Design Guidelines for Adaptive Multimodal Mobile Input Solutions

Bruno Dumas, María Solórzano and Beat Signer
Web & Information Systems Engineering Lab
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels, Belgium
{bdumas,msolorza,bsigner}@vub.ac.be

ABSTRACT
The advent of advanced mobile devices in combination with new interaction modalities and methods for the tracking of contextual information, opens new possibilities in the field of context-aware user interface adaptation. One particular research direction is the automatic context-aware adaptation of input modalities in multimodal mobile interfaces. We present existing adaptive multimodal mobile input solutions and position them within closely related research fields. Based on a detailed analysis of the state of the art, we propose eight design guidelines for adaptive multimodal mobile input solutions. The use of these guidelines is further illustrated through the design and development of an adaptive multimodal calendar application.

Author Keywords
Design guidelines; mobile multimodal interaction; user interface adaptation; context-aware systems.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces.

General Terms
Design; Human Factors.

INTRODUCTION

Multimodal interfaces make use of multiple input or output modalities to enhance the communication and interaction with a user. By their very nature, they are excellent candidates for adaptive interfaces [18]. Depending on a user's profile or some environmental conditions, a multimodal interface may switch from one modality to another modality. Furthermore, multimodal interfaces potentially support adaptation at the input as well as at the output level. While the automatic adaptation of multimodal interfaces on the output side has been explored by various researchers, the automated adaptation of input modalities has been less studied.

MobileHCI 2013, August 27–30, 2013, Munich, Germany. Copyright 2013 ACM xxxx

This lack of adaptive multimodal input solutions might be partly caused by the fact that until recently mobile devices did not offer a large enough range of input modalities. Nevertheless, the automatic adaptation of multimodal input in combination with advanced multimodal fusion algorithms has the potential to enhance the overall usability of mobile devices by providing multiple alternative modalities. In addition, interfaces can adjust to environmental conditions and thereby improve the individual user's experience and performance. The automatic adaptation of input modalities might further be instrumental in addressing security and safety issues. For example, a user driving a car should not use the touch screen of their smartphone but rather rely on input modalities which allow their attention to stay focussed on the road. Last but not least, the social context might favour or prevent the use of certain modalities. The development of these kind of adaptive multimodal systems not only requires detailed technical knowledge, but also clear design principles.

The new body of research on adaptive multimodal input is going to build on the existing research on automatic input adaptation over the last decade. We therefore start by presenting a review of existing work from the field of automatic multimodal input adaptation and position the research activities in this body of work. This is based on solutions in related fields such as the adaptation of output modalities, multimodal mobile interaction and context-aware applications. The existing work is first compared and then analysed in a more detailed way based on three dimensions: the *combination of modalities*, *context awareness* and the *automated adaptation of input modalities*. Based on this analysis, we propose a set of eight high-level design guidelines for the creation of context-aware adaptive multimodal mobile applications. Finally, the development of an adaptive multimodal calendar application is discussed to illustrate the application of the aforementioned design guidelines.

We start by introducing some general concepts including multimodal interaction, adaptation to user and context as well as three related theoretical frameworks. This is followed by a presentation of the research papers that have been analysed as part of our review. The analysis of existing work starts by summarising a set of characteristics for each project. Three detailed analyses about the combination of modalities, context influence and the automatic adaptation of input modalities are presented. After outlining some promising future research directions, we provide some general conclusions.

BACKGROUND

The field of automatic input adaptation based on contextual information is positioned at the border of a number of different research fields. We provide an overview of the most important fields and introduce three theoretical frameworks: the CARE properties, the ISATINE framework and the Cameleon reference framework.

Multimodal Interaction

The strength of multimodal interfaces lies in the potential combination of different modalities. Beyond the improved expressive power by using multiple modalities, as illustrated by Richard Bolt's seminal "put-that-there" application in which a user combines speech and gesture to express complex commands, multimodal interfaces also improve the usability [2,17]. Input modalities are generally continuous probabilistic sources such as speech or gesture recognisers, as opposed to standard deterministic sources including keyboard or mouse input. The management of continuous probabilistic sources asks for parallel processing. Probabilistic sources and parallel processing in turn demand for distributed and time-sensitive software architectures [11].

CARE Properties

On a modelling level, the CARE properties defined by Coutaz and Nigay [6] illustrate how modalities can be composed in a multimodal application. Note that CARE stands for *complementarity*, *assignment*, *redundancy* and *equivalence*. Complementary modalities will need all modalities to reach a new application state, assigned modalities allocate one and only one modality to a particular operation, redundant modalities imply that all modalities can lead to the same operation and equivalent modalities assert that any of the modalities can be used to move to a new application state. The difference between redundancy and equivalence lies in how cases with multiple modalities occurring at the same time are dealt with. In the case of automatic input adaptation based on context information, equivalent or redundant modalities are essential and the appropriate modality for a given context will be selected or favoured.

Adaptation to User and Context

While adaptation has been investigated for traditional WIMP interfaces [5] or context-aware mobile interfaces [1], the work devoted to the automatic adaptation of input based on contextual information is rather limited. When considering user adaptation, different ways of adapting the user interface can be considered. On the one hand, the user interface can be created in such a way that it will adapt automatically and without any user intervention which is also the focus of this paper. On the other hand, we see solutions where users are given the possibility to modify the user interface in a proactive way. More formally, Malinowski et al. [16] proposed a complete taxonomy to describe user interface adaptation. Their taxonomy is based on four different stages of adaptation in a given interface, namely *initiative*, *proposal*, *decision* and *execution*. Each of these adaptations at the four different stages can be performed by the user, the machine or both. López-Jaquero et al. [15] proposed the ISATINE framework with seven stages of adaptation including the *goals for user interface adaptation*, the *initiative for adaptation*, the *specification of adaptation*, the *application of adaptation*, the *transition with adaptation*, the *interpretation of adaptation* as well as the *evaluation of adaptation*.

