DESIGN SPACE FOR MULTIMODAL INTERACTION

Laurence Nigay
CLIPS-IMAG,Université de Grenoble 1 38000 Grenoble, France
Email Laurence.Nigay@imag.fr

Abstract: One trend in Human Computer Interaction is to extend the sensory-motor capabilities of computer systems to better match the natural communication means of humans. Although the multiplicity of modalities opens a vast world of experience, our understanding of how they relate to each other is still unclear and the terminology is unstable. In this paper we present our definitions and existing frameworks useful for the design of multimodal interaction.

Key words: Multimodal UI, I/O devices, Interaction Languages, Combination.

1. INTRODUCTION

The area of multimodal interaction has expanded rapidly and since the seminal “Put that there” demonstrator (Bolt 1980) that combines speech, gesture and eye tracking, significant achievements have been made in terms of both modalities and real multimodal systems. Indeed, in addition to more and more robust modalities, conceptual and empirical work on the usage of multiple modalities is now available for guiding the design of efficient and usable multimodal interfaces. As a result, real multimodal systems are now being built in various application domains including medicine (Oviatt et al. 2000) and education.

Recent progress achieved in the miniaturization of microprocessors and in wireless networks make it possible to foresee the disappearance of the "grey box" that is the personal computer, or at least to understand that it is no

longer the only place of interaction between people and the numerical world. This development is driven by the recent concepts of Ubiquitous Computing and Disappearing Computer and from the evolution occurring in the field of interaction modeling. Indeed the research is now gradually directed towards models of interaction in which the data-processing resources are distributed in a multitude of everyday objects with which users interact in explicit (active modalities) and implicit ways (passive modalities). This has given rise to several recent interaction paradigms (i.e., Augmented Reality, Ubiquitous/Pervasive Computing, Tangible Bits, and Embodied Multi-surfaces) that increase the set of possibilities for multimodal interaction. A good example of a recent type of modality is provided by the "phicons" (Physical Icons) that define new input modalities based on the manipulation of physical objects or physical surfaces such as a table or a wall that can be used for displaying information (output modality) in an ubiquitous computing scenario.

Although the multiplicity of modalities opens a vast world of experience, our understanding of how they relate to each other is still unclear and the terminology is unstable.

2. DEFINITION: MODALITY

2.1 Device and Language

In his theory of action, Norman structures the execution and evaluation gulfs in terms of semantic and articulatory distances that the user needs to cover in order to reach a particular goal (Norman 86). This user-centered approach pays little attention to the processing steps that occur within the computer system. Our Pipe-lines model makes these stages explicit (Nigay 1994). By so doing, we extend Norman's theory in a symmetric way within the computer system. Two relevant concepts emerge from this model: the notion of physical device and that of interaction language. Interestingly, these concepts cover the semantic and articulatory distances of Norman's theory.

A physical device is an artifact of the system that acquires (input device) or delivers (output device) information. Examples include keyboard, loudspeaker, head-mounted display and GPS. Although this notion of device is acceptable for an overall analysis of an interactive multimodal system, it is not satisfactory when one needs to characterize the system at a finer grain of

interaction. The design spaces of input devices such as that of Mackinlay et al. (Mackinlay et al., 1990) and of Foley et al. (Foley et al., 1994) are frameworks that valuably refine a physical device. A review of these taxonomies are presented in (Nigay et al., 1996).

An interaction language is a language used by the user or the system to exchange information. A language defines the set of all possible well-formed expressions, i.e., the conventional assembly of symbols, that convey meaning. Examples include pseudo-natural language, direct manipulation language. Three properties of an interaction language are introduced in the theory of output modalities (Bernsen et al., 1994): (1) Linguistic or non-linguistic (2) Analogue or non-analogue (3) Arbitrary or non-arbitrary.

The generation of a symbol or a set of symbols, results from a physical action. A physical action is an action performed either by the system or the user on a physical device. Examples include highlighting information (system physical actions), pushing a mouse button or uttering a sentence (physical actions performed by the user). The physical actions performed by the user can be either explicitly performed for conveying information to the system (explicit actions of the user towards the interactive system) or can be part of the user's tasks and is a source of information that is not explicitly expressed to the computer but is useful for the interaction ("perceptual user interfaces" (Turk et al., 2000)).

