A study of gestural feedback expressions

Jens Allwood
Department of Linguistics
University of Göteborg.
Sweden

Loredana Cerrato
Department of Speech Music and Hearing
KTH, Stockholm
Sweden
Swedish Graduate School of
Language Technology

Abstract

The aim of this study is to explore the realization of gestures related to verbal feedback expressions.
Distributional, functional, typological and acoustic analyses of verbal and gestural feedback were carried out on spontaneous dialogues between a travel agent and some customers recorded in a travel agency in Sweden.
The investigation addresses in particular the following research questions:

> Do gestural feedback expressions always co-occur with vocal/verbal feedback expressions? Or can they occur on their own?
> What is the function of feedback gestures?

The results show that feedback is mostly expressed simultaneously by vocal/verbal and gestural means. The gestures accompanying verbal and vocal feedback expressions can be broadly categorized according to their function in the given context. Single head nods or jerks are the most frequent head movements accompanying feedback expressions in the Swedish dialogues we analysed.

1 Introduction

Several studies of feedback strategies have been carried out in the past years, with different purposes, such as getting insight in the structures of human conversation and describing the different ways of producing feedback expressions (Clark & Schaefer 1989, Allwood 1988, Cerrato 1999) and recently to understand the strategies of communication between human and machines (Bell & Gustafson 2000, Cerrato 2002c).

Several definitions of feedback have been proposed 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 successfulness of communication.

This information can be exchanged by means of verbal and vocal expressions and by means of gestures.

Previous studies have shown that it is possible to categorize vocal feedback expressions according to their behavioural form, and/or to the function they accomplish (Allwood 1992, Cerrato 2002a, Cerrato 2002b). We believe that it is possible to categorize gestural feedback in the same way.

By gestural feedback we refer to those movements that are produced to show feedback. They can be categorized as follow:

- **Head movements** (nods, jerks, waggles, side-way turns)
- **Facial expressions** (eyebrow movements, gaze, smile)
- **Other gestures** (shoulder shrug, hand and trunk movements)

The two main research questions addressed in this investigation are:

1) Do gestural feedback expressions always co-occur with vocal feedback expressions? Or can they occur on their own, replacing the vocal feedback expression?
2) What is the function of feedback gestures?

2 Materials and method

2.1 Material

Video-recording of four real dialogues between four different customers (two women and two men) and a travel agent (always the same woman), in a travel agency in Gothenburg, Sweden, were selected from the Spoken Language Corpus of the Linguistics Department of Gothenburg University (Allwood 1999).

The dialogues can be classified as belonging to the activity type of "information seeking", 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.

The agent stands or sits behind a desk, and the customer stands on the other side.

The microphone and the video-camera were placed on the desk, in a position which allowed the recording of both dialogue participants by one side. The customers were informed of the presence of the recording apparatus and of the purpose of the recording.

Table 1 reports some information related to the dialogues. In dial1 and dial3 the customers not only get different information from the travel agent, but they book also their travel. Dial2 is very short because the customer only asks for a specific piece of information, while dial4 is the longest one since there occur some problems with the terminal during the booking of a flight.

Dialogue	**Customer**	**Number of utterances**	**Duration in minutes**
dial1	female	107	8.42
dial2	male	65	2.15
dial3	male	150	16.42
dial4	female	112	27.31

Table 1 Schema of the information related to the materials used in our study.

2.2 Technical support

Feedback has been analysed with the support of the Multitool package software, which simultaneously displays the video and the relative orthographic transcription of the dialogues (Allwood et al. 2002). For our study of feedback we use a multi-tier annotation, which consist of the following tiers displayed on the *score lines* of Multitool:

- **Text**: reports the transcription of the utterances per speaker.
- **Comments**: reports different kinds of comments of the transcribers, which are not part of the coding schema described below.
- **Function**: reports the function of the utterances under analysis. For our investigation we coded the function of those utterances which we labelled as feedback expression (FB) and the function of the utterances preceding them.
- **Gesture**: reports the visible gestures of the speakers which we considered related to feedback.
- **Gesture function:** reports the specific function of the gesture**.**
- **Gaze** : report the direction of the speakers' gaze using two values: mutual gaze, non-mutual gaze.

