In: Z. W. Ras and L. Saitta (eds.) Methodologies for Intelligent Systems, 3 Amsterdam: North-Holland, pp. 101-111.

An Intelligent Multimodal Interface

Wolfgang Wahlster

Computer Science Department
University of Saarbrücken
6600 Saarbrücken 11, Fed. Rep. of Germany

Abstract

In face-to-face conversation humans frequently use deictic gestures parallel to verbal descriptions for referent identification. Such a multimodal mode of communication is of great importance for intelligent interfaces, as it simplifies and speeds up reference to objects in a visualized application domain. Natural pointing behavior is very flexible, but also possibly ambiguous or vague, so that without a careful analysis of the discourse context of a gesture there would be a high risk of reference failure. The subject of this paper is how the user and discourse model of an intelligent interface influences the comprehension and production of natural language with coordinated pointing, and conversely how multimodal communication influences the user and discourse model. We briefly describe the deixis analyzer of our XTRA system, which handles a variety of tactile gestures, including different granularities, inexact pointing gestures and pars-pro-toto deixis. We show how gestures can be used to shift focus and how focus can be used to disambiguate gestures. Finally, we discuss the impact of the user model on the decision of the presentation planning component, as to whether to use a pointing gesture, a verbal description, or both, for referent identification.

1. Introduction

In face-to-face conversation humans frequently use *deictic gestures* (e.g. the index finger points at something) parallel to verbal descriptions for referent identification. Such a *multimodal* mode of communication can improve human interaction with machines, as it simplifies and speeds up reference to objects in a visual world.

The basic technical prerequisites for the *integration of pointing and natural language* are fulfilled (high-resolution bit-mapped displays and window systems for the presentation of visual information, various pointing devices such as mouse, light-pen, joystick and touch-sensitive screens for deictic input, the DataGlove™ or even image sequence analysis systems for gesture recognition). But the remaining problem for artificial intelligence is that explicit meanings must be given to natural pointing behavior in terms of a formal semantics of the visual world.

Unlike the usual semantics of mouse clicks in direct manipulation environments, in human conversation the region at which the user points is not necessarily identical with the region which he intends to refer to. Following the terminology of Clark, we call the region at which the user points *the demonstratum,* the descriptive part of the accompanying noun phrase *the descriptor* (which is optional), and the region which he intends to refer to *the referent* [Clark83]. In conventional systems there exists a simple one-to-one mapping of a demonstratum onto & referent, and the reference resolution process does not depend on the situational context. Moreover, the user is not able to control the granularity of a pointing gesture, since the size of the predefined mouse-sensitive region specifies the granularity.

Compared to that, natural pointing behavior is much more flexible, but also possibly ambiguous or vague. Without a careful analysis of the *discourse context* of a gesture there would be a high risk of reference failure, as a deictic operation does not cause visual feedback from the referent (e.g. inverse video or blinking as in direct manipulation systems).

The subject of this paper is how the user and discourse model of an intelligent interface influences the comprehension and production of natural language with coordinated pointing to objects on a graphics display, and conversely how multimodal communication influences the user and discourse model. Before we review previous research on the combination of natural language and pointing and describe some current approaches related to our work let us briefly introduce the basic concepts of user and discourse modeling.

This is a condensed and revised version of my paper 'User and Discourse Models for Multimodal Communication', which appears in 'Sullivan, J.W., Tyler, S.W. (eds.) Architectures for Intelligent Interfaces: Elements and Prototypes. Reading: Addison-Wesley 1988.' The research was partially supported by the German Science Foundation (DFG) in its Special Collaborative Programme on AI and Knowledge Based Systems (SFB 314).

Methodologies for Intelligent Systems, 3
Z.W.. Ras and L. Saltta, Editors

2. User Models and Discourse Models

A reason for the current emphasis on user and discourse modeling [Wahlster86] is the fact that such models are necessary prerequisites in order for a system to be capable of exhibiting a wide range of intelligent and cooperative dialog behavior. Such models are required for identifying the objects which the dialog partner is talking about, for analyzing a non-literal meaning and/or indirect speech acts, and for determining what effects a planned utterance will have on the dialog partner. A cooperative system [Wahlster84] must certainly take into account the user's goals and plans, his prior knowledge about the domain of discourse, as well as misconceptions a user may possibly have concerning the domain.