Cameleon Reference Framework

Calvary et al. [4] presented a model-based user interface framework for multi-target interfaces, allowing to describe design and run-time phases without taking into account specific implementation requirements. *Multi-target* refers to multiple contexts, whereas *context* denotes the context of use of an interactive system described in terms of a user, platform and environment model. The user model contains information about the current user, such as user preferences or limitations of disabled users. The platform model describes physical characteristics of the device the system is running on, including the size of the screen or the processor speed. Finally, the environment model contains information about social and physical attributes of the environment in which the interaction is taking place.

MULTIMODAL MOBILE INPUT SOLUTIONS

We now present the articles that we selected for our review of multimodal mobile input solutions. Related work in the field of context-aware automatic input adaptation is discussed in the section on Automatic Multimodal Input Adaptation. As we show, only limited research has been carried out in that field. In order to have a more complete overview, we also include related work in the field of user-driven multimodal input adaptation. Some of these solutions also illustrate how multimodal composition, the selection of modalities and context management have progressed in pervasive applications.

User-driven Multimodal Input Adaptation

We start by discussing the state of the art in the field of multimodal mobile input with a focus on solutions where environmental properties are taken into account as parameters to influence the selection of modalities. This includes well-known work that has been carried out in the domain of user-driven (non-automatic) context-aware input adaptation.

Lemmela et al. [14] presented an iterative method to design multimodal applications in mobile contexts. An interesting contribution of their approach is that they identified which modalities and combinations of modalities best suit different mobile situations based on the user's sensory channel load. Speech input was assigned as the default interaction technique for the car environment whereas 2D gestures (finger strokes) and motion gestures (tilt gestures) were used in the walking environment. In both scenarios, users had to read and write SMS messages while performing other activities. It has been pointed out that users prefer to use the speech input modality while being in the car context. De Sá et al. [8] describe a set of techniques and tools to support designers during the development of multimodal mobile applications. One of these tools related to context and environmental influence is a conceptual framework called *Scenario Generation and Context Selection Framework*. The framework aims to facilitate the process of scenario selection and generation in

a mobile setting. It relies on the analysis of a set of variables that might affect user interaction. Ronkainen et al. [24] introduced a conceptual framework called *Environment Analysis Framework* to perform a systematic environment analysis. The framework was built based on the analysis of previous work regarding mobile usage and context influence. The authors claim that the output of the framework can be used to guide the design of adaptive and/or multimodal user interfaces and devices optimised for certain environments.

Some projects also addressed how the influence of context and environmental properties in particular affect the user interaction and preferred modalities in different application domains. For instance, the map-based application domain has been analysed by several authors. Doyle et al. [10] conducted a review and analysis of existent map-based multimodal systems. They further proposed and evaluated a new multimodal mobile geographic information system (GIS) called Compass. Parameters such as the effectiveness and efficiency compared to unimodal interaction were evaluated. Ramsay et al. [20] proposed using motion gestures such as the forward and backward tilting of an external device for map navigation. The study focussed on evaluating user preferences and their perception of new input interaction techniques in comparison to traditional keypad navigation. A tourist guide application was developed to conduct a user evaluation in the field.

Other typical mobile outdoor activities like shopping, accessing different kinds of web services and form-filling tasks were also explored. For example, the goal of Wasinger et al.'s work [30] was to explore the use of tangible interaction as a complementary input modality for speech, 2D gestures on touch displays and pen gestures. They developed a proof of concept application to measure the intuitiveness as well as to evaluate a user's preferred combination of modalities in a shopping context. On the other hand, as part of the SmartWeb project, Sonntag et al. [28] proposed the use of natural language and multimodality as an interface to intuitively retrieve different types of information from web services. An important aspect of their system is its context-aware processing strategy. All recognised user actions are processed with respect to their situational and discourse context. Finally, Reis et al. [22] investigated the preferred user modality in different mobile environments. The authors presented a mobile multimodal application that allowed users to answer questionnaires and fill in information. The test was conducted in four environments including the home, a park, the subway and a car. The evaluated input modalities were 2D and pen gestures as well as speech and keyboard-based input. Their results highlight that in quiet environments without the presence of strangers or other disturbing factors, users were eager to experiment with new modalities.

Automatic Multimodal Input Adaptation

We now outline the state of the art in the field of mobile multimodal input with a focus on solutions where environmental properties are taken into account as parameters that influence the selection of modalities. The main features of the work presented in this section are summarised and further classified in Table 1.

As part of the SmartKom [23] research project, Bühler et al. [3] presented a prototype of a mobile version of the SmartKom system. The relevance of this paper is based on the introduction of a conceptual framework that deals with the flexible control of the interaction modalities as well as a new architecture for flexible interaction in mobile settings. Users as well as the system can initiate a modality transition between the default modalities. The conceptual framework describes five combinations of input and output modalities based on the user's level of attention in a car and a pedestrian setting, namely *default*, *listener*, *silent*, *car* and *speech only*. The adaptation mechanism is based on a set of predefined rules. For instance, an automatic input transition in the driver environment allowed the pen gesture input modality to be automatically switched off when the mobile device was connected to the docking station in a car. Likewise, when high levels of background noise were sensed in the pedestrian environment, the speech input modality was disabled.

Using the same rule-based approach, David et al. [7] proposed a mobile middleware that facilitates the development and maintenance of adaptive multimodal mobile interfaces. The middleware is built on the Android framework and is composed of two layers including a *Services* and a *Programming Language* layer. One of the novelties of the approach is that this library is based on the context-oriented programming paradigm presented in [26]. An instant messenger prototype has been built to illustrate the approach. The application allowed users to read and write SMS messages by using the keyboard or speech. Thereby, the input modes were adapted depending on the user's movements in a stressful condition such as cycling. For example, when the user was riding a bike, the default input modality was automatically set to speech and when the user stopped biking, the speech modality was deactivated.