If we adopt Hemjslev's terminology (Hemjslev 1947), the physical device determines the substance (i.e., the unanalyzed raw material) of an expression whereas the interaction language denotes its form or structure.

2.2 Interaction Modality and Multimodality

In the literature, interaction modality is discussed at multiple levels of abstraction from both the user and the system perspectives. At the lowest level, a modality may refer to a human sensory capability or to a computer physical device such as a microphone, a camera, or a screen. At a higher level of abstraction, a modality is viewed as a representational system, such as a pseudo-natural language that the user and the system might share. Whereas the device level is related to the human sensory capabilities, the representational level calls upon cognitive resources. Clearly, the physical and the representational computer models are tightly coupled to the sensory and cognitive dimensions of human behavior. For this reason, in (Nigay et al., 95) we define a modality as the coupling of an interaction language L with a physical device d: <d, L>. Examples of input modalities while using a

PDA (Zouinar et al., 2003) include: <microphone, pseudo natural language>, <camera, 3D gesture>, <stylus, direct manipulation> and <PDA, 3D gesture> (Embodied user interface (Harrison et al. 1998)).

Within the vast world of possibilities for modalities, we distinguish two types of modalities: the active and passive modalities. For inputs, active modalities are used by the user to issue a command to the computer (e.g., a voice command or a gesture recognized by a camera). Passive modalities refer to information that is not explicitly expressed by the user, but automatically captured for enhancing the execution of a task. For example, in the "Put that there" seminal multimodal demonstrator of R. Bolt (Bolt 1980), eye tracking was used for detecting which object on screen the user is looking at. Similarly, in our MEMO system (Bouchet et al. 2004), "orientation" and "location" of the mobile user are two passive input modalities. The modality "orientation" is represented by the magnetometer (device) and the three orientation angles in radians (language), the other modality "localization" by the pair <Localization sensor, 3D location>. MEMO allows users to annotate physical locations with digital notes which have a physical location and are then read/removed by other mobile users.

In the literature, multimodality is mainly used for inputs (from user to system) and multimedia for outputs (from system to user), showing that the terminology is still ambiguous. In the general sense, a multimodal system supports communication with the user through different interaction modalities. Literally, "multi" means "more than one".

Our definition of modality and therefore of multimodality is system-oriented. A user-centered perspective may lead to a different definition. For instance, according to our system-centered view, electronic voice mail is not multimodal. It constitutes a multimedia user interface only. Indeed, it allows the user to send mail that may contain graphics, text and voice messages. It does not however extract meaning from the information it carries. In particular, voice messages are recorded and replayed but not interpreted. On the other hand, from the user's point of view, this system is perceived as being multimodal: The user employs different modalities (referring to the human senses) to interpret mail messages.

In addition our definition enables us to extend the range of possibilities for multimodality. Indeed a system can be multimodal without having several input or output devices. For example, a system using the screen as the unique output device is multimodal whenever it employs several output interaction languages. In (Vernier et al. 2000), we claim that using one device and multiple interaction languages raises the same design and engineering issues as using multiple modalities based on different devices.

3. COMBINATION OF MODALITIES

Although each modality can be used independently within a multimodal system, the availability of several modalities in a system naturally leads to the issue of their combined usage. The combined usage of multiple modalities opens a vastly augmented world of possibilities in user interface design. Several frameworks addressed the issue of relationships between modalities. In the seminal TYCOON framework (Martin 1997) six types of cooperation between modalities are defined:

1. Equivalence involves the option of choosing between several modalities that can all equally well convey a particular chunk of information.
2. Specialization implies that specific kinds of information are always conveyed by the same modality.
3. Redundancy indicates that the same piece of information is conveyed by several modalities.
4. Complementarity denotes several modalities that convey complementary chunks of information.
5. Transfer implies that a chunk of information processed by one modality is then treated by another modality.
6. Concurrency describes the case of several modalities conveying independent information in parallel.

