological Perspectives on Music and Emotion," in *Music and Emotion: Theory and Research*, P. N. Juslin, Ed. New: Oxford University Press, 2001, pp. 71–104.
- [79] L. B. Meyer, *Emotion and Meaning in Music*. University Of Chicago Press, 1961, p. 315.
- [80] P. Juslin and P. Laukka, "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening," *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, Sep. 2004.
- [81] J. A. Etzel, E. L. Johnsen, J. Dickerson, D. Tranel, and R. Adolphs, "Cardiovascular and respiratory responses during musical mood induction.," *International journal of*

psychophysiology : official journal of the International Organization of Psychophysiology, vol. 61, no. 1, pp. 57–69, Jul. 2006.

- [82] Y.-H. Yang, C.-C. Liu, and H. H. Chen, “Music emotion classification,” in *Proceedings of the 14th annual ACM international conference on Multimedia - MULTIMEDIA '06*, 2006, p. 81.
- [83] A. Gabrielsson, “Emotion perceived and emotion felt: Same or different?”
- [84] “Pandora,” 2012. [Online]. Available: <http://www.pandora.com>. [Accessed: 20-Aug-2012].
- [85] “Last.fm,” 2012. [Online]. Available: www.last.fm. [Accessed: 20-Aug-2012].
- [86] “Grooveshark,” 2012. [Online]. Available: www.grooveshark.com. [Accessed: 20-Aug-2012].
- [87] “8Tracks,” 2012. [Online]. Available: www.8tracks.com. [Accessed: 20-Aug-2012].
- [88] P. Rani, N. Sarkar, and J. Adams, “Anxiety-based affective communication for implicit human–machine interaction,” *Advanced Engineering Informatics*, vol. 21, no. 3, pp. 323–334, Jul. 2007.
- [89] G. Welch and G. Bishop, “An Introduction to the Kalman Filter,” Chapel Hill, NC, 1997.
- [90] R. J. Meinhold and N. D. Singpurwalla, “Understanding the Kalman Filter,” *The American Statistician*, vol. 37, no. 2, pp. 123–127, 1983.
- [91] E. A. Wan and R. Van Der Merwe, “The Unscented Kalman Filter for Nonlinear Estimation,” in *Proc. IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.
- [92] M. Maghami, R. A. Zoroofi, B. N. Araabi, M. Shiva, and E. Vahedi, “Kalman filter tracking for facial expression recognition using noticeable feature selection,” in *2007 International Conference on Intelligent and Advanced Systems*, 2007, pp. 587–590.

- [93] Z. Li, J. E. O'Doherty, T. L. Hanson, M. A. Lebedev, C. S. Henriquez, and M. A. L. Nicolelis, "Unscented Kalman filter for brain-machine interfaces.," *PloS one*, vol. 4, no. 7, p. 18, Jan. 2009.
- [94] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human Computing and Machine Understanding of Human Behavior: A Survey," in *Artificial Intelligence for Human Computing*, vol. 4451, T. S. Huang, A. Nijholt, M. Pantic, and A. Pentland, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 47–71.
- [95] N. Oliver, A. Pentland, and F. Bérard, "LAFTER: a real-time face and lips tracker with facial expression recognition," *Pattern Recognition*, vol. 33, no. 8, pp. 1369–1382, Aug. 2000.
- [96] E. M. Schmidt and Y. E. Kim, "Prediction of Time-Varying Musical Mood Distributions Using Kalman Filtering," in *2010 Ninth International Conference on Machine Learning and Applications*, 2010, pp. 655–660.
- [97] G. Kreutz, U. Ott, D. Teichmann, P. Osawa, and D. Vaitl, "Using music to induce emotions: Influences of musical preference and absorption," *Psychology of Music*, vol. 36, no. 1, pp. 101–126, Nov. 2007.
- [98] P. G. Hunter and E. G. Schellenberg, "Music and Emotion," in *Music Perception*, vol. 36, M. Riess Jones, R. R. Fay, and A. N. Popper, Eds. New York, NY: Springer New York, 2010, pp. 129–164.
- [99] G. C. Mornhinweg, "Effects of Music Preference and Selection on Stress Reduction," *Journal of Holistic Nursing*, vol. 10, no. 2, pp. 101–109, Jun. 1992.
- [100] J. Chung and G. S. Vercoe, "The Affective Remixer: Personalized Music Arranging," in *CHI '06 extended abstracts on Human factors in computing systems - CHI '06*, 2006, p. 393.
- [101] V. N. Salimpoor, M. Benovoy, G. Longo, J. R. Cooperstock, and R. J. Zatorre, "The rewarding aspects of music listening are related to degree of emotional arousal.," *PloS one*, vol. 4, no. 10, p. e7487, Jan. 2009.