We use the following definitions of user and discourse models [Wahlster88]: A *user model* is a knowledge source which contains explicit assumptions on all aspects of the user that may be relevant for the dialog behavior of the system. A *user modeling component* is that part of a dialog system whose function is to

- incrementally build up a user model and to maintain its consistency,
- to store, update and delete entries in it,
- and to supply other components of the system with assumptions about the user.

A *discourse model* is a knowledge source which contains the system's description of the syntax, semantics and pragmatics of a dialog as it proceeds.
A *discourse modeling component* is that part of a dialog system whose function is to

- incrementally build up a discourse model,
- to store and update entries in it,
- and to supply other components of the system with information about the structure and content of previous segments of the dialog.

While it seems commonly agreed upon that a discourse model should contain a syntactic and semantic description of discourse segments, a record of the discourse entities mentioned, the attentional structure of the dialog including a focus space stack, anaphoric links and descriptions of individual utterances on the speech act level, there seem to be many other ingredients needed for a good discourse representation which are not yet worked out in current discourse theory.

An important difference between a discourse model and a user model is that entries in the user model often must be explicitly deleted or updated, whereas in the discourse model entries are never deleted (except for forgetting phenomena). Thus according to our definition above, a belief revision component is an important part of a user modeling component.

3. Related Work on Deictic Input

Although in an intelligent multimodal interface the 'common visual world' of the user and the system could be any graphics or image, most of the projects combining pointing and natural language focus on business forms or geographic maps.

To the best of our knowledge, Carbonell's work on SCHOLAR represents the first attempt to combine natural language and pointing in an intelligent interface [Carbonell70], SCHOLAR, a tutoring system for geography, allowed simple pointing gestures on maps displayed on the terminal screen. NLG [Brown79] also combined natural language and pointing using a touch screen to specify graphics with inputs like (1).

(1) Put a point called Al here <touch>.

Woods and his coworkers developed an ATN editor and browser, which can be controlled by natural language commands and accompanying pointing gestures at the networks displayed on the screen [Woods79].

In SDMS [Bolt80] the user can create and manipulate geometric objects by natural language and coordinated pointing gestures. The first commercially available multimodal interface combining verbal and non-verbal input was NLMenu [Thompson86], where the mouse could be used to rubber band an area on a map in sentences like (2).

(2) Find restaurants, which are located here <pointing> and serve Mexican food.

All approaches to gestural input mentioned so far in our brief review were based on a simple one-to-one mapping of the demonstratum onto a referent and thus have not attacked the central problems of analyzing pointing gestures.

Recently, several research groups have addressed the problems of combining non-verbal and verbal behavior more thoroughly. Several theoretical studies and empirical investigations about the combination of natural language and pointing have been published [Hayes86, Hinrichs87,Reilly86]. Working prototype systems have been described, which explore the use of complex pointing behavior in intelligent interfaces.

For example, the TACTILUS subcomponent (designed and implemented by J. Allgayer) of our XTRA system [Kobsa86], which we will describe below in more detail, handles a variety of *tactile gestures*, including different granularities, inexact pointing gestures, and *pars-pro-toto deixis*. In the latter case, the user points at an embedded region when actually intending to refer to a superordinated region.

In the DIS-QUE system [Wetzel87] the user can mix pointing and natural language to refer to student enrollment forms or maps. The deictic interpreter of the T^3 system [Scragg87] interacts with a natural language interpreter for the analysis of pointing gestures indicating ship positions on maps and deals also with continuing or repeated deictic input. CTJBRICON [Neal88] is yet another system which handles simultaneous input in natural language and pointing to icons on maps, using language to disambiguate pointing and conversely.

