Towards the design of

usable multimodal command languages

Noëlle Carbonell

LORIA, CNRS & INRIA

Overview

► Context and motivation

► Specific design issues in multimodal (MM) HCI

► Empirical evidence sources: 3 studies

► Recommendations:

  ➢ Spontaneous or constrained speech?

  ➢ Method for the design of acceptable speech and gesture Uis

  ➢ Other recommendations

N.B.: multimodal(ity) = speech + 2D designation gestures

Context and motivation

▶ Speech associated with designation gestures:
usually considered as a "natural" human means of expression

↘ only design recommendation: "emulate human communication"

▶ However:
poor usability of oral and multimodal user interfaces

⬇

Design issues:

Spontaneous or constrained speech/MM?

How to design acceptable oral/MM user interfaces?

Specific design issues for multimodal HCI

▶ In the context of HCI:

Is spontaneous speech with or without designation gestures
preferable to controlled oral/MM languages?

▶ In the case of controlled languages:

How to design acceptable oral/MM command languages
i.e.
artificial interaction languages which are both:

❶ interpretable easily and accurately by present recognition systems

❷ easy to learn, remember and use, hence usable (cf. Nielsen)

Empirical evidence sources

Presentation of the three studies: E0, E1, E2

▶ Common features

▶ Methodology

    ➤ First study

    ➤ Second and third studies

▶ The Wizard of Oz technique

▶ Methods for the analysis of oral/MM protocols

Common features

Three empirical studies with similar overall setups:

- ➢ Interaction with an oral/MM user interface capable of:
  - • "understanding" speech/MM inputs (speech + 2D gestures)
  - • issuing fluent oral/MM outputs (speech + graphics)

- ➢ Simulated user interface (Wizard of Oz technique)
  two wizards + software assistance

- ➢ Three weekly sessions (half an hour each or so)

- ➢ Subjects from outside the academic world (general public)

- ➢ Video-recording of subjects, wizards and screens

- ➢ Analyses of oral/MM protocols (from written transcripts)

Methodology – First study

?Influence of the context on users' behaviours?
Human dialogue vs voice human-computer interaction

▶ Complex hybrid approach (experimental framework + observation):

  ➢ Realistic interaction context: Information Centre

  ➢ Two groups → same tasks, same operator (human operator)

  ➢ Different instructions: talking machine vs human operator

⇘ Implementation problems: constraints on the wizard's behaviour

  • Stability (expression, dialogue, reactions)

  • Ambiguity (interpretable either as human or machine-like)
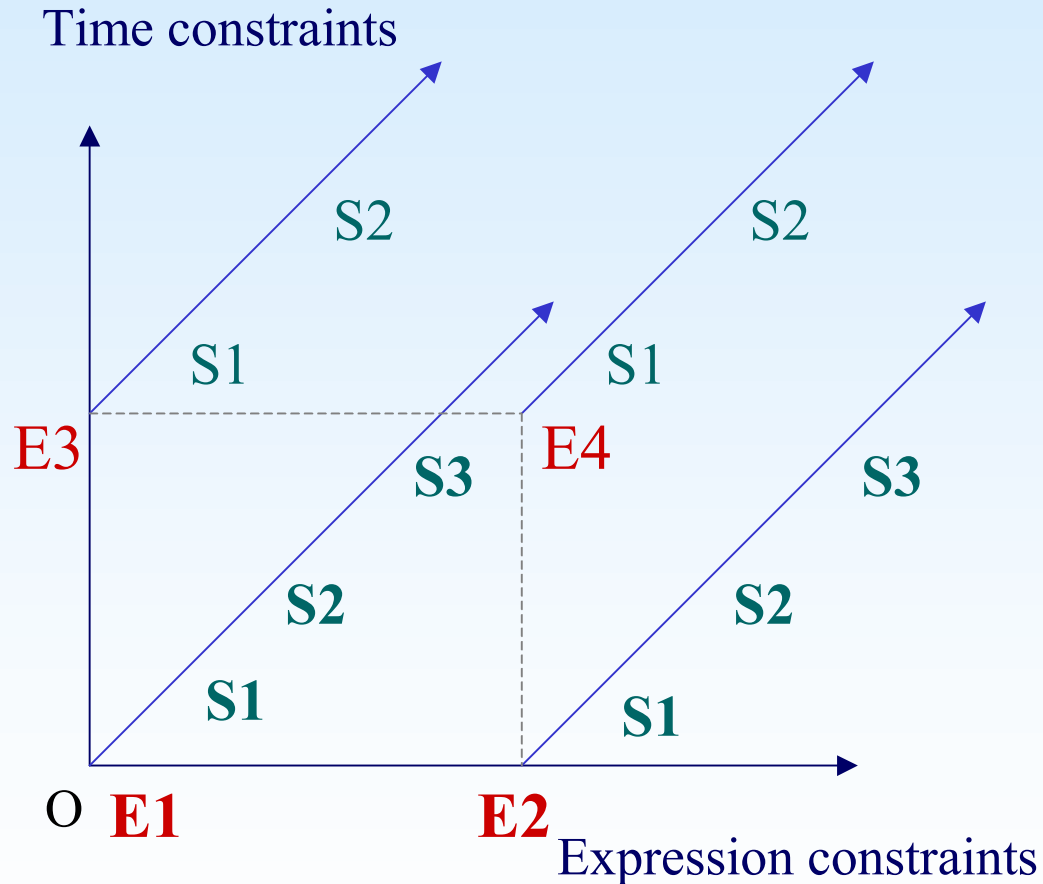
Protocol overall description – First study

