Communicative Competence and Adaptation in a Spoken Dialogue System

Kristiina Jokinen

University of Art and Design Helsinki and University of Helsinki

Kristiina.Jokinen@helsinki.fi

Abstract

One of the much discussed topics in building spoken dialogue systems is how to take the users into account when designing practical systems: given the more complex environment in which we have to interact with various automatic services, it is obvious that the systems are not only required to function impeccably in regard to their technical specification, but they should also fulfill requirements concerning appropriate user needs. In this paper, some usability issues related to human factors in interface design are discussed from the point of view of communicative competence and adaptation.

1. Introduction

Dialogue management technology has already reached a level of maturity where dialogue systems are commercially viable. Various systems that use speech interfaces offer solutions for application areas ranging from telecommunications (electronic yellow pages, call routing, banking) to various customer services (information providers on weather information, airline and train time-tables, tourist information), and voice control technology for cars and home appliances. On the other hand, complicated interactive tasks have also emerged as new application areas for spoken dialogue systems: negotiation in eCommerce, question-answering and navigation in the web, interaction with complex systems in a ubiquitous computing environment, game interfaces. Moreover, there is a growing need to deal with multilingual and multicultural communication in situations where the participants do not share the same language or cultural backgrounds but need to communicate in order to accomplish a particular task.

A new metaphor for human-computer interaction has thus been introduced: the computer is not only a tool that the user must learn to use, but it is (a software) agent that interacts with the user and mediates between the user and the application (e.g. [1]). This view presupposes that interactive systems exhibit conversational abilities and can also adapt themselves to a broader range of users than so far (expert and novice users, businessmen and ordinary citizens, children and elderly people, users with special needs). Consequently, system design needs to observe and adhere to user requirements, preferences and needs. The integration of human factors into system design usually follows one of the three different approaches: user-centred design, ergonomics, and user modelling. In this paper, a common basis for the different approaches is found to be in the normal practices of communication between the partners. Speech-based human-computer interaction systems should thus incorporate requirements for the system’s communicative competence in them. Including the ability to communicate explicitly in the system architecture and component processes, the system’s usability as an interactive agent can be evaluated.

The paper is structured as follows. Section 2 discusses different approaches to user-oriented design in human-computer interaction and dialogue management. Section 3 introduces the notion of communicative competence as the main design objective for interactive systems. Section 4 gives a short overview of the evaluation of spoken language systems and discusses adaptation and communicative competence from the view point of the spoken language interaction system, AthosMail, developed in the project DUMAS [2]. Conclusions are drawn in Section 5.

2. Human aspects in design

2.1. User-centred design

The aim of this approach is to make end-products that are usable in the specified context of use, and thus emphasis is placed on the quality of interaction between the users and the products. Computer applications should be effective, efficient and satisfying to use, and focus is on the need to develop usable software. The development proceeds through iterative refinements whereby the users take part in the design and testing of the prototypes, and the feedback is then used to modify the system [3].

User-centred design activities include understanding and specification of the context of use. This requires the software be designed with reference to such characteristics of the intended users as their knowledge, skill, experience, habits, and motor-sensory capabilities. Moreover, the user’s tasks and the environment need to be understood, related to a wider context of organisational, technical and physical factors [3]. The approach also links to the Design for all standards for accessibility. These standards respond to the need to shift viewpoint from the traditional approach of defining elaborate technical product standards to the establishment of user requirements which are more compatible with a fast changing information society: towards the universal right to access and use digital information, taking particular care of the users with special needs.