While the simultaneous exploitation of both verbal and non-verbal channels provides maximum efficiency, most of the current prototypes do not use truly parallel input techniques, since they combine *typed* natural language and pointing. In these systems the user's hands move frequently back-and-forth from the keyboard to the pointing device. Note, however, that multimodal input makes even natural language interfaces without speech input more acceptable (fewer keystrokes) and that the research on typed language forms the basis for the ultimate speech understanding system.

4. An Intelligent Multimodal Interface to Expert Systems

XTRA (eXpert TRAnslator) is an intelligent multimodal interface to expert systems, which combines natural language, graphics and pointing for input and output. As its name suggests, XTRA is viewed as an intelligent agent, namely a translator who acts as an intermediary between the user and the expert system. XTRA's task is to translate from the high-bandwidth communication with the user into the narrow input/output channel of the interfaces provided by most of the current expert systems.

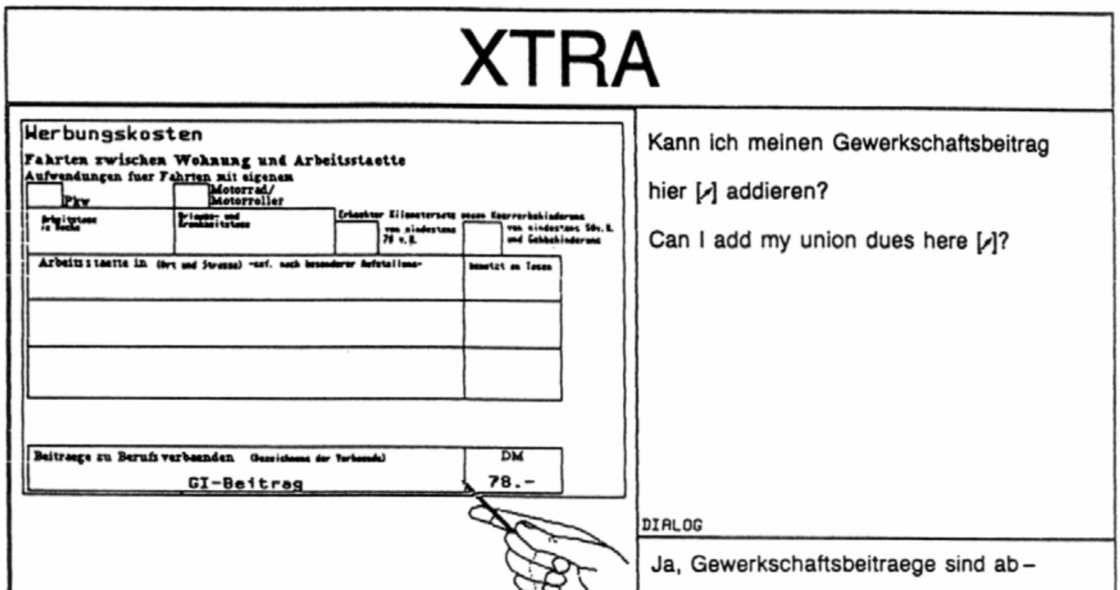

Figure 1: The Combination of Natural Language, Graphics and Pointing in XTRA

The present implementation of XTRA provides natural language access to an expert system, which assists the user in filling out a tax form. During the dialog, the relevant page of the lax form is displayed on one window of the screen, so that the user can refer to regions of the form by tactile gestures. As shown in figure 1, there are two other windows on the left part of the display, which contain the natural language input of the user (upper part) and the system's response (lower part). An important aspect of the communicative situation realized in XTRA is that the user and the system share a common visual field - the tax form. As in face-to-face communication, there is no visual feedback after a successful referent identification process. Moreover, there are no predefined 'mouse-sensitive' areas and the forms are not specially designed to simplify gesture analysis. For example, the regions on the form may overlap and there may be several sub-regions embedded in a region of the form.

In addition to the direct interpretation of a gesture, where the demonstratum is simply identical to the referent, TACTILUS provides two other types of interpretation. In a pars-pro-toto interpretation of a gesture the demonstratum is geometrically embedded within the referent. An extreme case of a pars-pro-toto interpretation in the current domain of XTRA is a situation where the user points at an arbitrary part (pars in Latin) of the tax form intending to refer to the form as a whole (pro toto in Latin). Another frequent interpretation of gestures is that the demonstratum is geometrically adjacent to the referent: the user points, for instance, below or to the right of the referent. Reasons for this may be the user's inattentiveness or his attempt to gesture without covering up the data in a field.

