MULTIMODAL INTERACTION WITH MOBILE DEVICES
Outline of a Semiotic Framework for Theory and Practice

Gustav Öquist
Department of Linguistics and Philology, Uppsala University, Sweden

Keywords: Mobile devices, Interaction, Multimodality, Framework, Semiotics.

Abstract: This paper explores how interfaces that fully uses our ability to communicate through the visual, auditory, and tactile senses, may enhance mobile interaction. The first step is to look beyond the desktop. We do not need to reinvent computing, but we need to see that mobile interaction does not benefit from desktop metaphors alone. The next step is to look at what we have at hand, and as we will see, mobile devices are already quite apt for multimodal interaction. The question is how we can coordinate information communicated through several senses in a way that enhances interaction. By mapping information over communication circuit, semiotic representation, and sense applied for interaction; a framework for multimodal interaction is outlined that can offer some guidance to integration. By exemplifying how a wide range of research prototypes fit into the framework today, it is shown how interfaces communicating through several modalities may enhance mobile interaction tomorrow.

1 INTRODUCTION

Mobile devices have inherited a large body of principles regarding how to interact with computers from the desktop paradigm. This is natural since most developers of mobile interfaces have a background in desktop interface design, whereas the users generally have thought of handheld computers and cellular phones as an extension of the office domain. Mobile devices do however differ in many respects from desktop computers. As a result we see that although the computational power of mobile devices are ever increasing, the two main constraints that reduce usability remains. The first is the limited input capabilities and the second is the limited output capabilities, both caused by a combination of the user's demand for small devices and the developer's reuse of desktop interaction methods. However, it is vital to see that small display and keyboard sizes are not elective; they are decisive form factors since mobile devices have to be small to be mobile.

Nonetheless, size is not all, and mobile devices do have many beneficial properties that may enhance interaction. For example, they usually have good information processing capabilities. Yet, the opportunities offered by having devices that can process and display information in ways that suit the small screen better, moreover with respect to whom is using it and where, have not been widely employed. Moreover, since mobile devices are strongly associated with cellular phones, they are much more socially acceptable to speak with than desktop computers. Yet, the opportunities offered by having a device that you can verbally interact with have not been widely put into practice. Furthermore, mobile devices lack physical confinement, probably the foremost reason for any customer to buy one in the first place. Yet, the opportunities offered by having a device that you can hold in your palm and freely interact with in space, and in relation to other devices, have not been widely put to use.

To be able to make more of these promising properties, and combine them into useful interfaces for multimodal interaction, we need models for reasoning around how to interact through several senses. A natural starting point is to look at how humans use multimodal communication and this is where we begin in the next section. Based on our observations we proceed by outlining a model of multimodal interaction based on three tiers: Communication circuit, semiotic representation, and sense applied for interaction. We then exemplify how a range of research prototypes fit into the framework. Finally, a discussion and a few concluding remarks wrap up the paper.

Öquist G. (2006).
MULTIMODAL INTERACTION WITH MOBILE DEVICES - Outline of a Semiotic Framework for Theory and Practice.
In *Proceedings of the International Conference on Wireless Information Networks and Systems*, pages 276-283

2 MULTIMODAL COMMUNICATION

Perception is both the key and keyhole for communication since it both enables and restrains acquisition and further interpretation of information. From a human centred perspective, a modality can essentially be seen as one of the senses we utilize to make ourselves aware of the world around us. If we stick with Aristotle's traditional categorization, we have vision, hearing, touch, smell, and taste. Each of these senses can be used to perceive information that is quite different in nature, but it is also often possible to perceive the same information through several senses at once.

The form of communication we are interested in is the interactive where the sender and receiver intentionally and actively transfer intelligible information between each other with the aim of achieving a mutually understood goal. Information can be seen as the raw material for message construction and the exchange of meaning. Some sort of coding is always a part of the creation of information since meaning cannot be delivered through any given medium in its pure form (that would equal mind reading).

