Applying Interruption Techniques from the HCI Literature to Portable Music Players

Amirhossein Mehrabian, Joanna McGrenere

May, 2011

Technical Report TR-2011-04

Department of Computer Science, University of British Columbia
2366 Main Mall, Vancouver, B.C., Canada, V6T 1Z4

Contents

1 Introduction

The advancement in computer hardware systems has made it possible to design small electronic devices with significant computational capacities. This led to the production of small electronic gadgets. These gadgets are prevalent in industrial and developing societies. One of the most common types of these electronic gadgets is the portable music player (PMP). There are two different scenarios for using these portable music players:

1. Listening to a portable music player is the primary task
2. Listening to a portable music player is a secondary task.

For illustrative purpose, a representative task for each scenario will be used in the rest of this section. Listening to an audio-book can be a representative task for the first scenario since it requires full attention of the user in order to understand the content of the book. It imposes a high cognitive load upon the user since he/she must listen to the voice and at the same time try to understand and rationalize the context. As a representative for the second scenario we choose listening to music while running. In this context the primary task is running and the user is not focusing on the music being played by the portable music player, although the music that is playing should meet user's expectations. These expectations could be matched by the current activity of the user. For example, when the user is running relatively fast then he/she may prefer to listen to fast-paced music and vice versa.

Since contemporary PMPs have computational power, they can utilize this extra power to provide a better experience for the user in both scenarios. Considering the computational power of PMPs, the design of PMPs as passive devices with simple play and stop buttons no longer is an acceptable solution. New designs might be developed for PMPs that are active devices capable of performing more complex tasks that allow interacting with the user, his/her preferences and so on. In the new design these PMPs would have access to more information about users than previous passive PMPs. The new design should make it possible to devise new interaction techniques. The new techniques should take into consideration the computational power and information about the user available on PMPs. They should achieve better results in term of user satisfaction.

In both scenarios the new design would be capable of more sophisticated methods of communicating information to the user than the previous passive design. These methods can be driven by several factors:

- The improved design may be result of some internal information about the device which needs user action, hence should be delivered to the user immediately. An example of such information could be a low battery signal that will allow the user enough time to properly react.
- The new design might also be based on the 'user affect'. This user affect could be decided with some degree of certainty by monitoring the user's biometric signals. As an

example the device may decide to change the current music track that is being played because the device believes the user doesn't like it.

- There may be times that a change in the environment may initiate the need for communication. Consider the first scenario where the user is listening to an audio-book and somebody starts talking to him/her. The PMP might immediately detect this change in activity in the environment, based on the reading of some signals (biometric or non-biometric) and perform an action. This action might be the creation a bookmark in the audio-book that the user was listening to, or simply pause the current track and start playing again as soon as the device senses that the conversation is finished. Another example for a second scenario might be when the user is jogging and then suddenly starts running. In this situation the device may decide to change the pace of the music being played on the device.

The communication that occurs in the above scenarios shares a lot of common problems with those problems that are discussed in the interruption literature. Although the interruption literature explores the problems in a desktop computing environment, they can also provide valuable insights into the problems in different environments. An overview of the interruption literature will help us to make an informed decision about those problems specific to PMPs. We can categorize the problems, which arise in the interruption context, into the following groups:

- When to interrupt:
    - Based upon consideration of the user's cognitive load.
    - Based upon external activities surrounding the user's environment.
    - Based upon the task structure in which the user is engaged.
- How to interrupt:
    - Considering the different means available to interrupt a user, in the context of PMPs we are most interested in non-visual means.
- Inputs to monitor for any decision making task:
    - Inputs that can be gathered from surrounding environment.
    - Inputs that can be gathered from user (e.g. biometric signals)

In designing an active PMP we also need to pay attention to other characteristics of such a device in order to make an informed decision. One of the characteristics of using PMPs is that the user usually prefers to not maintain eye contact with the device while using it. He/she tries to minimize the time he/she spends looking at the device and prefers to interact with the device without having to look at it. This calls for a better means of interruption. In this essay we will discuss the different means for interrupting users other than a visual interruption. The other characteristic of PMPs is that they are usually carried by the user so an overview of the ubiquitous computing literature might provide an insight.

This essay provides an overview of the literature in the HCI community related to a subset of problems that arise in the context of PMP-usage. The subset consists of those problems regarding the timing of interruptions, the inputs that can be used to inform the user about the earlier problems of timing these interruptions, and discovering a new means for interrupting users other than using the de-facto visual channels.

2 When to interrupt

A very important question in the interruption literature is when to interrupt a user. Choosing the appropriate time for interrupting a user is dependent on several factors. A portion of the research in this area is devoted to understanding how the human mind perceives and reacts to an interruption. The goal of our avenue of research is to gain a better understanding of the human mind. Results from this research would help provide an explanation and thus aid us in making predictions regarding interruptions. Another avenue of our research is devoted to the development of models for predicting the appropriate time to interrupt a user. These models take advantage of different information available to the system. This information could be the user's pattern of interaction with the system, or activities of the user within the system, or data from a sensor, such as a biological sensor. Recent investigations have considered the context of the user when using these devices. This trend is greatly influenced by the wide spread usage of mobile devices that can be used in many different environments and situations. We are going to provide examples of research in these areas in the rest of this section.

2.1 Task Resumption Models

Humans evolved to operate in environments with numerous interruptions. We have the innate ability to multi-task, that is, our minds are capable of working on several different tasks at the same time. We can keep track of a variety of tasks concurrently and be able to switch our attention among them. This ability requires us to suspend the task we currently are performing and then to resume that suspended task. In order to better understand how we can perform these tasks, we need to have a model of the manner in which our minds work. A number of different models have been developed that attempt to explain how the human mind performs such a task. Each model tries to provide a set of factors that are important in order to perform any tasks and then return to the task that was suspended upon interruption. One of the models that has received a lot of attention in interruption-research is the *goal-activation* model introduced in a paper by Altmann and Trafton in 2002. Their model focuses on memory and analyzes human behavior from this aspect. In their paper the authors discuss *goal* and define it in their paper as follows:

> "The term goal refers to a mental representation of an intention to accomplish a task, achieve some specific state of the world, or take some mental or physical action." (Altmann & Trafton, 2002)

In their model they show that our memory for storing goals is stored in the same place for storing events and facts. This is a contrary to what most researchers have believed from previous studies. Then Altmann and Trafton define three factors that have a direct influence over selection of a goal and its influence over that person's behavior. They use the *ACT-R* model which states that any goal, to be retrieved from memory, must be the most active trace in memory and become the center of a person's awareness. Once this trace is selected and retrieved from memory it can then drive a person's behavior. The authors found that three factors control the activity level of a goal in memory: the interference level, the strengthening constraint, and the priming constraint.

Unlike the memories we use in our computers, old items in one's human memory decay rather slowly. This slow decay causes a cluttered situation in memory where these old items make it difficult for us to correctly retrieve them from memory. Put another way this faulty retrieval creates an 'interference level'. According to Altmann and Trafton, if a memory is to be successfully retrieved from memory and alter a person's behavior, then the goal should first have an activation level higher than interference level. Figure 1 shows the time course of a newly activated goal in memory. Observe that, as time passes, the activation level of the goal decreases. The graph is defined by following equation:

$$m = \ln\left(\frac{n}{\sqrt{T}}\right)$$

In this equation m is the activation level of the goal in the memory and n is the number of the times that this goal has been accessed in memory. T denotes the time that the goal has been in memory. The more the goal is accessed in memory the higher its activation level will be. It also shows why the rate at which these goals decay in memory is slow.

When we want a goal to drive the behavior of a person we have to bring the activation level of that goal above the interference level. This will cause the goal to have a higher chance of being retrieved from memory and thus influence the behavior of the person. Although it seems straightforward to raise the activation level of a goal in order to pass the interference level, we have to be careful to not boost or reinforce this too much. We want to raise the activation level of our goal high enough to pass the interference level while keeping in mind that once we are done with the task we don't want it to provide a lot of interference for our next goal. A 'priming-constraint' specifies that once we restart a suspended task and bring its activation level above the interference level, we can take advantage of the cues related to the goal that we want to raise. These cues could be in relation to the task-environment or history of performing the task in the user's mind. The cues help a user to find the answer to the question "Now, what am I doing?". This implies that cues should be available in the environment before the interruption, in order to help a user find the answer to the question.

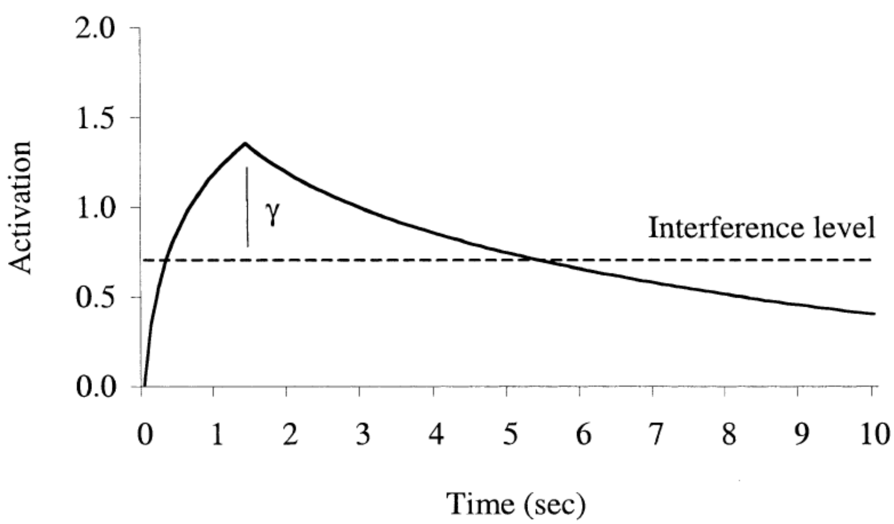

Figure 1 Time course of activation of a new goal (solid line) and the interference level due to old goals (dashed line). (Altmann & Trafton, 2002)

The goal-activation model was followed by experiments done by researchers to see whether empirical results derived from user studies agreed with the predictions of the model. We will provide examples of work done in this area and discuss their results in the rest of this section.