2.3 Coding of feedback

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 categorized as feedback if its primary function serves one of the following purposes, which Allwood refers to as "attitudes"(Allwood 2001, page 35-36):

- show continuation of contact (C): when the interlocutor wishes to show that s/he is willing and able to continue the interaction, either by letting the other speaker continue to talk or by getting the turn;
- show perception (P), when the interlocutors show awareness and discernment of expression of the message;
- show understanding (U), when the interlocutors show that they have understood;
- show behavioural and attitudinal reactions (R) towards the meaning conveyed; this includes assent, negation or contradiction, assertion surprise, disappointment and enthusiasm etc.

Perception and understanding are related to what Clark and Schaefer (1989) refer to as "acknowledgement", that is to describe a hierarchy of methods used by interlocutors to signal that a contribution has been understood well enough to allow the conversation to proceed.
All four functions are related to basic requirements of human communication. To communicate we need two participants who have to establish *contact* with each other. To be able to convey a message it is necessary that the message is *perceived* by the receiver, who must also be able and willing to *understand* it. The receiver then provides the main communicator (the sender) with information concerning how the message has been received.
Moreover feedback expressions are coded according to their "directional function type" (Allwood 2001, pages 35-36), which can be:

- Giving,
- Eliciting,
- Giving-Eliciting

The speakers give feedback when they wish to let the interlocutor understand that they are listening, paying attention, understanding or agreeing with what he/she is saying.

The speakers elicit feedback when they wish to know whether the interlocutor is listening, paying attention, understanding, or agreeing, disagreeing with what they are saying. We annotated also the turn-taking function carried out by the feedback expressions. The speaker who gives feedback can either produce feedback without wanting to interrupt the other speaker by getting the turn, or produce feedback to show the intention of taking the turn.
The visible movements that we considered "gestural feedback expressions" were annotated on the "Gesture" tier of the score line of Multitool.

These movements can be divided in three main groups: head movements, facial expressions and other gestures.
Head movements are coded as follow:

- **nod:** is a forward movement of the head going up and down, which can be multiple
- **jerk:** is a backward movement of the head which is usually single
- **shake:** is a left-right or right-left movement of the head which can be multiple
- **waggle:** is movement of the head back and forth left to right
- **swturn:** *side-way turn* is a single turn of the head left or right

Facial gestures include:

- **smile**
- **laughter**
- **eyebrows rising**
- **eye brow frowning**

Other gestures include:

- **movef:** *move forward* is a forward movement of the whole trunk
- **moveb:** *move backward* is a backward movement of the whole trunk
- **hand:** refers to hand/s movement
- **shrug** refers to shoulders shrug

The movements can be multiple, sequential and simultaneous.

Multiple movements: some movements can be multiple; in this case they are annotated with the addition of the number of times the movement is performed as in the following example: nod: 2

Sequential movements: sometimes more than one movement is annotated in the same coding, for example: -nods, waggle-means that the two movements are sequential.

Simultaneous movements: if the two movements are simultaneous we use: +

To code the specific function of the gestural feedback we adapted a classification proposed in Allwood (2001) and by Poyatos (2002, pp. 281-283) to describe the possible relationship between the "*nonverbal system and words*". Gestural feedback expressions co-occurring with vocal/verbal feedback expressions can modify the meaning of the vocal/verbal expression in the following ways:

- Reinforcement (R), when the gesture adds redundancy by giving the same information as the vocal message.
- Positive (P), when the gesture indicates a positive reinforcing attitude, like for instance showing enthusiasm.
- Negative (N), when the gesture weakens what has been said vocally, to indicate for instance lack of enthusiasm or interest.
- Contradicting (C), when the gesture contradicts what has been said vocally, to denote sarcasm, irony.