The user first chooses the granularity of the intended gesture by selecting the appropriate icon from the pointing mode menu or by pressing a combination of mouse buttons, and then performs a tactile gesture with the pointing device symbolized by the selected mouse cursor. The current implementation supports four pointing modes:

- exact pointing with a pencil
- standard pointing with the index finger
- vague pointing with the entire hand
- encircling regions

The deixis analyzer of XTRA is realized as a constraint propagation process on a graph which represents the topology of the tax form. A pointing area of a size corresponding to the intended granularity of the gesture is associated with each available pointing mode. A plausibility value is computed for each referential candidate of a particular pointing gesture according to the ratio of the size of the part covered by the pointing area to the size of the entire region. The result of the propagation process is a list of referential candidates consisting of pairs of region names and plausibility values.

Since pointing is fundamentally ambiguous without the benefit of contextual information, this list often contains many elements. Therefore, TACTILUS uses various other knowledge sources of XTRA (e.g. the semantics of the accompanying verbal description, case frame information, the dialog memory) for the disambiguation of the pointing gesture (see [Allgayer86] and [Kobsa86] for further details).

5. The Influence of Pointing Gestures on the Discourse Model

Pointing is not only used for referent identification but also to mark or change the *dialog focus*, i.e. to control or shift *attention* during comprehension. As we noted in section 2, focus is an important notion in a discourse model, since it influences many aspects of language analysis and production. For example, focus can be used to *disambiguate* definite descriptions and anaphora [Grosz81].
Figure 2 gives an example of the disambiguation of a definite description using a focusing gesture.

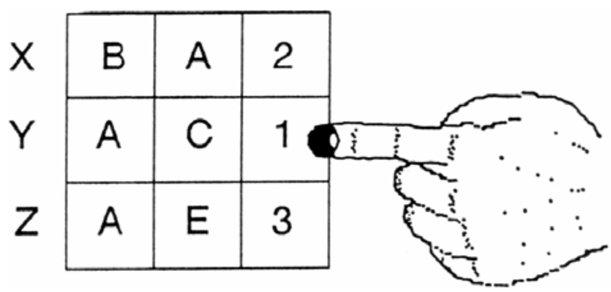

Figure 2: Focusing Gesture Disambiguating the Question 'Why should I delete the 'A"

Without focus the definite description 'the A' is ambiguous in the given visual context, since there are three objects visible which could be referred to as 'A' (one in each row of the table displayed in figure 2). Together with the pointing gesture at row Y, which marks this row as a part of the immediate focus, the definite description can be disambiguated, since there is only one 'A' in the focused row.

As in the case of gestures for referent identification, the effect of a focusing gesture can also be produced by a *verbal paraphrase*. For the example presented in figure 2, a meta-utterance like 'Now let's discuss the entries in row Y' would have the same effect on the discourse model and help to disambiguate the subsequent definite description.

As we noted earlier, without a discourse context most pointing gestures are ambiguous. In the example above, we have seen that a discourse context can be established not only by verbal information but also by gestures. Thus there is a twofold relation between gestures and focus. Gestures can be used to shift focus and focus can be used to disambiguate gestures.

From this follows that in *simultaneous pointing actions* two communicative functions of pointing can be combined: focus shifting and reference. The following two types of simultaneous pointing can be identified:

- One-handed input:
  - Focusing act: For example, the pencil is put down on the form, so that it points to a particular region on the form.
  - Referential act: A subsequent pointing gesture refers to an object in the marked region

- Two-handed input (see also [Buxton86]):
  - - Focusing act: For example, the index linger of one hand points to a region of the form.
  - - Referential act: The index finger of the other hand points to an object in the marked region.