One of the direct results induced from the goal-activation model is that the length of the interruption would have an effect on resumption of a suspended goal. In the goal-activation model, the activation level of a goal will decrease as time passes if the goal is not accessed in the memory. This reduction in activation will limit one's ability to retrieve the goal from memory and thus reduce the possibility of eventually directing human behavior. When a user is interrupted, he/she needs to suspend the goal related to a primary task and begin work on the goal related to the interruption task. While he/she is working on the interruption, the activation level for the goal related to primary task will decrease, since it is accessed less often in memory. Once the interruption task is finished it will be harder to resume the suspended task due the reduction in activation level. Another important factor that should be considered about an interruption task is the complexity of it. Complexity of a task indicates how much memory resources should be allocated in order to process and perform a task. The more resource allocated to the task the less the chance the user has to rehearse the primary task while he/she is attending to the interruption. Rehearsal of the primary goal helps to keep its activation level high. Users in some interruption scenarios have the chance to rehearse the primary goal while working on the interruption task. As an example consider a user working on a long document when he/she receives an electronic instant message. He/she will suspend working on the document and start chatting with the other person using instant messaging software. In this scenario the user has the chance to have a look at the document he/she is writing while he/she is waiting for his/her friend to type a reply message. This scenario provides the user with the chance to rehearse his/her primary goal while he/she is working on an interruption task. It is predicted that when a user has a chance to rehearse a primary goal it's easier to resume a suspended task.

Monk, Trafton, and Boehm-Davis performed a series of user studies to better understand the effect of interruption-duration and demand on the resumption of suspended goals (Monk, Trafton, & Boehm-Davis, 2008). First the authors wanted to show why previous work had failed to successfully observe the relationship between interruption-duration and resumption of suspended task. The authors claim that previous work has used measures that are not sensitive to the effect of interruption-duration on task-resumption. The authors used resumption-lag as a measure to study the effect of interruption-duration on suspended-task-resumption. Resumption-lag is defined as the amount of time that is needed for a user to resume work on a primary task after interruption. The authors also pointed out that, according to the activation level equation, the rate at which activation decreases is relatively much faster in its initial stage of decay than in later stages. In other words the decay-rate of the activation level slows over time. They claim that in previous work the duration of the interrupted task was such that the activation level of the primary goal in memory was at a stage that it was decaying at a very slow rate. And so they failed to detect an impact of the interruption length on performance. Because of this observation, they adjusted the interruption length so that the activation level of the primary goal in memory is at a stage with a faster decay rate. As a result they could observe the effect of interruption length on performance.

In their experiment they chose VCR-programming as the primary task and tracking as the secondary task (Monk et al., 2008). The primary task was interrupted for intervals of 3, 8 and 13 seconds. The results obtained from the experiment agreed with their goal-activation model. The authors also discovered that resumption-lag was relatively longer for long interruptions and relatively shorter for short interruptions. In a second experiment they extended the interval of interruption to one minute (by investigating interruptions of 28, 32, and 56 seconds in addition to the three in their first experiment) to show that the resumption-lag follows a logarithmic function. Their results from this second study showed that short interruptions (less than 25 seconds) are less disruptive than longer interruptions (Monk et al., 2008).

In their third experiment Monk et al. tried to study the effect of interruption demand on resumption-time (Monk et al., 2008). Based on a goal-activation model they expected to see longer resumption-lags for interruptions with a greater demand on memory resources. This greater demand on memory resources will leave fewer resources for the user to apply toward a primary goal that he/she is working on during the interruption task. Thus this greater demand on memory will cause a relatively faster decay of the activation-level of the primary goal in memory. In their experiment Monk et al. used three different types of interruptions, each with a different level of demand (none, medium and high). They also used three different interruption-durations, similar to their first experiment. Their experimental results demonstrated that those interruptions that placed a relatively higher demand on memory caused a longer resumption-lag for the primary task. They also showed that in the situation of a no-demand-interruption, the decay, due simply to the passage of time, remained dominant.

The results from the Monk et al. study (Monk et al., 2008) showed the effect of rehearsal in reducing the degree of resumption-lag. In another study, Trafton et al. studied the effect of warning users about incoming interruptions (Trafton, Altmann, Brock, & Mintz, 2003). In order to better understand the concept consider this scenario: you are working in your office and you know that you are about to be interrupted by a phone call. When the phone rings, it informs you of incoming interruption (the actual conversion over the phone). You then have some time to make a decision whether to pick-up the phone or not. You may choose to finish the task you were working on (such as composing a sentence) and then pick-up the phone, after the primary task is completed—or you can think about and perhaps plan what you want to do after the phone call is done. There are countless situations similar to this example where one has prior knowledge of an imminent interruption. Trafton et al. call the time interval between the onset of the warning of the interrupt and the actual interruption the *interruption-lag* (Trafton et al., 2003). In their study they asked the participants to use a protocol of talking-aloud in order to verbalize their thoughts before the interruption and also during performing the interruption-task. Trafton et al. used this talking-aloud protocol to understand more about the thought process of the users during interruption-lag and the interruption itself. This data helped the researchers to understand whether participants use this time to rehearse the primary goal. The primary task involved planning a route, with a variety of constraints; the secondary task involved object-classification. Trafton et al. predicted that participants would use this time in order to choose their next sub-goal or next step for this current sub-goal. They divided participants into two groups. One group received a warning 600ms before the interruption and the other group received the interruption immediately. The warning was chosen carefully to not engage the participants in any way that might distract them from rehearsing. After studying the resumption-lag and the verbal thinking of the participants, they showed that the participants used the interruption-lag time in the predicated manner and it helped the participants to reduce the resumption-lag.

These results encourage us (in designing a PMP) to train users to utilize the interruption-lag for rehearsing goals and preparing for resumption of the primary task. We can also observe people in scenarios with an MP3 player in order to learn about any strategies that they choose to speed up the task resumption process. For example, in such scenarios, when somebody approaches a user to start a conversation, it could be interpreted as presenting the user with a warning about an imminent interruption. Studying user behaviors in this situation will guide us to develop more efficient resumption strategies and also adapt our design to facilitate such behaviors.

While previous studies were concerned about the effect of interruptions with different duration and demands, there are some studies that observed the effect of frequency of interruption on the primary task. Monk designed an experiment with two different interruption frequencies (Monk, 2004). In both experiments the participants were asked to program a VCR but, in the first experiment, they were interrupted every 10 seconds and in the second experiment every 30 seconds. The interruption-duration was the same for both experiments. He used resumption-lag to measure the effect of the interruptions on the primary task. The results however were contrary to the expectation of the author. They showed that the resumption-lag was shorter for relatively more frequent interruption scenarios and the user finished the primary task in slightly less time than the scenarios with infrequent interruptions. One explanation that Monk provides for these unexpected results is that the participants had developed a more active method for rehearsing goals in the scenario where interruptions were more frequent. Previous studies have shown that there is a meaningful relationship between the rehearsing of the goal and the reduction in resumption-lag (Monk et al., 2008).

The introduction of the goal-activation model (and conducting of associated experiments in order to evaluate the different effects of interruption on resumption of the primary task based on this model) helps us to better understand and analyze human interruption. It guides us with insights into designing systems with sufficient cues that help the user to easily resume interrupted tasks. In the MP3-player scenarios we can take advantage of this knowledge and study the best possible ways to find or create cues for the user.

Recently a new model for task resumption is introduced by Salvucci (Salvucci, 2010) . Salvucci believes that all previous work in this area were focusing on the resumption task as a purely memory based process; he thinks these models do not succeed in explaining applied experimental results because they have reduced the task-resumption problem to a simple memory retrieval task. The author mentions two reasons for this shortcoming. First, resumption-lag in applied studies is longer than what it would have been in a purely memory-based process. Second, the context of the applied-task is usually more complex which make the memory-based approaches implausible. For example in an experiment, when the context of a task might only require remembering two or three items in memory, an applied domain-task (such as paper-writing or programming) certainly couldn't be reduced to a single memory retrieval.

Salvucci introduces a new framework for task-resumption that considers task-resumption as a reconstruction process. This model relies on ACT-R as a computational, cognitive architecture for describing and predicting behaviors. The ACT-R model introduces the *problem-state* as a mental resource to account for a task context. A *problem-state* is temporal information that is associated with the current task and required in order to perform the task. For example in a computation task like 4 (5+3), the problem-state could be the number 8 which is the result of addition of the numbers 5 and 3. A more complex task like writing a paper would have a more complex problem-state. Due to scarceness of mental resources a person could only maintain this problem-state for one task at a time. This fact implies that whenever a user is interrupted, he/she must store the current problem state in memory and replace it with the new problem-state. Upon resumption he/she must retrieve the problem state from memory. Now consider the situation in which the user fails to recall the problem-state. According to ACT-R model, the memory decay will happen as time goes by and it will make recalling the problem-state very hard or even impossible. It is a very common situation in applied domains when we have interruptions with very long durations. It is also possible that the failure to retrieve a problem-state is caused by the complexity of the problem-state itself. In both situations, failure to recall the problem-state will lead to reconstruction of the problem-state. According to the ACT-R model it will take 250ms to create a new problem-state. Once a new problem-state is created, the user needs to fill it with the information required to replicate the lost problem-state. This process is very dependent on the task domain. Each domain may involve different steps in gathering information and also filling the problem-state with relevant information. Since those interruption scenarios that we are interested in are more likely to cause the user to forget the problem-state, this framework should help us to design a better system for listening to an audio book using MP3 player. We need to study this task and identify the steps needed to reconstruct a problem-state. We could then model this information using the ACT-R technique. The constructed computational model for the audio-book listening task will help us to better understand the process and make predictions about a user's behavior.

2.2 Task structure

Task-resumption models help us to have a better understanding of the effect of interruption on users. They enable us to make predictions about a user's behavior at the time of interruption. They also help us to design better interruption signals in order to reduce the interruption-cost for the user. However we still need models that enable us to decide about the time of interruption. Research in psychology and social behavioral has shown that the effect of interruption on the user and the existing mental load have a linear relationship. The relationship suggests that interrupting a user during moments of a lower mental load should decrease the interruption-cost. An explanation could be that user's attention is a limited resource and, at the time of increased mental load, there is insufficient attention to be allocated to an interruption, while during moments of decreased mental load there is a relatively greater resource available to be allocated to the interruption task.