Where possible we have also annotated if the interlocutors where looking at each others (MU: mutual gaze).

The annotation of gaze could be useful to verify the hypothesis that a high frequency of listeners' non-verbal feedback expressions is produced when there is visual contact between the interlocutors.

3 Results

3.1 Distributional analysis

Figure 1 displays the percentage of utterances including at least one feedback expressions, across dialogues and speakers. So for instance in dial1, the travel agents produces 25 utterances, which include at least one feedback expression, if we relate this number to the total number of utterances the agent produces in the dialogue (53) we obtain a percentage of

49% of utterances including a feedback expression. The number of utterance including at least one feedback expression does not necessarily correspond to the number of total feedback expressions produced, since an utterance can include more than one single feedback expression.

In the four dialogues the agent is always the same woman while the customers are four different people.

The customers tend to produce a higher number of feedback expressions compared to the agent. Given the particular communicative situation, it is mostly the customer that has the role of "listener" while the travel agent has the role of "speaker". The 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 agent produces longer utterances, while the customers produce a great number of short utterances including or consisting of feedback, in order to show continuation of contact, perception and understanding.

It is striking that the percentage of utterances including at least one feedback expression in dial2, is so high. This dialogue is the shortest of all (only 65 utterances) and the high number of utterances including feedback is due to the fact that the customer, after having asked his question to the travel agents, produces mainly short feedback expressions in reply to the travel agent information.

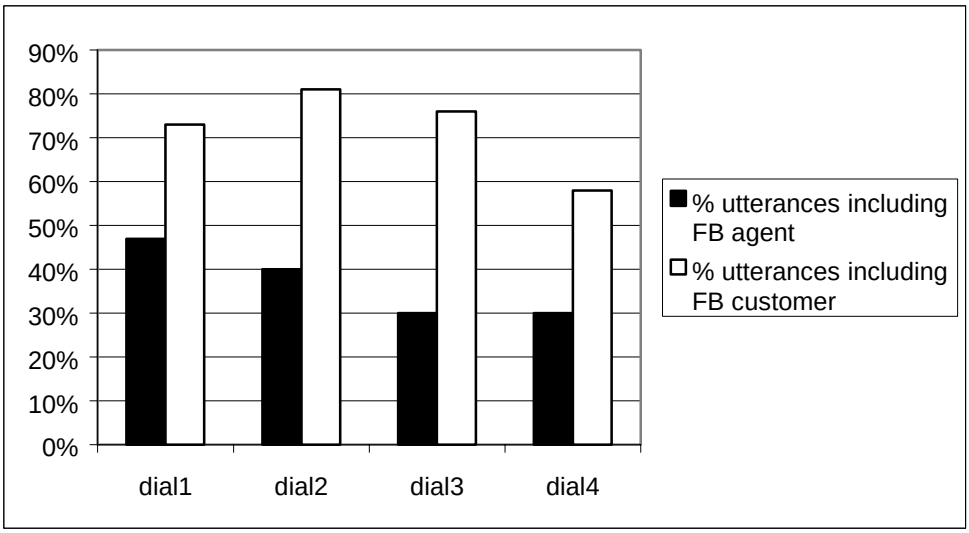

Figure 1: Percentage of utterances including at least one feedback expression.

Feedback expressions can be: verbal, verbal+gestural, gestural. In the four dialogues we analysed gestural feedback almost always co-occur with vocal feedback expressions, there are in fact only four examples of gestural feedback expressions produced without accompanying a verbal feedback expressions: three in dial1 and one in dial4.

