Proceedings of the Institute of Acoustics

COHERENCE AND REDUNDANCY IN HUMAN-COMPUTER DIALOGUE

M Karpiński

Institute of Linguistics, Adam Mickiewicz University, Poland

1. INTRODUCTION

During the years of the personal computer boom, most of the research concerning human-computer dialogue (HCD) was devoted to technological issues. Now, Graphical User Interface (GUI) seems to be reaching the ceiling of its evolution. The quality of speech synthesis has become astounding and the efficiency of restricted-vocabulary continues speech recognition has reached an impressive level (at least for English language). Most of the researchers agree, however, that there is still much to do to improve human-computer interaction - the weakest point of the global information processing chain. Many view speech as a powerful future tool of HCD (see: Arons & Mynatt [1]). Nevertheless, some issues limit the widespread of Speech User Interfaces (SUI). First of all, speech c a n n o t be used just anywhere (e.g., noisy offices and shops) and s h o u l d n o t be used just anywhere (e.g., the places where silence should be respected or the situations in which privacy is a matter of importance). Secondly, SUI is not very well developed yet. It needs more than high quality speech synthesis and accurate recognition to work efficiently (Zue [2], Karpiński [3]). The speech channel in HCD is frequently regarded as merely an addition to what GUI can offer.

The aim of the present research was to find whether computer users employ any of the communication strategies and patterns commonly met in natural languages and to look for important features of their communicative behavior. Two similar experiments were prepared to check for these strategies and patterns in interaction with GUI as well as with SUI. The number of subjects was limited because the research was conceived only to determine the direction of future, more detailed studies. Nevertheless, the observations suggest that HCD is influenced by two strong factors: users' need for coherence and for redundancy. The factors in question are not merely linguistic entities, because they encompass almost all the components of the communication process.

2. COHERENCE AND REDUNDANCY IN NATURAL LANGUAGE

Many researchers perceive coherence and redundancy as crucial features of natural languages. Coherence seems to be a crucial issue of discourse analysis (e.g., Schiffrin [4], Blakemore [5], and Stubs [6]). In fact, it can be regarded as one of the main defining feature of discourse. Some insightful remarks on coherence viewed as a psychological process can be found in Gernsbacher & Givón [7] or Kintsch [8]. The psychological nature of coherence and its extremely wide basis makes it an universal notion which can be used to analyze HCD.

Redundancy in various language subsystems has been studied by quantitative linguists in detail (e.g., Frick & Sumby [9], Altman & Lehfeldt [10], Stanley [11]). Let us notice that the level of redundancy may indicate the style of a piece of text. Surprisingly, in formal texts, it can be very high as well as very low. The redundancy in natural language is a multidimensional phenomenon: low redundancy in one dimension may be compensated by higher redundancy in a certain other dimension (a low volume level of an utterance may be compensated by repeating it twice, extremely clearly, or by using vocabulary of an unique and easily recognizable phonetic form).

Certainly, there exists a strong but complex relation between the two factors in question. Some linguists do attempts to generalize it. Givón [12] describes the repetition as the case of maximal redundancy (maximal redundancy = maximal coherence), but he also claims that interpretable texts "must fall somewhere in the middle between total redundancy and utter incoherence". Indeed, the texts comprised of multiple repetitions are rare and of doubtful practical meaning. It seems reasonable that there should exist a point of balance between redundancy and coherence in natural language utterances, but much more research is needed to transform such statements into specific testable hypotheses and to make a practical use of them.

One may expect that people speaking natural languages do some effort to regulate the force of coherence and the dose of redundancy so as to make their utterances most efficient from the viewpoint of the communication process. Various situations demand to set the point of this balance at different places, i.e., the level of redundancy and the force and quality of coherence must be contextually adequate. Linguistic and communicative competence obviously include certain mechanisms that allow speakers and hearers obtain this adequacy in a co-operative process.