A common approach for measuring the mental load is to monitor some biological signals like Electrocardiograph (ECG) and Electroencephalography (EEG). In Section 2.3.3 the reader can find examples of how EEGs and ECGs can be useful for predicting the appropriate moment to interrupt a user. Monitoring these signals usually is done using special hardware that requires probes to be placed on specific places on a user's body. Due to these technical problems it is impractical to use such signals to monitor and predict a user's work-load and hence find the appropriate moment for an interruption.

Given the challenges of using physiological signals to determine the mental load of users, researchers have been motivated to devise new methods for predicting a user's mental load and his/her 'interruptibility'. One line of research has tried to find a connection between task structure and the changes in mental load of the user. In this approach researchers try to define a structure for the task that is going to be performed by a user and then find the appropriate moments for interrupting the user based on the task structure. Adamczyk and Baily designed an experiment in order to investigate effect of interruption at different moments in task execution (Adamczyk & Bailey, 2004). In their experiment they used three tasks: editing a document, watching a video clip and writing a summary of the clip, and finding some information on the web and saving it in a document. They provided video clips of instances of the three tasks to 25 subjects in order to find break points in executing these tasks. As an example a sample of coarse and fine break points for checking email was provide to the subjects in order to help them differentiate between coarse and fine break points. Figure 2 shows the sample of coarse and fine break points in email checking provided to the subjects. The researchers believed that an interruption at a coarse break point in task execution would have less interruption-cost than fine break points.

Coarse-unit Descriptions	**Fine-unit Descriptions**
Moves to Start Menu	Moves to Start Menu Moves to Apps folder
Selects email application	Selects email application Types in username Types in password
Opens email	Selects email function Goes to message index Moves through messages
Opens message	Opens particular email Scrolls through message Selects text
Copies text from message	Copies text Exits email function
Closes application	Logs off Closes application

Figure 2 Coarse and Fine Break point in email checking task (Adamczyk & Bailey, 2004)

In their experiment the interruption task was to select a title for a news wire among provided options. They interrupted the user at the presumed best and worst moment for an interruption and also interrupted the user at random. The triggers for interrupting users at the presumed best and worst moment for each task in their experiment are shown in Figure 3.

Task	Interruption	Interruption Trigger
Edit	Presumed Best	Upon completing an edit – spelling mistake or comment
	Presumed Worst	During an edit – typing, selecting, or confirming
Media	Presumed Best	After completing summary, but before the save process
	Presumed Worst	During the viewing of the Video clip
Search	Presumed Best	After copying the citation, but before the save process
	Presumed Worst	During the search process, on the publications page

Figure 3 Interruption trigger by task (Adamczyk & Bailey, 2004)

The researchers used resumption lag and a modified version of The NASA-TLX survey in order to measure the effect of an interruption at different moments of task execution. The NASA-TLX is a subjective assessment tool for measuring user's mental load ("NASA TLX Homepage," n.d.). The author chose this tool because it had a short length and a continuous scale. Results from the study showed that interruption at the presumed best-timed interruption caused less annoyance and frustration for the subject.

Results from Adamczyk's (Adamczyk & Bailey, 2004) work encouraged other researchers to investigate the task structure and further explore its connection to mental work load. In 2008 another study conducted by Bailey and Iqbal explored the relationship between task-structure and the user's mental load. The goal of their study was to show that mental workload changes during execution of a task. Based on this fact they wanted to find moments in a task structure during which a user would have a reduced mental workload and during these moments could allocate greater attention to the interruption task. They used pupil size as a measure of mental workload. It has been shown that pupil dilation is a valid indicator of mental workload under controlled conditions (Bailey & Iqbal, 2008). Controlled conditions include the control of illumination and any other factors which might result in a change of pupil size, for example, the brightness of a computer screen. Bailey makes use of the fact that an increase in pupil size correlates with an increase in mental workload. This observation holds true for different tasks and different individuals.

In order to study the changes in mental workload Bailey and Iqbal defined three tasks for users: route planning, document editing and email classification. In the route planning task users had to work with an interactive map in order to find the best possible route between two cities on the map. There were two possible routes between cities and each route consists of three segments.

For each segment the information about the distance and the fare associated with that segment was available to the user. The user had to calculate the distance and the fare along each route by entering information about each segment into a table and performing a mental calculation. The difficulty of the task was controlled by manipulation of the fare and distance values.

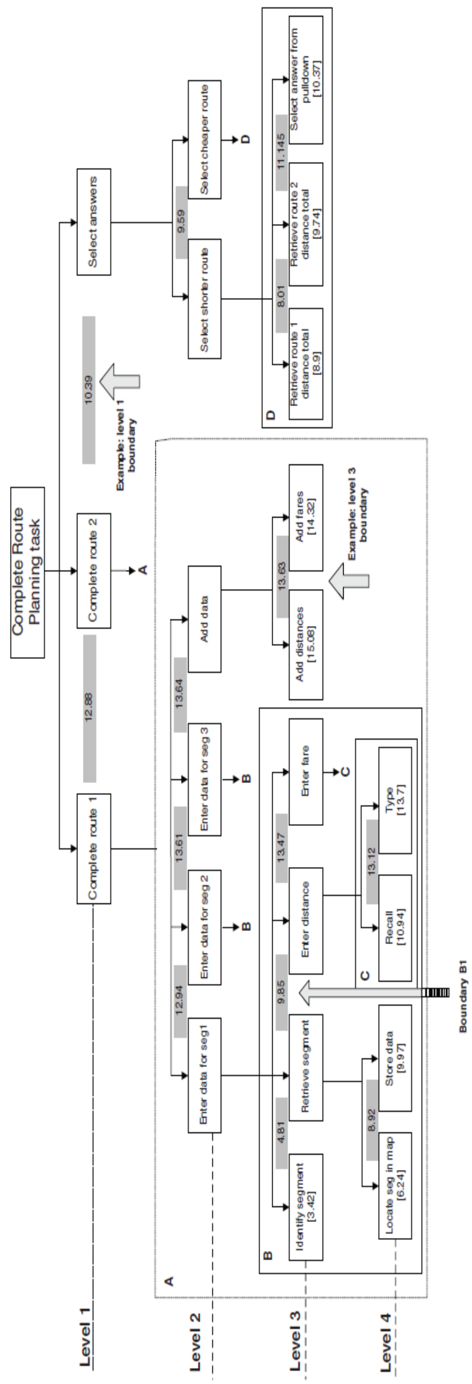

Figure 4 Task Model for route planning. The interior nodes represent user goal nodes and the leaf nodes represent operators. Time moves from left to right of the diagram. (Bailey & Iqbal, 2008)

Bailey and Iqbal defined a task-model for each of the three tasks involved in the experiment using an iterative process; with each iteration they refined each task-model with the information extracted from video clips of users performing a particular task. Figure 4 shows a task-model for a route planning task at different levels. Any node in the model represents a subtask and subtask boundary refers to the period between adjacent subtasks. The level of a boundary between two adjacent subtasks is the depth of their shared ancestor plus one in the model. For example if a user finished the “locate seg in map” task and moved on to the “Store data” task then it is a level four boundary, since their shared ancestor, “Retrieve segment” is at level three.

In the next step Bailey and Iqbal aligned the data about the changes in pupil size of the users with the task model. The result of the combination of the pupil size data and the task model reflect the correlation between pupil dilation and changes in metal workload during task-execution. It showed that a user’s allocation of his/her attention resources was dynamic during task-execution and these resources were allocated and freed at different moments during the task-execution. Their results also indicated that a users’ mental workload decreased in the transition between subtask boundaries and further their results also showed that the changes in mental load were not significant in transitions between subtasks at lower levels (4, 3) but the decrease in mental load is more significant at higher levels (1, 2).

One must always keep in mind that the result of these studies was based on a specified task structure and alternate task structures may produce different results. It is important to use the results from this study as a general guideline only in designing systems which require interrupting a user.

Previous studies have relied on a task-model that suggested that an interruption should occur at coarse break points during the task execution since these create less disruption for the user. Later research corroborated these earlier claims with data indicating the mental workload of the user is linked to pupil dilation. The data agreed with the previous results and provided general guidelines for the design of similar systems. In a study conducted by Iqbal and Bailey, they constructed a statistical model to predict the cost of an interruption based on a selected feature from their task-model (Iqbal & Bailey, 2006). They used resumption-lag as an indicator for the cost of interruption. The authors defined three primary tasks for their experiment: video editing, route planning and document editing. They defined these task-models for the primary task in order to extract features for predicting the cost of interruption. Task-models are similar to what is depicted in Figure 4. Then, after they had selected a set of subtask boundaries for interrupting the user, Iqbal & Bailey performed a *Wizard of Oz* study to collect data about the cost of interruption at different subtask boundaries. Next they installed a series of filters in order to find the most predictive feature of their task model. Subsequently the authors selected the following features for constructing their final statistical model that predicted the cost of interruption in three categories: the level of subtask boundary, the carry-over of data at subtask boundaries, and the difficulty of the next subtask. Their final model used these three features in order to detect among the categories the cost of interruption: high, medium and low. The result of their study showed that information about the task structure could be used in a statistical model in order to improve the prediction about the cost of interruption.

After reviewing the research on task structures and the effect of interruption during different moments of task-execution, one problem seems to be predominant in all the research: we always need to have a reliable model of the task at hand in order to apply research results. Developing a reliable task-model could be very difficult, and sometimes impossible, for some categories of task, for example free-form tasks (Iqbal & Bailey, 2007). On the other hand such models could be very useful for tasks that have a fairly stable sequence of actions to be performed.

The results from these studies could be applied to some of the scenarios discussed in the Introduction of this essay. For example listening to an audio-book could be modeled as a task and finishing each chapter or a paragraph could be a subtask in this task model. Another example could be listening to a playlist consisting of several tracks. The whole listening experience could be considered a task and listening to each track could be a subtask. It seems reasonable to assume that the limited number of tasks that could be performed by an MP3 player enables us to apply the results from this research to this domain. In situations where there is some flexibility over the timing of interruption, we can apply these models to decide an appropriate moment for interrupting or communicating with the user.

2.3 Statistical Models

Detecting an appropriate moment for interrupting a user is one of the most challenging topics in the interruption literature. An appropriate moment is one that imposes the least interruption-cost while still benefiting the user. The previous section provided an overview of some approaches which tried to select a moment based on the structure of the task. While interrupting the user based on the task structure information is beneficial, it doesn't always achieve the least interruption cost for the user. In addition, we don't always have access to the task structure to find the appropriate moment for interrupting users.