[Amalberti, Carbonell, Falzon 93]

Influence of the dialogue context, H-H vs H-M interaction,
on users' expression, dialogue and problem solving strategies?

➤ Information requests to:

- A human operator, situation R

- A simulated Information Centre or "talking system", situation E

➤ Two groups of subjects, one per situation

➤ Six subjects per group

➤ Information retrieval scenarios of varying complexity

Methodology – Second and third studies

Time constraints

S2   S2

S1   S1

E3   E4

S3   S3

S2   S2

S1   S1

O **E1**   **E2**

Expression constraints

Four related empirical studies
(observation)

→ *multimodal interaction*

An experimental
research program

Variables:
- expression constraints (2)
- time constraints (2)
- experience (3)

Protocol overall description – 2nd & 3rd studies

[Carbonell, Dauchy 99] [Robbe *et al.* 00]

Definition of acceptable constraints on users' MM expression

↘ Comparative study:
  - E1, free expression
  - E2, controlled predefined MM language

➢ Free choice between speech, gestures and the combination of both

➢ Interaction with a graphical application → simple design tasks

➢ Eight subjects per study

➢ E2: use of a speech recognizer on the market
    → speech recognition + syntax checking (CF grammar)

Wizard of Oz paradigm

Wizard of Oz paradigm

▶ an efficient rapid prototyping method for simulating:

- ➤ NL understanding and
- ➤ "intelligent" dialogue control

▶ but difficult to implement:

- ➤ training
- ➤ software support
- ➤ coordination (2 wizards)

Wizards

E1 and E2 setups

11

Analysis of oral/MM protocols

▶ Quantitative analyses *per* session, *per* group, …:

  ➢ based on the coding of written transcripts of the protocols

  ➢ using standard and *ad hoc* taxonomies

▶ Qualitative analyses subject *per* subject:

  ➢ content analysis → protocols (ethnomethods)

  ➢ questionnaires

  ➢ debriefing interviews

N.B.: In new contexts of use, new interaction situations, qualitative empirical results are to be sought for, first and foremost, → the formulation of realistic hypotheses on users' behaviours, which will be validated later by quantitative experimental results

Recommendations for the design of acceptable oral/MM interaction

▶ Inferred from E1 and E2 mainly, but

▶ Motivated by, and based on the main scientific finding of E0:

The use of "natural" modalities of expression,
such as speech and designation gestures,
varies greatly according to the communication/interaction context:

Human communication/dialogue, or
Human-computer interaction

Spontaneous or controlled speech? – 1

▶ Spontaneous speech is more difficult to interpret reliably than controlled speech, since it encourages:

- ➤ Prosodic diversity, hence inter- and intra-speaker variability

- ➤ Wording/enunciation errors and slips

   cf. comparisons between E1 and E2 [Robbe *et al.* 97]

▶ The usability of spontaneous speech is questionable:

- ➤ Personalized telegraphic written style is preferred [Borenstein 86]

- ➤ Guidance for the wording of commands is requested (free speech)
   cf. E1 [Carbonell, Mignot 94]

   Spoken or written NL = an additional cognitive workload?

