# Positive and Negative User Feedback in a Spoken Dialogue Corpus

*Linda Bell and Joakim Gustafson*

Centre for Speech Technology, Royal Institute of Technology, Stockholm, Sweden

## ABSTRACT

This paper examines feedback strategies in a Swedish corpus of multimodal human–computer interaction. The aim of the study is to investigate how users provide positive and negative feedback to a dialogue system and to discuss the function of these utterances in the dialogues. User feedback in the AdApt corpus was labeled and analyzed, and its distribution in the dialogues is discussed. The question of whether it is possible to utilize user feedback in future systems is considered. More specifically, we discuss how error handling in human–computer dialogue might be improved through greater knowledge of user feedback strategies. In the present corpus, almost all subjects used positive or negative feedback at least once during their interaction with the system. Our results indicate that some types of feedback more often occur in certain positions in the dialogue. Another observation is that there appear to be great individual variations in feedback strategies, so that certain subjects give feedback at almost every turn while others rarely or never respond to a spoken dialogue system in this manner. Finally, we discuss how feedback could be used to prevent problems in human–computer dialogue.

## 1. INTRODUCTION

As conversational speech interfaces become more advanced and human-computer dialogues appear more "natural", we may expect users of spoken dialogue systems to integrate a larger number of human discourse features into their speech. In human-human conversation, dialogue participants continuously give each other positive and negative feedback as a way of showing attention, recognizing the intention what the other conversant is saying or to signal nonunderstanding or misunderstanding. In the present paper, we examine a broad range of feedback phenomena observed in a multimodal dialogue corpus. The multimodal AdApt system is designed to provide users with information about apartments in downtown Stockholm, and for the purposes of the present study a semi-simulated version of the system was employed. Despite the fact that this system never gave the subjects any explicit acknowledgements in the course of the dialogues, positive and negative feedback occurs in a surprisingly large number of user turns.

## 2. BACKGROUND

Clark's theory of grounding [5] describes discourse as a joint activity in which participants continuously work at establishing a common ground. An acknowledgement or repetition of the dialogue partner's previous contribution hardly moves the conversation forward, since such a turn contributes little or no new information. These utterances have been categorized as a subgroup of the "informationally redundant utterances" [12]. According to the theory of grounding, however, dialogue participants use acknowledgements and feedback to signal understanding and nonunderstanding throughout the discourse. These cues often carry important information about the grounding process and the state of the dialogue. Clark and Schaefer [6] suggest that there are a number of ways in which a dialogue participant can demonstrate that he has understood a discourse contribution. "Acknowledgement" is placed in the middle of a scale ranging from "continued attention" to "display". In dialogue, an acknowledgement is often expressed by a nod or a "yeah", "uh huh" or something similar. Brennan and Hulteen [2] present a list of acknowledgement strategies that is partly based on Clark and Schaefers' scale, and emphasize the importance of feedback for coordinating the user and systems' knowledge states in a dialogue system and for facilitating problem solving. In a study based on tutorial dialogues, Brandle [1] divides acknowledgements into several subgroups according to their function in the grounding process. In this classification scheme, explicit acknowledgements are distinguished from implicit ones. In a study of cues used for tracking initiative in dialogue, Chu-Carrol and Brown [3] use the term "prompts" for similar phenomena. In a recent publication, Ward and Heeman [13] report that acknowledgements are used to a rather large extent when subjects interact with a telephone-based automated service system. Even though this system did not explicitly encourage the use of feedback, it provided opportunities for and responded to acknowledgements. Ward and Heeman report that about half of the subjects of their study used acknowledgements at least once during their interaction.

## 3. MOTIVATION

### 3.1 Error handling

In human–computer dialogue, frequent occurrences of errors threaten to make users frustrated and may result in a premature closure of the interaction. Errors are inevitable in human–human as well as human–computer dialogue, but in human–human dialogue refined strategies for dealing with problematic interactions have been developed. Clark [4] has suggested that conversants begin by trying to prevent foreseeable but avoidable problems, then warn partners about predictable but unavoidable problems and lastly resort to repairing those problems that have already arisen. In Clark's view, we should expect human–human dialogue participants to prefer preventatives to warnings, and warnings to repairs [4]. The reason for this is the relative high cost of repairing problems that have already arisen in a dialogue, compared to the relative low cost of an extra (perhaps unnecessary) dialogue turn. As reported by Smith and Gordon [10], there is a similar problem in human–computer interaction. Here, developers of dialogue systems have to consider the trade-off between being terse and risking being misunderstood on the one hand, and being overinformative and repetitive on the other. In a study of errors in a spoken dialogue system caused by misrecognition, Smith and Hipp [11] proposed that verification subdialogues should be used selectively to recover from errors. The context of the utterance is shown to be helpful in selecting which utterances to verify.

