

A Framework for the Combination and Characterization of Output Modalities

F. Vernier and L. Nigay

CLIPS-IMAG, BP 53, 38041 Grenoble cedex 9, France
Tel. +33 4 76 51 44 40, Fax: +33 4 76 44 66 75
Email: {Frederic.Vernier, Laurence.Nigay}@imag.fr

Abstract. This article proposes a framework that will help analyze current and future output multimodal user interfaces. We first define an output multimodal system. We then present our framework that identifies several different combinations of modalities and their characteristics. This framework assists in the selection of the most appropriate modalities for achieving efficient multimodal presentations. The discussion is illustrated with MulTab (**M**ultimodal **T**able), an output multimodal system for managing large tables of numerical data.

1 Introduction

The use of multiple modalities such as speech and gesture opens a vast world of possibilities in user interface design. The goal of multimodal interfaces is to extend the sensory-motor capabilities of computer systems to better match the natural communication means of human beings. The purpose is to enhance interaction between the user and the computer by utilizing appropriate modalities to improve:

- ? the information bandwidth between the human and the computer; that is the amount of information being communicated;
- ? the signal-to-noise ratio of conveyed information; that is the rate of information useful for the task being performed [20].

Although the potential for innovation is high, the current understanding of how to design, build, and evaluate multimodal user interfaces is still primitive. The power and versatility of multimodal interfaces result in an increased complexity that current design methods and tools do not address appropriately. This problem is exacerbated by the proliferation of new input and output modalities, such as the phycons [11] or ambient modalities [15].

In this paper, we focus on the design of output multimodal interfaces and we define a framework for characterizing output modalities and their combinations. This framework provides a better understanding of modality characteristics and of their combinations, and as such represents a step towards achieving the potential gain of multiple output modalities. Our unified framework coherently organizes the elements useful for the two key design issues of multimodal output user interfaces: the selection of mo-

dalities based on their characteristics and the combination of modalities for the design of a coordinated output interface. Combinations and characteristics of output modalities are useful for eliciting design rules, for classifying existing output systems and for evaluating the usability of a system.

The structure of the paper is as follows: First, we clarify the notion of modality using the concepts of interaction language and physical device. Indeed, as pointed out in [6], differences of opinion exist as to the meaning of the term "multimodal". Having defined an input/output modality, we present the main steps in the design of an output multimodal interface. After positioning our study in this design process, we present the two spaces of our framework: combinations of modalities and characteristics of a modality. The discussion will be illustrated with MulTab, a multimodal system that we developed.

2 Output multimodality

Multimodality has mainly been studied for input (from user to system) interfaces [22, 23, 25], by utilizing multiple input devices for exploiting several human sensory systems. The "put that there" paradigm which emphasizes the synergistic use of speech and gesture is one such attempt. In addition to the fact that fewer studies focus on output multimodality, the related studies mainly investigate a single output modality including speech synthesis, natural language text generation and network diagram generation. There is consequently a crucial need for a model of output multimodal user interfaces. Indeed such output interfaces are very complex and nowadays their design and implementation rely on empirical skills of the designers and developers. Moreover, we believe that output multimodality is a more difficult problem to address than input multimodality. For the case of multiple input modalities, the user decides which modalities to employ and their function (complementary or redundant use) based on his expertise and the context. For outputs, the designer or the system itself must be able to perform such choices and combinations based on knowledge of the concepts to be presented, interaction context, available output devices as well information about the user.

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 modalities such as voice, graphics, and text [7]. Literally, "multi" means "more than one" and the term "modal" may cover the notion of "modality" as well as that of "mode".

1. Modality refers to the type of communication channel used to convey information. It also specifies the way an idea is expressed or perceived [8].
2. Mode refers to a state that determines the way information is interpreted for conveying meaning.

In a communication act, whether it is between humans or between a computer system and a user, both the modality and the mode will come into play. The modality defines the type of data exchanged whereas the mode determines the context in which the data is interpreted. Thus, if we take a system-centered view, output multimodality is the capacity of the system to communicate with a user along different types of communication channels and to convey meaning automatically. We observe that both multimedia and multimodal systems use multiple communication channels. But in addition, a multimodal system is able to model the content of the information at a high level of abstraction. A multimodal system thus strives for meaning.