- [102] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–83, Dec. 2008.
- [103] D. Yang and W. Lee, "Disambiguating music emotion using software agents," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR04)*, 2004, pp. 52–58.
- [104] T. Eerola, O. Lartillot, and P. Toiviainen, "Prediction of Multidimensional Emotional Ratings in Music from Audio using Multivariate Regression Models," in *Int. Soc. Music Information Retrieval Conf.*, 2009, no. Ismir, pp. 621–626.
- [105] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *International Journal of Computer Vision*, vol. 3, no. 3, pp. 209–238, Sep. 1989.
- [106] P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. Norwood, MA: Artech House, 2008, p. 518.
- [107] G. W. Brown and M. T. Cliff, "Investor sentiment and the near-term stock market," *Journal of Empirical Finance*, vol. 11, no. 1, pp. 1–27, Jan. 2004.
- [108] T. F. Bewley, Ed., *Advances in Econometrics: Volume 1: Fifth World Congress*. Cambridge University Press, 1994, p. 321.
- [109] S. Julier and J. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in *Proc. Symp. Aerospace/Defense Sensing, Simulation and Controls*, 1997.
- [110] D. Simon, *Optimal State Estimation: Kalman, H [infinity] and Nonlinear Approaches*. John Wiley and Sons, 2006, p. 526.
- [111] B. Stenger, P. R. S. Mendonca, and R. Cipolla, "Model-based Hand Tracking using an Unscented Kalman Filter," in *British Machine Vision Conf.*, 2001, pp. 63–72.
- [112] S. Dambreville, Y. Rathi, and A. Tannenbaum, "Tracking deformable objects with unscented Kalman filtering and geometric active contours," in *2006 American Control Conference*, 2006, p. 6 pp.

- [113] M. C. Vandyke, J. L. Schwartz, and C. D. Hall, “Unscented Kalman Filtering for Spacecraft Attitude State and Parameter Estimation,” in *Proc. AAS/AIAA Space Flight Mechanics Conf.*, 2004.
- [114] S. Gannot and M. Moonen, “On the Application of the Unscented Kalman Filter to Speech Processing,” in *Proc. Int. Workshop Acoustic Echo Noise Control*, 2003, p. 27.
- [115] D. Labarre, E. Grivel, M. Najim, and N. Christov, “Dual H_{∞} Algorithms for Signal Processing— Application to Speech Enhancement,” *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5195–5208, Nov. 2007.
- [116] I. A. Essa and A. P. Pentland, “Coding, analysis, interpretation, and recognition of facial expressions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757–763, Jul. 1997.
- [117] O. Lartillot, P. Toiviainen, and T. Eerola, “MIRtoolbox.” [Online]. Available: <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>. [Accessed: 12-Jul-2012].
- [118] O. Lartillot and P. Toiviainen, “A Matlab Toolbox for Musical Feature Extraction from Audio,” in *Proc. Int. Conf. Digital Audio Effects*, 2007, pp. 1–8.
- [119] O. Lartillot, *MIRtoolbox 1.4 - User’s Manual*. 2012, pp. 164–170.
- [120] C. Harte, M. Sandler, and M. Gasser, “Detecting Harmonic Change in Musical Audio,” in *Proc. ACM workshop on Audio and music computing multimedia*, 2006, p. 21.