Certain linguistic markers are often used to signal understanding, nonunderstanding or misunderstanding in dialogue. However, in a given context the significance of such markers can be difficult to assess. In a study describing a French dialogue system, Derriks and Willems [7] show that negative feedback cues exhibit ambiguity, so that for instance the word "pardon" can be given six different interpretations. Similarly, in the corpus presently analyzed, some feedback cues were found to be inherently difficult to interpret. In some cases, these cues could be given both positive and negative interpretations. Contextual and prosodic cues can often help resolve such ambiguities. If correctly interpreted, a positive or negative linguistic marker can be used by a spoken dialogue system as an indication of the dialogue status. When a positive feedback turn has been recognized, and a problem occurs later on in the dialogue, it is reasonable to assume that the dialogue was fine at least up until that time. If a verification subdialogue is initiated by the system at a later stage, it does not have to go back further than necessary. Negative user feedback can be interpreted as a sign of discontentment, as a warning of an upcoming problem or as a reaction to an error that has already occurred. If rapidly identified by a dialogue system as a problem or a warning, these negative feedback utterances could be used to facilitate error handling and perhaps avoid a longer error sequence.

## 3.2 Feedback in the August corpus

Part of the motivation for the present study came from observations made in the previously developed August system. This experimental spoken dialogue system, whose animated talking head was modeled after the Swedish author August Strindberg, was used to collect speech data from members of the general public. The August database consists of more than 10,000 utterances of spontaneous computer–directed speech from around 2,500 users, and is described in [8]. Because of the high levels of background noise in the public location where August was displayed, a push-to-talk mechanism was used for speech recording. The system itself used no explicit acknowledgements or feedback, nor were its users encouraged to do so. Nonetheless, analyses of the August corpus indicated that the users quite frequently gave the system feedback on previous turns. Since some of the human–computer dialogues in the August database were very short, a subsection of the corpus with only those interactions that went on for three or more user turns was extracted. The total number of users in this subsection was 1206, and out of these 18% gave the system positive or negative feedback at least once. The total number of utterances was 6876, out of which 6% contained feedback to the system. In 89% of these cases, the feedback appeared in a turn of its own. This figure can probably be explained by the fact that the users had to push to talk, and thus tended to convey one speech act at the time to the system. These preliminary figures, obtained in the analysis of the August corpus, inspired us to perform a more exhaustive study of user feedback strategies in the AdApt system.

# 4. DATA

## 4.1 The AdApt corpus

AdApt is a Swedish conversational multimodal dialogue system which can be used for accessing information about apartments for sale in downtown Stockholm. Figure 1 shows the system's graphical user interface. It consists of the animated talking agent Urban, an interactive map of Stockholm and a table for displaying textual information. The AdApt corpus comprises 50 dialogues with 33 subjects, all collected in a series of Wizard-of-Oz experiments. The total number of utterances in the corpus is 1845. The subjects were given pictorial tasks that involved finding one or several apartments in Stockholm that fulfilled certain criteria. To solve these tasks, the subjects were asked to take their time to look around, and to compare different apartments in order to find a suitable one. The tasks were deliberately designed to be vague, so that the subjects' linguistic behavior would be as natural and unconstrained as possible. In the course of these experiments, an open microphone was used to facilitate the integration of speech and


Figure 1. A user interacting with the AdApt System and a closeup of the interactive map with apartment icons.

graphical input to the system. A pointing device was used to carry out graphical operations, namely selecting a position or an apartment icon indicated on the map or marking an area on the screen.