Our definition of output 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 order to support our definition of output multimodality, we define an output modality as the coupling of a physical device d with an interaction language L : $\langle d, L \rangle$ [23].

- ? A physical device is an artifact of the system that delivers information. Examples of output devices include the loudspeaker and screen.
- ? An interaction language defines a set of well-formed expressions (i.e., assembly of symbols according to some convention) that convey meaning. The generation of a symbol, or a set of symbols, involves actions on physical devices. Examples of interaction languages include pseudo-natural language and graphical animation.

Our definition of an output modality enables us to extend the range of possibilities for output multimodality. Indeed a system can be multimodal without having several output devices. A system using the screen as the unique output device is multimodal whenever it employs several output interaction languages. 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.

Having defined an output multimodal system, we can now describe the different stages for achieving efficient multimodal presentations.

3 Output multimodal interface design

The design of output multimodal interfaces requires the selection and the combination of multiple modalities. Such selection of atomic or composite output modalities can be performed:

1. by the designer while designing the system,
2. by the user while using the system,
3. by the system while running.

In case 2, we refer to the system as being *adaptable*. Case 2 must be related to case 1 because *adaptability* implies that the designer has previously selected a range of candidate modalities. In case 3 we call the system *adaptive (adaptivity)*.

Our discourse here is general and we present the steps for achieving a multimodal presentation. Three main steps are traditionally identified. These steps can be performed through design (by the designer) or through generation (by the system). These steps are:

1. content selection, which identifies what to say,
2. modality allocation, which identifies in what modalities to say it,
3. modality realization, which identifies how to say it in these modalities.

Within the design process or generation process, our framework is dedicated to the modality allocation step: i.e., the selection of an atomic or composite modality. In particular, our framework identifies a set of combinations of modalities and a set of characterizations of a modality. While the combination space enables the definition of new composite modalities, the characterization space helps in the choice of a modality, either atomic or composite. In the following paragraphs, we first present our combination space of output modalities and then our characterization space. We then illustrate our framework using our MulTab system.

4 Combination space

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 TYCOON framework [18], six types of cooperation between modalities are defined, a modality being defined as a process receiving and producing chunks of information:

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.

Each of these six types of cooperation is studied according to the usability criterion that it helps achieve. Such usability criteria are therefore called "goals of cooperation". The CARE properties 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. The notions of equivalence, assignment (or specialization), redundancy, and complementarity were primarily introduced by Martin [18]. We define these four notions as relationships between devices and interaction languages and between interaction languages and tasks. In [9], we formally define the CARE properties and showed how these properties affect the usability of the interaction. Finally, in our multi-feature system design space [23] and in our MSM framework [8], we emphasized the temporal aspects of the combination, a dimension orthogonal to the CARE properties.

Our combination space encompasses the types of combination presented in TY-COON, CARE and MSM, and identifies new ones. Our space is organized along two axes. The first axis considers the aspects that are combined. We first identify four aspects that can be combined:

1. Time: temporal combination
2. Space: spatial combination
3. Interaction language: syntactic combination
4. Semantic: semantic combination

Temporal and spatial combinations have been studied for combining output modalities [24]. We then introduce one aspect of combination, namely syntactic combination, which is based on our definition of a modality: an output modality being the coupling of an interaction language L with a physical device d : $\langle d, L \rangle$. These three first aspects of combination (i.e., temporal, spatial and syntactic) focus on a modality as a vehicle of information. The last aspect that must be considered while combining modalities is the relationship between the meaning of information conveyed by the composite modalities. This last aspect is called semantic combination.

The second axis ranges over a set of combination schemas, as presented in Figures 1-5. These schemas use the five Allen relationships [1] to provide a means of combining multiple modalities into a composite modality.

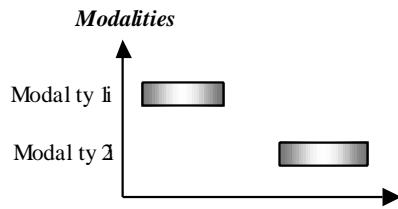


Fig. 1. Schema for distant modality combination.