In turn, Kong et al. [13] proposed a framework based on human-centric adaptation. In contrast to rule-based approaches, they quantify the user's average preference for a modality in a given interaction context. For instance, a dark environment might reduce a user's preference score for modalities related to the visual display. Adaptation can therefore be seen as searching for an optimal set of modalities with the highest preference score for a given scenario. The adaptation algorithm also verifies that the selected modality does not exceed the system resources. The application design and development process according to the proposed framework can be summarised in three steps. First, determine the tasks and available input/output modalities for a given device type. Then, the interaction scenarios as well as the interaction context should be determined. Last but not least, the designers have to evaluate the average preference score of a modality for a specific interaction context. To obtain this value, a survey with end users must be conducted. The results of the survey are finally used as input for a heuristic algorithm.

Zaguia et al. [31] presented an interesting approach for the detection of the optimal modality. They explored the domain of web service access using a context-based modality activation approach. The important part of their approach is that

Name	Modalities	Interaction Techniques	Interaction Sensors	Devices	Output Influence	CARE Properties	Environment Conditions	FMA	Adaptation	Application Domain
Buhler et al. 2002 [3]	- speech - pen gestures	- natural dialogue - pointing	- microphone - stylus	wireless computer (iPAQ) and laptop	**yes** - graphical - audio	**complementarity** natural dialogue + pointing (information about a place) **assignment** speech mode: natural dialogue silent/listener mode: pointing **equivalence** default mode: natural dialogue or pointing	**driving** social conditions: stress location: car **walking** physical conditions: high noise level social conditions: stressful social interaction location: street	**F**	**ME:** driving speed, state of brakes, location, noise level **AP:** rule based **AM:** switch default input mode	**Map & GIS** SmartKom
David et al. 2011 [7]	- speech - 2D gestures	- dictation - typing	- microphone - touch display	smartphone (Android)	**yes** - graphical - audio - vibration	**equivalence** typing, dictation (write message, check answer)	location: GPS location social conditions: stress (cycling)	**M**	**ME:** GPS speed **AP:** rule based **AM:** switch modalities	**Communication** SMS application
Kong et al. 2011 [13]	- speech - 2D gestures - pen gestures	(N/A)	- microphone - touch display - stylus	smartphone	**no** - audio - vibration	**assignment** library: 2D gestures **equivalence** outdoors: speech, 2D gestures shopping mall: 2D gestures, pen gestures	location: shopping mall, outdoors, library physical conditions: light level (bright, medium, dark), weather (sunny, cloudy, rainy), noise level (noisy, quiet)	**F**	**ME:** ambient light, noise level, weather, location, temperature **AP:** human centric (preference score) **AM:** switch current set of modalities	**Entertainment** social networking application
Zaguia et al. 2010 [31]	- speech - 2D gestures - indirect manipulation	- voice commands - single tap - typing - key press	- microphone - touch display - keyboard	wireless computer and laptop	**yes** - graphical - audio	**equivalence** select fields: key press, voice commands, single tap data entry: voice commands, typing	location: home, work, on the go physical conditions: light level (bright, dark, ...), noise level (noisy, quiet)	**A**	**ME:** noise level, location, light level **AP:** rule based **AM:** enable/disable preferred modality	**Services** flight reservation
Porta et al. 2009 [19]	- speech - 2D gestures - motion gestures	- natural dialogue - single tap - shake	- microphone - touch display - accelerometer	smartphone (iOS)	**no** - graphical - audio	**complementarity** single tap + natural language (find details of an item) **assignment** speech: ask for information shaking: undo task	location: relative location (device-mouth), outdoors social conditions: stress, time critical tasks	**A**	**ME:** device-mouth proximity **AP:** rule based **AM:** enable speech	**Services** B2B system
Turunen et al. 2009 [29]	- speech - motion gestures - indirect extra gestures - manipulation	- voice commands - tilt - touch item - key press	- microphone - accelerometer - NFC reader - keypad	smartphone (Symbian)	**yes** - graphical - vibration	**complementarity** vertical tilt + key press (move selection), horizontal tilt + key press (zoom options) **equivalence** speech, extra gestures (menu options)	location: relative location (device-mouth) physical conditions: light level (very dark)	**A**	**ME:** device-mouth proximity **AP:** rule based **AM:** enable speech modality	**Entertainment** media centre

Table 1. Summary of research on context-sensitive automatic input adaptation. The Output Influence column lists all output modalities together with information whether the output is adapted. In the column Adaptation, 'ME' stands for monitoring entity, 'AP' for adaptation policy and 'AM' for adaptation mechanism.

it ensures that the invoked modalities are suited for a user's current situation. The system, or more specifically a dedicated context interaction agent, is in charge of the detection of context information as well as the selection of the optimal modality. This value is calculated based on the evaluation of factors related to the interaction context. Interaction context refers to *user context* (e.g. regular user, deaf, mute, manually handicapped or visually impaired), *environmental context* (e.g. noise level or the brightness at the workplace) and *system context* (e.g. capabilities of the computing device). The approach was further illustrated with a ticket reservation system. The application allowed users to reserve a ticket based on the optimal modalities provided by the system and potential transitions were depicted by Petri net diagrams.

Porta et al. [19] investigated the use of multimodal input in the domain of decision support in order to ease the access of information in time critical situations. A business-to-business (B2B) system that supported 2D gestures, motion gestures as well as speech input modalities was developed. In this application, the speech and 2D gesture modalities were used in a complementary way to search information, whereas motion gestures (e.g. shaking the device) were used for implementing an undo operation. In this approach, the system automatically switched on the speech modality when the user moved the arm holding the device closer to their mouth.

Within the entertainment domain, Turunen et al. [29] presented a multimodal media centre interface. The interface allowed users to interact using speech input, extra gestures (tangible interaction) and motion gestures. In the same fashion as Porta et al., the speech modality was automatically switched on when a user moved the device close to their mouth.

ANALYSIS

Our analysis is based on the articles introduced in the previous section. Three different dimensions are observed: the *combination of modalities*, *context awareness* and *automatic adaptation*. The first two dimensions serve as a conceptual basis to understand common multimodal composition patterns as well as to analyse the suitability of modalities according to context variations. These aspects are paramount during the design phase of an effective context-aware adaptive multimodal application. Note that, to a lesser extent, Gentile et al. [12] also explored features of adaptive systems.