Improvements in machine-learning have led researchers to develop systems to predict the appropriate moment for interrupting a user based on different sources of data about the user and his/her interaction with the system. This data could be from interaction of the user with the system or a reading from different kind of sensors in the environment or worn by the user. In this approach the system use the statistical model to make prediction about interruptibility state of the user. It is very common practice to refer to such data as 'evidence' or 'feature' in the literature. We will use the term evidence in the rest of this section in order to refer to such data.

Using statistical models in making decision about interrupting a user introduces a new dimension to the problem: uncertainty. Uncertainty is an inseparable part of using statistical models in making decision about actions performed by various systems. Horvitz (1999) provides a set of guidelines for decision-making with uncertainty. He suggests that the user should have control over the decisions made by the system and be able to modify these decisions. Horvitz, in 1999, outlined a framework for decision-making under uncertainty. He provides examples of the usage of the guideline in *LookOut* (an intelligent software agent that helps users of the *Outlook* application review calendar dates and arrange appointments). A subset of related guidelines from Horvitz are presented here:

- **Allowing direct invocation and termination**. Systems sometimes make poor decisions under uncertainty. It is desirable for the user to be able to have control in these situations. The user should be able to terminate or invoke action directly.
- **Continuing to learn by observing**. An automated system should continue to learn about the user. This will help the system to make better decisions as time goes by.
- **Maintaining working memory of recent interactions**. The System should keep a history of the recent interactions of the user with the system. System should let the user make reference to this recent history efficiently.
- **Employing socially appropriate behavior for system-user interaction.** The System should choose a method of interaction which matches the best social expectation of the user.
- **Providing a mechanism for an efficient agent-user collaboration to refine results.** We should design the system with the assumption that a user may wish to refine the analysis provided by the system.
- **Inferring ideal action in light of costs, benefits, and uncertainties.** Each action performed by the system is associated with its context-dependent cost and benefits. A system should take into consideration these parameters in the form of the expected value of taking actions.

The guidelines chosen above are important in the context of the MP3-player communication. Since an MP3-player must rely upon limited sources of data in making a decision, there is probably a considerable amount of uncertainty.

In the next section we will discuss a framework for decision-making which was introduced by Horvitz and his collaborators (Horvitz & Apacible, 2003; Horvitz, Jacobs, & Hovel, 1998).

2.3.1 Decision Making under Uncertainty

Horvitz, in 1998 and later, in 2003 with others (Horvitz & Apacible, 2003) introduces an approach for making decisions under uncertainty. Horvitz defines a utility function for interruption that takes into account cost and benefit of the interruption in the user's context. (For a more detailed discussion of context we direct the reader to Section 2.3.3). The same approach was introduced by Horvitz, regarding system actions and user goals in the context of the *LookOut* agent in the *Outlook* program (Horvitz, 1999). A similar approach has been used by others to predict human 'interruptibility' using sensors (Fogarty et al., 2005).

In this section we describe this framework. The framework defines the basis for reasoning and decision-making under uncertainty. It is flexible and can be easily modified to reflect the needs of the application.

We begin by explaining the framework. Suppose we have a set of evidence, E, about the state of the user in the system. The evidence could come from different sources like interaction of the user with the system, personal data about the user (for example, his/her appointments), or readings from different sensors in the system. We define a set of interruption states for the user;

F_i. The statistical model will assign a probability, p, to each state F_i based on the evidence available to the system, $(F_i|E)$. Based on the probabilities for each interruptibility state we can define the expected cost and benefit for each action.

In the next step we can define the expected cost of action, A, when the user is in state F_i using the following equation:

$$ECA = \sum_i C(A, F_i)\, p(F_i|E)$$

In the equation above ECA is the estimated cost of interruption to the user when the system believes the user is in the interruptibility state F_i. We should point out that the most common interruption cost used by most researchers in this area is resumption-lag. We believe that this belief is influenced by the Altman's work in introducing memory for goal-model (Altmann & Trafton, 2002).

Similarly, we can define an expected benefit for an action using the same equation by simply replacing the cost function with the benefit function. The benefit of an action could be defined by the information contained in the action. Horvitz and colleagues (Horvitz, Kadie, Paek, & Hovel, 2003) employ classifiers in order to determine the urgency, hence the benefit, of the messages in the system. It should be mentioned that defining a 'benefit function' is a domain-dependent function. Each application can have different criteria for assigning benefit to the interruption. This assignment can be influenced by factors related to user's context, relevancy of the interruption-information to the task, and the priority of the interruption.

Once the ECA value and its expected benefit of action A have been defined, a utility value for action, A, regarding the state of interruptibility could be defined. The utility function takes into consideration the cost of interruption and the benefit that action A will provide to the user. The utility function enables us to provide a balance between the cost of the interruption and the usefulness of the information or action to the user. The utility function can be defined as the difference between the expected cost of action, A, and the expected benefit that action, A, will provide for the user.

This approach enables us to divide the interruptibility state of the user into different segments. Defining different costs and benefits for actions in each segment will provide us finer control over interruptions.

2.3.2 Detecting Interruptibility State

In this section we will provide an overview of the different statistical models that have been used in literature in order to detect a user's interruptibility state.

Bayesian networks and Dynamic Bayesian networks are one of the most used statistical models in predicting interruptibility of the user. They have been used by Horvitz and his collaborators to predict user interruptibility based on various evidence available in the system (Horvitz & Apacible, 2003; Horvitz et al., 2003).

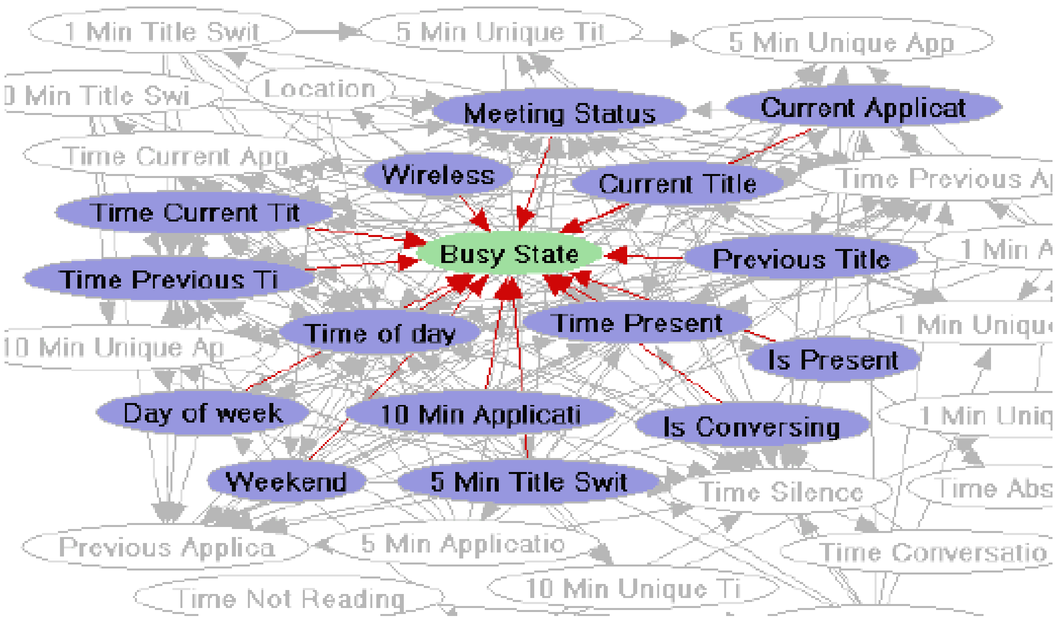

Figure 5 Bayesian Network used for predicting user interruptibility (Eric Horvitz, Koch, & Johnson Apacible, 2004)

Figure 5 show a portion of the Bayesian network used by Horvitz for detecting a user's interruptibility state (Horvitz, Koch, & Apacible, 2004). The highlighted nodes represent the most influential types of evidence for determining the probability that the user is in a busy state. In his experiment Horvitz uses the user's self-defined monetary cost as the interruption cost. It represents the amount of money each user is willing to pay in order not to be interrupted in that state.

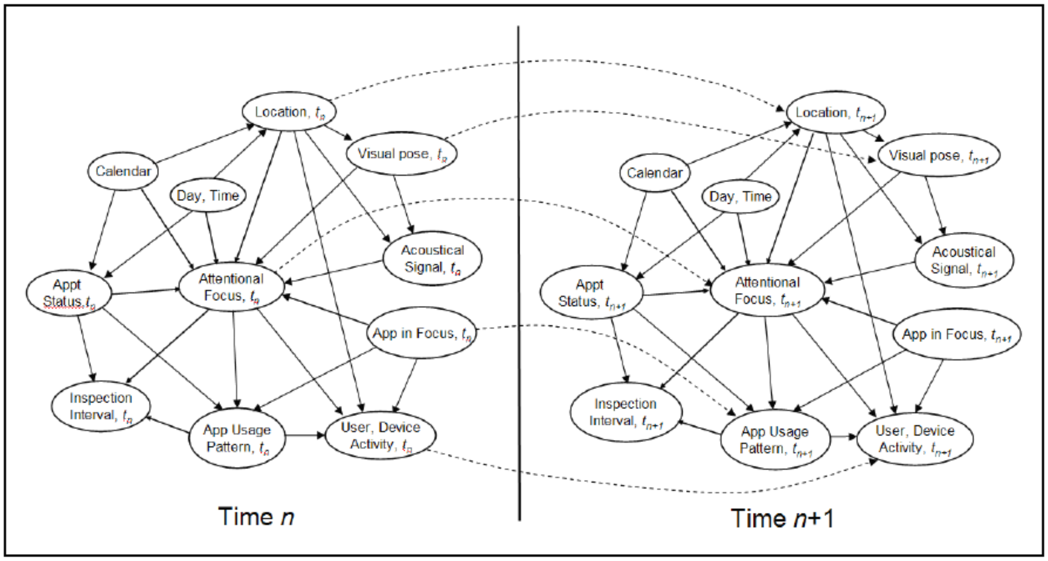

Figure 6 Dynamic Bayesian Networks for predicting user interruptibility (Horvitz et al., 2003)

Figure 6 shows the dynamic Bayesian network used to predict a user's interruptibility state (Horvitz et al., 2003). Dynamic Bayesian network models take into account that some of the evidence in the environment changes over time. It enables this model to monitor and predict a user's state over time. This property makes it possible for the system to adapt itself to the changes in the environment. The Dynamic Bayesian network model achieves this goal by creating a series of Bayesian networks that are connected together. Each Bayesian network represents the state of the system in a specified time (a time slice). Each new time slice captures the changes in the environment and is influenced by the values of the nodes in previous time slices. A new time slice is created either based on some predefined time slice or triggered by some actions in the environment. Dynamic Bayesian networks are more complex than static Bayesian networks and take more space and time in order to process evidence. However, they are more accurate and have the ability to adapt over time to changes in the environment. Both models have the flexibility to represent different interruptibility states for the user in order to provide finer grained control over interruption.