Appendix A – Physiologically-Triggered Bookmarking Experiment

Materials

A1 Participant Consent Form



THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science

2366 Main Mall

Vancouver, B.C. Canada V6T 1Z4

tel: (604) 822-3061

fax: (604) 822-4231

(RESEARCHER'S COPY CONSENT FORM)

Project Title: PHLo: Physiological Haptic Loop

(UBC Ethics #B01-0470)

Principal Investigator: Associate Professor K. MacLean

Co-investigators: Prof. E. Croft, Assoc. Prof. J. McGrenere

The purpose of this study is to examine the role of that some of your physiological responses can play in controlling a computer application's behavior; and of the ability of haptic (touch sense) feedback to reveal to you how the application's behavior is changing as a result. You will be asked to wear external (i.e., non-invasive) sensors that collect some basic physiological information such as the heart rate, respiration rate, some muscle activity, and perspiration. We will also display simple haptic signals to you, e.g. vibrations on your hand or arm. Please tell the experimenter if you find the sensors or haptic display uncomfortable, and adjustments will be made. For some parts of the experiment, you will be asked to listen to a taped lecture or audiobook. You will be asked to answer questions in a questionnaire and an interview as part of the experiment. This session will be videotaped. The contents of these videotapes will be used for analysis purposes. No parts of the videotapes will be publically presented without your further consent. You have the option not to be videotaped. No compensation will be provided for participation in this experiment.

If you are unsure about any instructions, do not hesitate to ask.

TIME COMMITMENT: ½ -1 hour session

CONFIDENTIALITY: *Your results will be confidential: you will not be identified by name in any study reports. Test results will be stored in a secure Computer Science account accessible only to the experimenters.*

You understand that the experimenters will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you

have received a copy of this consent form for your own records, and consent to participate in this study.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

You hereby CONSENT to participate in this study and acknowledge RECEIPT of a copy of the consent form:

NAME _____ DATE _____
(please print)

SIGNATURE _____

A2 Pre-Experiment Questionnaire

This questionnaire was provided to participants to fill out prior to running Experiments 1-3.

**Please fill out the questionnaire below regarding the experiment.
The comment sections are optional.**

- 1) What is your age? _____
- 2) What is your gender? _____
- 3) What is your dominant hand (circle one)? **Right / Left / Ambidextrous (Both Hands)**
- 4) Have you been diagnosed with any attention disorders or equivalent (e.g., Obsessive compulsive or attention deficient disorders)? **Yes / No**
- 5) If you answered yes to question 4, which disorder? Are you currently taking medication for this? _____
- 6) How often do you listen to audiobooks or podcasts (circle one)?
Almost Everyday / Once Every 2-3 Days / Once a Week / Once a Month or Less / Never
- 7) How hard do you think it is to concentrate on audio books?

Not at all		Sometimes		Very Hard
1	2	3	4	5
- 8) How frequently do you feel distracted by disruptions in the external environment?

Not at all		Sometimes		Very Hard
1	2	3	4	5
- 9) Are you comfortable with an experiment patting or tapping your shoulders, upper arms or upper back (circle one)? **Yes / No**
- 10) Are you comfortable with having non-invasive contact sensors being placed on your fingers (circle one)? **Yes / No**
- 11) Do you have any experiment with wearing sensors on your body before? **Yes / No**

A3 Experiment 2 and 3 Experimenter Instructions

Set-up Procedure:

Set up video behind where the human subjects will be sitting.

Set up laptop with the customized MP3 player (with the Free Chapter 1+2 mp3 file open) and speaker on.

- to open the playlist of MPC: ctrl+7

Experimenters will sit behind the camera and the subject.

Instructions to the Subject:

0.0 Put the GSR sensor on the subject's non-dominant hand, and tell the person to not play around with it.

0.1 Ask the person whether the subject is familiar with the kind of mp3 player, and have him/her get used to the player if needed. Play an mp3 file other than Free for this optional session.

0.2 Instruct the person that he/she will be told when to start playing the mp3 player

0.3 Instruct the person to pay close attention to the audiobook and aim to understand it fully.

0.4 Tell the person that we will be asking him/her some content related questions.

0.5 Instruct the person to NEVER pause or stop the audiobook, but to always use rewind button if they lost track of content etc. (or with the customized mp3 player, we the person won't be able to do this)

0.6 Tell the person that he/she will be asked to review the video with the experimenters after the session to go over some of the highlights.