Combination of Modalities

All of the fourteen articles except one, explored different equivalent and complementary combinations of modalities. The remaining article, Ronkainen et al. [24], investigated the assignment of one specific input modality to perform a particular task. For instance, the authors studied the use of the

	Speech	2D Gestures	Motion Gestures	Pen Gestures	Indirect Manipulation	Extra Gestures
Speech		E [7,8,13,14,22,31], C [19,30]	E [29]	E [3,8,10,22], C [3,8,10,28,30]	E [8,22,31]	E [29], C [30]
2D Gestures	E [7,8,13,14,22,31], C [19,30]		E [14]	E [8,13,22,30]	E [8,22,31]	E [30]
Motion Gestures	E [29]	E [14]			E [20], C [29]	
Pen Gestures	E [3,8,10,22], C [3,8,10,28,30]	E [8,13,22,30]			E [8,22]	E [30]
Indirect Manipulation	E [8,22,31]	E [8,22,31]	E [20], C [29]	E [8,22]		
Extra Gestures	E [29], C [30]	E [30]		E [30]		

Table 2. Summary of the combination of modalities with the corresponding citations ('E' stands for equivalence and 'C' for complementarity)

double tap gesture at the back of the device as an alternative technique to silence the device or to start speech input recognition. It is important to notice the influence of the current input modality in the selection of the output modality. The discussion of output modalities is out of the scope of this analysis, however this specific relationship has been reviewed. The results showed that in seven out of fourteen articles [3,7,8,14,24,29,31] the selected input modality influenced the output modality. It is also worth highlighting that only in Wasinger et al.'s [30] work the use of redundancy was observed. A summary of the combinations of different modalities is provided in Table 2.

Interestingly, the results show that speech input modality is present in around 86% of the articles with the exception of [20,24]. Voice commands consist mainly of short phrases containing a few words and matched against a grammar of rules. This interaction technique has been mainly used in conjunction with pointing-based techniques. In this way, it was possible to ask the system for information related to an element pointed to by the user. Twelve out of the fourteen research projects made use of the touch displays of recent smartphones or PDAs. Modalities associated with touch displays are 2D gestures and pen gestures, with an appearance of approximately 64% and 50% in the reviewed papers. The results further reveal that the interaction techniques used by the two modalities are very similar. In general, they were used for pointing tasks, specifically to select items in the interface. Gestures based on gesture recognition techniques are mostly used to perform navigational or atomic tasks. For instance, navigation gestures like tilting the device up, down, left or right have been observed in three out of four articles [14,20,29]. In turn, atomic actions are mostly performed using a "shake gesture". For instance, Porta et al. [19] used this gesture to specify an undo action. The advantage of this type of gestures is the support for one-handed interaction. Moreover, the level of attention to the device is also reduced. Last but not least, the usage of extra gestures through the interaction with tangible objects has been only explored by Wasinger et al. [30] and Turunen et al. [29]. In the former, RFID tags have been attached to products in a store and *pick up* and *put back* user actions were evaluated to detect whether the objects were either in or out of the shelf. In the latter, Turunen et al. mapped the main options of the system to an A3 control board tablet that—via labelled RFID tags—stored links to menu options. In both articles, user evaluations indicate a good acceptance by participants.

Context Awareness

Let us now address the suitability of the input modalities according to specific context settings. This analysis heavily relies on the articles that conducted user studies in real-world settings [10,14,20,22]. Likewise, studies conducted in laboratory settings [13,29,30] are also taken into account, because they evaluate the preferability or suitability of modalities in specific contexts. Finally, the context analysis performed in four articles [3,8,24,31] are an additional source for the presented analysis.

In general, users feel comfortable using all modalities in private places, where social interaction as well as noise levels are low. This was observed in Zaguia et al.'s [31] work, where all modalities were categorised as appropriate for the semantic location *home* with low levels of noise and a well illuminated environment. However, in public environments the results were different. In Reis et al.'s work [22], a study performed in the wild, specifically in the four real-world settings of a home, a park, the subway and a car showed that in the home environment, users are eager to experiment with modalities they found interesting. The results in this setting highlighted that all evaluated users applied voice interaction for selecting menu options as well as for data entry. Note that the behaviour was not the same in the other three environments.

When ideal environmental conditions change, a user's preferred modality varies as well. Table 3 provides a summary of the suitability/preferability of each modality under specific environmental settings. The observed environment variables are taken from the environment model described in the Cameleon reference framework [4].

Environment Variables		Speech	2D / Pen / Indirect Manipulation	Motion Gestures	Extra Gestures
physical condition	brightness	B/D [31]	B [31]	B/D [29]	N/A
	noise level	L [10,31]	L/H [31]	N/A	N/A
social condition	stress	H [3,14,22]	L [8,22,24]	M [14,20]	N/A
	social interaction	L [22]	M/H [14,20,22]	M [14,20]	N/A
location	semantic location	I [10,22,30,31], O [3,13,14,22]	I [13,14,24,30,31], O [13,20,30]	I [14], O [20]	I [29,30], O [30]

Table 3. Suitability of modalities for different environmental conditions ('B' stands for bright, 'D' for dark, 'L' for low, 'M' for medium, 'H' for high, 'I' for indoors and 'O' for outdoors)

Automatic Adaptation

Based on the references presented in Table 1, an analysis of the automatic adaptation core features is presented. First, we provide an analysis of the *monitoring entity* component. Therefore, we reviewed the entities that monitor and start the adaptation mechanism in the reviewed articles. Relying on the Cameleon framework's environment property classification (physical conditions, location and social conditions), it was possible to analyse which specific environment variable was used for which task. Furthermore, we analysed the type of *adaptation policy* and *adaptive mechanism* technique observed in different applications. Table 4 provides a summary

of the investigated articles in terms of the core features for automatic adaptation.

Monitoring Entity

This section focusses on the conditions these types of systems adapt to and the used environment variables. From Table 4, one can observe that only physical and location-based variables were used as adaptation triggers. Interestingly, none of the reviewed papers uses methods for social cue detection based on built-in mobile sensors or other techniques able to detect social interaction. In the following, detailed information about the *physical conditions* and *location* variables that influenced the adaptation is provided.