Figures 3 and 4 illustrate the use of focusing gestures for the disambiguation of referential gestures. Note that in both situations displayed in figures 3 and 4 the index finger points at the same location on the form and that the utterances combined with these referential gestures are identical. The cases shown in both figures differ only in the location of the pencil which is used for focusing.

Let us explore the processing of these examples in detail. Since the referential gesture with the index finger is relatively inexact, TACTILUS computes a large set of possible referents. The head noun 'numbers' of the verbal description accompanying the pointing gesture imposes two restrictions on this set of possible referents. Since there are only four numbers displayed on the part of the form shown in Figs. 3 and 4, the semantics of the noun restricts the solution space to the power set of {3,4,7,5}, and the plural implies that only sets with at least two elements are considered in this power set. Finally, the position of the index finger on the form makes the interpretations {3,7,5}, {3,4,5}, {3,4,7} and {4,5,7} implausible, so that the resulting set of plausible referential readings becomes {{3,4}, {4,5}, {3,4,7,5}}, where {3,4,7,5} is a typical example of a pars-pro-toto reading.

This means that there remain three possible interpretations before we consider the focusing gesture. It is worth noting that this is one of the cases where the combination of verbal and nonverbal information in one reference act does not lead to an unambiguous reading. Here information from the discourse model helps to disambiguate. In figure 3 the pencil points at the row beginning with 'XYZ', so that this row and all its parts become focused. Now the intersection of the set of plausible referents and the currently focused objects results in the unique interpretation {3,4}. Similarly, in figure 4 the pencil is pointing at the block of columns called 'C3', so that the intersection of the focused elements with the results of the referential analysis is again a unique interpretation, namely {4,5}, but it differs from the set of referents found for the gestural input shown in figure 3. These examples once again emphasize the basic premise of our work, i.e. that pointing gestures must be interpreted in a highly context-sensitive way and that all approaches supposing a one-to-one mapping of the demonstratum onto the referent will fail in complex multimodal interactions.

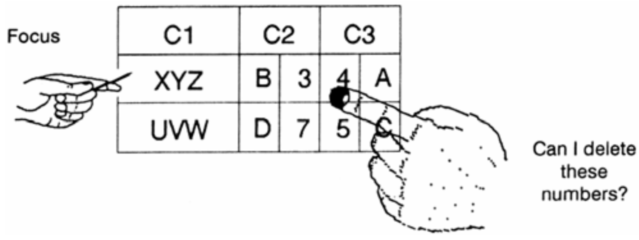

Figure 3: Simultaneous Pointing Gestures

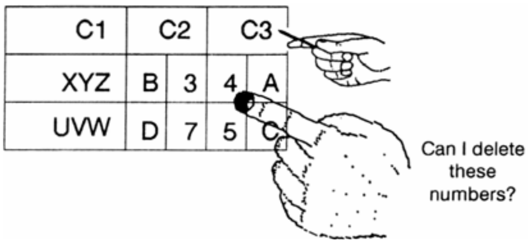

Figure 4: Simultaneous Pointing Gestures with Different Focus

6. User Modeling for Presentation Planning

As we noted at the outset, an intelligent interface should not only be able to analyze multimodal input, but also to generate multimodal output. The design of XTRA's generator allows the simultaneous production of deictic descriptions and pointing actions [Reithinger87]. Since an intelligent interface should try to generate cooperative responses, it has to exploit its user model to generate descriptions tailored to users with various levels of expertise.

One important decision which a multimodal presentation planner has to make, is whether to use a pointing gesture or a verbal description for referent identification. Let us explore the impact of the user model on this decision using an example from our tax domain.