2.2. Ergonomics

Ergonomics looks at the design process from the point of view of the human sciences of anatomy, physiology and psychology. Its goal is to apply scientific information to design objects, systems and environment so that the user’s capabilities and limitations are taken into account. Through understanding the demands placed on the user’s memory by the particular tasks, it is possible to design robust systems that are fit for use.
In human-computer interaction, ergonomics is not only related to the design of displays and keyboards, but also to the user’s cognitive load. The work on interface design has especially focused on system prompts which would successfully elicit required user input as an answer to the system question. The prompts should be clear and unambiguous, and provide the user with transparent information about the task as well as about the system’s capabilities. For instance, the 20 Laws of Interface Design, proposed by [4] introduce such aspects as linguistic clarity, simplicity, predictability, accuracy, suitable tempo, consistency, precision, forgiveness and responsiveness, to make the interface easy and transparent to use. The aim is to take human cognitive limitations into consideration, to fit to the user’s work and thinking, and to help the user to feel in control of the application. On the other hand, [5] advocates user-friendly interaction with an intelligent and rational system so that it is the system’s intelligence that provides the good interface and ergonomy of the service. Also [6] calls for a system that would adapt itself to the users’ behaviour by learning appropriate response patterns through interaction.

2.3. User modelling

The aim of conversational dialogue management is to improve human-computer interaction by studying human communicative capabilities and building models for their computational treatment. Previous research has focused on such aspects as the speaker intentions, speech and language errors, response planning, mutual beliefs, shared information, and rational cooperation. Individual users can be taken into account by constructing user models, explicit representations of the properties of a particular user. User models can vary from simple lists of user preferences to sophisticated models of knowledge intensive reasoning, and they are usually associated with system adaptation. E.g. [7] distinguishes static adaptation, which refers to options that the users can choose from when making decisions on interface aspects such as colour or sound, and dynamic adaptation, which refers to on-line adaptation through interaction, e.g. clustering of users on the basis of their similar navigation choices. Static adaptive features can be listed in personal profile files, but dynamic adaptation requires that the system is capable of observing user behaviour, and can learn from it. For instance, the user’s familiarity with the system functionality can be traced on-line, and the system responses adapted to the suitable competence level [8]. A realistic on-line adaptation would further require dynamic updates in the system’s knowledge-base, i.e. knowledge to be modeled as clusters of pieces of information that can change according to the interactions with the user.

In general, user models are assumed to enhance effectiveness and usability of software systems, and two opposite perspectives have been prevalent in developing them: one with the goal to endow computers with human-like abilities (human imitation), and the other where interaction should exploit the asymmetry between humans and computers and develop new interaction possibilities (human inspiration). As discussed above, fundamental asymmetry between humans and computers does not seem to be so clear any more, and the computers are more likely to be assigned properties that relate their functioning to human properties such as cooperation, cognitive and ethical considerations, and trust. However, asymmetry is encountered in dialogue participant roles: an information seeker vs. an information provider, a novice vs. an expert. Each role carries with it certain requirements on the competent participant, and rights and obligations with regard to the communication and other kinds of actions available for the participant [9]. The participants also possess attitudinal information and different knowledge about the activity, tasks and entities involved. This kind of asymmetry influences the type of dialogue acts the participants use, as well as their vocabulary. In other words, even though asymmetry between human users and computers does not seem to hold when considering interaction control, the communicative situation is always asymmetric.

2.4. Communication as a human factor

The different approaches all concern human factors and aim at building systems that would make the system to do the “right” thing at the “right” time in the “right” way [10]. Due to their historical roots, the approaches have focused on different aspects in the design process, and resulted in somewhat different practical activities. However, all of them emphasise interaction between the user and the software application, and place demands on the appropriate feedback that the user should get from the system so as to successfully exchange information with it.

Simultaneously to the need to integrate speech technology and human factors, also adaptation has become a visible notion in the interface design. Usually adaptation is considered to take place in the user who adapts to the system properties. The main goal of the system design is thus to make this kind of accommodation most feasible and easy for the users. Adaptation can also take place the other way round: the system observes the user and the environment, and adapts its behaviour accordingly. In natural language communication, both partners adapt: mutual adaptation is part of their communicative competence.

In the interface design, the main human factor thus seems to be the quality of communication, especially the quality of feedback that would allow both the user to adapt herself to the system, and the system to adapt its responses to the user. In addition to the step-by-step evaluation of the various system components, a holistic view of the success and evolution of communication is thus needed. In fact, interactions can be understood as long feedback chains, successive reactions to the changing environment where the system itself constitutes part of the varying context. If interface design is to facilitate communication between the user and digital information, the design process is an on-going learning activity between the user, the designer and the system, and its success is measured by the success of communication.

