Workplan
(CLIPS contribution)

Domains:
1. Tourism
2. Help desk

Languages:
1. English
2. German
3. Italian
4. French

Work packages:

1. Management
2. Requirements
3. Show-cases Development
4. Human Language Technology (HLT) Modules
5. Testing and Evaluation
6. Dissemination

Partners:

ITC-irst. For Italian language and both domains:
Management
Requirements
CMU. For English language and both domains:
?
UKA. For German language and both domains:
HLT modules
CLIPS. For French language and both domains:
Responsible ? Dissemination
APT (Trentino tourism association):

AETHRA:
Show-cases development
ALCATEL:
- ?

In the following, the term *scientific partners* will indicate ITC-irst, CMU, UKA and CLIPS.

Work packages

Management

CLIPS : 4 man x months

Requirements

Tourism Domain Definition

CLIPS : The task of tourism negotiation consists in :

〈 transport accommodation (long distance and short distance for local visit),
〈 hotel reservation and restaurant presentation (by multimodal menus)
〈 tourist event information (event program presentation, reservation, accommodation, etc.)

Subtasks:

- Defining the multimodal requirements for the user interactions,
- Structuring the database (input and output),
- Recording the multimedia database (texts, audio and images) concerning the transport area, hotels and restaurants, tourist tours, visit programs.

4 man x months

Help-desk Domain Definition

The users are the partners having the necessary knowledge to define the domains in detail. They will provide all the material necessary to a complete conceptual description of the domain.

CLIPS : Negotiative or cooperative task execution among humans involves sharing perspectives about how to do the task, and about the task itself. Based on communication theories where the participants jointly build a common perspective, we will design a experiment involving human negotiative task-execution using multimodality. The corpus analysis will yield a taxonomy of the actors' actions (involving spoken utterances, gestures and combinations of both) and a dialogue model. Based on these two elements, we propose a representation of the user's perspective.

Subtasks:

- Eliciting the domain-based knowledge (static and dynamic)
- Eliciting the task knowledge (know how) GEOD : 2 man x months
- Definition of multimodal user interactions and structuring the database (input and output),
- Recording the multimedia database (texts, audio and images) concerning the help-desk domain.

6 man x months

Intermediate Representation Definition

The design of the IR is critical for the project success, as it determines the set of all and only the things which are to be considered. If it requires too complete a representation, then it will induce a failure every time something is missing, due to either disfluencies in the input or an error in the system. On the other hand, if something can not be represented in the IR, then there is no way of taking it into account. Fortunately, all the scientific partners have a successful preceding common experience which will surely be a very good starting point, both in positive for the things which worked and in negative, for the detected problems.

A point to be considered is whether we should include in the IR some multimodal information, what and in which way.

CMU has the best experience to guide the IR design. Nevertheless, all the scientific partners need to be involved in it, in order to have the best results.

CLIPS : will participate to the definition of the IR,

- At the semantic level,
- At the multimodal level (representing anaphora and deictics and more generally solving the reference problem related to gesture pointing).

6 man x months

Show-case development

The overall system hardware architecture needs to be designed in the perspective of the future real application of the system. Nevertheless, the show-cases architecture can accomplish the developments requirements, provided that an easy way to pass from the show-case design to the future system one is given.

The final system will consider a service provider, corresponding to the system user, that is the tourist destination in the first scenario, and the help-desk provider in the second one. Of course, it will be necessary that all the system is in the service provider site, as the requirements the potential client is required to fulfill should be minimum. The client need only to have a personal computer, WEB access and a browser in order to access the service.

On the other hand, as long as the system is under development, it is advisable that every module is as close as possible to the site which developed it and which is therefore responsible for its debugging and maintenance. Therefore, during development it is better to have a distributed architecture, where the HLT modules are in the site which developed them and only the IR and the material supporting the negotiation are exchanged.

First Show-case (tourism)

Second Show-case (tourism + help desk)

Hardware architecture

Software architecture

HLT modules

User study

In order to determine the requirements for the system, it is very important to deeply study the way in which the user performs the task the system should manage to perform. While the domain definition is independent from the language, the user study will focus on the language dependent aspects. This will permit both the design of the functional requirement of the system and the show-cases HLT modules evaluation. This study includes both interviews with the users and the study of data collected with their help, if possible directly on the field. If on the field collection will not be possible, then simulations will be planned, by strictly following their indications.

Each scientific site will collect and annotate the data needed for the development and evaluations of the modules it is responsible of.