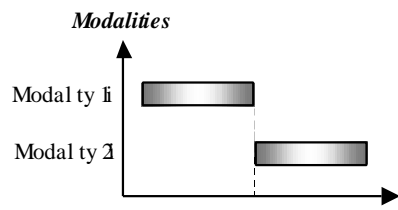


Fig. 2. Schema for modality combination with one point of contact.

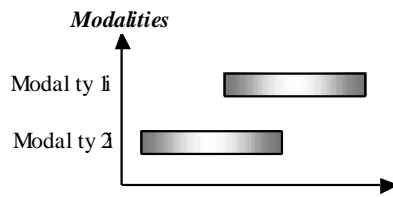


Fig. 3. Schema for modality combination with a non-empty intersection.

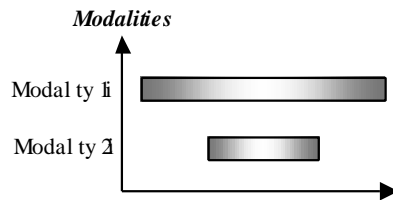


Fig. 4. Schema for modality combination with inclusion.

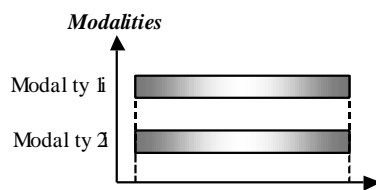
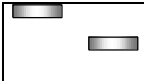
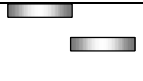
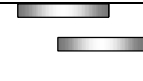
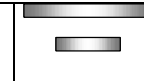



Fig. 5. Schema for modality combination with the same characteristics.

While the combination schemas define how to combine several modalities, the combination aspects determine what to combine. Our two axes (schemas and aspects of combination) are orthogonal: Table 1 names each type of combination obtained by blending these two axes. In the following paragraphs, we detail these combinations, our argumentation being based on the four identified aspects (i.e. (temporal, spatial, syntactic and semantic)).

Table 1. Applying the five combination schemas to the four combination aspects (temporal, spatial, syntactic and semantic).

		<i>Combination schemas</i>				
						
<i>Combination aspects</i>	Temporal	Anachronism	Sequence	Concomitance	Coincidence	Parallelism
	Spatial	Separation	Adjacency	Intersection	Overlaid	Collocation
	Syntactic	Difference	Completion	Divergence	Extension	Twin
	Semantic	Concurrency	Complementary	Complementary and Redundancy	Partial Redundancy	Total Redundancy

4.1 Temporal combination of modalities

As shown in Table 1, sequential and parallel combinations are two types of temporal combination that have been studied in the literature [8, 18, 22]. For example in [22], one dimension of the design space, called "use of modalities" primarily covers the absence or presence of parallelism at the user interface. We identify here three new temporal combinations: anachronism, concomitance and coincidence.

Two modalities are combined anachronously if there is a temporal gap between their usage. Anachronism and sequence are distinguished by the size of the temporal window between the usage of the two modalities. This size is defined by the designer. For example the designer can consider that two modalities are used anachronously if the temporal gap is longer than the perception time for causality (longer than a second). Concretely two sonic messages are perceived as independent if there is more than a second between the end of the first message and the beginning of the second one.

Two modalities are concomitant when one modality replaces another one with a time interval during which the two modalities coexist. As for anachronism, the designer must define the size of the time interval that we describe as transition time. This combination is important in helping the user understand the transition between two

modalities and delegating part of the cognitive load to the perceptual human process. A concomitant combination implies that the two devices corresponding to the two modalities can function in parallel.

Finally, the coincidence of two modalities is when one modality is only used in the context of another one. Such a combination is necessary to implement a modality that can only be used with another one. If the main modality can be terminated, two design solutions are possible:

1. The main modality cannot be stopped if the included modality is in use.
2. Terminating usage of the main modality implies terminating usage of the included one.

For example let us consider a form, defining one modality, and a dialog box, defining a second modality. The dialog box is opened from the form. Two design solutions are possible. On the one hand, the form cannot be closed if the dialog box is not closed first. On the other hand, if the form is closed, all the dialog boxes opened from this form are automatically closed.