A spoken dialogue system's way of providing feedback affects the users' manner of interacting with that system. The AdApt system did not explicitly acknowledge that the subjects' input to the system was being processed or had been correctly recognized. However, indirect visual cues were conveyed through the system's animated talking head. While speech input was being processed, the talking head appeared to be "listening", and as soon as a user had finished speaking, the head indicated that the spoken input was being interpreted by responding with a "thinking" gesture. Furthermore, by "understanding" most of what was being said, the system indirectly encouraged the subjects' conversational behavior. In the course of the dialogues, the system offered implicit evidence of understanding. A translated example from the AdApt database illustrates this:

**System 1**: Where in Stockholm would you like to live?

**User 1**: I want to live in the Old Town.

**System 2**: How many rooms do you need?

In the above example, the subject's input is indirectly acknowledged. The system's next dialogue turn is relevant, and no repetition of the user's previous utterance is requested. A few turns later, when the system has found a selection of apartments in the Old Town and they are displayed on the screen, the user will know for certain that this turn was correctly interpreted. If the system had used an explicit acknowledgement strategy instead, the system's response to User 1 would for example have been: *"The Old Town. Is that correct?"*. If this sort of explicit prompt had been employed, user feedback strategies would probably have been different. Intermediate strategies, where the system's acknowledgement is part of the next turn, are also possible.

## 4.2 Annotation of data

The AdApt corpus was manually transcribed and the subjects' utterances were individually labeled for feedback, taking into account the context of the system's previous utterance and the dialogue history. For example, when "no" was used as a way of signalling dissatisfaction or disagreement in the dialogue, it was marked as feedback. Conversely, when "no" occurred as response to a question posed by the system, it was not labeled as feedback. Those parts of the user utterances that had been marked as feedback were then tagged with respect to the following three parameters:

| # | System | User | Feedback | | |
|---|---|---|---|---|---|
| 32 | This building was constructed in 1680 | **Yes that is very old indeed ....**when did you say the white one was built | Pos | Exp | Atti |
| 33 | This building was constructed in 1861 | **Yes yes that's right…..is** there a tiled stove there too | Pos | Exp | Attn |
| 34 | This apartment has a fireplace | **Yes that's all right too ……..** how high is the building | Pos | Exp | Atti |
| 35 | This apartment is on the first floor | **Okay ….** and I see it is close to the German church there | Pos | Imp | Attn |
| 36 | I don't know anything about such things | **Well okay…………….** yes but I think I'm happy with that | Neg | Exp | Atti |

Table 1 A translated excerpt from the AdApt corpus. The part of the user utterance that has been labeled as feedback is in boldface, and the type of feedback — *positive/negative, explicit/implicit, attention(Attn)/attitude(Atti)* — is in the table to the right.

**Positive/Negative**

Positive feedback typically include expressions like "good", "yes", and "thank you". Examples from the negative feedback category include "no", "well" and "too bad". Since some expressions, such as "okay", function as either a positive or negative cue, all sound files were individually assessed. Prosodic or contextual cues indicated whether an utterance was intended by the subject as a positive or negative response to the system's previous utterance.

**Explicit/Implicit**

In some of the feedback utterances the subjects literally expressed what they meant, so that for example a presentation of a new apartment would get the response "that's great" or "very good Urban". These were labeled as explicit, while those utterances where the feedback was conveyed in a less direct way were labeled as implicit. Implicit feedback was often expressed through cues like "mhm", and "aha, all right". Again, some cases were ambiguous.

**Attention/Attitude**

Attention was interpreted as an indication from the user that the system's message has been received. Typical examples include "I see", and "No bath tub ". Attitude, on the other hand, was seen as an indication of the user's attitude toward the system or the previous turn in the dialogue. Positive and negative value judgements occur frequently in this category. Examples from the corpus include "that's good", "great, Urban", "thanks", "that was quite expensive" and "too bad".

All feedback utterances were categorized along these three axes, resulting in a total of eight groups. As previously observed, some expressions in the corpus turned out to be inherently ambiguous. The word "okay", for instance, was labeled as belonging to all of the categories depending on the context in which it appeared. Table 1 shows part of an annotated dialogue sequence in which examples of most of these labeling categories are included. In this excerpt, the user gives the system feedback at every turn. Most of the feedback was labeled as positive. The single instance of negative feedback from the user, turn 36 in Table 1, appears as a response to a system turn where no information was conveyed.