This is where semiotics or the theory and study of signs come into play, since information is what we decipher from signs (Chandler, 2001). However, we do not really produce signs, we produce stimulus, neither do we perceive signs, we perceive stimulus. By encoding meaning into signs, the sender shapes information into the form of stimulus that the receiver is assumed to perceive and decode as the signs conveying the original meaning. This means that the sender must be able to anticipate what the receiver is going to recognize the stimuli as, an anticipation that is based on situational, social, and cultural conventions.

Since humans can communicate more efficiently through several senses, it seems straightforward that humans should communicate more efficiently with mobile devices through several senses. Yet, there are several questions that arise in the wake of this assumption. Do mobile devices really have what it takes? How should sight, hearing, and touch be combined in a way that actually enhances interaction? In order to find some answers to these questions, we will now have a closer look at what multimodal interaction implies by outlining a framework. The framework in itself is not limited to mobile devices but in the scope of this paper, we will focus on how we can interact with mobile devices using several senses.

3 A FRAMEWORK MULTIMODAL INTERACTION

The challenge with multimodal interaction is to channel the right information through the right sense in the right way. We will attempt to offer some guidance to this by structuring the communicative situation where a user interacts multimodally with a device into three tiers. Each tier can be thought of as a level of reasoning that is interleaved with the others, thus it is not a layer in a strict sense as such can be pealed off and viewed in isolation. A tier is more like something that binds other things together, and in the case, they bind our model of multimodal communication together.

The first tier is the circuit of communication that defines how the information can flow between the sender and receiver through interaction. The second tier is the form of information that governs on how meaning is represented as signs used in the interaction. The third, and last, tier is the mode of interaction that categorizes the communication depending on the modality that is used for transfer of the information. Let us now examine each tier in turn and then see how they fit together.

3.1 Circuit of Communication

The first tier has to do with the relation between the participants in the interaction and how information is communicated between them. Interaction implies at least two communicative participants that we will refer to as the sender and the receiver. Since we are interested in interaction with mobile devices, we can safely assume that either the sender or the receiver is a mobile device. We can also assume that there is a channel of communication established between them based on a mutual understanding of the purpose with the interaction.

Consider the following brief scenario: "A user reads an e-mail on a mobile phone by paging down with a joystick". It is quite evident that the user interacts with the device by pushing down the joystick, whereas the device interacts with the user by presenting more of the e-mail on the screen. Do both the user and the device intentionally and actively transfer intelligible information between each other? Yes. Do both the user and the device produce and perceive stimulus? Yes. Since both the user and the device simultaneously produce and perceive stimulus we have two separate circuits of communication (Table 1).

Table 1: Circuits of communication.

Circuit	Sender	Receiver
Forward	Produces stimulus	Perceives stimulus
Reverse	Perceives stimulus	Produces stimulus

The interaction in the reverse circuit is what we think of as feedback. What the sender or receiver perceives of its own stimulus production is also feedback, but that has more to do with the senders inert communication skills than interaction. Throughout this paper, the mobile device will be referred to as the receiver and the user referred to as the sender. Input is thus when the device perceives the user and output is when the device stimulates the user. However, we also have a feedback loop, where the input is when the user perceives the device and the output is when the user stimulates the device. At this point, it should be clear that the sender and receiver reciprocally transfers information in one forward and one reversed circuit during interaction. Now we have a frame for the communication, next we will turn to the form of information.

3.2 Form of Information

The second tier has to do with the properties of the information that is transferred during interaction. The constructs of information, or meaning representations that can be communicated, are usually called signs. Signs do not convey any meaning in themselves, only when meaning is adhered to them do they become signs. Analogously, anything can be a sign as long as someone interprets it as signifying something, e.g. referring to or standing for something other than itself, "Nothing is a sign unless it is interpreted as a sign." (Peirce cited in Chandler, 2001). The Swiss linguist Ferdinand de Saussure and the American philosopher Charles Sanders Peirce developed the two currently dominant models of what constitutes a sign around a century ago.