4.2 Spatial combination of modalities

Spatial combination is important for output modalities, especially when considering multiple graphical modalities on screen. Multiple modalities, using the same device (screen or loudspeaker, etc.) and sharing the same location, are possible design solutions because human perception is able to acquire several pieces of information in parallel using a single human sense (sight, hearing, etc.). Different sounds can be played in parallel and distinguished by the user [4]. Nevertheless for the user the perceptual space (source of the sounds and their propagation space) is the same. Likewise transparency is one mechanism for combining two graphical modalities sharing the same space on screen.

On the one hand, two modalities can be used separately, and consequently do not share a common space. Thus the user will perceive the two modalities as separate. On the other hand, the four other spatial combinations of modalities will likely be perceived as related, as explained in the psychological guide of perception presented in [19]. Adjacent modalities share one point or one edge in space. Intersected, overlaid and collocated modalities define three types of transparency. A magic lens [26] defines one modality overlaid on another modality used for displaying the background. Two magic lenses intersecting on top of a background illustrate the intersected combination. Finally our mirror pixel mechanism [27] is an example of collocated combination: Here two modalities are used at the same place (the full screen) to display a document and the video of the user (a camera pointing to the user).

While separate modalities are likely to be used as the vehicles of independent information, the four other spatial combinations will imply some dependencies between the conveyed information. The later combinations are also useful for saving space (limited space of the screen for example) but will also engender perceptual problems.

In particular, visual continuity [24] is an ergonomic criterion that must be carefully studied when considering such spatial combinations.

4.3 Syntactic combination of modalities

In paragraph 2, an output modality is defined by the couple (physical device, interaction language). Syntactic combinations consider the interaction language of the modality (i.e., the logical form of the modality).

Two combined modalities can nevertheless have different syntaxes. For example in WIP [2], one modality is based on the English grammar (pseudo natural language) whereas another one is graphically depicted. An example of this syntactic combination (named difference) is seen in the following scenario: "The on/off switch is located in the upper left part of the picture" displayed above a picture.

Two modalities complete each other at the syntactic level when their corresponding syntaxes are combined to form a new syntax. The following generated sentence is one example of such a completion: "the date is 04/09/2000". Here two modalities are used, one based on pseudo natural language and one dedicated to displaying dates. The two corresponding syntaxes are combined to form a new syntax.

Two modalities are divergent when their corresponding interaction languages partially share the same syntax. For example speech synthesis and textual natural language generation correspond to two modalities that can be combined in a syntactically divergent way. Indeed the syntax of the two interaction languages is nearly the same, but spoken language is more informal than written language.

The syntactic combination named extension corresponds to two combined modalities where one modality has the syntax of its interaction language related to the syntax of the interaction language of the second modality. For example in [3], the generation of natural language text is combined with a text formatting modality, such as bullets corresponding to a sequence in the generated text.

Finally two syntactic twin modalities have interaction languages sharing the same syntax. This is the case of two modalities based on pseudo natural language: One of the modalities is related to the screen and the generated sentence is displayed while the other one is linked to the loudspeaker and the sentence is spoken. The CUBRICON system illustrates such a combination [21].

4.4 Semantic combination of modalities

The most studied aspect of combination is the semantic one, where one considers the meaning of the conveyed information along the modalities. The most common combi-

nations are those of complementarity and redundancy. One example of complementarity can be seen in the following sentence, which is displayed above a picture: "The on/off switch is located in the upper left part of the picture". The meaning conveyed by the textual modality and the graphical modality are complementary. One example of redundancy consists of the same text, displayed on screen and vocally (speech synthesis) [21]. In contrast to complementarity and redundancy, concurrent combination of modalities implies that the conveyed information has no related meaning. In addition to the well-known complementarity, redundancy and concurrency, we introduce two new types of combination, namely Complementarity-Redundancy and Partial redundancy.

"Complementary-redundant" modalities convey information that is partially redundant and complementary. Multiple graphical views often use such a combination to display two different attributes of the same piece of information. One part is redundant to help the user understand the semantic link between the visual presentations. For example in the MagniFind system [16], two views of a hierarchy of folders and files are displayed on screen: one view (one modality) displays the folders and files as lists and sub-lists, whereas the second one depicts them in the form of a hyperbolic tree. The two modalities are redundant because they both display the same list of folders in different ways at the highest level of the hierarchy. But the modalities are also complementary because one displays the precise information about a subpart of the hierarchy while the other one displays the full hierarchy without details.

