A COMPARISON BETWEEN FEEDBACK STRATEGIES IN HUMAN-TO-HUMAN AND HUMAN-MACHINE COMMUNICATION

Loredana Cerrato

Center for Speech Technology, Dept.of Speech Music and Hearing
KTH, Stockholm, Sweden
Graduate School of Language Technology, Sweden
loredana@speech.kth.se

ABSTRACT
The results of a comparative analysis of feedback strategies used in human-human communication and human-machine communication in Swedish are reported in this paper. The aim of this study is twofold: to provide a categorization of feedback expressions based on contextual information and to verify the hypothesis that acoustic characteristics of feedback expressions can be regarded as cues to the interpretation of the function they carry out and the communicative intention they convey.

1. INTRODUCTION
With the recent development in speech technologies, conversational speech interfaces are becoming more advanced and more and more users expect to be able to interact with their computer systems in the way they do with other people. This means that users of speech-based interfaces tend to integrate a larger number of human discourse features into their speech and expect the interface to be able to produce human-like behavior.

Examples of human discourse features used during human-machine interactions are feedback expressions. In human-human conversation, dialogue participants continuously inform each other on the state of the communication. This is done by means of different kind of verbal and visual expressions that have different functions and convey different attitudinal reaction depending on the context in which they occur.

Feedback strategies have been studied to get insight about conversational human behavior both in human-human communication [1,2,3] and in human-machine communication [4,5,6].

Different definitions of feedback have been given and even if they differ somehow in their formulation, they all seem to agree on the fact that feedback strategies are used as a “cooperative” way of exchanging information about the efficiency of the communication.

In the framework of virtual conversational systems feedback strategies are generally grouped in two main classes: positive and negative. Positive feedback is given when the user wishes to show that s/he understands and/or agrees with what the system says while negative feedback is produced when the user wishes to signal some problems of understanding and/or disagreement. Recent analyses of human-computer interaction have shown that users of multimodal dialogue systems produced a surprisingly large number of feedback expressions, despite the fact that the virtual agents of the systems never explicitly elicited nor gave them acknowledgements in the course of the dialogues [5].

This paper reports the results of a comparative analysis of feedback strategies in four human-human interactions and in four human-machine interactions. Notwithstanding the different communicative situations it was possible to observe common ways of expressing feedback.

Feedback expressions were categorized both on the basis of contextual information and acoustical characteristics. The results of the acoustical analysis supported the hypothesis that acoustic characteristics of feedback expressions, such as duration and pitch contour, can be regarded as cues to the interpretation of the function and the communicative intention conveyed by feedback.

2. MATERIAL
Video-recording of four real dialogues between four different customers (two women and two men) and a travel agent (always the same woman), were selected from The Gothenburg University Spoken Language Corpus [7]. These dialogues can be described as “information seeking” exchanges, where the customer asks the travel agent for information about time tables, visas, hotels and so on and the travel agent provides the information required.

Four human-machine interactions were selected from the Adapt database [8], which consists of a total of 50 dialogues produced by 33 users recorded by means of the Wizard of Oz technique.

Adapt is a Swedish conversational multi-modal dialogue system, able to provide information about real estate in Stockholm. The dialogues in the Adapt database can also be described as “information seeking” or “task oriented” since the users were given the task of finding apartments in Stockholm that fulfilled certain criteria.

The results of the analysis of positive and negative users’ feedback in the Adapt database [5] showed that 94% of the users used feedback at least once in their interaction with the system, even if large individual variations were noted (i.e. some users gave more feedback than others). For this study the four users (three male and one female) were randomly selected from those dialogues in which the number of utterances labeled as including feedback was at least 10%.
3. METHOD

In order to be able to categorise feedback expressions it is necessary to take into account contextual information. In this study feedback expressions are interpreted and categorised in terms of reactions to the previous communicative act. Thus an expression is categorised as feedback if its primary function serves one of the following purposes:

1. show continuation of contact: when the interlocutor wishes to show that s/he is willing and able to perceive the message and continue the interaction.

2. Acknowledge comprehension: when the interlocutor wishes to show that s/he is able to understand the message. Acknowledgement is here meant in the same sense proposed by Clark and Schaefer (1989) that is to describe a hierarchy of methods that interlocutors might use to signal that a contribution has been understood well enough to allow the conversation to proceed.