Saussure offered a two-part model of the sign. He defined the sign as being composed of the signifier and the signified, where the signifier is the form a sign takes whereas the signified is the concept it represents. The sign itself is the result of an association between the signifier and the signified. The association is purely arbitrary and there is no one-to-one relation between the signifier and the signified; signs have multiple rather than single meanings and the meaning of a sign depends on its context in relation to other signs. Peirce on the other hand formulated a model of the sign composed of three parts. He defined the sign as consisting of a representamen, the form of the sign, an interpretant, the sense made of the sign, and an object, what the sign refers to.

Whereas Saussure did not offer any typology of signs, Peirce offered several. Peirce's categorization of signs also provides a richer context for understanding how representations convey meaning. The most general categorization is based on three kinds of signs. Firstly, there are indications, or indices; that show something about things because of their being physically connected with them. Secondly, there are likenesses, or icons; that serve to convey ideas of the things they represent simply by imitating them. Thirdly, there are symbols, or general signs, that have become associated with their meanings by usage (Chandler, 2001) (Table 2).

Table 2: Forms of information.

Form	Definition
Indexical	Sign is directly connected to the object
Iconic	Sign is analogously connected to the object
Symbolic	Sign is arbitrarily connected to the object

Indexical signs can be thought of as all representations and actions that directly connect the mobile device with the user and the environment. Examples of indexical signs are an alarm signal indicating an alarm, pointing the device at something, or tilting the device. For mobile appliances, there is also a close relation between indexical signs and instances of context awareness (Kjeldskov, 2002). Iconic signs can be thought of as all representations and actions that resemble something else. Examples of iconic signs are a battery icon indicating battery status, a picture of lifted phone indicating the connect call function, or a tone resembling a popular pop song. Symbolic signs can be thought of as all representations and actions that have to be learned, including all instances of language used in an interface.

3.3 Mode of Interaction

The third tier has to do with how signs can be stimulated and perceived in interaction. This is where multimodality come into play as signs can be expressed through several senses. As mentioned, we mostly use the visual, auditory, and tactile modalities for interaction. Each of these modalities has unique properties for conveying information that is very different in nature. There is also a difference in how the same information can be expressed through the different modalities. There is obviously no point in designing interfaces that interact through modalities in ways that the mobile devices cannot perceive. However, most mobile devices actually do have means to use all three modalities. Not every device may have all means for input and output, but it is likely that most devices will feature several of them, if not primarily for multimodal interaction, at least for multimedia content delivery (Table 3).

Table 3: Modes of interaction.

Mode	Input	Output
Visual	Camera, IR sensor	Screen, LED's
Auditory	Microphone	Speaker, Headphones
Tactile	Buttons, Tilt sensor	Buzzer, Gyro

The most commonly used input modality for mobile devices, as for computers at large, is the tactile, and generally in the form of button presses. One could argue that the audible input channel is more commonly used on mobile phones, given that most people use them to talk in, but the information transferred then is not really aimed for the mobile device. What the most commonly used output modality is for mobile devices is a little harder to decide on. For a majority of the users it is probably the visual via the screen, but for people who only use mobile devices for telephony, it may just as well be the audible for call notification. Yet, when communicating actively with the device, the main output modality is the visual.

3.4 Bringing the Framework Together

Interaction presupposes a forward and reverse circuit of communication. Through each of these circuits, information can be represented in the form of indexical, iconic, or symbolic signs. Depending on the modality that is used the information can be expressed in a visual, auditory, or tactile mode. If we map these instances against each other, we get a graph with nine multimodal information types. Each type corresponds to a certain type of information that corresponds to the combination of semiotic form and interaction mode. In multimodal communication each information type can be independently and concurrently communicated (Figure 1).