3. Communicative Competence

To construct communicatively adequate, user-oriented systems we need to answer at least the following questions:

- What are the requirements for such a system, i.e. definition of communicative competence?
- How can we operationalise the definition and find quantifiable measures for it which system properties are identified with the given definition?
- What kinds of evaluation measures are necessary to assess and compare different systems?

We answer the first question here, and the others in Section 4.
It has been suggested that the dialogue system's desired behaviour can be grounded on the notion of cooperative and rational communication [9,11]. The desiderata for a communicatively competent, spoken language interaction system include the following:

- Physical feasibility of the interface
- Efficiency of reasoning components
- Natural language robustness
- Conversational adequacy

3.1. Interface

Communication has fundamental constraints on physical and biological levels (the user's eye-sight and hearing, capability to use hands to type or move the mouse, etc.), and they form the first important requirement for flexible, user-friendly interaction systems: enablements for communication should be available. Usually these factors are linked to ergonomics of the interface, or gathered under the design principles of usability (the system is easy to use, easy to learn and easy to talk to) and transparency (system capabilities and limitations should be evident to the user from experience). Different input and output modalities also contribute to the system’s robustness, especially from the Design-for-all point of view [12]. In the DUMAS project [2], the AthosMail e-mail application is built for mobile phones and desktop use, and the design of high-quality speech input and output pays special attention to visually impaired users.

3.2. Efficiency

Dialogue studies show that even if the users acknowledge the naturalness of system responses, they regard long response times as the main factor in characterizing differences between computer dialogues and human-human conversations. Given that feedback is essential in smooth communication but the dialogue partner fails to reply in a reasonably short time, the speaker normally starts to think that contact has been lost. In computer interaction, the situation may cause unnecessary turn takings, and usually ends up with deep frustration. To allow real-time operation and maintain the system's reliability and user-friendliness, various processing heuristics are used besides looking for faster and more efficient algorithms: predicting most likely next utterances, classifying different discourse phenomena as frequently or rarely occurring and providing a fast response method for the former, while allowing more processing for the latter, etc. A new solution is to use means of narration and story telling to design interaction patterns where the necessary input processing is interleaved with presentations to the user.

3.3. Natural language robustness

Linguistic sophistication has often been questioned as esoteric and rather unimportant in interface design: the number of words used in human-computer interactions is rather limited, and in some cases natural language enhancements have been considered as obscuring the task achievement. Negative conclusions may be due to constrained underlying tasks and/or simple recognition technology. Considering elaborated interactive systems and their communicative capability, quality language interpretation and generation components are of course essential. They increase the system’s usability by reducing such unintended problem solving situations like “what would be the best way to put the questions so that the system would understand me”.

Research on spoken language interfaces has also identified several sources of linguistic variability, and experimented with models accounting for these phenomena (e.g. disfluent speech and hyperarticulate pronunciation [13]). Extensive use of dialogue information also suggests that the role of a natural language front-end is to be redefined: much of the interpretation of the user input, as well as generation of the system output, takes place in the reasoning components distributed in the system’s dialogue and user models.

3.4. Conversational adequacy

Conversational adequacy manifests itself mainly in the contributions that the system produces to clear up vagueness, confusion, misunderstanding, or lack of understanding occurred in user contributions. An intelligent dialogue agent is capable of conducting mixed-initiative dialogues, negotiating correct meanings for references and vague expressions, and of producing helpful, cooperative responses that would enable partners to take steps towards the fulfillment of the underlying task. In DUMAS, the focus has also been on the adaptation to the user’s skill levels so as to formulate appropriate responses.

The requirement calls for development in computational techniques. E.g. in DUMAS, software agents and a distributed architecture are used to provide implementation tools, and machine-learning techniques are applied to build models for conversational phenomena. Integration of the techniques into practical dialogue systems is an active research area and more research is needed to determine which parts of dialogue management will benefit most from the results obtained through the new models. A better understanding of human conversation is also necessary, especially to catch relevant features in adaptation. As already argued in [6], one of the main issues in the information society is the problem of knowledge acquisition, and thus the need for on-line learning and adaptation becomes relevant: the systems should learn through interaction, update their knowledge bases, and adapt their output to new situations appropriately.