3. Express a point of view: when the interlocutor wishes to show her/his opinion by agreeing or disagreeing with what the speaker says and by expressing some emotion, like surprise or disappointment.

Positive and negative answers to yes/no questions are labelled as feedback expressions only if their main function serves one of the above mentioned purposes.

To these 3 groups it is possible to add different subcategories deriving from the way in which feedback is expressed. For instance under the category of acknowledgement of comprehension we could have the following subcategories:

- short expressions: monosyllabic morphemes like ja, ne;
- repetition or reformulation of part or all of the speaker’s last utterance
- gestures and facial expressions: like head nods, head shakes, shoulder and hands shakes, smiles, gaze, blinking.

The acoustic characteristics of selected feedback expressions produced in the dialogues were measured using the software package Wavesurfer (Sjölander & Beskow 2000).

4. RESULTS

4.1. Distribution of feedback expressions in human-human interactions

In Figure 1 the distribution of feedback expressions per dialogue and per interlocutors (namely, A: travel agent and C: customer) is reported in percentage to the number of total individual turns. For the travel agent also the expressions used to elicit feedback are reported on the graph.

It is striking that the number of feedback expressions of the customers reaches more than 50% in all the dialogues. In particular in dialogue 2, which was the shortest of all, 87% of the customer’s turns consisted of feedback expressions!

Given the particular case of communicative situation, it is mostly the customer that has the role of the listeners and the travel agent that has the role of the speaker. The travel agent has in fact the right to maintain the turn until she supplies the information appropriate to the needs of the customer. As a consequence the customers/listeners use a considerable amount of feedback to show continuation of contact. This is because they wish to show their “active” participation to the interaction.

However when the roles change and the customer gets the turn to formulate a request or to make a comment, the travel agent also gets the opportunity to give feedback. Otherwise the travel agent uses strategies to elicit feedback.

The most common strategy adopted by the travel agent to elicit feedback is to ask an explicit verification question, like for instance: “did you say you wanted to travel on Monday?”

Another way consists of repeating or reformulating what the customer has said in her/his last turn, as we can see in the following example:

A: how many of you are travelling?

C: we are two

A: you are two

C: yes

Fig. 1 Distribution of feedback per interlocutors in human-human dialogues

4.2. Distribution of feedback expressions in human-machine interactions

In Figure 2 is reported the distribution of feedback expressions of the 4 users per dialogue.

Fig. 2 Distribution of feedback per user in human-machine dialogues

The amount of feedback produced by the users is variable, it ranges from 16% to 60%, and however these figures are striking if we consider that the virtual agent never elicited nor gave explicit feedback.

---

1 The examples reported here are translated from the original transcription in Swedish. The original transcription conventions adopted in the corpus is not faithfully reported here.
4.3. Categorization of feedback expressions

The categorization of feedback expressions in the 3 above-mentioned categories provides an interpretation of feedback expressions in terms of type of reaction to the previous communicative act. This categorisation is applicable both to the feedback expressions produced in the human-human dialogues and to those produced in human-human interactions.

In Figures 3 and 4 the feedback expressions produced by the customers and the travel agent are grouped according to the 3 categories.

These expressions will be referred to as "minimal non-intrusive responses" since they occur on a turn of their own and do not seem to provide new information to the conversation, nor to express the points of view and the attitudinal behaviour of the interlocutors, they rather serve the purpose to assure the speaker that the information has been conveyed successfully.

As suggested in [1] to interpret these minimal non-intrusive responses it is important to consider how the "polarity" of the preceding communicative acts affects their function. In the following example, extracted from dialogue 2, we can see that the function of the expression "nä" is to acknowledge the acceptation of the information, since the preceding statement has a negative polarity:

A: then I don’t know how the prices are going to be next year
C: nä

For the customers these short expressions occur 90% of the time in a turn of their own, for the travel agent in 80% of the time.

The customers use more expressions to show continuation of contact (category 1) whereas the travel agent uses more expressions to acknowledge the comprehension of the message (category 2).

It is easy to interpret this result in terms of cooperation strategies: the customers show active listening when the agent gives them information and the travel agent acknowledges the understanding of the message when the customer request some more information, or provide their personal information for the booking.