The labels that are used, e.g. image, sound, or signal, should not be interpreted literally; they are denotations for a certain combination of sign form and modality mode. The types are not unconditional, nor are they unambiguous, since it in most cases is hard to draw a line between different sign forms. The intention with the categorization is to give us a richer framework for reasoning around how different types of information are used in multimodal interaction. Furthermore, by categorizing between the different types we get a more specific view of how information is worked with in different interfaces.

There are other frameworks and typologies that are similar to the one outlined here. Bernsen (1994) presents a typology based on a generic approach to the analysis of output modality types. There are two main differences between Bernsen's typology and this framework. Firstly, we consider input and output inseparable in interaction whereas Bernsen mainly focus on output. Secondly, our framework is based on semiotic theory whereas Bernsen categorizes properties of multimodal interfaces,

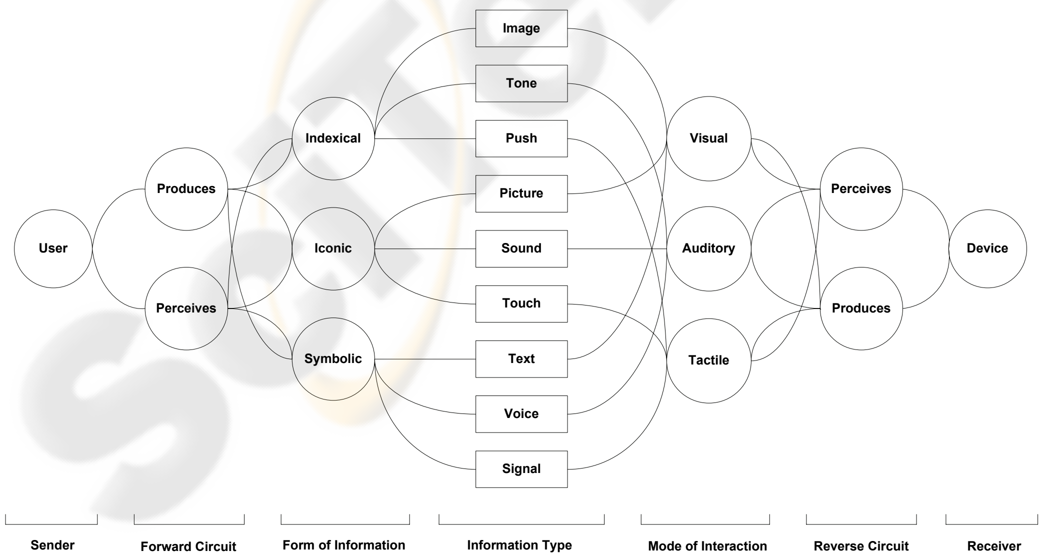

Figure 1: Framework for multimodal interaction where the information types correspond to the mapping between indexical, iconic, and symbolic forms over the visual, auditory, and tactile modes.

resulting in no less than 48 more or less atomic types. Our framework is less detailed, but also more expressive. Nigay and Coutaz (1993) presents a design space for multimodal systems that is complementary to the framework outlined here in the sense that it primarily focus on the distinction between sequential and parallel use of modalities and their combination. In our framework, we do not make this distinction although we do allow for both concurrent processing and data fusion.

4 MULTIMODAL INTERACTION WITH MOBILE DEVICES

We will now show how each information type can be interacted with for both output and input by providing examples from previous research. We have chosen to group the examples together according to the sign forms, mostly to make it apparent how similar the information becomes although different interaction modes are used. Although most examples only use a certain sign form or mode of modality it is shown how the different information types may be used for interaction. When we have looked at all the examples, we will turn to a concluding discussion about how different interaction techniques may be integrated into useful interfaces for multimodal interaction.