Partially redundant modalities describe two combined modalities where one conveys a subpart of the information that the second one conveys. This is for example the case of a thumbnail view combined with a global view.

4.5 Aspects and schemas of combination: a unified framework

The four identified aspects and five schemas that we have identified define a unified framework that encompasses the existing frameworks, including TYCOON, CARE and MSM. Each combination of modalities can be characterized in terms of the four aspects and the five schemas.

The combination of modalities gives birth to new composite modalities. We now need to characterize an atomic or composite modality in order to be able to select the most appropriate one. The next paragraph presents our characterization space.

5 Characterization space

Characterization of atomic or composite modalities is necessary to be able to select them for an efficient multimodal presentation. As explained in paragraph 3, such selection is either performed by the designer or by the system itself (adaptativity). One

characterization space of output modalities has been proposed in [5]. Four boolean properties, defined as modality profiles, are presented:

- Static or dynamic
- Linguistic or non-linguistic
- Analogue or non-analogue
- Arbitrary or non-arbitrary

Static/Dynamic property refers to the articulatory level (the physical form of the modality: the device d , part of a modality $\langle d, L \rangle$) while the Linguistic/Non-linguistic property corresponds to the syntactic level (the logical form of the modality: the interaction language L , part of a modality $\langle d, L \rangle$). Analogue/Non-analogue and Arbitrary/Non-Arbitrary properties are related to the interpretation process and therefore the semantic level. We introduce three new characteristics.

- Deformed or non-deformed
- Local or global
- Precise or vague

The deformed/Non-deformed property is related to the syntactic level. Indeed a deformed modality is a modality that must be combined at the syntactic level as an extension of a non-deformed modality. For example let us consider the written sentence "r u happy?". This defines a modality that is based on a pseudo natural language modality "are you happy?" and a deformation modality, i.e. two modalities syntactically combined by extension.

The two other properties are related to the semantic level. For a given set of information to be presented, the Local/Global property refers to the range of information conveyed at a given time using the modality. If the user perceives all the pieces of information, the modality is global. If the user perceives only a subset of the information, the modality is local. In the software "PowerPoint", the slide by slide view is local while the slide sorter view is global. For each piece of information to be presented, the second property, Precise/Vague, characterizes the precision of a modality. If the modality conveys all the information about one element necessary for the task to be performed, it is a precise one, otherwise it is a vague modality. For the editing task, the slide by slide view is precise whereas the slide sorter view is a vague modality.

As shown by the seven properties, a modality can be characterized at three levels, the articulatory, syntactic and semantic levels (power of expression). One problem that we have still not addressed is the characterization of composite modalities: for example the characterization of an arbitrary modality combined with an analogue modality. Nevertheless, the seven properties define a starting point for characterizing modalities, in order to define design rules for their selection.

Neither our combination nor characterization spaces directly provide guidelines for the design of an efficient multimodal presentation. For example a semantic complementary combination is not better than a semantic redundant combination, and an analogue modality is not better than an arbitrary modality. Our two spaces are the foundations for defining design rules. To identify design rules, we base our approach on ergonomic criteria [13].

For example let us consider the ergonomic observability criterion: Because of the limited size of the screen, observability of a large set of elements is impossible in its entire scope and detail. One interesting solution to the problem is to make observable one subset of the elements in detail while maintaining the global set of elements observable without detail, using compression procedures: This approach is called "Focus + Context". It involves a combination of a local/precise modality with a global/vague one. The combination is defined as: (Temporal-Parallelism, Spatial-Adjacency, Syntactic-Difference, Semantic-Complementary).

Another example of design rules, related to the ergonomic insistence criterion, consists of an (Syntactic-Difference, Semantic-Redundancy) combination of modalities: The same information is conveyed twice by both modalities, which are based on different interaction languages. This design rule is closely related to the urgency rule defined by [14].

Having presented our combination and characterization spaces, we now illustrate them using our MulTab system. MulTab (Multimodal Table) is an output multimodal system for managing large tables of numerical data.