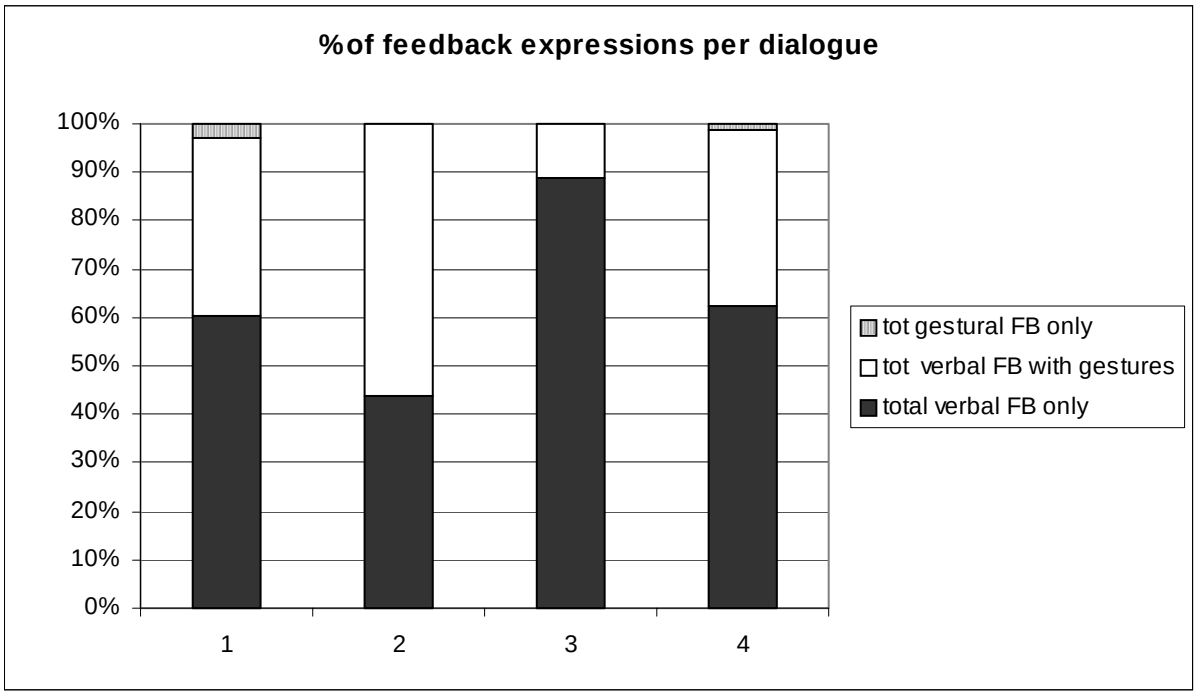

Figure 2: Distribution of feedback expression per dialogue

Figure 2 reports the distribution of the kind of feedback expressions per dialogue. Each column displays the percentage of verbal feedback not accompanied by gesture (the black area), the verbal feedback accompanied by gesture (the white area) and feedback, which is only gestural (the patterned area, only in dialogue 1 and 4, because only in these dialogues we found some feedback expressions which were only gestural).

Figure 2 shows that there is not a clear trend in the percentages, which varies across the four dialogues.

In dial3 in particular, we can notice that the production of feedback expressions accompanied by gesture is very low. This can be interpreted as a peculiarity of the customer, who might be considered as a "low producer" of gestural feedback or can be explained as depending on the fact that the customer in this dialogue was standing in a peculiar position: bent towards the agent with his elbow on the desk and his hand under his chin. This position is likely to have prevented him from moving his head freely.

Figure 3 reports the distribution of feedback expressions accompanied by gesture across dialogues and speakers. The customers tend to produce more gestures than the agent (except in dial3). It is difficult to understand if this depends on the personal predisposition or tendency of the speakers to produce gestures, or on the communicative situation, which probably encourages the customer to produce more feedback and therefore more accompanying gestures.