4.1 Indexical Interaction: Image, Tone and Push

The first information type is the indexical visual ***image***. By image we mean visual information that is directly connected to a specific context. For input of images the digital camera, more or less standard on mobile phones supporting MMS, is probably what comes first to mind. However, the input image could also be something only the device uses, for example using a sensor to monitor the light in the surrounding environment, or using a camera to monitor if the user is looking on the screen or not as in the SmartBailando browser (Öquist, 2002). For output of images almost all new mobile devices, as PDAs or multimedia enabled phones, offer high-resolution colour screens. If the screen is not large enough, a solution may be to display images on a device with a larger screen in the vicinity as exemplified in the Pick-and-Drop interface (Rekimoto, 1987), another possibility is to use a head-mounted display (www.virtualvision.com).

The indexical auditory type is referred to as the ***tone***. By tone we mean audible information that is directly connected to a specific context. For input of tones in the form of sounds we need a microphone. One possibility with using tones as auditory input is exemplified in the Tuneserver (Prechelt and Typke, 2001), where sounds were transformed into an indexical representation and matched against templates of musical scores to find the name of a song or melody. Another, more mobile specific example, is to monitor the loudness level in the surrounding and adapt interfaces to that (Mäntyjärvi and Seppänen, 2001). For output of tones we always have the ring tone as an example, but there are others that are more interesting. Earcons were introduced by Brewster as a substitute for graphical elements when navigating a hierarchy of nodes in an interface. Earcons are abstract, synthetic tones constructed from motives using timbre, register, intensity, pitch, and rhythm (Brewster, 1998). By using a pair of headphones, it is possible to index sounds in three dimensions and for example create audible interfaces for menu selection (Lorho et al, 2002), or directing the user's attention to objects that are outside the visual area of the screen, as exemplified in the Fishears interface (McGookin and Brewster, 2001).

The indexical tactile information type is referred to as the ***push***. By push we mean tactile information that is directly connected to a specific context. For input of push we need some form of tactile sensor, it can be a button, touch screen, or accelerometer (for sensing degree of tilt). Examples of using tilt as indexical tactile input for navigation, e.g. tilting up/down, or left/right, has been exemplified in very small interfaces, as in the Hikari interfaces (Fishkin et al, 2000). Physically pointing the device at objects as an interaction method has been explored in the mobile Direct Combination interaction technique (Rekimoto, 1987). For push output, the most common example is the tactile feedback you get when pressing buttons. For device initiated tactile output we need some form of tactile generator, most mobile phones do also have a vibrator for unobtrusive call notification. A more elaborate, yet straightforward, example is the TactGuide (Sokoler et al, 2002) that literary points the user to a target location by using subtle tactile directional cues.

4.2 Iconic Interaction: Picture, Sound and Touch

The visual iconic information type is referred to as the ***picture***. By picture we mean visual information that connects to an object or entity since it looks like it. For input of new pictures we would need some form of pad or touch screen to draw on, but more common is to have predefined pictures to choose from, such as inserting a graphical smiley emoticon in a text message, or combining pictures with each other, as in the direct manipulation paradigm (Schneiderman, 1982). One another possibility is to use something similar to the Bitpict program (Furnas, 1991), where a matrix of pixels served as a blackboard for a picture production. However, the most common use of pictures is for output, in the form of icons and metaphores used in the graphical user interface. Not only in minituarized desktop interfaces, but also when content is viewed on mobile devices. An example is the Smartview browser (Milic-Frayling and Sommerer, 2002), it displays geometrically sectioned miniaturized representations of web pages as they would have been displayed in a full size screen, by selecting a section it is possible to view that portion of the page in isolation

The iconic auditory type is referred to as the ***sound***. By sound we mean audible information that connects to an object or entity since it sounds like it. The most commonly used sound input is probably the voice activated calling function, the user has then added a sound profile to a contact in the phone book. By saying the “magic word”, the contact is called. This should not be confused with speech recognition that typically concerns continuous speech (Gold and Nelson, 1999). Just as for pictures, sounds are most commonly used for output on mobile devices. The most common example on mobile phones is probably to turn pop songs into monophonic ring tones, then being a metaphor of the actual song. However, as more and more devices get polyphonic sound playback capabilities, these sounds are likely to be exchanged for sound effects instead. The addition of nomic auditory icons (Gaver, 1986), e.g. straight depictions like sound effects in a movie, to self-paced reading of text on a mobile device has been found to significantly increase the feeling of immersion while reading (Goldstein et al, 2002).