Naïve Bayesian networks have also been used in order to predict a user's interruptibility (Fogarty et al., 2005; Hudson et al., 2003). Naïve Bayesian networks are much simpler compared to dynamic Bayesian networks, yet good results have been achieved in predicting a user's interruptibility state interruptibility (Fogarty et al., 2005; Hudson et al., 2003). The basic assumption in designing Naïve Bayesian networks is that different types of evidence in the environment are independent and thus one does not affect another. Although this assumption is not always true this model has produced good results in practice.

Hudson et al. also uses ***Decision Trees*** **and** ***Support Vector Machine*** in order to predict human interruptibility (Hudson et al., 2003). They use WEKA (Witten & Frank, 1999) in order to implement these statistical models. WEKA is a popular toolkit that provides a lot of statistical models implemented in the Java programming language.

2.3.3 Detecting user context

Until recently most of the work on interruptions has considered the user in an office setting while the user is interacting with a desktop computer. There has been little effort in considering the user activity or context in the past. Hudson et al. have tried to use data from sensors in an office environment in order to predict a user's interruptibility (Hudson et al., 2003). They have considered in their model the time of the day, audio sensors, telephone, keyboard, and mouse usage. This work is among first attempts to include a user's context and activity in the models for interruption prediction. In our scenario a user may be engaged in various activities while, in the background, he/she is using a mobile MP3 player. In order to develop better models for predicting a user's interruptibility we need to have a more concise definition of user's context. In the first part of this section we will provide examples and definitions of user's context to establish a better understanding of the term. After that we will provide examples of how the concept was used in designing models for interrupting users in various environments.

There is a need for consideration of user's context in making decisions about the timing of mobile device communication with the user. Unlike the scenario with a desktop computer, the user's primary focus is not on the device. There also might be other competing devices in the environment that try to capture the user's attention. Context can be a combination of the activity of the user and the environment that surrounds him/her. In order to better visualize the problem we provide the following examples.

Imagine a situation, where an employee is in a meeting with his/her boss. In this situation s/he prefers not to be interrupted. In the past some systems have accounted for these situations in their statistical models by considering evidence that would help them understand the user's context. Horvitz et al. (Horvitz & Apacible, 2003) and Hudson et al. (Hudson et al., 2003) have taken into consideration these situations by including such evidence in the model. This information can be obtained from calendar programs and sensors in the office environment. This problem becomes more challenging because we don't have access to such information and the user has high mobility. To provide a better picture of the mobile scenarios we will provide two examples that shed light on the problem.

Imagine a student sitting in a library, and regardless of the interruptibility state of the student, certain kinds of interruptions should not be applied to his/her situation, such as auditory interruptions. This would create a disturbance. However, in this same situation if the system recognizes that the student is not studying and is just relaxing then the system may send him/her notification messages that can be perceived only by the student.

Imagine a second situation where the user is in a café. This café environment is a relaxed environment that doesn't impose any constraint on the interruptibility of the user, and so the state of the user's interruptibility is the key in making any decisions. If the user is walking around inside the café, then the system may produce an auditory interruption to get the user's attention. While in another configuration, when s/he is sitting down and relaxing, a calmer interruption method is preferred.

The term *context* has been defined and widely used in the context-aware computing literature. However, there is not a well establish definition of what context is and what constitutes context. In most of the literature the focus is on developing applications that behave appropriately in relation to context. However in the literature some researchers have tried to identify the important aspects of a context-aware computing task. From the examples provided it is apparent that in these context-aware scenarios both the device and the user are mobile and the environment as well is constantly changing and so we have to be prepared to react to these changes. Some of the environment characteristics that have been studied in the context-aware computing literature are related to connectivity, power consumption, lighting in the environment, noise, et cetera. While these characteristics are mostly concerned with the operational aspects of the context-aware computing, there are other factors that take into account the social aspects of context-aware computing. Shilit, Adams, and Want define three important aspects of context as: who you are with, where you are, and what resources are nearby (Schilit, Adams, & Want, 1994). Answering these questions will help us to define a user context and to design systems that adapt to user context. Although this kind of definition for context will guide us in designing a context aware computing environment, we still need to have a more concrete definition. This definition of the context should help us to select information that forms the context of using an application. Dey defines context as “Any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves” (Dey, 2001). This definition helps us to decide whether to include a piece of information in user context. He also defines context-aware as “A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user’s task” (Dey, 2001). Using these formal definitions and the problem at hand (in this case an MP3 player) we can decide what information in the environment we want to include in our context and what service, presentation methods and automation are desired. From the definition, it is obvious that for different tasks and sensors we would have a different definition of context and behaviors.

Kern et al. address this problem by considering the interruptibility space in two dimensions (Kern, Antifakos, Schiele, & Schwaninger, 2004). In this two- dimensional space one axis is used to provide information about the user’s interruptibility and the other axis is used to provide an index into social interruptibility that reflects the environmental constraints about the interruptibility of the user. This two-dimensional space is divided into different segments. Each segment suggests a different method for interrupting the user. They incorporate information regarding the user’s coordinates in this two-dimensional space into their statistical model. Figure 7 shows a sample of such division in interruptibility space where each axis is divided into three segments, each with a given label for suggested context. It provides representative activities and shows where they each fit in the two dimensional space.

Kern et al. have used different types of sensors in order to determine user’s context in the environment (Kern et al., 2004). Information from different sensors can help us to have a better understanding of the user while performing a task. In the rest of this section we will provide an overview of the work in this area.

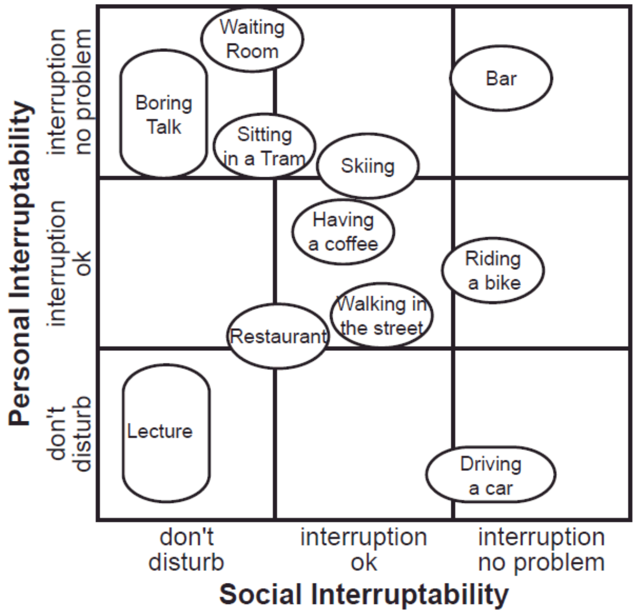

Figure 7 Two dimensional interruptibility space used by Kern et al. (Kern & Schiele, 2003)

Using Sensors

Sensory information from different sources could be combined to provide a good prediction of user's context. These sensors could provide information about different aspects of user's context, like movement, position, biological signals, and environmental properties. In the following sections we will provide an overview of the different types of sensors that have been studied in order to provide a prediction about user's context.

Location sensors

One of the most popular sensors for providing information about a user's position is Global Positioning System (GPS). GPS sensors can provide information about the location of a user with an accuracy of 10 to 20 meters (Chen & Kotz, 2000). There have been several research projects which have utilized this information and created adaptive systems that provide information to the user based on location.

Zheng et al. developed an algorithm for the prediction of user-motion from their GPS logs (Zheng, Li, Chen, Xie, & Ma, 2008). In this research they identify a set of features based on GPS logs to predict a user's motion mode (walking, driving a car, biking). They also come up with two different post processing algorithms for enhancing their predictions. The work done by Zheng and colleagues is different since the focus of their work is not determining the location of the user but how he/she is moving among locations. Systems can be trained to distinguish between various modes of travel.

Marmasse and Schmandt try to develop a location-aware information delivery system (Marmasse & Schmandt, 2000). Different types of information can be related to a location. A simple example can be a to-do list that reminds the user of the task that he/she is supposed to do at that location.

They exploit the fact that the GPS signal is lost when a user enters a building and develop an algorithm to find locations of the interest of the user.

Kern and Schiele use information from wireless LAN sensors in order to determine the location of the user and incorporate this information into their context model (Kern & Schiele, 2003).

Accelerometer

Accelerometer sensors are small inexpensive sensors that are available on most current handheld devices. They provide information about device movement along two or three axes. In several studies they have been used in order to determine a user's activity in the environment. Information from accelerometers enables us to distinguish among simple activities like walking, running, standing, sitting and so on.

Kern, in his study, gathers information about his user's activity from twelve accelerometers fixed to various locations on his subjects' bodies (Kern & Schiele, 2003). Ho and Intille also fixed two wireless accelerometers on their user's body in order to predict the activity of the user in the environment (Ho & Intille, 2005). In this work they distinguish between three activities: sitting, standing and walking. They propose a model to detect the transition among these three states in real time. Then they suggest delaying non-critical interruptions until a physical transition intervenes a user's activities. They believe that this model of delivering interruptions will reduce the burden of informational overload on users. In both models they use sensors that, unlike handheld mobile devices (like an MP3 player), can be secured in fixed positions to a user's body. In the MP3 player scenario, the researchers have no control over the location of the device on the user: The user might place the device in his/her front or back pocket. We require an algorithm that accommodates any location the user might choose for the device. Such algorithms should provide a useful prediction about the user's activity in the environment regardless of the location of the device on the user.