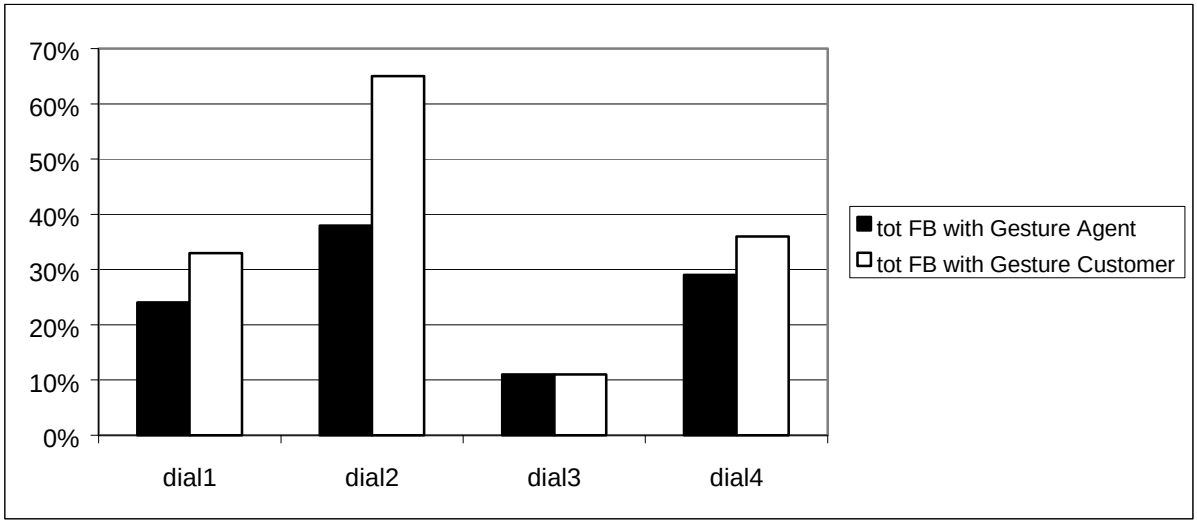

Figure 3: Percentage of feedback expressions accompanied by a gesture per dialogue.

The hypothesis that the interlocutors tend to adapt to each other and they tend to produce head movements in response to the production of a head movement by the interlocutor is confirmed by the results.

Looking at the videos it is easy to detect this “mimicking” phenomenon, which is quite evident especially when the interlocutors look at each other.

In dial2 the percentage of gestures produced by the customer is quite predominant; this speaker produces more than 80% of his utterances as short feedback and more than 60% of the feedback expressions he produces are accompanied by gestures.

In dial3 there is a low percentage of feedback accompanied by gesture, this might depend, as we noticed before, on the fact that the customer assumed a position that might have prevented him from producing head movements. As a consequence of the low gesture production by the customer side, the agent seems to keep the production of accompanying gestures quite low.

On the basis of these preliminary observations it is possible to suppose that some accompanying head-movements are produced unconsciously, while other are produced intentionally to modify the meaning of the verbal expression they are accompanying. This assumption is mainly based on the observation that in order to convey a specific intentional meaning a movement has to be seen-received by the interlocutor, otherwise its communicative effect fails. If a speaker wishes to add some more information to his/her vocal production by means of an emphasizing gesture for instance, he/she does it in a conscious way and supposing that the listener is looking at him/her and able to interpret his/her signal.

This hypothesis is based not only on common sense, but also on results of analysis like those carried out for instance by McClave (1998), who reports that a higher number of head movements is produced when the interlocutors have eye contact.

In our material we have noticed that often small head movements such as nods and jerks are produced even when there is no eye contact between the speakers. These results is not the first of this kind, since previous studies have shown that head movements accompanying feedback are produced even when the interlocutors do not or cannot look at each other eyes, as for instance in telephone conversations (Nivre & Richthoff 1988).

Besides gestures produced with feedback function, it is possible to notice other kind of gestures, mostly head movements, hand movements (pointing gestures, iconic gestures) produced during the dialogues. However a detailed analysis of these gestures is currently beyond the scope of our investigation.

3.2 Typological analysis

The most common gestures accompanying feedback expressions consist of head movements. Table 2 reports a list of the most common gestural feedback expressions.

Type of gesture	
multiple nod	51
single nod	16
jerks	14
shake	9
movef	4
laughter	3
swturn	2
moveb	1
Tot	100

Table 2: Number of occurrences of gestures produced with feedback function

Head nods and head jerks are the most frequent in our material. Single nods are produced usually together with vocal feedback expression having the main function of showing continuation of contact perception and understanding (FBCPU). Head nods co-occur frequently with short expressions such as "*ja*" and "*mh*".