The iconic tactile information type is referred to as the ***touch***. By touch we mean tactile information that connects to an object or entity since it resembles the feeling of it. Pirhonen et al. (2002) investigated the use of metaphorical gestures to control an MP3 player. For example, the “next track” gesture was a sweep of a finger across the screen left to right and a “volume up” gesture was a sweep up the screen, bottom to top. For output of iconic tactile information, there are to the authors’ knowledge no stimuli generators for mobile devices yet. However, there are some under development. Immersion Inc. (www.immersion.com) claims that their engineers have developed a device that would make it possible to create tactile sensations that resemble how surfaces feel and how a certain action feels in three dimensions. Research on touch output has otherwise mostly been about medical equipment and robotics, however more recently a number of researchers have reported improvements in interaction with tactile feedback (Oakley et el, 2002).

4.3 Symbolic Interaction: Text, Voice and Signal

The visual symbolic information type is referred to as the ***text***. By text, we mean visual information that has a connection to an object or entity that has to be learned. The most widely used input figure is of course the characters in language. Language is extremely expressive, but you have to learn how to use it. Text input on mobile devices is hard to get efficient because of the devices small form factors. A multitude of solutions have been devised, among those that use pure symbolic input we have different forms of character recognition as text is written on a touch sensitive screen (either as regular characters or as short forms), or when characters are entered on a soft-keyboard on the screen (MacKenzie and Soukoreff, 2002). Nonetheless, if we thought entering text was cumbersome, output can be even worse. Since text, and other figures such as graphs or tables, in a document usually has a spatial layout, problems arise when you are attempting to read it on a screen with the size of your palm. It gets even worse if you want to view additional content, such as images as well. A few different solutions have been proposed; one similar to the predictive text input interface is Adaptive RSVP (Öquist and Goldstein, 2003), where the text is broken up in smaller units that are successively displayed on the screen for durations that are assumed to match the processing time.

The auditory symbolic information type is referred to as the ***voice***. By voice we mean auditory information that has a connection to an object or entity that has to be learned. The prime form of vocal input is naturally speech recognition, an input method that offers great promises, but is extremely difficult. Especially in mobile environments even more difficult because of additional sounds in the surroundings. Recognition of fluent speech on mobile clients is a major research topic and there are many issues that need to be resolved until we can rely on it for interaction. However, limited speech recognition is not far fetched, and there is work in progress on how to define limited vocabularies and at least attain limited speech interaction (von Niman et al, 2002). For auditory symbolic output, there is of course speech synthesis, somewhat easier to accomplish than recognition, but similarly hard to get natural. It is also hard to get interaction with speech synthesis efficient since listening to speech is half as fast as reading text (Williams, 1998). In order to achieve conversational interface there are also several other components, besides those for speech recognition and synthesis, that must be integrated into a system that can sustain a fruitful dialog (McTear, 2002).

The tactile symbolic information type is referred to as the ***signal***. By signal we mean tactile information that has a connection to an object or entity that has to be learned. For entering text there are several tactile interfaces. Typing on buttons is probably the most commonly used although most mobile devices do not have a proper keyboard. There are several solutions for text entry on mobile devices without a proper keyboard. The smarter of the methods are those similar to Tegic T9 (www.tegic.com) or LetterWise (MacKenzie et al, 2001) that use linguistic knowledge to achieve single-tap, instead of multi-tap typing. There are also a few interfaces for tilt based typing as exemplified by the Unigesture prototype (Sazawal et al, 2002), where different characters are added to words by tilting the device in different directions. A quite different solution to text entering is Dasher (Ward et al, 2002) where characters slides across the screen and are selected by indicating them by tilt selection or gaze detection. The only form of symbolic tactile output the authors could come to think of was the Braille printer for blind people that is based on six pegs that are raised in different combinations that can be interpreted as characters.