XPod is a context aware music player which uses the accelerometer sensors, heat flow, Galvanic Skin Response (GSR), and song meta-data (genre, artist, album, title, beats-per-minute) to predict user's rating for a song in different context (Dornbush, Fisher, McKay, Prikhodko, & Segall, 2005). They use the data form these sensors to train different statistical models for predicting the user's rating for each song. Their results show the potential of using this approach for developing a context-sensitive music player, however the accuracy of their model is relatively low.

Hinckley and Horvitz use three sensors on a handheld device in order to predict user's current activity (Hinckley & Horvitz, 2001). They use a touch sensor to detect whether a user is holding the device or not. They also have an infrared sensor that senses the proximity of the device to the user and also has a two-axis accelerometer.

Kawahara, Kurasawa, and Morikawa developed a model that predicts a user's activity based on data from only one accelerometer sensor embedded in the handheld device (Kawahara, Kurasawa, & Morikawa, 2007). They consider the fact that users tend to keep devices on different parts of the body. For example some people tend to keep their mobile device in their pant pocket while other people tend to keep their mobile devices in their chest pocket. They develop an algorithm which first determines the position of the device on the user's body and then, based on the position of the device on the user's body the authors predict the user's activity in the environment. They can distinguish between four different user-postures (sitting, walking, standing and running).

Yi et al. combined the data from a single accelerometer sensor with information about lightning of the environment in order to distinguish between different user activities (Yi, Choi, Jacko, & Sears, 2005).

Microphones

A user in an environment is always surrounded by different sounds that can be studied in order to provide a better prediction about the user's context. This surrounding sound can hint at the place the user is located. It also can provide useful information about the activity in which the user is engaged. For example we may understand the user is on the street and talking to someone based simply on the surrounding sound.

Kern uses auditory classification in order to distinguish between four different situations: street, restaurant, lecture, conversation (Kern & Schiele, 2003). A very simple approach was also used by Hudson (Hudson et al., 2003) to detect a phone ring and conversation in an office environment for predicting a user's interruptibility. A more sophisticated algorithm using Support Vector Machines (SVM) is developed in (Lu, Zhang, & Li, 2003) to classify sound. They successfully classified sound in four differ categories: music, background sound, pure speech, non-pure speech.

EEG and ECG

Signals from both the brain and heart can be used to construct probabilistic models about a user's cognitive load. An electrocardiogram (ECG) can provide a strong signal that can be used to detect a user's cognitive load level (Chen & Vertegaal, 2004). An electroencephalogram (EEG), which provides an image of brain activity, is a relatively weaker signal but can still provide information about the motor activities of the body. In HCI research these signals have been used in determining a user's cognitive load. Readings from these signals will help us to determine when a user is relaxed or isn't engaged in a task.

Chen and Vertegaal use a combination of EEG and ECG for developing a model to distinguish between four user-states (Chen & Vertegaal, 2004). Then, they integrate a cell phone with this model to enable a user to have different preferences for each state.

Lee and Tan used EEGs to determine the task a user is performing based on the cognition-load of the user (Lee & Tan, 2006). They trained a Bayesian network to classify tasks. Their classifier, in a controlled environment, was able to classify between rest, mental arithmetic and mental rotation (a user was asked to imagine an object with as much detail as possible and then rotate that object in his/her mind). In other work by Shenoy and Tan, an EEG is used to detect when a user is looking at a facial image (Shenoy & Tan, 2008). This work is interesting because the subjects of the study were not asked to detect a facial image but the system itself detected the image being shown to the user merely based on specific EEG signals received among the entire group of EEG signals.

Although these signals provide very useful information about a user's mental and physical activities, they are very intrusive and the equipment to monitor these signals is usually bulky and uncomfortable to use for the subject. Figure 8 shows the placement of sensors on a subject in the study performed by Chen and Vertegaal (Chen & Vertegaal, 2004).

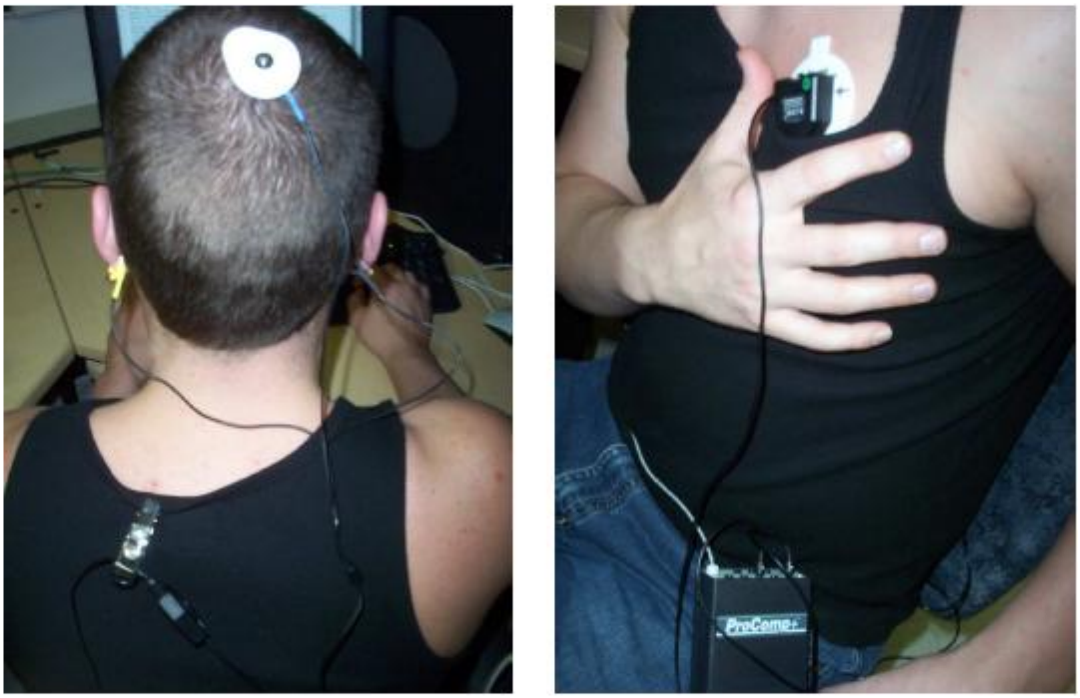

Figure 8 Placement of the sensors for EEG and ECG (Chen & Vertegaal, 2004)

2.4 Summary

We started this part of the essay with an introduction of the task resumption models. These models enable us to make predictions about interruptions. They predict the effect of interruption duration, mental demand, and task rehearsal on task resumption. The models use resumption-lag as a measure to study the cost of interruption. Several experiments were introduced which studied the predictions made based on these models. The results show that long interruptions and higher mental demand will increase task resumption-lag. These models help us to design systems that facilitate task resumption for users.

Although these task resumption models enable us to have a better understanding of the interruption handling process, we still need to decide when to interrupt a user. The second part of this section of the paper was devoted to models that enabled us to select an appropriate time for interrupting the user. At first we studied models based on the task structure. These models choose the most appropriate time for interrupting the user based on task structure. They predict that the interruption cost would be minimized if the user was interrupted between subtask boundaries.

However, task structure information is not always available for every task and it is also very hard to define task structure for some type of tasks. Statistical models choose the appropriate time for interruption regardless of the task structure. They use information available in the environment about a user's interaction with the system and also use data from various sensors to make a decision about interrupting the user.

User's context is very important in making decisions about time of interruption. While the scenario of detecting and defining a user's context when using a desktop application is challenging, it is much more complicated for mobile devices. We provided a definition of context from the mobile computing and context-aware computing literatures in this essay. We introduced examples of how user's context was incorporated into models for predicting an interruption time.

Sensor data is the major part of detecting context. We provided examples of different sensors that can be used to detect context. A designer can choose among the available sensor data for detecting context, based on the definition of context provided earlier.

3 How to Interrupt

In the life cycle of a system there are inevitable moments that the software needs to contact the user. These communication moments can be seen as the system interrupting the user. In previous sections we studied the effect of interruptions on user performance. We studied methods for reducing the interruption cost to make them less intrusive. We also studied the theory behind how interruptions affect mind and memory. We learned about the processes involved in resuming an interrupted task and how we can improve these processes. In this section we want to cover different methods available for interrupting users. This introduction will help us to choose the best method for interrupting the user based on the task and information we want to convey to him/her. We also cover research that has tried to find the right balance between the interruption information and the method for interruption.

We also cover some of the recent approaches to interrupting users that are not based upon using the visual modality. This is especially relevant in our MP3 scenario since user have minimal visual contact with the device.

3.1 Methods of interruption

There are different methods for interrupting humans when working with a system. McFarlane and Latorella defined four different methods for coordinating interruption (McFarlane, 2002).

- Immediate interruption: sometimes it is required for the user to interact with the interruption immediately and he/she has no option for postponing the interaction with the interruption. This may cause a problem when the user decides to resume the interrupted task. Research has been done in terms of providing guidelines for UI design in order to make the task resumption task easier and more effective.
- Negotiated Interruption: humans usually negotiate human-human interruption. It also has been shown that humans can handle a lot of information from their background environment while working on a specific task. Consider these facts in negotiated-interruptions unlike immediate interruption we give user a chance to:
    - Handle it immediately
    - Notify the interrupting person and handle it later
    - Do not handle it
    - Ignore it by not interacting with the interrupting person
- Mediated interruption: in an office setting all of the received information (interruptions) goes through a secretary and the secretary decides when and how to inform the boss that new information has arrived. Introducing a mediator into the process adds the overhead of supervising the mediator. One way to reduce this overhead is to automate the mediator but this introduces a whole new problem in user-modeling and detecting the appropriate interruption time (read Section 2.3).
- Scheduled interruption: The idea behind this method is to let the user know in advance when he is going to be interrupted. It helps to make different types of interruptions less intrusive and make them more of a scheduled task. This method requires studying a user’s time management and organization.