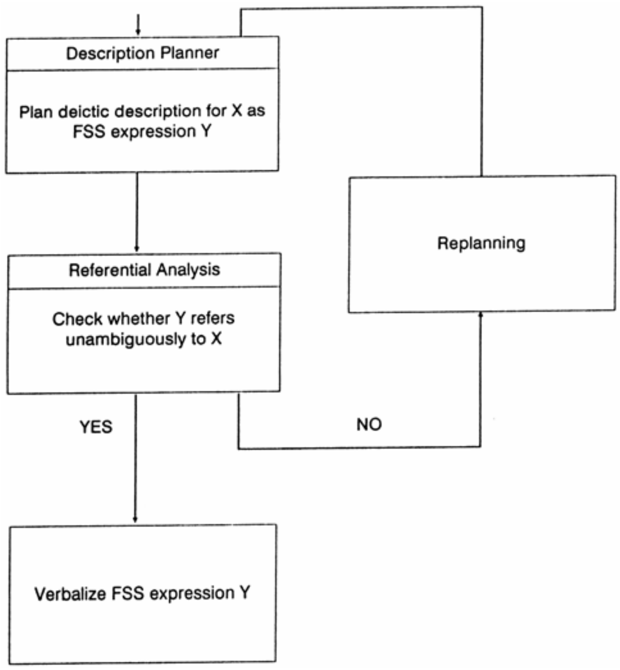

Figure 5: An Anticipation Feedback Loop for Presentation Planning

Suppose the system knows the concept 'Employee Savings Benefit' and an entry in the user model says that the current dialog partner seems to be unfamiliar with this concept. When the system plans to refer to a field in the tax form, which could be referred to using 'Employee Savings Benefit' as a descriptor, it should not use this technical term but a pointing gesture to the corresponding field. This means that in the conversational context described (3) would be a cooperative response, whereas (4) would be uncooperative.

(3) You can enter that amount here [↗] | in this [↗] field.

(4) You can enter that amount as employee savings benefit.

To summarize that point, if the system knows that a technical term which could be used to refer to a particular part of the tax form visible on the screen is not understandable to the user, it can generate a pointing gesture, possibly accompanied by a mutually known descriptor.

In the following, we discuss a particular method of user modeling, called *anticipation feedback*, which can help the system to select the right granularity of pointing when generating multimodal output. Anticipation feedback loops involve the use of the system's comprehension capability to simulate the user's interpretation of a communicative act which the system plans to realize [Wahlster86]. The application of anticipation feedback loops is based on the implicit assumption that the system's comprehension procedures are similar to those of the user. In essence, anticipation on the part of the system means answering a question like (5).

(5) If I had to analyze this communicative act relative to the assumed knowledge of the user, then what would be the effect on me?

If the answer to this question does not match the system's intention in planning the tested utterance, it has to replan its utterance, as in a generate-and-test loop.

Figure 5 shows an extremely simplified version of a multimodal description planning process with an anticipation feedback loop for user modeling. Let us assume that the generator decided to plan a deictic description of an object X, which the systems intends to refer to. The result of the description planning process is a an expression Y of the functional-semantic structure (FSS) together with a planned gesture. The FSS is a surface-oriented semantic representation language used on one of the processing levels of the how-to-say component of XTRA's generator.

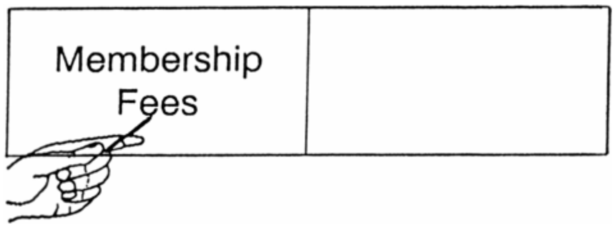

Figure 6: Planned Pointing Gesture

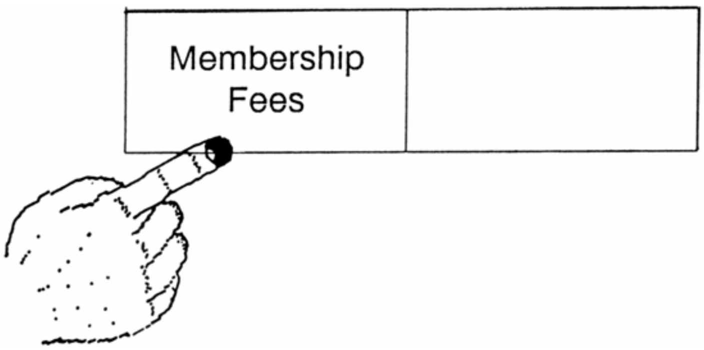

Figure 7: Pointing Gesture after Replanning

This preliminary deictic description is fed back into the system's analysis component, where the referent identification component together with the gesture analyzer TACTILUS try to find the intended discourse object. If the system finds that the planned deictic description refers unambiguously to X, the description is fed into the final transformation process before it is outputted. Otherwise, an alternative FSS and/or pointing gesture has to be found in the next iteration of the feedback process (see figure 5).