5 DISCUSSION

The main contribution of this paper is that it offers a framework for reasoning around multimodal interaction with mobile devices in a structured manner. We offer a design space that encapsulates all of the interaction possibilities using a multimodal interface. Empirical usability evaluations are always necessary to validate hypothesizes about usability, but finding a common ground to compare and discuss results are equally important. Allwood (2002) has presented a framework for bodily communication that is similar to the one presented here in the sense that it also rests on Peirce's indexical, iconic, and symbolic, signs. It does however mainly concern bodily interaction, in addition to voice and writing, and is thus not fully comparable to the framework outlined here. Yet, it raises one very interesting question. Will the inclusion of more expressive descriptions of communication support or complicate our understanding? Allwood argues that this is not likely to be without problems, but hopefully the reward "will consist in an increased understanding of human communication" (2002:20). The intention with the framework outlined in this paper is similarly to provide a richer context for understanding multimodal interaction.

Interfaces in which users are able to choose between using different modalities are already in use. As more integrated interfaces appear, users will not have to select the modality to use, they will be able to switch seamlessly from one to another. Multimodal interfaces will allow the mobile user to interact through the modality that bests suit them and the environment where they are. Integrated multimodal interfaces will allow users to make use of their ability to work with multiple modes of interaction in parallel. Eventually, multimodal interfaces may let users interact with mobile devices in the way humans normally do with each other: by looking, talking, and touching, all at the same time. As functionality gets more sophisticated, interaction gets more natural. This represents a challenge today, but it also represents the promise of multimodal interaction with mobile devices for the future.

6 CONCLUSION

We have shown how interfaces that utilizes our ability to communicate through the visual, auditory, and tactile senses, may enhance mobile interaction. As we have seen mobile devices are apt for multimodal interaction, and we have raised the question of how we may coordinate information communicated through several senses in a way that promotes interaction. A framework for integration of multimodal interaction has been outlined by mapping information over communication circuit, semiotic representation, and sense applied for interaction. By exemplifying how different research prototypes fit into the framework today, we have shown how interfaces can be interacted with through multiple modalities. The foremost benefit of the framework is that it can support our reasoning around how to make the best of these possibilities tomorrow.

REFERENCES

Allwood, J. (2002). *Bodily Communication - Dimensions of expression and Content*. In B. Granström, D. House, and I. Karlsson (Eds). Multimodality in Language and Speech Systems, 7-26. Kluwer Academic Publishers.

Bersen, O. (1994). *Foundations of multimodal representations. A taxonomy of representational modalities*, Interacting with Computers, 6(4), 347-371.

Brewster, S.A. (1998). *Using non speech sounds to provide navigation cues*. ACM Transactions on Computer-Human Interaction, 5(3), 224-259.

Chandler, D. (2001). *Semiotics: The Basics*. New York: Routledge.

Fishkin, K. P., Gujar, A., Harisson, B. L., Moran, T. P. and Want, R. (2000). *Embodied user interfaces for really direct manipulation*. Communications of the ACM, 43(9), 75-80.

Furnas, G.W. (1991). *New graphical reasoning models for understanding graphical interfaces*. In Proceedings of ACM CHI'91 Conference (New Orleans, LA), 71-78. New York, NY: ACM Press.

Gaver, W. (1986). *Auditory icons: Using sound in computer interfaces*. Human-Computer Interaction, 2, 167-177.

Gold, B., and Nelson, M. (1999). *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. New York, NY: John Wiley & Sons.