Next, after identifying these approaches, we must find which of the above methods is the best approach for interrupting users. In a study performed by McFarlane four different UIs were implemented, where each represented one of the above approaches to interruption. Game-playing is the main task and object-matching is the interrupting task. The results show that there is no single best interruption method and, based on the situation and the system, each of the above approaches can be used. They provide a set of guidelines based on the result of the experiment. These guidelines are quite specific and one should be careful in using them. They are as follows:

- If accuracy is important then continuous task negotiation method is best; scheduled method is worst
- If the number of tasks is important then switching between the main and interruption task schedule is best and immediate the worst
- If completing an intermittent task is important then immediate method is the best approach
- If promptness of response is important then the immediate method is best.

An important part of designing a system is to choose the most appropriate interruption method. The selected interruption method should provide the desired level of 'attention-grabbing' while effectively conveying information to the user. Gluck, Bunt and McGrenere conducted a study to evaluate the effect of matching the utility of interruptions with their attentional draw (Gluck, Bunt, & McGrenere, 2007). They defined attentional draw for a notification signal as the time that it takes for a user to notice the signals presence after it was generated by the system. They performed a first study to rank the notification signals based on attentional draw, and in a second study they try to match the attentional draw of a signal to utility of its content (amount of information a notification signal provides for helping complete the primary task) compared to a uniform approach which uses only one notification signal with medium attentional draw. Their results showed that matching attentional draw with utility results in a reduction of annoyance and an increase in performance. It should be further noted that their results need more investigation in realistic situations and perhaps need some other infrastructure for evaluating the utility of a notification in general cases.

The approach taken by Gluck, Bunt and McGrenere is particularly interesting in the scenario of using an MP3 player. Because the number of messages the device can communicate to the user is limited, they can be evaluated in terms of their utilization and, consequently, the best interruption method can be determined based upon the utilization estimation. These results are also interesting when one considers them alongside the statistical methods introduced in section 2.3 Statistical Models. With statistical methods the most important concern is the decision when to interrupt the user, yet we can go further and choose the right approach to interrupt the user simply by considering the user's context. Some interesting work in this area is presented in the next section.

Another approach that has been studied by Horvitz, Apacible and Subramani is "notification deferral policies" (Horvitz, Apacible, & Subramani, 2005). The main idea in this approach is that when an alert arrives and if the user is also busy, then the system will wait for a pre-specified amount of time before informing the user. Thus it is a trade-off between user-awareness of the events that are happening in the system and the disruption of the user. In their experiment they analyzed information about work habits of their subjects in order to find patterns in the transition from high-cost interruptions to low and medium-cost interruptions. Based on the observed values they defined a maximum amount of time that a notification can be delayed. They also let users define different policies for messages with different priorities. This is a good example of the "mixed-initiative approach" that includes the user in decision-making process.

These methods are good examples of work in in the field of human-computer interaction. It is as important to choose the right method for interrupting a user as when to interrupt him/her. A similar approach might also be successfully applied to the MP3-player scenario.

3.2 Non Visual Approaches

Most of the work in the interruption literature is concentrated around visual signals and visual interruptions. This, we believe, is because researchers consider interruption scenarios revolving mainly around information workers in an office setting. However there has been some recent research in interruption scenarios that considers new devices such as mobile phones and user context. Nevertheless, we could not find many papers on new methods of interrupting users in a mobile context. Most of these articles are centered on finding the appropriate time for interrupting a user that rely upon existing methods for interrupting the user. Since our MP3-player scenario has a great deal in common with mobile computing and context-aware systems, we used a considerable amount of information from these fields in preparing the previous sections.

For the remainder of this section we will provide examples of some non-visual approaches for interrupting users. These should help us design systems that require less visual attention and communicate with user through the other, non-visual senses.

3.2.1 Auditory interrupts

Audio signals have been used in computers for signaling events for a long time. They have usually been used to provide notification about tasks and events (receiving an email or notification, finishing a file copy). They are also very powerful in grabbing the immediate attention of humans and because of that they have been used in the design of alarm systems and other types of systems that require immediate attention. We will provide two examples of systems that use audio signals for interrupting and providing information to the user.

The first example is work done by Garzonis and associates (Garzonis, Jones, Jay, & O'Neill, 2009). In their work they utilize auditory icon we call 'earcons' that provide information about availability of a mobile service to the user. Before we talk about the results we need to provide a definition for auditory icon and earcon. Earcons are a non-verbal audio messages employed in a computer interface to provide information to the user about some computer object, operation or interaction. This is different than an auditory icon which is a sound that establishes a relationship between the event and some background information. For example, the sound of glass shattering may be used to indicate that a file has been dropped into a computer's trash. In their study they grouped mobile services into ten groups and assigned each group an earcon and an auditory icon. The goal of the study was to compare earcons and auditory icons in terms of learnability, memorability, intuitiveness, and user preference. It was expected that the auditory icons would have better learnability and memorability because of their definition. Results were in line with expectations, showing that auditory icons have better learnability and memorability and intuitiveness. They also have a lower rate of forgetability compared to earcons. Results from this study have the implication that the design of the user-interface of devices such as MP3-players can effectively communicate with the user through audio signals alone. Auditory signals may in fact be the preferred method for notification about events.

Audio signals can also be used as a means for conveying background information. Humans have evolved the ability to constantly monitor environmental sounds and extract useful information about the surrounding environment. We can use an office setting to provide an example of this information processing: a person, while sitting in his/her office, may hear other people in the hallway and may anticipate these people entering his/her office. In another example, when that office worker passes a colleague's office he/she may hear that person talking and thus be aware about his/her presence.

Mynatt et al. studied this extraordinary trait in designing audio auras (Mynatt, Back, Want, Baer, & Ellis, 1998). The authors detail how audio auras are created to provide an implicit way to provide information to the users about their surrounding environment. In their study the users wore headphones and the system would provide various bits of information to the user through headphones. The researchers also received information about the users' locations in an office via sensors placed in the office environment. For example when the users were in the office kitchen (to get a coffee) the audio aura would convey information to the users about how many new emails they have received. In designing their audio aura the researchers used a combination of auditory icons, earcons, music, and verbal statements. Each different aura would provide the user with a different level of information and engaged the user at various cognition levels. For example verbal statements provide users with the most information and also quickly grab the attention of the user. A similar approach can be used to design the MP3 player system. Some of the work related to location aware information delivery systems can be found in Section 2.3.3.

3.2.2 Thermal displays

In scenarios related to MP3-players most of the time the device is around the arm or hidden within reach, often in a user's pocket. Its proximity opens the possibility of using touch to communicate. More specifically, a thermal display can communicate some basic information from the device to the user in a quick and easy way.

Husin-Ni and Jones have developed a thermal display that enables a user to distinguish between different surfaces using thermal cues (Ho & Jones, 2007). This thermal device is interesting because of its potential for conveying discrete information to the user. An example scenario could be the genre of the music he/she is listening to, or the average rating of other people for the music. Another approach in using thermal displays for conveying information is used by Wettach et al. (Wettach, Behrens, Danielsson, & Ness, 2007). In their work they have designed a thermal display that can represent up to five discernible temperatures. This kind of display can be used to convey information about ratio or quality of information. An example would be an alarm system for mobile phones. Imagine you are in a very important meeting and don't want to be disturbed but you don't want to miss important calls either. So you have set your mobile phone to 'silent' mode and this phone also features a thermal signaling device. If the person who is calling is your boss then the device's temperature may go relatively quite high, informing you that this is a very important call. On the other hand, if the caller is your mother, the phone may stay relatively cool and so not disturb you.

3.2.3 Haptic feedback

Humans have used touch to draw one another‘s attention long before the development of language. Hands, of course, are obvious for such tasks, like gently squeezing another's hand or tapping somebody on the back. A variation in position and speed of these contacts conveys different meanings. Haptic research taps into such language and communicates with users using the same signals. Baumann et al. has used this idea in emulating human behavior in attention getting by evaluating various haptic devices (Baumann, MacLean, Hazelton, & McKay, 2010). They succeeded in designing a wrist band that can emulate human behavior in getting attention. The importance of such research in our MP3-player scenario is that we often attach these devices to our bodies. Imagine a scenario in which you have your MP3-player device around your arm while running. The device can communicate with the person invisibly, noiselessly, just by using the haptic technology explored by Baumann et al. (Baumann et al., 2010).

3.3 Summary

Let us review the many methods we have discussed for interrupting a user. The first section categorized various interruption techniques based on their coordination methods. A coordination method for an interruption defines the way that interruption will interact with a user. However within each level of coordination an interruption might have varying intensity in terms of getting a user's attention. An appropriate interruption method for any level is the method that has the best balance between attention-getting and the information it conveys to the user. We provided an overview of some of the approaches that have defined such a balance.

The last section was devoted to non-visual methods for interrupting a user. Non-visual approaches are of particular interest in the MP3-player scenario since the user has limited visual contact with the device. Because the device is out of the user's sight most of the time the interruption method should rely on other senses to alert the user. We briefly discussed some of the audio, thermal and haptic methods for communicating with the user.

4 Conclusion

This essay gathered information in regard to the MP3-player scenario introduced in the introduction of this essay. Our research might provide the foundation of a solution to the MP3 player scenario. In Section 2 we briefly reviewed the current literature in relation to the question of when to interrupt the user. We very briefly studied different models of the human mind. These models provide an explanation of what happens when a user is interrupted. Particularly, we studied the ‘goal- activation’ model that enables us to make predictions about interruptions. This model predicts that the duration of an interruption and the demand it makes upon memory would cause difficulty for a user in the resumption of an interrupted task. This information could be helpful.

Another important issue is when is the best time to interrupt a user. We studied different models that employ various bits of information about the user and the task at hand in order to predict the best time for interruption. The first models we studied were based on the task structure. These models break the task that a user is performing into different layers and segments. They suggest that interrupting a user at the time of transition between subtasks at higher levels would cause less disruption for the user. Although these models may provide useful information about the best time to interrupt a user, the models also require substantial knowledge about a user’s task. This information is not always available to the designer and extracting this information manually for each task is time consuming. These problems make this model impractical for dynamic and complex tasks. Models that are agnostic of the task structure perform the prediction task based on knowledge about the user’s interaction with the system and sometimes data from different type of sensors.

The sensors could provide information about the surrounding environment or the user itself. Biological sensors have been used to gather information about the user in some models. The installation or intrusiveness of these sensors introduces some problems in using them in real applications. Different probabilistic methods were introduced that have been used in constructing such models. These models are more appropriate for dynamic and complex tasks, since they require no knowledge about task structure.