The CARE properties (Coutaz et al., 1995) define another framework for reasoning about multimodal interaction from the perspectives of both the user and the system: These properties are the Complementarity, Assignment, Redundancy, and Equivalence that may occur between the modalities available in a multimodal user interface. We define these four notions as relationships between devices and interaction languages and between interaction languages and tasks. In addition, in our multifeature system design space (Nigay et al., 1995) we emphasized the temporal aspects of the combination, a dimension orthogonal to the CARE properties. Finally in (Vernier et al., 2000), we present a combination framework that encompasses and extends the existing design spaces for multimodality. The combination framework is comprised of schemas and aspects: While the combination schemas (Allen's relationships) define how to combine several modalities, the combination aspects determine what to combine (temporal, spatial, syntactic and semantic).

4. CONCLUSION

In this article, we have provided an overview of a number of definitions and frameworks useful for the design of multimodal user interfaces. To do so

we have focused on the definition of a modality and then on the composition of modalities.

ACKNOWLEDGEMENTS

This work is partly funded by French DGA under contract #00.70.624.00.470.75.96. The author also acknowledges the support of the SIMILAR European FP6 network of excellence (http://www.similar.cc).

REFERENCES

Bolt, R., (1980), Put-that-there: Voice and gesture at the graphics interface, in Proceedings of the 7th annual conference on Computer graphics and interactive techniques, pp. 262-270.

Bouchet, J., Nigay, L., 2004, ICARE: A Component-Based Approach for the Design and Development of Multimodal Interfaces, *in Proceedings of CHI'2004*, ACM Press.

Bernsen, N., (1994), A revised generation of the taxonomy of output modalities, *Esprit Project AMODEUS Working Paper RP5-TM-WP11*.

Coutaz, J., et al., 1995, Four easy pieces for assessing the usability of multimodal interaction: The CARE properties, in Proceedings of Interact'95, Chapman&Hall, pp. 115-120

Foley, J., Wallace, V., Chan, P., (1984), The Human Factors of computer Graphics interactiontechniques*, IEEE computer Graphics and Applications*, 4(11), pp. 13-48.

Harrison, B., et al., 1998, Squeeze me, Hold me, Tilt Me ! An exploration of Manipulative User Interface, in Proceedings of CHI'98, ACM Press, pp. 17-24.

Hemjslev, L., 1947, Structural Analysis of language, *Studia Phonetica*, Vol. 1, pp. 69-78.

Mackinlay, J., Card, S., Robertson, G., (1990), A Semantic Analysis of the Design Space of Input Devices, *Human Computer Interaction*, Lawrence Erlbaum, 5(2,3), pp. 145-190.

Martin, J. C., 1997, TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. Intelligence and Multimodality in Multimedia Interfaces, AAAI Press.

Norman, D. A., 1986, Cognitive Engineering, *User Centered System Design, New Perspectives on Computer Interaction*, Lawrence Erlbaum Associates, pp. 31-61.

Nigay, L., 1994, Conception et modélisation logicielles des systèmes interactifs : application aux interfaces multimodales, *PhD dissertation University of Grenoble 1*, 315 pages.

Nigay, L., Coutaz, J., 1995, A Generic Platform for Addressing the Multimodal Challenge, *in Proceedings of CHI'95*, ACM Press, pp. 98-105.

Nigay, L., Coutaz, J., 1996, Espaces conceptuels pour l'interaction multimédia et multimodale, *TSI,* 15(9), AFCET&Hermes Publ, pp. 1195-1225.

Oviatt, S. et al., 2000, Designing the user interface for multimodal speech and gesture applications, *Human Computer Interaction*, Lawrence Erlbaum, 15(4), pp. 263-322.

Turk, M., Robertson, G., (ed.), 2000, Perceptual user Interfaces, *Communications of the ACM*, 43(3), ACM Press, pp. 32-70.

Vernier, F., Nigay, L., 2000, A Framework for the Combination and Characterization of Output Modalities, *in Proceedings of DSV-IS2000*, LNCS, Springer-Verlag., pp. 32-48.

Zouinar, M., et al. , 2003, Multimodal Interaction on Mobile Artefacts, *Communicating with smart objects-developing technology for usable pervasive computing systems*, Hermes Penton Science, ISBN 1-9039-9636-8.