Programming languages and most of the languages and codes used to communicate with computers are marked by extremely low redundancy. Any deviation from their rules may lead to misunderstanding. Keeping the redundancy low may ensure highly effective communication, but leaves no place for errors. The developers of UIs should provide ways to obtain higher levels of redundancy and coherence, because this would make computer users feel more comfortable and self-assure in HCD.

3. HCD USING ICONS AND SPEECH: EXPERIMENTS

Two groups, each of 25 intermediate and semi-advanced computer users, were asked to take part in the experiments. While the groups were not identical, their profiles concerning relevant features remained very similar (they were obtained using selection inquiries). The task in both the experiments was to key in and edit a short passage of text. All the participants were Poles, the text and the spoken commands were in the Polish language.

In Experiment I, subjects used icons to convey commands to the program. There were no menus visible on the screen. In Experiment II, speech was the means of conveying main commands to the system and the screen was deprived of icons and menus. The mouse and the keyboard were available in both the experiments, but the key-sequences for editing functions were disabled.

The environment for the experiments was developed within the MS Windows operating system, using a relatively popular, highly configurable word processor and a simple, commercial voice-recognition and speech-synthesis program. The quality of the synthesized speech was average, but none of the subjects had any difficulties understanding it. The speech recognition engine offered only speaker-dependant, isolated-phrase recognition of average accuracy. The set of commands available was strictly defined by the experimenter. Subjects read them twice aloud when teaching the speech recognition software. The list of these commands was visible on the screen all the time, as were the icons in Experiment I. During the experiment, the system produced three basic reactions to subjects' utterances: (a) repeating the command as it was understood (which frequently differed from what was actually uttered); (b) asking the subject to repeat the command if it was assessed as not clear enough; (c) no reaction at all (this happened when commands or other utterances were too quiet).

4. MAIN FINDINGS

4.1. General strategies: gaining more feedback

In Experiment I, more skilled subjects tended to employ a 'step strategy', making extensive use of block-marking. They keyed in longer strings of the text and then proceded with editing. This allowed them to obtain higher feedback, although the number of elementary operations they needed to complete the task was higher then in a 'linear strategy'. The latter was used mostly (but not only) by less advanced subjects. One may expect that more advanced users realized that there were alternative ways of completing the task and, in their choices, they were guided not only by the level of simplicity.

This was not the case in Experiment II, in which subjects were offered additional spoken feedback. After an uttered command was processed by the recognition system, the synthesizer repeated it in the form in which it was identified. For example, when the command "switch bold" was recognized by the machine as "switch italics", then the synthesizer generated the latter utterance. This occurred to be an important source of feedback and probably it strongly influenced the strategies employed by the subjects.

The number of subjects was to small to implement more advanced statistic methods (e.g., factor analysis) to determine in more detail the relevant factors which influenced the strategies applied by them.

4.2. Strategies to deal with recognition errors

The participants were relatively aware of what the speech recognition system offered. They were instructed to speak naturally and not to artificially separate words in spoken commands when teaching it as well as when performing their editing tasks.

The experimenter prepared the set of available commands so as to lower the actual speech recognition accuracy level of the program. It was achieved by making some basic commands sound relatively similarly (e.g., first three of five syllables were identical). Therefore, the situations of misrecognition were frequent (although a few voices were recognized very well). Subjects faced the problem of pronouncing the phrase in a way which would make the computer 'understand' it. Five types of subjects' reactions to misrecognition and rejection of commands were observed:
(a) Multiple repetition of the same command (i.e., the same command was repeated a few times without any meaningful changes).
(b) The command was repeated in another form (i.e., its lexical components were changed; for example, 'activate italics' was changed to 'switch italics').
(c) The command was repeated with a certain change in prosody. This change could be conscious as well as unconscious. Some subjects intenentionally varied intonation, while others did it spontaneously, e.g., out of impatience or anger. The most common reaction of this type was speaking louder and slower.
(d) Some keen subjects tried to make use of the fact that the computer's errors occurred in certain patterns.
For example, if the command 'switch bold' was initially interpreted by the computer as 'change margins', they intentionally tried to make use of the former to obtain the result of the latter.
(e) Some subjects gave up one course of achieving their aims and looked for other, more efficient or simply more plausible methods.