In the MP3-player scenario user’s context plays an important role in selecting an appropriate time for interruption. We addressed this issue by providing an overview of some mobile-computing and context-aware computing literature. We provided a definition of user’s context that would help designers in selecting the appropriate information from the environment. This information would help us analyze user’s context and help us make better predictions about the optimal interruption time. A very useful source of information for analyzing user’s context is information about a user’s surrounding environment and a user’s activity. Sensors can provide us with this information. We provided examples of sensors employed to detect a user’s location, or sounds in the environment, and a user’s movement. Sensors provide a valuable resource for understanding user’s context and also for constructing models for interrupting a user.

The second part of the essay addressed several approaches employed in interrupting a user. An important issue in finding the interruption method is to find the right balance between intrusiveness and the information that the interruption provides for the user. The remainder of the essay investigates a number of non-visual approaches for interrupting a user, including audio, thermal, and haptic methods.

We have discussed some interesting approaches to the design of effective communication between mobile devices and users.

5 Challenges

We have discovered a number of challenges and opportunities with regard to MP3-player design. First these challenges and opportunities are related to timing of interruption. As the goal-activation model predicts, resumption-lag increases as the length of interruption increases. In our MP3-player scenario one type of interruption is external, such as starting a conversation with a friend. It can be assumed that the duration of the interruption would usually be long; hence we have to design helpful cues for resuming a task. Since the majority of the tasks performed by an MP3-player have a simple structure we can utilize "Salvucci's theory" in designing the cues. We can study the task and identify the important steps in reconstructing it and then use information from the user and the environment to select and save important information about the task as cues.

An opportunity exists for designing models for interrupting a user because generally the tasks performed with an MP3-player have a simple structure. We can use this simple structure and research on interruptions based on task structure in designing interruption models for MP3 player. It is important because these models are generally simpler than statistical models and can be implemented on low-cost devices with less computational power. However since an MP3 device may be used in a dynamic environment while a user is mobile we still need to use statistical models. The need for statistical models introduces two challenges that are related to each other. The first challenge is to select the best statistical model for interrupting user. Because dynamic Bayesian networks have been used by many researchers (and described in many research papers) for deciding the optimal interruption time, this makes them a very good choice for this task. However, selecting the best evidence and assigning probabilities in the model is a time consuming and difficult task. The other important factor that has a direct influence over which statistical model to choose is the selection of data. As we discussed earlier, data from sensors can provide valuable information about user's context. Since different sensors can provide different types of data (continuous or discrete) with varying levels of noise and error, we also have to take into consideration these facts in selecting the best statistical models or appropriate filters.

Sensor data identification and selection is another big challenge before us. We provided a formal definition of user's context. The definition specifies that context is influenced by application. The challenge is to define the best possible set of sensor data and user data that identifies the user's context for each application. The selection should consider the constraints on power consumption (different sensors have different power consumption) of the device and also the level of accuracy required by the model.

The second group of challenges and opportunities are related to various methods of interruption. Our MP3player scenario demands a non-visual communication channel. We investigated some of non-visual approaches for communicating with the user. The challenge before us is to select the best combination of signals and design appropriate signals. Audio and haptic methods seem promising and there are some good examples of using these signals to communicate with the user.

6 Bibliography

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when? *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 271-278. doi:10.1145/985692.985727

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive science*, 26(1), 39–83.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction*, 14(4), 1-28. doi:10.1145/1314683.1314689

Baumann, M. A., MacLean, K. E., Hazelton, T. W., & McKay, A. (2010). Emulating human attention-getting practices with wearable haptics. *Proceedings of Haptics Symposium, 2010 IEEE*, 149-156. doi:10.1109/HAPTIC.2010.5444662

Chen, D., & Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive user interfaces. *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, 1513-1520. doi:10.1145/985921.986103

Chen, G., & Kotz, D. (2000). A Survey of Context-Aware Mobile Computing Research. Dartmouth College Technical Report TR2000-381. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.3717

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, 5(1), 4-7. doi:10.1007/s007790170019

Dornbush, S., Fisher, K., McKay, K., Prikhodko, A., & Segall, Z. (2005). XPod a human activity and emotion aware mobile music player. *Proceedings of the International Conference on Mobile Technology, Applications and Systems*, 1–6.

Fogarty, J., Ko, A. J., Aung, H. H., Golden, E., Tang, K. P., & Hudson, S. E. (2005). Examining task engagement in sensor-based statistical models of human interruptibility. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, 331-340. doi:10.1145/1054972.1055018

Garzonis, S., Jones, S., Jay, T., & O'Neill, E. (2009). Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. *Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, 1513–1522. New York, NY, USA: ACM.
doi:10.1145/1518701.1518932

Gluck, J., Bunt, A., & McGrenere, J. (2007). Matching attentional draw with utility in interruption. *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '07*, 41–50. New York, NY, USA: ACM.
doi:10.1145/1240624.1240631

Hinckley, K., & Horvitz, Eric. (2001). Toward more sensitive mobile phones. *Proceedings of the 14th annual ACM symposium on User interface software and technology - UIST '01*, 191-192.
doi:10.1145/502348.502382

Ho, H.-N., & Jones, L. A. (2007). Development and evaluation of a thermal display for material identification and discrimination. *ACM Transactions on Applied Perception (TAP)*, 4 (2), Article 13.
doi:10.1145/1265957.1265962

Ho, J., & Intille, S. S. (2005). Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, 909-918. doi:10.1145/1054972.1055100

Horvitz, E., Apacible, J., & Subramani, M. (2005). Balancing awareness and interruption: Investigation of notification deferral policies. In L. Ardissono, P. Brna & A. Mitrovic (Eds.), Lecture Notes in Computer Science: Vol. 3538. *User Modeling 2005*, (pp. 433–437). Berlin, Germany: Springer-Verlag.

Horvitz, Eric. (1999). Principles of Mixed-Initiative User Interfaces, *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '99,* 159–166.

Horvitz, Eric, & Apacible, Johnson. (2003). Learning and reasoning about interruption, *Proceedings of the 5th international conference on Multimodal interfaces -ICMI '03,* 20-27.

Horvitz, Eric, Jacobs, A., & Hovel, D. (1998). Attention-Sensitive Alerting, *Proceedings of UAI '99, Conference on Uncertainty and Artificial Intelligence*, 305–313.

Horvitz, Eric, Kadie, C., Paek, T., & Hovel, D. (2003). Models of Attention in Computing and Communication: From Principles to Applications. Communications of ACM, 46 (3), 52-59. Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.2435

Horvitz, Eric, Koch, P., & Apacible, Johnson. (2004). BusyBody. *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, 507-510. doi:10.1145/1031607.1031690

Hudson, S. E., Fogarty, J., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, Johnny C, et al. (2003). Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '03*, 257-264.

Iqbal, S. T., & Bailey, B. P. (2006). Leveraging characteristics of task structure to predict the cost of interruption. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, 741-750. doi:10.1145/1124772.1124882

Iqbal, S. T., & Bailey, B. P. (2007). Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, 697-706. doi:10.1145/1240624.1240732

Kawahara, Y., Kurasawa, H., & Morikawa, H. (2007). Recognizing User Context Using Mobile Handsets with Acceleration Sensors. *Proceedings of the IEEE International Conference on Portable Information Devices, PORTABLE07*, 1-5. doi:10.1109/PORTABLE.2007.12

Kern, N., Antifakos, S., Schiele, B., & Schwaninger, A. (2004). A model for human interruptability: experimental evaluation and automatic estimation from wearable sensors. *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, 158-165. doi:10.1109/ISWC.2004.3

Kern, Nicky, & Schiele, Bernt. (2003). Context-Aware Notification for Wearable Computing. *Proceedings of the 7th international symposium on wearable computing*, 223-230.

Lee, Johnny Chung, & Tan, D. S. (2006). Using a low-cost electroencephalograph for task classification in HCI research. *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*, 81-90. doi:10.1145/1166253.1166268

Lu, L., Zhang, H.-J., & Li, S. Z. (2003). Content-based audio classification and segmentation by using support vector machines. *Multimedia Systems*, 8(6), 482-492. doi:10.1007/s00530-002-0065-0

Marmasse, N., & Schmandt, C. (2000). Location-aware information delivery with comMotion. In P. Thomas & H-W. Gellersen (Eds.), Handheld and Ubiquitous Computing Second International Symposium, HUC 2000 Bristol, UK, September 25–27, 2000 Proceedings (pp. 157--171). Berlin, Germany: Springer-Verlag

McFarlane, D. (2002). Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human-Computer Interaction, 17(1)*, 63-139. doi:10.1207/S15327051HCI1701_2

Monk, C. A. (2004). The effect of frequent versus infrequent interruptions on primary task resumption. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 48*, 295–299.

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied*, 14(4), 299-313.

Mynatt, E. D., Back, M., Want, Roy, Baer, M., & Ellis, J. B. (1998). Designing audio aura. *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI ’98*, 566–573. doi:http://dx.doi.org/10.1145/274644.274720

NASA TLX Homepage. (n.d.). Retrieved December 8, 2010, from http://humansystems.arc.nasa.gov/groups/TLX/

Salvucci, D. D. (2010). On reconstruction of task context after interruption. *Proceedings of the 28th international conference on Human factors in computing systems, CHI ’10*, 89–92. doi:10.1145/1753326.1753341

Schilit, B., Adams, N., & Want, R. (1994). Context-Aware Computing Applications. *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on*, 85-90. doi:10.1109/WMCSA.1994.16

Shenoy, P., & Tan, D. S. (2008). Human-aided computing: utilizing implicit human processing to classify images. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, 845–854. doi:10.1145/1357054.1357188

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58(5), 583–603.

Wettach, R., Behrens, C., Danielsson, A., & Ness, T. (2007). A thermal information display for mobile applications. *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services, MobileHCI '07*, 182–185.doi:10.1145/1377999.1378004

Witten, I. H., & Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (1st ed.)*. Massachusetts, United States: Morgan Kaufmann.

Yi, J. S., Choi, Y. S., Jacko, J. A., & Sears, A. (2005). Context awareness via a single device-attached accelerometer during mobile computing. *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services - MobileHCI '05*, 303–306. doi:10.1145/1085777.1085839

Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W.-Y. (2008). Understanding mobility based on GPS data. *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, 312–322. doi:10.1145/1409635.1409677