Jerks are usually produced together with positive feedback words like: "*ja, aha jasså*" which often show a reaction of surprise.

Multiple nods and other movements (jerks, move forward or backward) are mostly produced to modify the function of the feedback word.

3.3 Functional analyses

Two different functional analyses were carried out. The first aimed at identifying the specific sub-functions of the vocal/verbal feedback expressions. The identification was done according to the coding system proposed in Allwood (2001), in which a feedback expression can be produced to:

- show contact perception and understanding — CPU
- acknowledge of the information received — cpu(A)
- show agreement, refusal, desire for more information, or show other attitudinal reactions — cpu (R)
- require clarifications, further explanation, further indications — cpu (B)

These different functions that feedback can carry out, can be expressed in different ways:

- by means of different type of verbal expressions: short expressions, such as *mh, ah*, or extended expressions, like questions, repetition or reformulation of an utterance;
- by means of different prosodic cues, such as specific F0 contours, duration and energy
- by means of different accompanying gestures.

The second analysis we carried out aimed at identifying the specific function of the gestures accompanying the vocal/verbal feedback expressions. To be able to identify the function of the gesture we used the classification proposed above in paragraph 2.3.

Table 3 reports a schema of the possible specific function that the gestures accompanying verbal feedback expressions can have.

Function of the vocal/verbal feedback expressions	**Most frequent accompanying gesture**	**Most frequent function of the gesture**
CPU	Nod, Jerk	R (reinforcing)
cpu (A)	Nod, Jerk, Smile	R (reinforcing), P(positive)
cpu (R)	Nod, Shake, Rising Eyebrows	N (negative)
cpu (B)	Shakes, Smile, Rising Eyebrows	P((positive)

Table 3: Specific function of the gestures accompanying feedback expressions.

When the function of these gestures is neutral (non-marked) the gesture can be minimal. This is consistent with the hypothesis that some gestures are produced unconsciously since they do not have a marked function. In our material head nods and jerks are often produced together with very short feedback expressions, such as "*m*" "*mm*" and "*ja*". These short expressions have usually the main function of showing contact, perception and understanding in a minimal intrusive way. This means that the speaker who produces these short feedbacks does not wish to interfere with what the interlocutor is saying and does not wish to show the intention to interrupt and get the turn, but wants only to show an active listening attitude.

It is not possible to establish a one-to-one correspondence between a specific head movement and a specific function. Different movements can in fact be produced to convey the same meaning, for instance both nods and a jerks are used to show feedback with continuation of contact perception and understanding. However it has to be mentioned that the meaning of the vocal/verbal feedback expression can at the same time be modified by means of prosodic cues: variation in F0 contour, duration and tone of voice can be used to express emphasis, de-emphasis, irony and sarcasm.

3.4 Acoustic analysis

It is quite uncontroversial that acoustic cues can be used with the purpose of marking information structure at the discourse level. Phonetic correlates for different function of short expressions have been found for a variety of languages: Italian (Cerrato 2002b), Swedish (Cerrato 2002a), English (Hirshberg & Nakatani 1996), Japanese (Ward & Tsukahara 2000), Dutch (Casper 2002).

Unfortunately due to the quality of the recording it was not possible to carry out accurate F0 analysis of the feedback expressions in our corpus, however it was possible to notice the following characteristics:

Feedback expressions having short duration and flat pitch contour generally show:

active listening attitude in non-intrusive way
intention to let the other speaker go on;

Feedback expressions having longer duration and rising pitch contour generally show:

active listening attitude in an intrusive way,
intention to take the turn.

Feedback expressions having longer duration and rising-falling pitch contour generally show:

acknowledgement of comprehension
desire for more information

3.4.1 On the relationship between prosody and gestures

An additional hypothesis that we preliminary tested is that when short verbal expressions having a feedback function are accompanied by head movements, such as nods and jerks, they might undergo lengthening phenomena.