A general tendency to keep the prosodic features of commands unchanged was observed especially among male subjects. Only one of them declared that he consciously tried to modulate the prosody of his commands. He was also the only person who tried to speak faster to make the computer recognize his instructions. The rest of the few female and male subjects who used conscious modulation seemed to be certain that only slower pronunciation was effective. However, the time-normalizing mechanism of the speech recognition engine was relatively inflexible and speaking too slowly also resulted in recognition errors.

The modulation of prosodic features was especially impressive in some female voices. After several unsuccessful attempts to make the computer understand a given command, a few subjects began to loose their patience. This found its expression in the intonation change. Two female subjects produced very peculiar intonation contours. The first syllable of the second word in two-word commands was extremely stressed, which is atypical in this kind of Polish phrases, unless the speaker shows high emotionality or particularly wants to stress a certain word.

As mentioned before, the subjects probably realized that any deviation from the intonation contour they applied when teaching the speech recognition system might result in a recognition error. Therefore, the reasons which made them modulate the intonation were seemingly strong. Acounting for this fact only by emotional factors seems to be an oversimplification. A more elaborate explanation may be that prosodic features play a role in ensuring the coherence of the mental model of discourse. They may appear to be redundant when it comes to speech recognition, but necessary to facilitate certain mental processes.

4.3. Additional utterances: In search of coherence and redundancy?

Many subjects went beyond the limited set of available commands and pronounced 'additional utterances' of various size and content. Since they had it clearly explained to them that the computer could 'understand' only the given set of commands and that any other word pronounced aloud may prompt the machine to perform an undesirable action, they must have had compelling reasons to do so (as was the case in intonation changes). Their 'additional utterances' varied in character. Generally, in many cases they may be viewed as very close to 'discourse markers' (as defined by, e.g., Schiffrin [13] and discussed by many others, e.g., Lenk [14], Hovy [15]; see also Craig & Tracy [16]). Some of them expressed impatience, anger or other emotions, while others seemed to play a syntactic role. In several cases subjects used them as if to achieve coherence of their utterances/action (certain sequences of actions in HCD were treated as utterances). For example, some participants in the study combined their actions with words like 'or' and 'or no'. Such phrases were clearly redundant from the viewpoint of the UI. Moreover, they posed a danger of unpredictable actions of the computer, because they might be interpreted as commands. However, the subjects needed them for some reasons. These 'joining phrases' seemed to be produced to keep the coherence of some larger structures of subjects' utterances. Coherence should be regarded here as a mental entity or process, which is found in the mental representation of discourse, but not (or at least not fully) in the text of a conversation.

Another interesting phenomenon was observed in two female subjects. In several first turns, after each command and its repetition by the machine, they added the word 'Accept' or 'Abort', e.g.:

> Subject: Switch margin width.
> Computer: Switch margin width
> Subject: Accept.

or

> Subject: Switch margin width.
> Computer: Switch bold.
> Subject: Abort.

This could be viewed as a way of obtaining - by the means of increased redundancy - a higher certainty of being 'understood' by the machine. The exchanges listed above display much similarity to typical three-turn exchanges in natural language conversation (e.g., as described by Sinclair & Coulthard [17] or Tsui [18]). They are typical for certain formal situations and teacher-student interactions.

5. FINAL REMARKS

The juxtaposition of the results of these two experiments leads to the conclusion that the presence of speech in UI influences the process and the structure of HCD. Speech channel added to standard GUI offers not only one more way to control the computer: it is also a source of additional feedback.

It is clear that computer users try to employ familiar conversational devices in their interaction with PCs (Karpiński [3], [19]). They frequently make both conscious and unconscious use of prosodic change. Many contemporary speech recognition systems treat most of the prosodic features as redundant. However, the amount of information conveyed by the means of prosody can be impressive. The fact of being allowed to freely make use of it may be of high psychological importance. Antypically modulated prosodic features in utterances addressing computers seem to be interesting for several further reasons. Firstly, they indicate that natural language mechanisms are readily moved up to the platform of HCD. Secondly, they show that these mechanisms are in some ways adapted to match - according to subjects' sense - a new communicative situation. Thirdly, as mentioned above, the research concerning these changes seems to be of great practical importance.