Physical Conditions: The monitoring entities in this category are mostly built-in sensors that constantly sense changes in physical conditions like noise and light level, weather or acceleration. Large variations in these values lead to a possible input modality adaptation. For instance, in three articles the usage of noise level sensing [3,13,31] was observed. However, this information is mostly used in conjunction with information derived from other sensors. Apart from sensing noise level variations, Bühler et al. [3] and David et al. [7] sensed acceleration variations to identify when an object stands still or is moving. Finally, only in the framework proposed by Kong et al. [13], variables such as temperature, weather and time were considered as possible adaptation triggers. The authors mentioned that any change in the values of these variables trigger the adaptation algorithm.

Location: The monitoring entities observed in this category are clustered in three groups, namely *relative*, *absolute* and *semantic location*. Relative location refers to the location of the device or user in relation to another point of reference. On the other hand, an absolute location represents the exact geographic position consisting of latitude and longitude coordinates obtained from the GPS sensor. As previously mentioned, only David et al. [7] used this information when the speed value from the GPS (Global Positioning System) was not available. Finally, three articles [3,13,31] use semantic locations as adaptation triggers. Since semantic location does not change frequently, dynamic variables such as the noise or light level can be associated with each location in order to be used in the adaptation process. The use of the location in combination with the noise level can, for example, improve the usability and user acceptance.

Adaptation Policy and Adaptive Mechanism

We now describe how the input adaptation takes place in terms of the *adaptation policy* and the *adaptation mechanism* components. With regards to the adaptation policy component, we analysed which decision inference mechanism (rule-based or heuristic algorithm) was used among the reviewed articles. Regarding the adaptation mechanism, we analysed the possible modifications that end users could perceive after an input modality adaptation was triggered. Hence, we reviewed whether the modifications occurred by enabling and disabling a modality or by switching between a set of modalities.

The rule-based approach refers to simple logical rules that indicate when the adaptation has to take place as well as which kind of adaptation should take place. One can easily notice that all reviewed articles with the exception of Kong et al.'s work [13] followed this approach. On the other hand, three articles [3,13,31] used the *switching technique* adaptation mechanism. For instance, in Bühler et al. [3] this technique was used to switch between the five input interaction modes. Each interaction mode encompassed a set of allowed and suitable modalities for specific contextual situations. When a rule was satisfied, the suitable interaction mode was activated. For instance, one rule specified that if the user's current location was different from car, the default interaction mode that relied on speech and pen gestures should be activated. In turn, when the noise level was sensed too high, the system switched to a *listener* mode and deactivated the speech input modality. Similarly, Zaguia et al.'s work [31] relied on a more complex set of rules to obtain the set of suitable modalities for a particular context and device. Similarly, the heuristic algorithm presented by Kong et al. [13] selects the set of modalities that achieves a maximum preference score.

As mentioned before, Kong et al. [13] was the only article that reviewed another approach in regard to the adaptation policy component. The authors argued that rule-based approaches face some issues, such as not covering all interaction scenarios or rules that are conflicting with each other. They therefore proposed a human-centric adaptation approach using a heuristic algorithm. Hence, in this context, the objective of the adaptation algorithm was to find a set of modalities that achieve a maximum preference score.

GUIDELINES

During the analysis phase, we noticed that the field of automatic input adaptation has barely been addressed. The analysis of user-driven adaptation showed that existing research efforts provide the necessary conceptual basis to systematically design a context-aware adaptive application. However, these concepts have only been taken into consideration by two articles [13,31].

Based on the reviewed articles and the results outlined in the analysis section, we therefore propose a set of guidelines that can be used to design context-aware adaptive multimodal mobile applications. Note that the presented guidelines are further based on work in the field of mobile context-aware systems by Reeves et al. [21], Schiefer et al. [27], Dey et al. [9] as well as Rothrock et al. [25]. These guidelines unify the key aspects and stages identified in the investigated articles. It is also worth mentioning that our guidelines are targeting the automatic adaptation of input channels. We organised the guidelines into three main phases: *context and modality suitability analysis*, *multimodal task definition* and *adaptation design*.

Context and Modality Suitability Analysis

Prior to the multimodal and adaptive design, the influence of environmental factors should be evaluated. In this phase, two

Monitoring Entity			Adaptation Policy		Adaptive Mechanism	
Physical Conditions	**Location**	**Social Conditions**	**Rule Based**	**Heuristic Algorithms**	**Enabling**	**Switching**
acceleration [3,7], noise level [3,13,31], light level [13,31], temperature [13], weather [13], time [13]	relative position [19,29], absolute location [7], semantic location [3,13,31]	(N/A)	[3,7,19,29,31]	[13]	[3,7,19,29]	[3,13,31]

Table 4. Core features for automatic adaptation

activities are important: conduct a context analysis and define suitable modalities for each semantic location.

(G1) Conduct a context analysis: It is important to define in advance the semantic locations (e.g. park, car, street or office) where the user is mostly going to interact with the application. After defining these locations, a context analysis should be conducted to have an overall picture of the environment factors that influence each location. In particular, for each location, the influence of environment variables such as noise level, social interaction or stress should be analysed. Then, qualitative values like *low*, *medium*, *high* should be assigned to each location/environment variable pair. The exact quantification of *low/medium/high* is left to the designer, as fewer or more values can be used depending of the required granularity. This guideline is derived from the location part of the analysis section. It also supports the analysis of mobile scenarios as reviewed in a number of conceptual frameworks [8,13,14,24].

(G2) Define suitable modalities for each semantic location: Based on the physical and social conditions assigned to the semantic locations, the designer should specify which modalities are appropriate for each location. This will lead to the selection of the modalities that will be supported by the application. To achieve this, the usability of each location/modality pair should be evaluated. For example, if the context analysis of the semantic location *street* outputs the values *high* for the noise level, *medium* for the social interaction and *medium* for the stress level, then the speech modality is labelled with a *difficult* value and the modality is considered as not suitable. As a rule of thumb, the designer should strive to keep most modalities enabled at all times. Note that this guideline is again derived from the location part of the analysis section as well as Reeves et al. [21] and Schiefer and Decker [27].