6 The MulTab system

MulTab is dedicated to managing large tables (20 000 cells) of numerical data along several output modalities, all based on the same output device, the screen. One main modality displays the entire table, M1, as shown in Figure 6. Because the cells are too small, the numerical data cannot be displayed. Therefore this modality is global and vague. Another modality, M2, is used to color each cell according to the numerical data. This modality is again global and vague but less vague than the previous one, M1. In addition this modality is arbitrary, because it is based on an arbitrary mapping between the colors and the data values. A slider at the bottom of the table (Figures 6 and 7) enables the user to define which cells are colored. This slider is a non-arbitrary output modality, M3, that explains the mapping function between the colors and the numerical data. Let us now consider the combinations between these three modalities. The combination between M1 and M2 is defined as follows:

- Temporal-Parallelism
- Spatial-Collocation
- Syntactic-Difference
- Semantic-Complementary

The combination of M2 and M3 is described as:

- Temporal-Parallelism
- Spatial-Adjacency
- Syntactic-Twin
- Semantic-Complementary

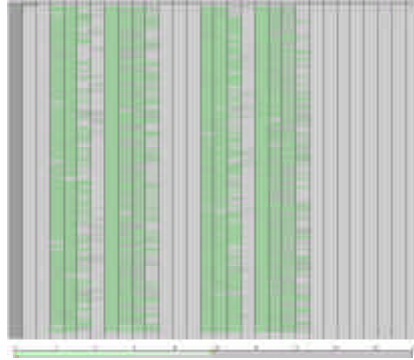


Fig. 6. Global view of the table and coloration of the cells.

In order to complement the vague modalities (M1 and M2), one local and deformed but precise modality is provided, as shown in Figure 7. This modality, M4, displays a part of the table with the numerical values of the cells. This modality is linguistic. The combination of modality M1 with modality M4 is described as:

- Temporal-Parallelism
- Spatial-Adjacency
- Syntactic-Extension
- Semantic-Complementary

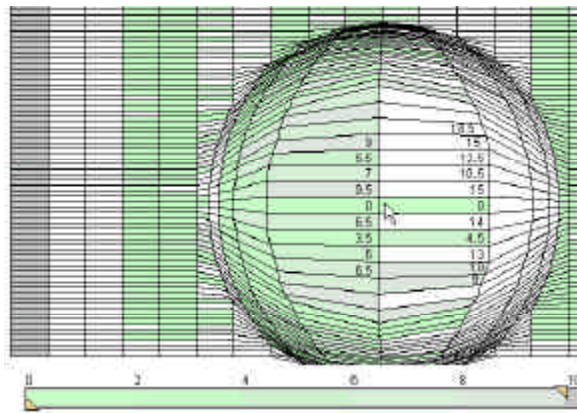


Fig. 7. A precise view (one modality) combined with the global view of the table (another modality).

Several precise modalities such as M4, can be used in parallel. As shown in Figure 8, multiple foci [17] within the table are useful for localization of a particular cell and comparison of the values of cells. A new focus is created from the main focus by direct manipulation. Each focus stems from the main focus. The foci move (lines and columns) as the main focus is moved. When the user moves the main focus, all the

related foci are automatically moved accordingly (spatial constraints). Such combination is described as:

- Temporal-Coincident
- Spatial-Variable, depending on the user
- Syntactic-Twin
- Semantic-Complementary

Variability in spatial combinations ensures coverage of all five schemas. This allows us to define the intersection spatial combination shown in Figure 9.

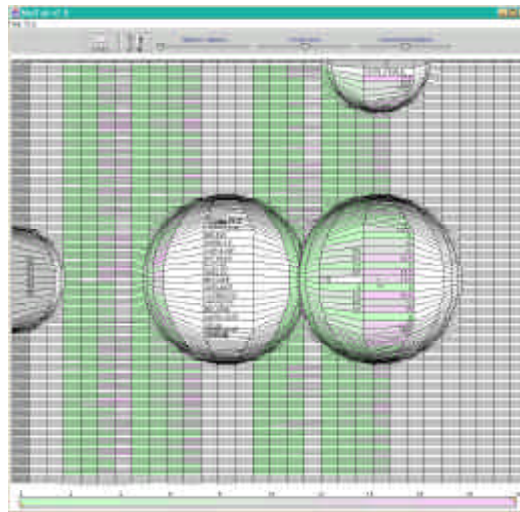


Fig. 8. Four foci within the table.