Experiment Procedure:

1. Start videotaping, have a sheet of paper to signal start of global timer on the experimenter computer.

2. Start temporary timer and wait 1 minute.

3. Bookmark (BM1) Ask the subject to start playing the mp3 file. Note the global time.

4. Throughout the experiment always jot down the global timestamp whenever the person rewinds the MP3.

5. Let the subject listen to the audio book for 2 minutes as baseline - just for the person to be engaged in the content of the book.

4. Ask the person the following questions:

4.1 (BM2) Short conversation "What is your name again?"

*** If the subject pauses, then instruct the person not to pause, let him/her listen wherever again and start over.

*** Back up question for when this happens: "Do you have any classes today?"

4.1.1 Give about 1 minute after the person starts listening to the book again.

4.2 (BM3) Medium conversation "Hi (subject name), What are you up to this weekend?"

4.2.1 Give about 1 minute after the person starts listening to the book again.

4.3 (BM4) Short conversation "What time does the laptop say?"

4.3.1 Give about 1 minute after the person starts listening to the book again.

4.4 (BM5) Long conversation "(subject name), where are you from?" "What program are you in?" "How old are you?" "Do you like XX?" "What projects are you working on?"

4.4.1 Give about 1 minute after the person starts listening to the book again.

4.5 (BM6) Long conversation "(subject name), how is giving out something for free related to Jello?" "What was the problem of selling Jello?"

4.5.1 Give about 1 minute after the person starts listening to the book again.

4.6 (BM7) Medium conversation "Is there a particular part that you liked more than others?"

4.6.1 Give about 1 minute after the person starts listening to the book again.

5. Stop recording, and ask the subject to stop the player.

6. Stop recording the GSR.

7. Using the global timestamps of mp3 rewind actions observed throughout the experiment, replay the video at these timestamps and ask the following questions:

7.0 Did you find the content interesting?

7.1 Why did you have to scroll back?

7.2 Did you lose the content of the audiobook because of an interruption (or something else)?

7.3 Why did you scroll back to the particular place?

7.4 Were you mostly on par in re-starting the mp3 audio to the last point where you were concentrating on the content? If not, describe why.

A4 Post-Experiment Questionnaire and Semi-structured Interview Questions

Overall about the experiment:

1. Overall, I felt relaxed during the experiment.

Strongly Disagree Disagree Neutral Agree Strongly Agree

2. Overall, I was definitely interrupted during the experiment.

Strongly Disagree Disagree Neutral Agree Strongly Agree

3. Overall, the interruption(s) was annoying.

Strongly Disagree Disagree Neutral Agree Strongly Agree

4. Overall, it was hard to be relaxed right after the interruption.

Strongly Disagree Disagree Neutral Agree Strongly Agree

5. Overall, I was mind wandering intensively and not concentrating on listening to the audiobook.

Strongly Disagree Disagree Neutral Agree Strongly Agree

6. Overall, the audiobook was engaging.

Strongly Disagree Disagree Neutral Agree Strongly Agree

7. How uncomfortable the sensors were in this experiment?

Not at All Somewhat Uncomfortable Very Uncomfortable

Comments: _____

Interruption: Knocking

8. I was interrupted while the interruption occurred. **Yes / No** Comment: _____

9. It was hard to refocus on the audiobook right after the interruption.

Strongly Disagree Disagree Neutral Agree Strongly Agree

10. If given the opportunity, I would have liked to have been able to rewind to place in the audiobook where the interruption had occurred.

Strongly Disagree Disagree Neutral Agree Strongly Agree

11. I tend to move my body (or either legs, head, arms, fingers, etc) right after this type of interruption. **Yes / No / Can't Remember**

Comments: _____

Interruption: Tapping on the shoulder

12. I was interrupted while the interruption occurred. **Yes / No** Comment: _____

13. It was hard to refocus on the audiobook right after the interruption.

Strongly Disagree Disagree Neutral Agree Strongly Agree

14. If given the opportunity, I would have liked to have been able to rewind to place in the audiobook where the interruption had occurred.

Strongly Disagree Disagree Neutral Agree Strongly Agree

15. I tend to move my body (or either legs, head, arms, fingers, etc) right after this type of interruption. **Yes / No / Can't Remember**

Comments: _____

Interruption: Oral Instruction from the Experimenter

16. I was interrupted while the interruption occurred. **Yes / No** Comment: _____

17. It was hard to refocus on the audiobook right after the interruption.

Strongly Disagree Disagree Neutral Agree Strongly Agree

18. If given the opportunity, I would have liked to have been able to rewind to place in the audiobook where the interruption had occurred.