Multimodal Task Definition

After defining the modalities that the application will support, this phase encompasses the design of the multimodal input channels. It includes two guidelines, namely the selection of multimodal tasks and the definition of equivalent input modalities. Both guidelines are derived from the combination of modalities part of the analysis section.

(G3) Select tasks that will offer multimodal behaviour: It is important to specify which tasks in the application will support a multimodal behaviour. Thereby, ideal multimodal tasks are frequently used tasks, error-prone tasks or tasks that involve a complex process [21,27].

(G4) Define equivalent modalities for the multimodal tasks: For any given multimodal task, an interaction technique must be specified for each available modality. In this way, the user can perform the same task using any of the supported modalities [21,27].

Adaptation Design

Based on the context and modality analysis information obtained from the previous stages, the designer should now specify the design of the adaptation process. In this phase, three aspects have to be taken into account, namely the definition of adaptation triggers and monitoring entities, the definition of the adaptation policy and modalities and context status feedback. These guidelines are derived from the monitoring entity and adaptation policy and adaptive mechanism parts of the analysis section.

(G5) Define adaptation triggers and monitoring entities: In this step, it must be clear for the designer which type of environment variables will influence the adaptation (physical conditions, social conditions, location). Likewise, it should be specified how the environment data related to the variables is going to be captured. Moreover, it should be specified how this sensor data is mapped to meaningful information for the application [9,25].

(G6) Define adaptation policy mechanism based on context analysis: Independently of the selected adaptation policy mechanism (e.g. rule based or using an heuristic algorithm), the assigment of input modalities according to the changes in the environment values have to be defined. These design decisions should take the context analysis and multimodal task definition of the previous stages into account [25].

(G7) Define adaptation mechanisms: In this type of adaptation, the designer should decide between two possible adaptation mechanism: enable/disable one specific modality or automatically switch a set of modalities [25].

(G8) Provide modality and context status feedback: The designer has to provide means to display the available modalities and the current context status. This contextual status and the set active modalities should be visible for the user at any time in an unintrusive manner [9,21,27].

USE CASE APPLICATION

The Multimodal Adaptive Agenda (MAA) is a proof of concept application that was developed to illustrate the guidelines. MAA is a mobile calendar application built on Android offering adaptive multimodal functionality based on form-filling tasks. We first describe the application itself and then illustrate how our guidelines were used to define the application behaviour of.

Application Description

The interface of the MAA application is shown in Figure 1. The application looks and feels like a standard calendar application, with the possibility to add new events, modify and

cancel them through classic Android GUI elements. However, a field has been added on the lower section of the screen, allowing the user to input commands through gestures on the screen. For example, a gesture comparable to the iPhone's "slide to unlock" gesture can be used to quickly cancel an appointment which otherwise would have been wrongly created by the user. Similarly, the signal from the accelerometers is used to allow the user to navigate between days and months in the calendar by tilting the phone to the left and right. Finally, Near Field Communication (NFC) tags can be used to create a new event (e.g. for a particular meeting). The tangible tag can be passed to other people in order that they can seamlessly create the same event in their calendar. Please note that all these functions can also be achieved through the GUI buttons and widgets of the interface.

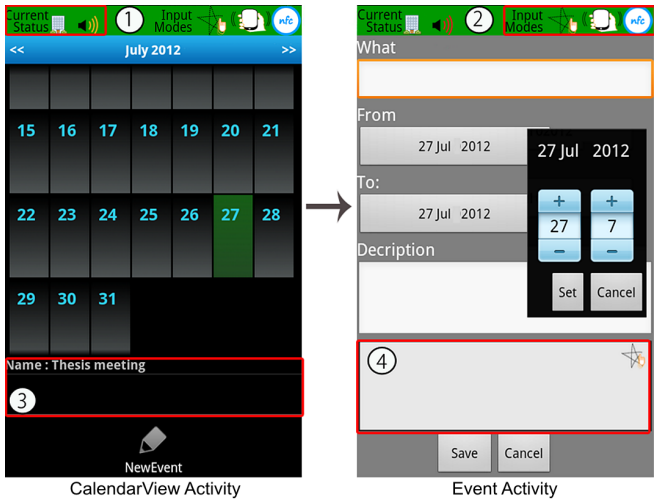

Figure 1. MAA user interface

As the permitted modalities are adapted on the fly, based on information from the surroundings, the interface also displays the detected environment along with a list of enabled and disabled modalities. Figure 2 displays four different cases of suitable modalities for the semantic indoor location and different variations of the noise level. As can be seen in the last case, when the system detects that the user is walking, it is no longer possible to interact using motion gestures, which helps to avoid wrong input due to movement.

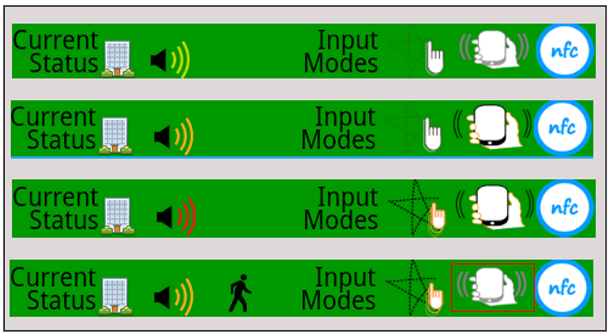

Figure 2. Suitable modalities for different indoor noise levels

Analysis and Design

Following the three phases described in the our design guidelines, in the following we describe the supported input modalities, interaction techniques and adaptation rules.

Context and Modality Suitability Analysis

The first step in the design of the MAA was the context analysis of the locations where users would most likely use the application. Two indoor locations (i.e. home and work) as well as an outdoor location (i.e. street) were evaluated. Then, the influence of three environment variables for each location was investigated. The environment variables *noise level*, *social interaction* and *stress level* were selected for evaluating each location. During the analysis, it was noticed that the variations in these three parameters influenced the use of a modality in a specific setting. In consequence, the level of influence was assigned using a high/medium/low interval scale. Table 5 illustrates the results of this analysis, with the values providing an overview of the influence of context.