Now let us follow the feedback method as it goes through the loop, using a concrete example. Suppose that the system plans to refer to the string 'Membership Fees' in the box shown in figure 6. Also assume that the presentation planner has already decided to generate an utterance like 'Delete this [↗]' together with the pointing gesture shown in figure 6.

For a punctual pointing gesture the system first chooses the pencil as a pointing device. In this case, the exact position of the pencil was selected according to XTRA's default strategy described in [Schmaucks88]: the pencil is below the entry, so that the symbol does not cover it.

When this pointing gesture is fed back into the gesture analyzer of the referent identification component, the set of anticipated reference candidates might be {'Fees', 'e', 'Membership Fees'} containing only elements which can be 'deleted' (the current version of TACTILTJS does not deal with characters or substrings of a string). Since the system has detected that the planned gesture is ambiguous, it starts replanning and then selects the index finger icon as a pointing gesture with less granularity (see figure 7). This time, the result of the feedback process is unambiguous, so that the system can finally perform the pointing action.

7. Conclusions

We have shown how the user and discourse model of an intelligent interface influences the comprehension and production of natural language with coordinated pointing to objects on a graphics display, and conversely how multimodal communication influences the user and the discourse model.

First, we described XTRA as an intelligent interface to expert systems, which handles a variety of tactile gestures, including different granularities, inexact pointing and pars-pro-toto deixis, in a domain- and language-independent way. Then we discussed several extensions to the XTRA's deixis analyzer and presented our approach to generating multimodal output.

We showed how gestures can be used to shift focus and focus can be used to disambiguate gestures, so that simultaneous pointing actions combine two communicative functions: focus shifting and reference. We explored the role of user modeling for presentation planning and described how the user model can be exploited to generate multimodal descriptions tailored to the user's level of expertise.

Finally, we discussed anticipation feedback as a particular method of user modeling, which can help the system to select the right granularity of pointing when generating multimodal output.

References

[Allgayer86] Allgayer, J. and Reddig, C. 1986. Processing Descriptions containing Words and Gestures - A System Architecture. In Rollinger, C.-R. (ed.), *Proc. of GWAI/ÖGAI1986,* Berlin, Springer.

[Bolt80] Bolt, R.A. 1980. Put-That-there: Voice and Gesture at the Graphics Interface. *Computer Graphics*, 14, pp. 262-270.

[Brown79] Brown, D.C., Kwasny, S.C., Chandrasekaran, B., Sondheimer, N.K. 1979. An Experimental Graphics System with Natural Language Input. *Computer and Graphics,* 4, pp. 13-22.

[Buxton86] Buxton, W. and Myers, B.A. 1986. A Study in Two-Handed Input. *Proc. CHI'86 Human Factors in Computing Systems,* ACM, New York, pp. 321-326. [

[Carbonell70] Carbonell, J.R. 1970. *Mixed-Initiative Man-Computer Dialogues.* BBN Report No. 1971, Bolt, Beranek and Newman, Cambridge, MA.

[Clark83] Clark, H.H., Schreuder, R. and Buttrick, S. 1983. Common Ground and the Understanding of Demonstrative Reference. *Journal of Verbal Learning and Verbal Behavior*, 22, pp. 245-258.

[Hayes86] Hayes, P.J. 1986. Steps towards Integrating Natural Language and Graphical Interaction for Knowledge-based Systems. *Proc. of the 7th European Conference on Artificial Intelligence,* Brighton, Great Britain, pp. 436-465.

[Grosz81] Grosz, B. 1981. Focusing and Description in Natural Language Dialogues, in Joshi, A., Webber, B., Sag, I. (eds.), *Elements of Discourse Understanding.* New York: Cambridge Univ. Press, pages 84-105.