Strongly Disagree Disagree Neutral Agree Strongly Agree

19. I tend to move my body (or either legs, head, arms, fingers, etc) right after this type of interruption. **Yes / No / Can't Remember**

Comments: _____

Interruption: Cell Phone Ringing

8. I was interrupted while the interruption occurred. **Yes / No** Comment: _____

9. It was hard to refocus on the audiobook right after the interruption.

Strongly Disagree Disagree Neutral Agree Strongly Agree

10. If given the opportunity, I would have liked to have been able to rewind to place in the audiobook where the interruption had occurred.

Strongly Disagree Disagree Neutral Agree Strongly Agree

11. I tend to move my body (or either legs, head, arms, fingers, etc) right after this type of interruption. **Yes / No / Can't Remember**

Comments: _____

Semi Structured Interview Questions

1. How realistic/relevant do you think these interruptions are compared to those in your daily life?
2. In your daily life how are you being interrupted most? How is it different than listening to audiobook?
3. How did you shift your focus back to the audiobook? What was the hard thing about it?
4. Would you find it useful if the system bookmarked the time where you were interrupted when you listen to an audiobook? Why?

Appendix B - Haptic Feedback during Bookmarking Experiment

Materials

B1 Participant Consent Form



THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science

2366 Main Mall

Vancouver, B.C. Canada V6T 1Z4

tel: (604) 822-3061

fax: (604) 822-4231

(RESEARCHER'S COPY CONSENT FORM)

Project Title: Physiologically-Based Bookmarking

(UBC Ethics #B01-0470)

Principal Investigator: Associate Professor K. MacLean

Co-investigators: Prof. E. Croft, Assoc. Prof. J. McGrenere

Student Investigators: Matthew Pan

The purpose of this study is to examine the role that some of your physiological responses can play in controlling a computer application's behavior; and of the ability of haptic (touch sense) feedback to reveal to you how the application's behavior is changing as a result. You will be asked to wear external (i.e., non-invasive) sensors that collect some basic physiological information such as the heart rate, respiration rate, some muscle activity, and perspiration. We will also display simple haptic signals to you, e.g. vibrations on your hand or arm. Please tell the experimenter if you find the sensors or haptic display uncomfortable, and adjustments will be made. For some parts of the experiment, you will be asked to listen to a taped lecture or audiobook. You will be asked to answer questions in a questionnaire and an interview as part of the experiment. This session will be videotaped. The contents of these videotapes will be used for analysis purposes. No parts of the videotapes will be publically presented without your further consent. You have the option not to be videotaped. No compensation will be provided for participation in this experiment.

If you are unsure about any instructions, do not hesitate to ask.

TIME COMMITMENT: 1½-2 hour session

CONFIDENTIALITY: *Your results will be confidential: you will not be identified by name in any study reports. Test results will be stored in a secure Mechanical Engineering account accessible only to the experimenters.*

You understand that the experimenters will ANSWER ANY QUESTIONS you have about the instructions or the procedures of this study. After participating, the experimenter will answer any other questions you have about this study.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in this study.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

You hereby CONSENT to participate in this study and acknowledge RECEIPT of a copy of the consent form:

NAME _____

DATE _____

(please print)

SIGNATURE _____

B2 Participant Instructions

Thank you for participating in this study.

In this experiment, we are looking at trying out a new device interaction for audiobook and podcast listening.

We have developed an algorithm which uses the galvanic skin response (a measurement of the electrical conductance of the skin) to detect when you've been interrupted. With this algorithm, we've developed a method of automatically placing 'bookmarks' in media streams at points where you've been interrupted. The main goal of this experiment is to determine the best way to present these bookmarks to the user.