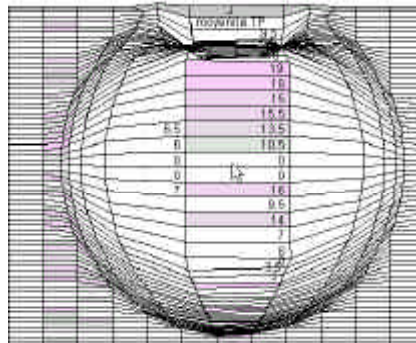


Fig. 9. Spatial intersection of two foci (two modalities).

In the design of MulTab we also studied the shape of the region of focus. The spherical shape of the focal region illustrated in Figures 7 and 8 is based on the fish-eye view study of [12]. In [10], two shapes have been experimentally compared in the context of the fovea system, shown in Figure 10. One focal region was circular (on the left in Figure 10) and one was rectangular (on the right in Figure 10). Results of the

experiment showed that the users preferred the circular shape but were more efficient using the rectangular shape. In our system VITESSE [24], we also experimentally studied the deformed shapes. One main result is that the users preferred analogue deformation such as the spherical shape instead of non-analogue ones such as our cartesian modality that does not correspond to an existing shape in real life.



Fig. 10. Circular and rectangular focal regions in the Fovea system.

Such experimental results prompted us to display different shapes of the region of focus in MulTab and to let the user select the one of his choice. Using a slider, the user can smoothly change the shape of the focal region from spherical to rectangular. In Figure 11, we present three implemented shapes: spherical and pyramidal shapes respectively on the left and on the right, and a hybrid shape in the middle.

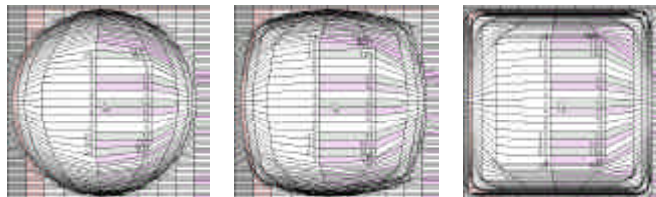


Fig. 11. Three shapes of the focal region in the MulTab system.

7 Summary of contribution and conclusions

We studied output multimodal interfaces from two points of view: the combination of modalities and the characterization of modalities. Our unified framework organizes in a coherent way the elements useful for the two key design issues of a multimodal output user interface: the selection of modalities based on their characteristics and the combination of modalities for the design of a coordinated output interface. Our framework is composed of two spaces. The first space, the combination space, is comprised of schemas and aspects: While the combination schemas define how to combine several modalities, the combination aspects determine what to combine. The second space, the characterization space, organizes the characteristics of a modality along three levels, articulatory, syntactic and semantic.

One contribution of our framework is to encompass and extend the existing design spaces for multimodality. However our combination and characterization spaces do not directly provide guidelines for the design of an efficient multimodal presentation. Our two spaces are the foundations for defining design rules. To identify design rules, we base our approach on ergonomic criteria. We have provided two design rules to illustrate our approach. Our future work will involve developing a coherent set of design rules based on our framework.

8 Acknowledgements

This work has been partly supported by the French Telecom-CNET Contract COMEDIR and by the French National Scientific Research Center (CNRS) with the SIIRI contract. Thanks to Emmanuel Dubois for his help in developing the MulTab system. Special thanks to G. Serghiou for reviewing the paper.