Holland, S., Morse, D.R., and Gedenryd, H. (2002). *Direct Combination. A new user interaction principle for mobile and ubiquitous HCI*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 108-122. Berlin: Springer.

Kjeldskov, J. (2002). *"Just-in-Place" information for mobile device interfaces*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 271-275. Berlin: Springer.

Lorho, G., Hiipakka, J., and Marila, J. (2002). *Structured menu presentation using spatial sound separation*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 419-424. Berlin: Springer.

MacKenzie, I. S., and Soukoreff, R. W. (2002). *Text entry for mobile computing: Models and methods, theory and practice*. Human-Computer Interaction, 17, 147-198.

MacKenzie, I. S., Kober, H., Smith, D., Jones, T., and Skepner, E. (2001). LetterWise: *Prefix-based disambiguation for mobile text input*. In Proceedings of UIST'01 (Orlando, FL), 111-120. New York, NY: ACM Press.

McGookin, D.K., and Brewster, S.A. (2001) *Fishears – The design of a multimodal focus and context system*. In Proceedings of IHM-HCI'01, Vol. II (Lille, France), 1-4. Toulouse: Cépaduès-Editions.

McTear, M.F. (2002). *Spoken dialogue technology: enabling the conversational interface*. ACM Computing Surveys, 34(1), 90 – 169.

Milic-Frayling, N., and Sommerer, R. (2002). *SmartView: Flexible viewing of web page contents*. In Proceedings of WWW'02, (Honolulu, USA).

Nigay, L., and Coutaz, J. (1993). *A design space for multimodal interfaces: concurrent processing and data fusion*. In Proceedings of InterCHI'93, (Amsterdam, The Netherlands), 172-178.

Mäntyjärvi, J., and Seppänen, T. (2002). *Adapting applications in mobile terminals using fuzzy context information*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 95- 107. Berlin: Springer.

Oakley, I., Adams, A., Brewster, S.A., and Gray, P.D. *Guidelines for the design of haptic widgets*. In Proceedings of BCS HCI 2002 (London, UK), 195-212. London: Springer.

Öquist, G., Goldstein, M., and Björk, S. (2002). *Utilizing gaze detection to stimulate the affordances of paper in the Rapid Serial Visual Presentation Format*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 378-381. Berlin: Springer.

Öquist, G., and Goldstein, M. (2003). *Towards an improved readability on mobile devices: Evaluating Adaptive Rapid Serial Visual Presentation*. Interacting with Computers, 15(4), 539-558.

Pirhonen, A., Brewster, S.A., and Holguin, C. (2002). *Gestural and audio metaphors as a means of control for mobile devices*. In Proceedings of ACM CHI'02 (Minneapolis, MN), 291-298. New York, NY: ACM Press.

Prechelt, L., and Typke R. (2001). *An interface for melody input*. ACM Transactions on Computer-Human Interaction, 8(2), 133-194.

Rekimoto, J. (1987). *Pick and Drop: A direct manipulation technique for multiple computer environments*. In Proceedings of UIST'87, 31-39. New York, NY: ACM Press.

Sazawal, V., Want, R., and Borriello, G, (2002). *The Unigesture approach*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 256-270. Berlin: Springer.

Schneiderman, B. (1982). *The Future of Interactive Systems and the Emergence of Direct Manipulation*. Behaviour and Information Technology, 1, 237-256.

Sokoler, T., Nelson, L., and Pedersen, E.R. (2002). *Low-resolution supplementary tactile cues for navigational assistance*. In Proceedings of Mobile HCI 2002 (Pisa, Italy), 369-372. Berlin: Springer.

Ward, D. J., Blackwell, A. F., and MacKay, D. J. C. (2002). *Dasher: A gesture-driven data entry interface for mobile computing*. Human-Computer Interaction, 17, 199-228.

Williams, J. R. (1998). *Guidelines for the use of multimedia in instruction*. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (Chicago, IL), 1447-1451. Santa Monica, CA: HFES.