Location	Noise Level	Social Interaction	Stress Level
home (sitting)	low	low	low
work (sitting at desk)	medium	medium/high	medium/high
street (walking)	medium/high	medium	medium

Table 5. Context analysis for the design of MMA

This information serves as a basis to evaluate how easy or difficult it will be to use a specific modality at a particular location. The modalities that should be analysed are the ones supported by the device on which the application will run. Since we decided that the application should be used on modern smartphones, the available input modalities are *speech* through the built-in voice recognition engine, *2D gestures* using multi-touch displays, *motion gestures* using the built-in accelerometer sensors and *extra gestures* using the built-in NFC reader of the smartphone.

The suitability level of each input modality was evaluated for each semantic location. For this analysis, a three level qualitative scale (*easy*, *medium*, *difficult*) was used. If the interaction appeared to be too difficult for a particular modality/location pair, the modality was considered as not suitable. In turn, when the modalities were evaluated with an *easy* value, the modality was considered as suitable. This information was used to take decisions regarding the modalities and interaction techniques to be used. For instance, we noticed that the use of speech was only appropriate within the home environment. We therefore decided not to include it as a supported modality in our proof of concept application. Finally, based on this analysis, we decided that the application should support 2D gestures (single tap and symbol drawing), motion gestures and extra gestures through the NFC reader.

Multimodal Task Definition

The functional requirements specify all the functionality supported by a system. Taking into account guideline G3, we selected the five tasks shown in Table 6 to support multimodal behaviour. The table further highlights which modalities can be used to interact with a given task as required by guideline G4. Please note that at least two equivalent modalities have been defined for each task.

Task	2D Gestures		Motion Gestures	Extra Gestures
	Single Tap	Symbol Drawing		
create new event	'new' button		'shake' gesture	bring an NFC tag close
save event	'save' button	'checkmark' symbol	'shake' gesture	
cancel event creation	'cancel' button	'line' symbol	'face down' gesture	
move between calendar months	'left/right' button		'flick left/right' gesture	
change day (date dialogue)	'+' or '-' button		'flick up/down' gesture	

Table 6. Supported input modalities and interaction techniques

Adaptation Design

As shown in Table 6, the user is able to interact with the application using four interaction techniques. However, users might feel overloaded by having all the input modalities available at once and having to decide which one to use. To address this issue, the application defines a default modality (2D gestures) and incorporates supporting modalities according to the context conditions. As recommended in guideline G5, the possible adaptation triggers for our application were identified. The location of the user was used as a first adaptation trigger. Noise level variations were also considered, as an indication of how stressful the environment is. Finally, taking into consideration guideline G6, the different transitions from one modality to the other were specified. This information is the conceptual basis to define context rules. Based on guideline G7, we settled on activating and disabling individual modalities. Finally, guideline G8 has been addressed by displaying the available and unavailable modalities at the top of the screen, as illustrated in Figure 1. With the tasks, their assigned modalities, the transitions between modalities and the full set of context rules defined, the concrete implementation of the MAA application followed a clear path.

The multimodal calendar application was designed from the ground to illustrate the application of our design guidelines for adaptive multimodal mobile applications. As such, its functionality is rather limited and it has to be seen as a test case to demonstrate the feasibility of the guidelines. The fact that modalities are either completely enabled or disabled can be seen as too radical. In a real-world setting, the complete disabling of a modality mainly makes sense for security, safety or legal reasons. In other cases, the application might favour some modalities over others instead of completely disabling them. While the presented calendar application clearly offers room for improvements, it serves as a first step towards a more detailed evaluation of the guidelines that we plan to conduct.

CONCLUSION

We have outlined how the automatic adaptation of input can lead to enhanced usability by providing alternative equivalent modalities, adjusting these modalities to environmental conditions and thereby improving an individual's performance in using the mobile user interface. The class of adaptable user interfaces is tightly linked to advances in multimodal interaction, including the fusion and temporal combination of input modalities.

Our detailed analysis of the existing body of work led to a set of core features that are necessary for automatic input adaptation, including multiple available equivalent modalities (or combination of modalities), rich metadata about the given context, user and device, well-designed feedback for the user as well as user control mechanisms. Our exploration of the field of automatic input adaptation helps to combine new modalities as well as managing richer contextual information about the user and their environment. In turn, these core features have been exploited to propose eight design guidelines for automatic adaptation in multimodal mobile applications. Last but not least, we have illustrated the application of the guidelines in the design and development of an adaptive multimodal mobile agenda application

ACKNOWLEDGMENTS

Bruno Dumas is supported by MobiCraNT, a project forming part of the Strategic Platforms programme by the Brussels Institute for Research and Innovation (Innoviris).

REFERENCES

1. Arhippainen, L., Rantakokko, T., and Tähti, M. Navigation with an Adaptive Mobile Map-Application: User Experiences of Gesture- and Context-Sensitiveness. In *Proc. of UCS 2004* (Tokyo, Japan, November 2004), 62–73.
2. Bernsen, N. O., and Dybkjær, L. *Multimodal Usability*. Springer, 2009.
3. Bühler, D., Minker, W., Häussler, J., and Krüger, S. The SmartKom Mobile Multi-Modal Dialogue System. In *Proc. of ITRW 2002* (Irsee, Germany, June 2002).
4. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., and Vanderdonckt, J. A Unifying Reference Framework for Multi-Target User Interfaces. *Interacting with Computers 15*, 3 (2003), 289–308.
5. Cesar, P., Vaishnavi, I., Kernchen, R., Meissner, S., Hesselman, C., Boussard, M., Spedalieri, A., Bulterman, D. C., and Gao, B. Multimedia Adaptation in Ubiquitous Environments: Benefits of Structured Multimedia Documents. In *Proc. of DocEng 2008* (São Paulo, Brazil, September 2008), 275–284.
6. Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., and Young, R. M. Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties. In *Proc. of INTERACT 1995* (Lillehammer, Norway, June 1995), 115–120.
7. David, L., Endler, M., Barbosa, S. D. J., and Filho, J. V. Middleware Support for Context-Aware Mobile Applications with Adaptive Multimodal User Interfaces. In *Proc. of U-Media 2011* (São Paulo, Brazil, July 2011), 106–111.
8. de Sá, M., and Carriço, L. Lessons From Early Stages Design of Mobile Applications. In *Proc. of MobileHCI 2008* (Amsterdam, The Netherlands, 2008), 127–136.
9. Dey, A. K., and Häkkilä, J. Context-Awareness and Mobile Devices. *User Interface Design and Evaluation for Mobile Technology 1* (2008), 205–217.