To test this hypothesis we measured the duration of all the instances of "*m*", "*{j}a*" and "*jaha*" produced with and without an accompanying head moment by the agent in all the dialogues. The agent is always the same female speaker. The measurements were carried out with the support of the analysis tool Wavesurfer (Sjölander & Beskow 2001), the results are reported in figure 4.

When the short feedback expressions "*m*", "*{j}a*" and "*jaha*" are produced together with a head movement (usually a head nod) they show a longer duration ranging between 20% and 40% longer. This result might depend on the fact that somehow the co-ordination of the gesture with the articulation of the verbal feedback expressions makes the articulation slower. Another explanation can be related to the fact gestures often co-occur with intonational cues (Bertrand et al 1995) which might means that feedback words produced with head movement might carry some focus. Previous studies (Casper 2002) have in fact shown that prosodically marked items have a longer duration compared to the same items without prominent prosodic characteristics. However we have not carried out a tonal analysis to verify this hypothesis.

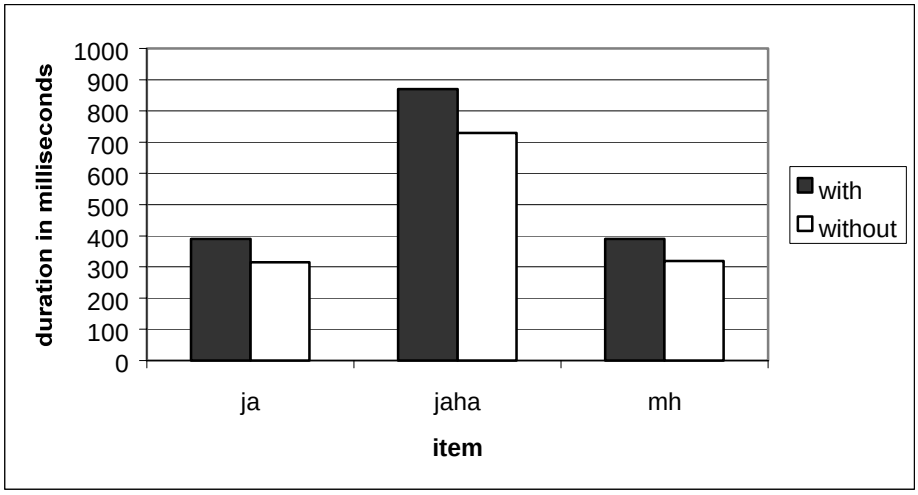

Figure 4: Duration of three feedback expressions produced with and without an accompanying head movement.

4 Conclusions and discussion

The aim of this study was to explore the realization of gestures accompanying verbal feedback expressions, with the intention of answering the two following research questions:
Do gestural feedback expressions always co-occur with vocal feedback expressions? Or can they occur on their own?

What is the function of feedback gestures?

At the light of the results it seems like feedback is mostly expressed simultaneously by vocal/verbal and gestural means. In our material we found that very rarely gestural feedback expressions are produced on their own.

It seems to be possible to make a broad categorization of feedback movement, which is related to their function.

Single head nods or jerks are the most frequent head movements accompanying feedback expressions, often short expressions, like “mm” “{j}a” produced in a non-intrusive way.

When the function of a short verbal/vocal feedback expression is mainly to show continuation of contact, perception and understanding, the accompanying gesture is likely to be a minimal head nod with the a neutral function. When the function of the verbal/vocal feedback expression is implicitly of showing continuation of contact, perception and understanding, and explicitly to add some attitudinal reactions (like agreement, disagreement, surprise, disappointment) the accompanying gestures is likely to be more extensive and its function is usually to add some extra information or to emphasize or contradict what it has been said.

Complex feedback expressions, like repetitions, reformulation, which acknowledge/refuse an information, ask for clarification and/or add some attitudinal reaction are often accompanied by more complex head movements, like multiple nods or sequences of expressions (for instance nods and smile). Moreover the attitudinal reaction seems to be often marked by some other phonological and prosodic phenomena, like lengthening and variation of pitch contour.