Instructions

During this experiment, you will be asked to listen to two podcasts. Please pay very close attention to these podcasts as you will be tested on their content. At several points during the experiment, you will be interrupted by the experiment to answer some questions that he will have for you. Answering these questions will be your top priority, which means you will have to divert your attention away from the audiobook. **DO NOT STOP OR PAUSE THE AUDIOBOOK.** You'll have to let it keep on playing as you're speaking with the experimenter. Don't worry, you'll get to scroll back to listen to any content you might have missed right afterwards. To help you navigate through the podcast after attending to an interruption, we will be providing automatically generated bookmarks based on your Galvanic skin response. These bookmarks will be presented in two forms: Visually (indicated on the screen by a green line and a black diamond icon denoting where the bookmark is in the audiobook) and haptically (felt as a buzz through a wrist mounted vibro-tactile motor).

Four conditions will be tested in this experiment (order will be randomized):

- One where visual bookmarks are present
- One where haptic bookmarks are present
- One where both bookmark displays are present; and,
- One where no bookmarks are displayed

Please note: because the bookmarking algorithm utilizes a physiological signal, there are chances where false positive bookmarks may be generated before or after a correctly-placed bookmark.

Thus, in the case where you do not recognize content at one of the later bookmarks, try scrolling to an earlier bookmark and see if that's the place where you left off.

Visual Bookmark Display

The visual display, as you will see, is quite Spartan. The most important feature will be the scrollbar which indicates the position and duration of the currently playing audiobook. As mentioned earlier, lines and black diamond icons indicating bookmarks will be overlaid on this scrollbar to show you where bookmarks have been placed. When you reach a bookmark while navigating through the podcast after an interruption, the black icon will turn green.

Wrist-Watch Device

You will be asked to wear a wrist watch device which has three main components:

- A black button which can be double-tapped to play/pause the audiobook
 - The black button should NEVER be pressed unless the experimenter asks you to at the beginning and end of the experiment.
- A soft potentiometer used for navigating through the audiobook
 - Run your finger along the potentiometer clockwise to scroll forward in the podcast
 - Run your finger along the potentiometer counter-clockwise to scroll backwards in the podcast
 - For best performance, try to keep your finger within the white area and run your finger along the circular track in one smooth motion (as opposed to jerky fragmented motions)
- A vibrotactile device to let you know where a bookmark has been placed during navigation

Questions

As mentioned previously, answering any experimenter questions is top priority. The podcast will still be playing, which means you may have to remove your headset to answer questions. You will be told when you can start listening to the podcast, at which point you can re-don the headset and begin scrolling through the audiobook to acquire the place where you last left off. If you're uncomfortable answering any of the questions, mention this to the experimenter. You can even lie in your responses – we're just using conversation to ensure distraction away from the audiobook.

Notes

- You will be video recorded during this session.
- Sometimes the audiobook will skip and jump randomly. This is caused by noisy measurements from the potentiometer. Just continue listening if this occurs.
- Bookmarks are placed 10 seconds prior to an interruption in the podcast.
- You will get a break in between the two podcasts presented.
- Compare this experience to listening to audiobooks or podcasts over your own device (Mp3 player, iPhone, iPod etc.). Please offer the experimenter some feedback based on this comparison.