CLIPS : CLIPS manages a platform, named MultiCom, for the design and the evaluation of user interfaces and more generally of interactive systems or products dedicated to the end user. MultiCom is a kind of "usability lab" which contains various distributed resources on network and dedicated software to capture and analyse the human working with a computer. It can be notice that very few generic usability labs exist now in the world. The methodology of MultiCom is based on the *cycle life* used in software engineering domain, by having multidisciplinary approach, unifying several methods on sociology, ergonomics, and computing engineering. The cyclelife starts from the original idea of product or system to need and ends by the evaluation after the complete exploitation by the client. The results show that it is very important to design all innovative product or service by integrating the end user in the "development loop" of the system. This point of view implies that the design process have to start from the situated work where a subject is observed in order to capture is behaviour. The data resulting of the capture stage are analysed in terms of usability factors defined a priori for the diagnostic and evaluation. Then, the results allows the designer to fix the precise system functionality. After the prototype realisation, a lot of users is observed again, and so on, then the prototype is modified until a convergence. At the end of the development cycle a complete evaluation can be provided not only on the final product but also on the design process itself.

Subtasks :

- Defining task scenarios,
- Selecting human subjects (at least ten people),
- Capturing their behavior on the MultiCom platform, in particular the usage of the multimodality,
- Analyzing the utterances produced and providing the vocabulary and the language model,
- Providing the activity tree and the task model.

6 man x months

HLT Modules

The system architecture is based on an Intermediate Representation (IR) independent from the source and target languages. The translation from language A to language B is then performed in two steps. First of all, the IR is built corresponding to the input utterance, in language A. Then, from the IR, the output utterance in language B is generated.

For every language, two chains need to be built: an *analysis chain* for building the IR from the input utterance, and a *synthesis chain*, for generating the speech utterance from the IF. The two chains are referred to as HLT Modules. Every scientific partner will use different approaches for realizing these modules: therefore, also their comparison will be taken into account.

Every site has the responsibility of building these modules for the language it considers: English for CMU, German for UKA, Italian for ITC-irst, French for CLIPS.

CLIPS will develop the analysis chain on Janus III for the speech recognition and ARIANE Y or and other tool for the translation from French. For the synthesis chain, CLIPS will develop a prosody generation 'dialogue centered' based on a commercial text-to-speech software.
One of the key challenges for robust recognition and translating is to cope with irregularities in the use of spoken communications and to cope with acoustic conditions. In particular, robust parsing is needed to deal with speech recognition errors occurring in spoken dialogue systems. Because of the restricted application domain of such systems, it is expected that a robust architecture can interpret a misrecognized utterance. When no parse can span the whole recognized utterance, a standard strategy consists in finding the best parses through the different hypotheses produced by the speech recognizer system. A special attention will be paid to the signal re-processing level in order to provide to the system a relative robustness in adverse conditions (bad quality of the microphone, user movements, cocktail party effect, etc.).

Subtasks:
- Adapting the acoustic models and language models for the application domains to Janus III, for French,
- Studying a signal pre-processing for the speech enhancement, based on source deconvolution,
- Integrating a bottom up robust parsing strategy based on Lexicalized Tree Grammars (when the parsing failed to span the whole utterance, the partial structures are combined through local rules in order to find the best),
- Integrating a semantic analysis based on case-grammar (see ATR)
- Integrating a multimodal grammar,
- Developing a IR to French generation module,
- Developing a prosody generation module based on the domain conversation rules,
- Integrating a commercial text-to-speech synthesizer.

24 man x months

Testing and evaluation

CLIPS : see above (the MultiCom platform)

Subtasks :
- Defining test scenarios,
- Selecting human subjects (at least ten people),
- Capturing their behavior on the MultiCom platform,
- Evaluating and inspecting their performances.

4 man x months

Dissemination

CLIPS :
- Publications in international congress (Eurospeech, ICSLP, IEEE, Coling, ACL)
- Publications in national congress (JEP, TALN, IHM)
- Publications in large diffusion journals (Science et Vie, La Recherche)
- Publications in scientific journals (SpeechCom, IEEE, ACM)
- Participation to the demonstration stands (ErgoIA, CHI, Interact)

- Publicity on the Web
- Publicity in the industrial world

Temporal plan

(number of month - task)

CLIPS : to complete after discussion with the partners
Total : 54 man x months

First showcase for tourism:

1.
2.
3. Requirements design (Tourism domain definition, user study and IR for tourism)
4.
5.
6.
7.
8.
9. Implementation (HLT modules and system integration)
10.
11.
12. **Testing/Evaluation** of the first tourism showcase

Second showcase for a more general tourism application and help desk:

13.
14.
15. **Requirements design** (Tourism domain extension definition, help desk domain definition, user study and IR for tourism extension and for help desk definition)
16.
17.
18.
19.
20.
21. **Implementation** (HLT modules and system integration)
22.
23.
24.
25.
26.
27.
28.
29.
30. **Testing/Evaluation** of the second complete showcase