To sum up, the designers of UI for PCs should take into account that computer users are prone to apply various methods of increasing and balancing redundancy and coherence in their utterances addressed to computers. As a sequel to the findings described above, the author decided to design a more detailed study, centered on the prosody of the speech directed to computers, which - hopefully - will result in at least a partial explanation of its peculiarities.

The participants were asked to shortly comment on the experiment. Although there were many enthusiasts among them, their critical remarks addressed the crucial problems faced by the developers of SUIs, e.g.,
'The task would be easier if the computer was able to understand commands pronounced in various ways (various tone, various words).'
'Such an experiment proves that computers are not perfect yet, i.e. we cannot treat them as partners.'

6. REFERENCES

[1] B ARONS & E MYNATT, 'The Future of Speech and Audio in the Interface', *SIGCHI Bulletin* 26 p44-48 (1994).
[2] V ZUE, 'Conversational Interfaces: Advances and Challenges', *ESCA, Eurospeech97*, KN9-KN18 (1997).
[3] M KARPIŃSKI, 'Psycholinguistics Aspects of Human-Computer Communication'. (unpublished Ph.D. thesis available at the Institute of Linguistics, Adam Mickiewicz University; 1998).
[4] D SCHIFFRIN, *Approaches to Discourse*. Cambridge-Oxford: Blackwell (1994).
[5] D BLAKEMORE, 'The Organization of Discourse', In: F Newmeyer (Ed.), *Linguistics. The Cambridge Survey*. Cambridge, p229-250 (1988).
[6] M STUBS, *Discourse Analysis*. Chicago: University of Chicago Press (1983).
[7] M A GERNSBACHER & T GIVÓN (Eds.), *Coherence in spontaneous text*. Philadelphia-Amsterdam: John Benjamins Publishing Co. (1995).

[8] W KINTSCH, 'The Psychology of Discourse Processing'. In: M A GERNSBACHER (Ed.) *Handbook of Psycholinguistics*. San Diego - N.Y. - Boston: Academic Press, pp. 721-737 (1994).
[9] F C FRICK & W H SUMBY, 'Control Tower Language', *Journ. Acoust. Soc. Amer.* 24 p595-596 (1952).
[10] G ALTMAN & W LEHFELDT *Einfuhrung in die Quantitative Phonologie*, Bochum (1980).
[11] R STANLEY, Redundancy rules in phonology. *Language*, 43 p393-436 (1967).
[12] T GIVÓN, 'Coherence in text vs. coherence in mind', In: M. A. Gernsbacher, T. Givón (Eds.), *Coherence in spontaneous text*. Philadelphia-Amsterdam: John Benjamins Publishing Co. (1995).
[13] D SCHIFFRIN, 'Discourse markers', In: W Bright (Ed.) *International Encyclopedia of Linguistics*. New York- Oxford: Oxford University Press (1992).
[14] U LENK, 'Discourse markers and conversational coherence', In: B Warvik, S K Tanskanen & R Hiltunen (Eds.) *Proceedings from the Turku Conference*, p341-352 (1995).
[15] E HOVY, 'The Multifuctionality of discourse markers', Paper presented at Workshop on DM, Emgmond am Zee, Holland (1995).
[16] R T CRAIG & K TRACY (Eds.), *Conversational coherence: Form, structure, and strategy*. Beverly (1983).
[17] SINCLAIR & COULTHARD, *Towards an Analysis of Discourse*. Oxford: University Press (1975).
[18] A B M TSUI, *English Conversation*. Oxford: University Press (1994).
[19] M KARPIŃSKI, 'Conversational devices in human-computer interaction using WIMP UI', In: R Murawski & J Pogonowski (Eds.), *Euphony and Logos*, Amsterdam-Atlanta: Rodopi B.V., p345-362 (1997).