B3 Post-Experiment Questionnaire

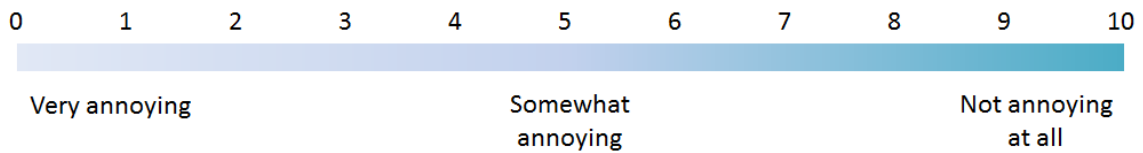
Podcast Content

Did you find the content engaging? Yes No

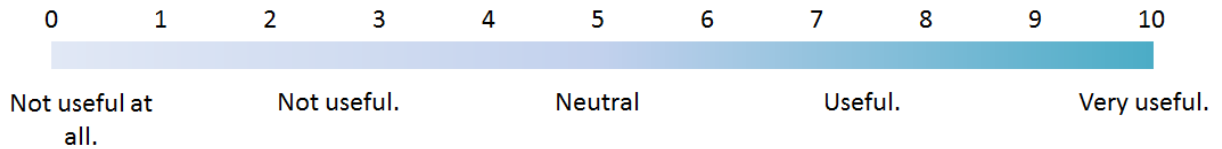
- 1) How many years has it been since Dave last been to the dentist?
 - A. 10 days
 - B. 10 weeks
 - C. 10 months
 - D. 10 years
 - 2) What is the name of the dentist?
 - A. Dr. O'Hagan
 - B. Dr. Tom
 - C. Dr. McDougall
 - D. Dr. Tong
 - E. Dr. Legg
 - 3) What is the name of the painkilling drug used to relieve the dentist's back?
 - A. Morphine
 - B. Xylocaine
 - C. Levorphanol
 - D. Wintergreen
 - E. Lidocaine
 - 4) How does the drug work?
 - A. It changes the electrolyte balance of the nerve
 - B. It changes the ability of the nerve to transmit pain
 - C. It is a general anesthetic changing inhibit excitatory functions of the nervous system
 - D. Both A and B.
 - E. None of the above.
 - 5) What type of school was Sam entering?
 - A. Preparatory School
 - B. Elementary School
 - C. Middle School
 - D. High School
 - 6) Which sports did Sam try out in prior to Field Hockey?
 - A. Soccer
 - B. Ice Hockey
 - C. Bowling
 - D. Curling
 - E. All of the above.
 - 7) Name as many members of Sam's field hockey team as possible.
-

Haptic Bookmark Display

How annoying are the vibrations used to notify you of **where** bookmarks were placed?



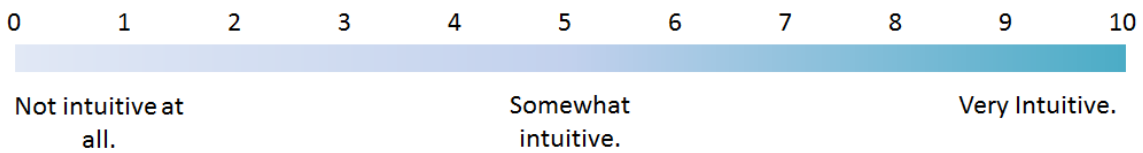
Judge the level of perceived usefulness of using vibration to notify **where** bookmarks have been placed.



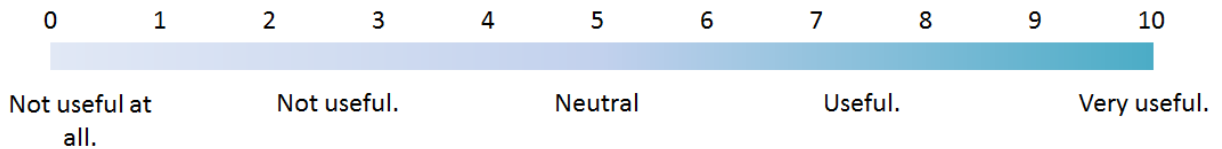
Do you have any comments on how the vibration used to notify **where** bookmarks have been placed can be improved or changed?

Circular Scrolling

Judge the intuitiveness of using the circular scroll device to navigate through the audio stream.



Judge the usefulness of using the circular scroll device to navigate through the audio stream.



Do you have any comments on how the scroll device that is used to notify navigate through the audio stream or how it could be improved or changed?

Overall System

Of all the bookmark display conditions presented to you, which condition do you prefer the most?

No Haptics No Visual	No Haptics With Visual	With Haptics No Visual	With Haptics With Visual
-------------------------	---------------------------	---------------------------	-----------------------------

Can you explain why?

Can you imagine yourself ever using an automatic bookmarking system with haptic notification and navigation for media streams? Why?

Additional Comments

Appendix C – Affect Estimation during Music Listening Experiment

Materials

C1 Participant's Consent Form



CONSENT FORM

Department of Computer Science
2366 Main Mall
Vancouver, B.C. Canada V6T 1Z4
tel: (604) 822-3061
fax: (604) 822-4231

Project Title: Physiological Responses to Music
(UBC Ethics # H10-00783)

Principal Investigators:

Prof. Karon MacLean, Department of Computer Science
Prof. Joanna McGrenere, Department of Computer Science
Prof. Elizabeth Croft, Department of Mechanical Engineering

Student Investigators:

Matthew Pan, Department of Mechanical Engineering

The purpose of this experiment is to examine a user's physiological and affective responses to music. Data collected in this experiment will be used to develop novel interaction techniques for portable audio players using haptic signals and human affect models.