References

1. Allen, J.: Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, Vol. 26, No. 11, (1983) 832-843
2. André, E., Finkler, W., Graf, W., Rist, T., Schauder, A., Wahlster, W.: WIP: The Automatic Synthesis of Multimodal Presentations. In Maybury, M.T. (ed.): *Intelligent Multimedia Interfaces*. AAAI Press/MIT Press, Cambridge, Ma. (1993) 73-90
3. Arens, Y., Miller, L., Sondheimer, N.: Presentation Design Using an Integrated Knowledge Base. In Sullivan J., Tyler, S. (eds.): *Intelligent User Interfaces*. Frontier Series. New York: ACM Press (1991) 241-258
4. Beaudouin-Lafon, M., Gaver, W.: ENO: Synthesizing Structured Sound Spaces. *Proceedings of UIST'94*. ACM Press (1994) 49-57
5. Bernsen N.: A revised generation of the taxonomy of output modalities. *Esprit Project AMODEUS*. Working Paper RP5-TM-WP11 (1994)
6. Blattner, M.M., Dannenberg, R.G.: CHI'90 Workshop on multimedia and multimodal interface design. *SIGCHI Bulletin* 22, 2, (Oct. 1990) 54-58
7. Byte. Special Issue on Computing without Keyboard, (July 1990) 202-251
8. Coutaz, J., Nigay, L., Salber, D.: The MSM framework: A Design Space for Multi-Sensori-Motor Systems. *Proceedings of EWHCI'93*. *Lecture Notes in Computer Science*, Vol. 753. Springer-Verlag, Berlin Heidelberg New York (1993) 231-241
9. Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., Young R.: Four easy pieces for assessing the usability of multimodal interaction: The CARE properties. *Proceedings of Interact'95* (1995) 115-120
10. Coutaz et al.: CoMedi: Using Computer Vision to Support Awareness and Privacy in Mediaspaces. *Proceedings (Extended Abstract) of CHI'99*. ACM Press (1999) 13-16
11. Fitzmaurice, G., Ishii, H., Buxton, W.: Bricks: Laying the Foundations for Graspable User Interfaces. *Proceedings of CHI'95*. ACM Press (1995) 442-449
12. Furnas, G.: Generalized fisheye views. *Proceedings of CHI'86*. ACM Press (1986) 16-23
13. Gram, C., Cockton G. (ed.): *Design Principles for Interactive Software*. Chapter 2. Chapman & Hall (1984) 25-51

14. Hovy, E., Arens, Y.: On the Knowledge Underlying Multimedia Presentations. In Maybury, M.T. (ed.): Intelligent Multimedia Interfaces. AAAI Press/MIT Press, Cambridge, Ma. (1993) 280-306
15. Ishii, H., Ullmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. Proceedings of CHI'97. ACM Press (1997) 234-241
16. Lamping, J., Rao R., Pirolli P.: A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. Proceedings of CHI'95. ACM Press (1995) 401-408
17. Leung, Y., Apperley M.: A Review and Taxonomy of Distortion-Oriented Presentation Techniques. ACM Transactions on Computer-Human Interaction. Vol. 1, No. 2, (June 1994) 126-160.
18. Martin, J.C.: Six primitive types of cooperation for observing, evaluating and specifying cooperations. Proceedings of AAAI (1999)
19. May, J., Scott, S., Barnard, P.: Structuring Displays: a psychological guide. Eurographics Tutorial Notes Series. EACG: Geneva (1995)
20. Maybury, M.T. (Ed.): Intelligent Multimedia Interfaces. AAAI Press/MIT Press, Cambridge, Ma. (1993)
21. Neal, G., Shapiro, S. C.: Intelligent Multi-Media Interface Technology. In Sullivan J., Tyler, S. (eds.): Intelligent User Interfaces. Frontier Series. New York: ACM Press (1991) 11-43
22. Nigay, L., Coutaz, J.: A design space for multimodal interfaces: concurrent processing and data fusion. Proceedings of INTERCHI'93. ACM Press (1993) 172-178
23. Nigay, L., Coutaz, J.: A Generic Platform for Addressing the Multimodal Challenge. Proceedings of CHI'95. ACM Press (1995) 98-105
24. Nigay, L., Vernier, F.: Design Method of Interaction Techniques for Large Information Spaces. Proceedings of AVI'98. ACM Press (1998)
25. Oviatt, S.: Then myths of multimodal interaction. Communications of the ACM, Vol. 42, No. 11, (1999) 74-81
26. Stone, M., Fishkin, K., Bier E.: The Movable filter as a user interface tool. Proceedings of CHI'94. ACM Press (1994) 306-312
27. Vernier, F., Lachenal, C., Nigay, L., Coutaz, J.: Interface Augmentée par effet Miroir. . Proceedings of IHM'99. Cepadues (1999) 158-165