In the four human-machine dialogues the short expressions never occur in a turn of their own, they occur instead at the initial position of a longer utterance and their function is never to simply show continuation of contact, but mostly to acknowledge the reception of information.

4.4 Acoustical characteristics of feedback expressions

The "minimal non-intrusive responses" in the human-human dialogues show the following acoustic characteristics:

- a lower intensity, compared to the intensity of other utterances of the speaker;
- an average duration of 290 ms with a standard deviation within and across speakers not higher than 50 ms;
- a flat pitch contour.

This characteristic flat pitch contour might depend on the fact that these expressions do not convey any particular attitudinal information. Being non-intrusive minimal responses, one would
expect that they should overlap with the turn of the “main” speaker. On the contrary they are systematically produced in appropriate points of the production of the main speaker, namely in correspondence with a short pause and mostly at the end of a grammatical clause. If overlap occurs, it is only partial, because the feedback expression always starts before the start of the main speaker’s new clause.

When speakers wish to convey a particular point of view with their feedback expressions they do not simply produce short expressions, but they tend to:

- produce reduplicated expressions, such as: *mm, jaja, jaha.*
- add a “reinforcing” expression like *visst, precis, juste.*

When *mm* is produced with a falling-rising contour it indicates that the message has been understood and the customer also shows a surprised attitude. The same attitude can be conveyed by *jassu* and *jaha,* as in the following example, where the customer acknowledges the information of getting a paid stopover in Paris with surprise:

A: you have to change flights in Paris and there is a stopover, but you get that paid

C: jaha!

There were not enough realizations of these short expressions produced at the beginning of a longer utterance in the human-human dialogues, however those few ones showed the following characteristics:

- longer duration: average duration across speakers ranges from 450 to 650 ms
- rising pitch contour.

In the human-machine dialogues the short expressions were produced always at the beginning of a longer turn and they showed the following characteristics:

- average duration across users ranging from 450 to 750 ms
- rising pitch contour.

To sum up the results of the acoustical analysis we can interpret them as follow:

- short duration and flat pitch contour manifest an active listening attitude and show that listeners do not wish to interrupt the speaker to take the turn;
- longer duration and rising pitch contour show an active listening attitude and the listeners’ intention to take the turn.

5. DISCUSSION

Notwithstanding the differences between human-human and human-machine interactions, it was possible to find some common ways of expressing feedback; in particular the expressions used to acknowledge the reception of information appear to be used in both kinds of interactions. In human-human interactions they are produced by both the interlocutors, in the analyzed human-machine interactions the system did not produce any explicit acknowledgements.

A particular kind of short feedback expression which was referred to as “minimal non-intrusive responses” was produced in a large number of turns in human-human interactions, but never appeared in a turn of their own in the human-machine dialogues. The simplest explanation could be given in terms of the way in which the system was designed, in fact the system did not record what the users said while the agent was speaking, as a consequence it is not possible to find short overlapping feedback expressions in the recordings. Another explanation can be given taking into account the results of a more detailed analysis of the human-human dialogues [10] that showed that both interlocutors tended to produce more verbal and visual feedback expressions in response to visual contact or in response to a visual elicitation feedback. This kind of visual contact or visual elicitation is not possible to have with the embodied conversational agent in the dialogue system.

6. CONCLUSIONS

Further investigation on larger material is going to be carried out to verify these preliminary results and hopefully provide results that could be used in an effective way in the development of multimodal dialogue systems.

Among current speech-based interfaces there are some systems that can elicit feedback under the form of explicit or implicit verification questions, but none of the existing systems is able to give the users explicit acknowledgement signal after processing their utterances. This lack of acknowledgement can leave the users puzzled since they do not understand if the system is actually processing their request or is stuck.

Since feedback expressions have such an important role in communicative interaction and since both participants in a dialogue produce them in a high number, it is very important that speech-based interfaces, and in particular multi-modal dialogue systems displaying conversational agents, should be able both to recognize users’ feedback expressions and to produce them.

7. REFERENCES


8. ACKNOWLEDGEMENTS

I would like to thank Prof. Jens Allwood for letting me use the The Gothenburg Spoken Language Corpus.