Spontaneous or controlled speech? – 2

Other drawbacks to the implementation of spontaneous speech:

➢ NL encourages misrepresentations of the software capabilities:
"Permute X and Y" (E1)

➢ Do users spontaneously control their expression in HCI contexts?
[Kennedy *et al.* 88] ≠ E0 [Amalberti *et al.* 93]

⬇

Possible influence of the system oral output
on the user's verbal expression
via his/her "system image"
[Sutcliffe, Old 87]

Design of acceptable MM user interfaces – 1

We propose a method for designing acceptable MM command languages, the relevance of which has been assessed by using it in the context of E2:

➡ The resulting interaction language proved easy to process reliably: CF grammar with low static and dynamic branching factors (5.5, 2.6)

➡ Expression constraints were easy to comply with and well accepted:

- rapid learning through interaction, without specific initial training

- no significant difference between the behaviors of E1 and E2 subjects during the first session (cf. interaction, task execution)

- direct manipulation favored only during the third session due to a dramatic degradation of speech recognition accuracy (mono-speaker recognizer, pattern recognition technique)

[Robbe *et al.* 97, 00]

Design of acceptable MM user interfaces – 2

Description of the method:

▶ Hypotheses:

➢ H1: verbal exchanges between operators are limited to a restricted subset of NL, which forms an operative sub-language

➢ H2: synonymy and polysemy/ambiguity are excluded from such a sub-language, the semantics of which is flexible

▶ Method:

1. Collection of spontaneous oral or MM interactions from potential users of the considered software

2. Then, elimination from this subset of NL, of all synonymous or polysemic words, phrases, structures

N.B.: This procedure guarantees that the semantics of the language is adequate

Other useful design recommendations – 1

▶ Some users will not spontaneously resort to synergic MM, that is:
"Put this there." + 2 designation gestures ($\rightarrow$ the object, its position)
They need appropriate inducement (cf. E1)

▶ In cases when MM inputs cannot be continuously interpreted accurately,
a robust alternative input modality should be provided (cf. E2)

▶ Departures from standard NL syntax and semantics are frequent in both controlled and spontaneous speech
These deviations or "errors" should be taken into account in the design of MM user interfaces (cf. E2)

Other useful design recommendations – 2

We observed two types of 2D gestures spontaneously associated with speech (cf. E1):

➢ Manipulation gestures

➢ Pointing gestures

corresponding to two different mental representations of the system and the interaction:

➢ Manipulation of graphical representations of the application objects

➢ Communication/cooperation with the system

Then, designers can influence users' representations of the system through the type of gestures they choose to implement

Subjects' adaptation to speech constraints – 1
First session of E2 – Quantitative results

Percentages of utterances including a given type of "error"

Enunciation slips:

➢ Hesitations   (50%)* 13,5%

Syntactic "errors":

➢ Syntactic structures $\notin$ L   5%

➢ Syntactically incorrect utterances   (26%)* < 3%

Semantic "errors":

➢ Vocabulary $\notin$ L   12%

➢ Words used with an inaccurate meaning   < 5%

High inter-individual diversity

* first session of E1

Subjects' adaptation to speech constraints – 2
First session of E2 – Qualitative results

▶ Incorrect utterances: How can they be remedied?

  ➢ Syntactic structures:
    VP + Prep. + NP (often in reformulations of unrecognized utterances)

  ➢ Semantics (inaccurate acceptation of a word/phrase mainly):
    "Move the armchair on (against) the wall."

▶ Utterances ∉ L: Easy to remedy

  ➢ Syntactic structures:
    "Move it (slightly) to the North slightly"

  ➢ Vocabulary: → use of synonyms in specific contexts
    "replace" vs "permute" (frequency of occurrence, familiarity)
    "kitchen, bedroom" vs "room" (use of hyponyms → precision)
    "lit" vs "lit-simple" (use of hyperonyms in anaphoric NPs)

Future research directions

The usability of "acceptable" oral or MM command languages could be increased by helping users to master their use during interaction

⬇

Three research directions are possible in order to achieve this goal:

➢ Extend the tolerance of present interpreters to "errors"

➢ Explore and exploit the potential of:

- alternative modalities ($\rightarrow$ error correction)
- and/or adaptability or adaptivity

➢ Implement "error" diagnosis and correction facilities into interpreters
= the most promising research direction, although the most difficult one