4. Adaptation and evaluation

Success in human-human communication is usually attributed to the shared satisfaction that the underlying goal of the communication was achieved. The goal may vary from a vague “keep the channel open”-type wish to chat with the partner, to a clear aim to get a particular task or activity completed. In any case, successful communication requires that the participants converse over a discussion topic and subtopics associated with it, and know how to apply their communicative competence to negotiate when the underlying goal has been fulfilled.

One of the spoken dialogue system evaluation frameworks is the PARADISE framework [14] which allows comparison of systems across tasks and domains. The goal is to evaluate user satisfaction, which includes maximizing task success and minimizing dialogue cost. The latter comprises efficiency measures and qualitative measures, both of which include subjective dialogue quality as an important performance factor alongside objective measures such as elapsed time, success rate, and the number of utterances needed to complete a task. When considering the quality of speech technology services, [15] points out that the measures
should take both the system design and the perception of the user into account. User satisfaction is not only based on the system's efficiency and error-free processing of the utterances, but also on the perceived qualities that increase the user's comfort and ease of communication. The evaluation taxonomy contains dialogue, communication, task and service related quality features.

A problematic aspect in assessing speech-based interaction is the evaluation of adaptation that is regarded as essential in the system's communicative competence. Adaptive systems usually perform better than non-adaptive systems concerning user satisfaction and task success [14]. However, assessment of the adaptation itself is not that clear: it is generally preferred that the system adapts to the user, but it is also considered necessary that the user has a transparent view of what aspects are included in the adaptation. Our experiments show that the users have fixed preconceptions of the system and how to use it, and they do not venture exploring the limits of the system, even though they are encouraged to do so [7].

In DUMAS, a three-level user model is used to experiment with the system adaptation to the user's expertise level [8]. The system monitors various parameters in its interactions with the user, and estimates the expertise level to be realized in the system response. The on-line component takes care of the immediate adaptation in the on-going dialogue, while the off-line component calculates long-term learning effects between the sessions. The levels are calculated for each system act separately, so as to allow the user be expert in some part of the system and novice in another part. In practice, when the perceived user expertise rises, the system removes unnecessary information elements from its responses, up to the level of giving initiative to the user.

Our user studies show that the model performs adequately, but to verify its functionality, a longer usage period is needed. Assessing the success of adaptive contributions is not a simple task. Communicative competence can, however, be related to the quality features as in [15]. The responses are justified by the requirements of efficiency and fluency, and common measures such as ASR confidence scores, mean concept understanding, cancels, interruptions, timeouts and help-requests are used. Also the dialogue strategy, or the way in which mutual knowledge is established, maintained, modified and exploited is available: system responses address communication problems explicitly, and the user's error type classification is used as a reference point for expertise modeling. Finally, the appropriate level of feedback can be attained by involving the user in the system design-development-evaluation cycle. We engage would-be and expert users to provide insights of the system design. Besides usability testing with a prototype, we allow walkthroughs and experiments with alternative system designs.

5. Conclusions
This paper has discussed various user-oriented requirements for adaptive spoken dialogue systems, and it advocates the view that interactive systems should be evaluated in terms of communicative competence. This includes assessment of the external enablements of communication, as realized in the physical feasibility of the interface, and the system's fast reaction times. Moreover, the system should address internal enablements for competent communication, namely robust natural language processing and conversational adequacy.

6. Acknowledgements
The work is partly sponsored by the EU Information Society Technologies Programme under contract IST-2000-29452, DUMAS, www.sics.se/dumas. Thanks to all project participants from KTH and SICS, Sweden; UMIST, UK; ETEx Sprachsynthese AG, Germany; and U. Tampere, U. Art and Design Helsinki, Connexor Oy, and Timehouse Oy, Finland.

7. References