# 5 RESULTS

Positive or negative feedback was found in 18% of all user utterances in the AdApt database. It is worth noticing that almost all subjects, 94%, used feedback at least once during their interaction with the system. In contrast to the August system, user feedback occurred in a separate turn in as few as 6% of all cases in the presently examined corpus. Instead, feedback typically occurred in the initial position of a longer user sequence, after which a silent pause was followed by a request for information. Turns 32 through 34 in the example dialogue in Table 1 provide examples of this phenomenon. In the AdApt database, 65% of all feedback utterances were judged to be positive. Two thirds of the feedback utterances were labeled as explicit, while one third were implicit. The groups of feedback tagged as attention or attitude were evenly sized.

When the function of the user feedback utterances was examined in a broader dialogue context, several interesting tendencies could be distinguished. The function of the largest group of utterances in the database was that of asking a direct question, for instance "finns det badkar" (*"is there a bathtub"*) In these cases, feedback turned out to be quite uncommon. Another frequently occurring type of utterance in the database was one where the user would define his or her preferences. For this group, feedback was provided in about one fourth of the utterances. Relatively speaking, feedback was very frequent in those utterances that were used for concluding the interaction with the system. An example from the database is: "okej, då tackar jag för hjälpen" (*"okay then, thanks for your help"*). The feedback provided indicates that the user wants to sum up before finishing the dialogue. Meta–utterances, that is, user comments about the AdApt system, remarks on the preceding dialogue and self–directed communication, were quite rare in the corpus. When they occurred, however, they often included feedback to the system.

The analysis of data also revealed large individual variations in feedback strategies. While some subjects gave the system positive and negative feedback in virtually every turn, others very rarely gave feedback at all. For the individual subjects, the number of utterances that were labeled as including feedback varied from 0% to 70%. Figure 2 shows that about one fourth of the subjects used feedback in half or more of their turns, while one fourth of the subjects very rarely or never used feedback. No correlations with the subjects' reported experience with computers in general or spoken dialogue systems in particular were found. It appears as if feedback to a spoken dialogue system, at least partly, is a matter of individual style.

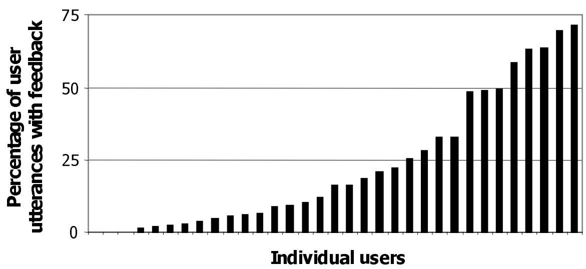


Figure 2. Distribution of feedback in user utterances.

The human–computer dialogue as a whole probably affected the way in which feedback was used in the multimodal dialogue system. To investigate feedback in the context of the discourse, the system's previous turn was correlated to the users' choice of strategy. As can be seen in Figure 3, the feedback categories attention and attitude appeared at different places in the dialogue. In the initial phase of the discourse, where the system took the initiative and inquired about the user's preferences, feedback was often used to signal attitude. When the system failed to fulfill the user's request, on the contrary, users merely signalled that they had understood what the system was saying.

| System answer to previous question | User feedback | +/- | Usable |
|---|---|---|---|
| The yellow building was constructed in 1890 | **Yeah, that sounds good** - does it have an tiled stove maybe | Pos | Yes |
| The apartment is on the second floor | **Okay** and are there any available two-room apartments on Östermalm. | Pos | No |
| This building was constructed in 1997 | **Ouch!** Is there an old building from the $19^{th}$ century | Neg | Yes |
| This apartment has a tiled bathroom | **Well that's not quite what I asked about Urban** does it have a bathtub | Neg | No |

Table 2. Translated examples of attitude feedback, marked for usability from the point of view of user preferences.

Figure 3 also indicates that when the system turned over the initiative by asking an open question ( e.g. *"Is there anything else you would like to know about the apartment?"*), the subjects responded with attitude feedback (*"Yes, I would like to know if the apartment has a balcony"*) It thus seems as if certain types of user feedback are likely to be provided in different phases in the dialogue. If, in a future system, it becomes apparent that difficulties often appear in a particular stage in the discourse, the system should anticipate negative user feedback. In this way, the user's warning to the system could prevent a more serious problem from occurring.