Further analyses are needed to verify our hypothesis and confirm our preliminary results. We are aware of the fact that our materials are quite limited both in their amount and in the fact that they represent just one cultural community.

5 References

Allwood, J. (1988) "Om det svenska systemet för språklig återkoppling" in P. Linell, V. Adelswärd, T. Nilsson, & P.A. Pettersson (Eds.) Svenskans Beskrivning 16, Vol.1. (SIC 21a) , University of Linköping, Tema Kommunikation.

Allwood J, Nivre J, Ahlsén E (1992) “On the Semantics and Pragmatics of Linguistic Feedback” in *Journal of Semantics* 64.

Allwood J (1999) “The Swedish Spoken Language Corpus at Göteborg University". In Proceedings of Fonetik 99, Gothenburg Papers in Theoretical Linguistics 81.

Allwood J (2001) “Cooperation and flexibility in multimodal communication”. In *Cooperative Multimodal Communication* Harry Bunt and Robbert-Jan Beun, editors Lecture Notes in Computer Science 2155 Springer Verlag, Berlin/Heidelberg. Also in *SALSA - Symposium About Language and Society - Austin.* Texas Linguistic Forum, Volume 44, nos 1 and 2. 2002. pp. 2 - 17 and in *Gothenburg Papers in Theoretical Linguistics* GPTL 84. Dept. of linguistics, University of Göteborg.

Allwood J, Grönqvist L, Ahlsén E, Gunnarsson M (2002):"Annotations and Tools for an Activity Based Spoken Language Corpus", Proc. of 2nd SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark

Bell L.,Gustafson J, (2000) “Positive and negative user feedback in a spoken dialogue corpus” in Proc. ICSLP 2000, Bejing, China.

Bertrand R., Boyer J. Cavé C., Guaitella I., Santi S. (1995) “Relationship between gestures and voice in verbal interaction: prosodic and kinesic aspects of back-channel signals”, in ICPhS 1995, Stockholm,vol 2. pp746-749

Casper J., (2002), “Prosodic characteristics of backchannels in Dutch map task dialogues: production and perception”, (accepted at Speech Communication)

Cerrato L. (2002a) “Some characteristics of feedback expressions in Swedish”**,** in TMH.OPSR Vol.43 Fonetik 2002 p. 101-104

Cerrato L., (2002b) "A Study of Verbal Feedback in Italian", in Proceedings of the NORDTALK Symposium on Relations between Utterances, Copenaghen 2002, pp. 80-97

Cerrato L. (2002c) A comparison between feedback strategies in Human-to-Human and Human-Machine communication, in proceedings of ICSLP-2002, Denver, Colorado USA.

Clark, H, and Schaefer, E, (1989) Contributing to Discourse. *Cognitive Science* 13, 259-294.

Hirshberg J., Nakatani C., (1996) "A prosodic analysis of discourse segments in direction-giving monologues". In Proc. ACL ,pp. 286-293 Santa Cruz, CA.

McClave E, (1998). "Cognitive and Interactional Functions of Head Movements in Conversation". In *Oralité et Gestualité*, Santi S. et al. Eds, pp.365-369

Nivre, J. and Richthoff, U. (1988) "Återkoppling och turtagning i två typer av samtal. En jämförande studie av samtal ansikte mot ansikte och telefonsamtal". Gothenburg Papers in Theoretical Linguistics S 10. Göteborg University, Department of Linguistics.

Poyatos F., (2002) Non-verbal communication across disciplines, J, Benjamin Publishing Company

Sjölander, K. and Beskow, J. (2000). "WaveSurfer - an Open Source Speech Tool". *Proceedings of ICSLP 2000*, Bejing, China, October 2000

Ward N and Tsukahara W. (2000) **"**Prosodic Features which Cue Back-channel Responses English and Japanese". *Journal of Pragmatics*, 23, pp 1177—1207.