10. Doyle, J., Bertolotto, M., and Wilson, D. A Survey of Multimodal Interfaces for Mobile Mapping Applications. In *Map-based Mobile Services*, Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg, 2008, 146–167.
11. Dumas, B., Lalanne, D., and Oviatt, S. Multimodal Interfaces: A Survey of Principles, Models and Frameworks. *Human Machine Interaction: Research Results of the MMI Program* (March 2009), 3–26.
12. Gentile, A., Santangelo, A., Sorce, S., Augello, A., Pilato, G., Genco, A., and Gaglio, S. Exploiting Multimodality for Intelligent Mobile Access to Pervasive Services in Cultural Heritage Sites. *Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability* (2009), 217–240.
13. Kong, J., Zhang, W. Y., Yu, N., and Xia, X. J. Design of Human-Centric Adaptive Multimodal Interfaces. *International Journal of Human-Computer Studies 69*, 12 (December 2011), 854–869.
14. Lemmelä, S., Vetek, A., Mäkelä, K., and Trendafilov, D. Designing and Evaluating Multimodal Interaction for Mobile Contexts. In *Proc. of ICMI 2008* (Chania, Greece, October 2008), 265–272.
15. López-Jaquero, V., Vanderdonckt, J., Montero, F., and González, P. Towards an Extended Model of User Interface Adaptation: The Isatine Framework. In *Engineering Interactive Systems*, vol. 4940 of *LNCS*. Springer, 2008, 374–392.
16. Malinowski, U., Thomas, K., Dieterich, H., and Schneider-Hufschmidt, M. A Taxonomy of Adaptive User Interfaces. In *Proc. of HCI 1992* (York, UK, September 1992), 391–414.
17. Oviatt, S. Multimodal Interactive Maps: Designing for Human Performance. *Human-Computer Interaction 12*, 1 (March 1997), 93–129.
18. Oviatt, S., and Lunsford, R. Multimodal Interfaces for Cell Phones and Mobile Technology. *International Journal of Speech Technology 8* (2005), 127–132.
19. Porta, D., Sonntag, D., and Nesselrath, R. New Business to Business Interaction: Shake Your iPhone and Speak to It. In *Proc. of MobileHCI 2009* (Bonn, Germany, September 2009).
20. Ramsay, A., McGee-Lennon, M., Wilson, G., Gray, S., Gray, P., and De Turenne, F. Tilt and Go: Exploring Multimodal Mobile Maps in the Field. *Journal on Multimodal User Interfaces 3* (2010).
21. Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J.-C., McTear, M., Raman, T., Stanney, K. M., Su, H., and Wang, Q. Y. Guidelines for Multimodal User Interface Design. *Communications of the ACM 47*, 1 (January 2004), 57–59.
22. Reis, T., de Sá, M., and Carriço, L. Multimodal Interaction: Real Context Studies on Mobile Digital Artefacts. In *Haptic and Audio Interaction Design*, vol. 5270. Springer, 2008, 60–69.
23. Reithinger, N., Alexandersson, J., Becker, T., Blocher, A., Engel, R., Löckelt, M., Müller, J., Pfleger, N., Poller, P., Streit, M., and Tschernomas, V. SmartKom: Adaptive and Flexible Multimodal Access to Multiple Applications. In *Proc. of ICMI 2003* (Vancouver, Canada, 2003), 101–108.
24. Ronkainen, S., Koskinen, E., Liu, Y., and Korhonen, P. Environment Analysis as a Basis for Designing Multimodal and Multidevice User Interfaces. *Human–Computer Interaction 25*, 2 (2010), 148–193.
25. Rothrock, L., Koubek, R., Fuchs, F., Haas, M., and Salvendy, G. Review and Reappraisal of Adaptive Interfaces: Toward Biologically Inspired Paradigms. *Theoretical Issues in Ergonomics Science 3*, 1 (2002), 47–84.
26. Salvaneschi, G., Ghezzi, C., and Pradella, M. Context-Oriented Programming: A Software Engineering Perspective. *Journal of Systems and Software 85*, 8 (August 2012), 1801–1817.
27. Schiefer, G., and Decker, M. Taxonomy for Mobile Terminals – A Selective Classification Scheme. In *Proc. of ICE-B 2008* (Porto, Portugal, July 2008), 255–258.
28. Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., and Reithinger, N. Smart Web Handheld – Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services. In *Artifical Intelligence for Human Computing*, vol. 4451 of *Lecture Notes in Computer Science*. Springer, 2007, 272–295.
29. Turunen, M., Kallinen, A., Sànchez, I., Riekki, J., Hella, J., Olsson, T., Melto, A., Rajaniemi, J.-P., Hakulinen, J., Mäkinen, E., Valkama, P., Miettinen, T., Pyykkönen, M., Saloranta, T., Gilman, E., and Raisamo, R. Multimodal Interaction with Speech and Physical Touch Interface in a Media Center Application. In *Proc. of ACE 2009* (Athens, Greece, October 2009), 19–26.
30. Wasinger, R., Krüger, A., and Jacobs, O. Integrating Intra and Extra Gestures into a Mobile and Multimodal Shopping Assistant. In *Pervasive Computing*, vol. 3468 of *Lecture Notes in Computer Science*. Springer, 2005, 323–328.
31. Zaguia, A., Hina, M., Tadj, C., and Ramdane-Cherif, A. Interaction Context-aware Modalities and Multimodal Fusion for Accessing Web Services. *Ubiquitous Computing and Communication Journal 5*, 4 (December 2010).