Those attitude feedback utterances that occurred after the system had supplied the user with information about some feature of an apartment, could be used to gain knowledge about the users' preferences. Instead of explicitly asking what kind of apartment the user would prefer, the system could attempt to interpret the user's feedback. For example, when a user asks: "*What can you tell me about this apartment?*", the system could present the apartment's most distinguishing feature(s). If the user provides the system with feedback, this could be used to decide which apartments to present later on in the dialogue. A similar method has previously been implemented in a text-based dialogue system [9]. In Table 2, four examples of attitude feedback are presented. In two of the examples, the feedback might be used to model user preferences. In general, negative attitude feedback appeared to contain more information and be more useful than positive attitude feedback. For instance, the feedback utterance in the last example in the table could be used to detect that a problem has occurred in the dialogue, and that the user wishes to correct the system's interpretation.

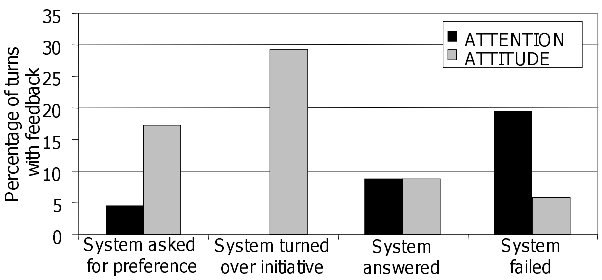


Figure 3. Number of turns with feedback depending on the previous system turn.

In the present study, positive and negative user feedback cues were found to signal understanding and misunderstanding throughout the dialogues. Certain user preferences were also expressed in the feedback utterances. In a future system, positive feedback can be utilized as a way for the system to increase its knowledge about the user's preferences. Complicated correction subdialogues can thus be avoided. Negative feedback is sometimes used to warn the system of an upcoming problem. If these cues are correctly interpreted and handled by the system, serious errors can perhaps be prevented from occurring.

# 5. ACKNOWLEDGEMENTS


The authors would like to thank the other members of the AdApt group at the Centre for Speech Technology.


# 6. REFERENCES


1. Brandle, S. 1997. *Using Joint Actions To Explain Acknowledgments In Tutorial Discourse*. PhD thesis, Illinois Institute of Technology.
2. Brennan, S. E. and Hulteen, E. A. 1995. Interaction and feedback in a spoken language system: a theoretical framework. *Knowledge-Based Systems* 8: 143-151.
3. Chu-Carrol, J. and Brown, M. K. 1997. Tracking Initiative in Collaborative Dialogue Interactions. In *Proc. of ACL-EACL'97*, pp 262-279.
4. Clark, H. H. 1994. Managing Problems in Speaking. *Speech Communication* **15**, 243-250.
5. Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.
6. Clark, H. H. and Schaefer, E. F. 1989. Contributing to Discourse. *Cognitive Science* **13**, 259-294.
7. Derriks, B. and Willems, D. 1998. Negative feedback in information dialogues: identification, classification and problem-solving procedures. *International Journal of Human–Computer Studies* **48**, 577-604.
8. Gustafson, J. and Bell, L. 2000. Speech Technology on Trial: Experiences from the August System. In *Natural Language Engineering* (1) 2000, in press.
9. Jameson, A., Kipper, B., Ndiaye, A., Schäfer, R., Simons, J., Weis, T. and Zimmermann, D. 1994. Cooperating to be Noncooperative: The Dialog System PRACMA. *Proc. of the $18^{th}$ German Conference on Artificial Intelligence.*
10. Smith, R. W and Gordon, S. A. 1996. Pragmatic Issues in Handling Miscommunication: Observations of a Spoken Natural Language Dialog System. *Proc. of the AAAI '96 Workshop on Detecting, Repairing and Preventing Human–Machine Miscommunication*, pp. 21-28.
11. Smith, R. W and Hipp, D. R. 1994. *Spoken Natural Language Dialogue Systems: A Practical Approach*. New York: Oxford University Press.
12. Walker, M. 1992. Redundancy in Collaborative Dialogue. *Proceedings of COLING '92*, pp. 345-351.
13. Ward, K. and Heeman, P. 2000. Acknowledgements in Human–Computer Interaction. *Proceedings of the ANLP-NAACL 2000*. p. 281-287.