

A contextual analysis of referring gestures

Frédéric Wolff, Laurent Romary

► **To cite this version:**

Frédéric Wolff, Laurent Romary. A contextual analysis of referring gestures. Proceedings of the 4th International Conference on Intelligent User Interfaces - IUI'99, 1999, Los Angeles, CA, United States. inria-00098819

HAL Id: inria-00098819

<https://hal.inria.fr/inria-00098819>

Submitted on 14 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Contextual Analysis of Referring Gestures

Frederic Wolff
Loria
BP 239
54506 Vandoeuvre-les-Nancy
France
wolff@loria.fr

Laurent Romary
Loria
BP 239
54506 Vandoeuvre-les-Nancy
France
romary@loria.fr

REFERRING GESTURES IN MULTIMODAL INTERFACES

Multimodal interfaces offer to users different ways to express their intentions using the most appropriate means. For instance, to refer to the object involved in the mentioned action, the difficulty of naming determines the modality (speech or/and gesture) used to access the referents as shown by De Angeli. However analysis of referring gestures remains based on simple principles which constraint users' gestural expression to unnatural production rules. Our experimental study has shown that the variability of natural gestures is governed by visual context [1]. In this article, we present a contextual modelling of the analysis of spontaneous referring gestures used in multimodal interfaces. So it becomes possible to cope with the designations' diversity and offer a wider freedom of expression for the user.

REFERRING GESTURE AND VISUAL PERCEPTION

Various features of referring gestures are directly determined by the perceptual organization of the visual scene. Our corpus of multimodal dialogues recorded during Wizard of Oz experiments, presents a wider freedom of gestural expression than the usual pointing paradigm of classical interfaces [1].

Referring gestures used in coordination with verbal utterances build a taxonomy based on four categories as shown in figure 1. Whereas pointing still remains the most used gesture with 77% of all trajectories, circling movements appear with a frequency of 11%. 8% of all gestures defines the targeting category corresponding to a pointing covering a widespread area. The last type of gesture, scribbling, appears only seldom, 1%.

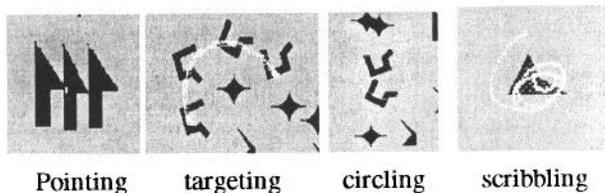


Figure 1: taxonomy of referring gestures

Each type of gesture is performed with a freedom in shape, precision, completeness and access strategy. For instance, it has been observed that the shape of referring

trajectories are produced on the one hand to fit the layout of the objects referred to, and on the other hand to avoid surrounding objects. User's movements are thus executed according to the perceptual environment in such a way that it becomes difficult to determine gesture type by considering only the trajectory shape. In the targeting example presented by figure 1, an out of (visual) context analysis of the gestural shape will lead to circling typing (and its corresponding referent searching heuristic) contrary to the user's intention. So we claim that a robust typing of gestures can only be performed by considering the visual context.

A CONTEXT DEPENDENT GESTURAL ANALYSIS

By considering the gestural trajectories and their position relative to the referred objects in our corpus, we observe two kinds of intention: gestures which aim the objects (elective gestures), and gestures which separate referents from the surrounding objects (separative gestures). Such alternatives reflect the use of visual space to access to referents. To determine the user's intention, we build a spatial partition of the visual scene as shown in figure 2. The first area is defined around each visual object, and corresponds to the elective part of visual space. The remaining inter-object space represents the visual area used by the user to carry out separative gestures like circling.

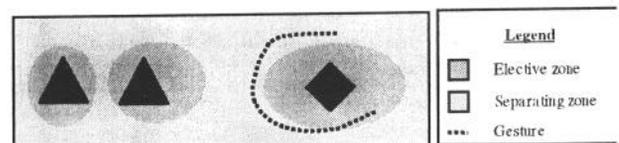


Figure 2: partitioning the gestural space

To determine the gestural intention of an entire trajectory, each point is typed according to its position in one of the two areas. The area most used by the point allows one to type the gesture and choose the heuristic adapted to retrieve the referents. For an elective gesture, the referred objects correspond to elective areas used by the user, whereas the referents of an separative gesture are found on the side of the trajectory which presents the greatest curvature.

REFERENCES

1. F. Wolff, A. De Angeli, L. Romary, 1998, Acting on a visual world: the role of perception in multimodal HCI, AAAI 98 Workshop on Representations for Multi-modal Human-Computer Interaction 98