In this experiment, you will be asked to provide ten (10) lyric-less songs that you enjoy listening to and examples of music genres you do not enjoy. A selection of musical clips will be played back in one-minute segments. During each segment, you will be asked to continuously rate how much you like or dislike the piece. A short survey will be conducted at the end of each song segment. You will be asked to wear external (i.e. non-invasive) sensors that collect some basic physiological information such as heart rate, respiration rate, some muscle activity, and perspiration. Please tell the experimenter if you find the sensors uncomfortable and adjustments will be made. Data will be collected by video and/or audio recordings, and by questionnaires and surveys. We will contact you for further permission before making any use of recordings of your participation in presentation of our results.

REIMBURSEMENT: NONE

TIME COMMITMENT: 2 Hours

CONFIDENTIALITY: *You will not be identified by name in any study reports. Data gathered in the sessions will be stored in a secure Computer Science account accessible only to the experimenters.*

You understand that the experimenter will ANSWER ANY QUESTIONS you have about the instructions or the procedures of these sessions. After participating, the experimenter will answer any questions you have about the sessions.

Your participation in these sessions is entirely voluntary and you may refuse to participate or withdraw at any time without jeopardy. Your signature below indicates that you have received a copy of this consent form for your own records, and consent to participate in these sessions.

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Info Line in the UBC Office of Research Services at 604-822-8598.

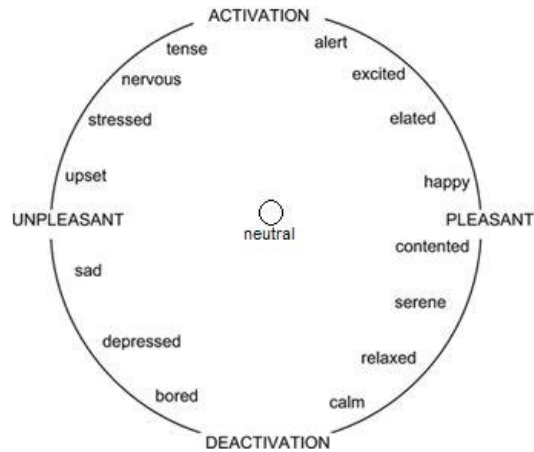
You hereby CONSENT to participate in this study and acknowledge RECEIPT of the consent form:

NAME _____

DATE _____

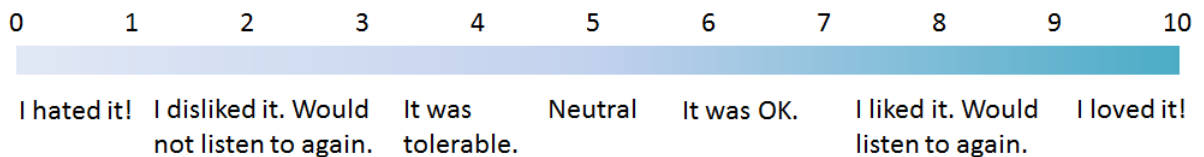
SIGNATURE _____

C2 Post-Trial Questionnaire



Did you recognize the song that was playing? **YES** **NO**

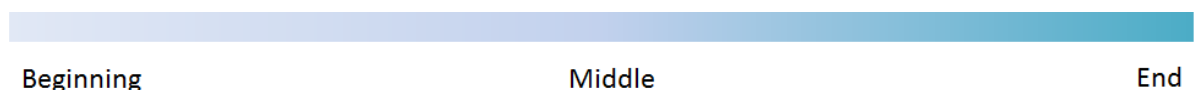
Overall, on a scale of 1 (not at all) to 10 (absolutely loved it), how much did you enjoy listening to the song?



Would you have changed the song if you had the choice? **YES** **NO**

Did you experience ‘chills’ while listening to the song? **YES** **NO**

If so, approximately at which point in the song did you experience the ‘chills’?



Do you feel you had emotional or cognitive associations with the song (e.g. every time you’re listening to this song, you remember a particularly memorable event)? **YES** **NO**

If so, could you describe this association (only if you’re comfortable doing so)?