[Hinrichs87] Hinrichs, E. and Polanyi, L. 1987. Pointing The Way: A Unified Treatment of Referential Gesture in Interactive Discourse. *Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting,* Chicago Linguistic Society, Chicago, pp. 298-314.

[Kobsa86] Kobsa, A., Allgayer, J., Reddig, C, Reithinger, N., Schmauks, D., Harbusch, K. and Wahlster, W. 1986. Combining Deictic Gestures and Natural Language for Referent Indentification. *Proc. of the 11th International Conf. on Computational Linguistics,* Bonn, West Germany, pp. 356-361.

[Neal88] Neal, J.G., Shapiro, S.C. 1988. Intelligent Multi-Media Interface Technology. In *Proc. of the Workshop on Architecures for Intelligent Interfaces: Elements and Prototypes.* Monterey, Ca., pp. 69-91.

[Reilly85] Reilly, R. (ed.) 1985. *Communication Failure in Dialogue: Techniques for Detection and Repair.* Deliverable 2, Esprit Project 527, Educational Research Center, St. Patrick's College, Dublin, Ireland.

[Reithinger87] Reithinger, N. 1987. Generating Referring Expressions and Pointing Gestures. In Kempen, G. (ed.) *Natural Language Generation,* Dordrecht, Kluwer, pp. 71-81.

[Retz-Schmidt88] Retz-Schmidt, G. (1988): Various Views on Spatial Prepositions. In *AI Magazine,* Vol. 9, No. 2, also appeared as: Report No. 33, SFB 314, University of Saarbriicken, Computer Science Department.

[Schmauks87] Schmauks, D. 1987. Natural and Simulated Pointing. In *Proc. of the 3rd European ACL Conference,* Copenhagen, Danmark, pp. 179-185.

[Schmauks88] Schmauks, D. and Reithinger, N. 1988. Generating Multimodal Output - Conditions, Advantages and Problems. To appear in *Proc. of the 12th International Conference on Computational Linguistics*, Budapest, Hungary. Also appeared as Report No. 29, SFB 314, Computer Science Department, University of Saarbruecken.

[Scragg87] Scragg, G.W. 1987. *Deictic Resolution of Anaphora.* Unpublished paper, Franklin and Marshall College, P.O.Box 3003, Lancaster, PA 17604.

[Thompson86] Thompson, C. 1986. Building Menu-Based Natural Language Interfaces. *Texas Engineering Journal*, 3, pp. 140-150.

[Wahlster84] Wahlster, W. 1984. Cooperative Access Systems. *Future Generation Computer Systems*, 1, pp. 103-111.

[Wahlster86] Wahlster, W. and Kobsa, A. 1986. Dialog-Based User Models. In Ferrari, G. (ed.) *Proceedings of the IEEE*, 74, 7, pp. 948-960.

[Wahlster88] Wahlster, W. 1988. Distinguishing User Models from Discourse Models, Report No. 27, SFB 314, Computer Science Department, University of Saarbruecken, Fed. Rep. of Germany, to appear in Kobsa, A. and Wahlster, W. (eds.) *Computational Linguistics*, Special Issue on User Modeling, 1988.

[Wetzel87] Wetzel, R.P., Hanne, K.H. and Hoepelmann, J.P. 1987. *DIS-QUE: Deictic Interaction System-Query Environment.* LOKI Report KR-GR 5.3/KR-NL 5, Fraunhofer Gesellschaft, IAO, Stuttgart, Fed. Rep. of Germany.

[Woods79] Woods, W.A. et al. 1979. *Research in Natural Language Understanding.* Annual Report, TR 4274, Bolt, Beranek and Newman, Cambridge, MA, USA.

[Zimmermann87] Zimmermann, T.G., Lanier, J., Blouchard, C, Bryson, S. and Harvill, Y. 1987. A Hand Gesture Interface Device. *Proc. CHI'87 Human Factors in Computing Systems*, ACM, New York, pp. 189-192.