A framework for analyzing embodied communicative feedback in multimodal corpora

Jens Allwood
jens@ling.gu.se, 1) 4)

Karl Grammer
karl.grammer@univie.ac.at, 2) 4)

Stefan Kopp
skopp@techfak.uni-bielefeld.de, 3) 4)

Elisabeth Ahlsén
eliza@ling.gu.se, 1), 4)

1) Department of Linguistics
Göteborg University, Box 200
SE-40530 Göteborg, Sweden

2) Ludwig Boltzmann
Institute for Urban Ethology, Althanstrasse 14,
1090 Vienna, Austria

3) Artificial Intelligence Group
Bielefeld University, P.O. 100131,
D-33501 Bielefeld, Germany

4) ZiF – Center for Interdisciplinary Research
P.O. 100131, D-33501 Bielefeld, Germany

ABSTRACT

Communicative feedback refers to unobtrusive (usually short) vocal or bodily expressions whereby a recipient of information can inform a contributor of information about whether he/she is able and willing to communicate, perceive the information, and understand the information. This paper provides a theory for embodied communicative feedback, describing the different dimensions and features involved. It also provides a corpus analysis part, describing a first data coding and analysis method geared to find the features postulated by the theory.

INTRODUCTION

The purpose of this paper is to present a theoretical model of communicative feedback, which is to be used in a VR agent capable of multimodal communication. Another purpose is to briefly present the coding categories which are being used to obtain data guiding the agent's behavior. Below, we first present the theory.

The function/purpose of communication is to share information. This usually takes place by two or more communicators taking turns in contributing new information. In order to be successful, this process requires a feedback system to make sure the contributed information is really shared. Using the cybernetic notion of feedback of Wiener (1948) as a point of departure, we may define a notion of communicative feedback in terms of four functions that directly arise from basic requirements of human communication: Communicative feedback refers to unobtrusive (usually short) vocal or bodily expressions whereby a recipient of information can inform a contributor of information about whether he/she is able and willing to (i) communicate (have contact), (ii) perceive the information (perception), and (iii) understand the information (understanding). In addition, (iv) feedback information can be given about emotions and attitudes triggered by the information, a special case here being an evaluation of the main evocative function of the current and most recent contributions (cf. Allwood, Nivre & Ahlsén 1992 and Allwood 2000, where the theory is described more in detail).

The central role of feedback in communication is underpinned already by the fact that simple feedback words like *yes*, *no* and *m* are among the most frequent in spoken language. A proper analysis of their semantic/pragmatic content, however, is fairly complex and involves several different dimensions. One striking feature is that these words involve a high degree of context dependence with regard to the features of the preceding communicative act, notably the type of speech act (mood), its factual polarity, information status and evocative function (cf. Allwood, Nivre & Ahlsén 1992). Moreover, when studying natural face-to-face interaction it becomes apparent that the human feedback system comprises much more than words. Interlocutors incessantly coordinate and exchange feedback information by nonverbal means like posture, facial expression or prosody. In this paper, we extend the theoretical account developed earlier to cover *embodied communicative feedback* and provide a framework for analyzing it in multimodal corpora.

DIMENSIONS OF COMMUNICATIVE FEEDBACK

Communicative feedback can be characterized with respect to several different dimensions. Some of the most relevant in this context are the following:

(i) Degrees of control (in production of and reaction to feedback)

(ii) Degrees of awareness (in production of and reaction to feedback)

(iii) Types of expression or modality used in feedback (e.g. audible speech, visible body movements)

(iv) Types of function/content of the feedback expressions

(v) Types of reception preceding giving of feedback

(vi) Types of appraisal and evaluation occurring in listener to select feedback

(vii) Types of communicative intentionality associated with feedback by producer

(viii) Degrees of continuity in feedback signal

(ix) Semiotic information carrying relations of feedback expressions

These dimensions and others (cf. Allwood 2000) play a role in all normal human communication. Below, we will describe their role for embodied communicative feedback. Table 1 shows how different types of embodied feedback behavior can be differentiated according to these dimensions. The table is discussed and explained in the 8 following sections (cf. also Allwood 2000, for a theoretical discussion).

Degrees of awareness and control and embodiment

Human communication involves multiple levels of organization involving physical, biological, psychological and socio-cultural properties. As a basis, we assume that there are at least two (human) biological organisms in a physical environment causally influencing each other, through manipulation of their shared physical environment. Such causal influence might to some extent be innately given, so that there are probably aspects of communication that function independently of awareness and intentional control of the sender. Other types of causal influence are learned and then automatized so that they are normally functioning automatically, but potentially amenable to awareness and control. Still other forms of influence are correlated with awareness and/or intentional control, on a scale ranging from a very low to a very high degree of awareness/control. In this way, communication may involve

1) innately given causal influence

2) potentially aware and intentionally controllable causal influence

3) actually aware and intentionally controlled causal influence.

Human communication is thus "embodied" in two senses, (i) since it always relies on and exploits of physical causation, (ii) because its physical actualization occurs through processes in a biological body. The feedback system as an aspect human communication shares these general characteristics.

Multiple levels of organization and parallel processing

The theory has a perspective on communication and feedback, which implies processes occurring on different levels of organization or put differently as can be seen in table 1 as implying processes that occur with different levels of awareness and control (intentionality). In addition to this, the theory also involves positing several qualitatively different concurrent processes.

Perceptual modality of feedback expression

Like other kinds of human communication, the feedback system involves two primary types of expression, (i) visible body movements and (ii) audible vocal sounds. Both of these means of expression can occur on the different levels of awareness and control discussed above. That is, there is feedback which is mostly aware and intentionally controllable, like the words *yes, no, m* or the head gestures for affirmation and negation/ rejection. There is also feedback that is only potentially controllable, like smiles or emotional prosody. Finally there is feedback behavior which one is neither aware of nor able to control, but that is effective in establishing coordination between interlocutors. For example, speakers tend to coordinate the amount and energy of their body movements without being aware of it.

Types of function/content of the expressions

Communicative feedback concerns expressive behaviors that serve to give or elicit information, enabling the communicators to share information more successfully. Every expression, considered as a behavioral feedback unit, has thus two functional sides. On the one hand it can evoke reactions from the interlocutor, on the other hand it can respond to

	Bodily coordination	Facial expression, posture, prosody	Head gestures	Vocal verbal
Awareness and control	Innate, automatic	Innate, potentially aware + controlled	Potentially/mostly aware + controlled	Potentially/mostly aware + controlled
Expression	Visible	Visible, audible	Visible	Audible
	C, P, E	C, P, E	C, P, U, E, A	C, P, U, E, A
Type of reception	Reactive	Reactive	Response	Response
Type of appraisal	Appraisal, evaluation	Appraisal, evaluation	Appraisal, evaluation	Appraisal, evaluation
Intentionality	Indicate	Indicate, display	Signal	Signal
Continuity	Analogue	Analogue, digital	Digital	Digital
Semiotic sign type	Index	Index, icon	Symbol	Symbol

Table 1. Types of linguistic and other communicative expressions of feedback.

the evocative aspects of a previous contribution. Giving feedback is mainly responsive, while eliciting feedback is mainly evocative. Each feedback behavior may thereby serve different responsive functions. For example, vocal verbal signals (like *m* or *yes*) inform the interlocutor that contact is established (C) that what has been contributed so far has been perceived (P) and (usually also) understood (U). Additionally, the word *yes* often also expresses acceptance or agreement with the main evocative function of the preceding contribution (A). Thus, four basic responsive feedback functions (C, P, U and A) can be attached to the word *yes*. In addition to these functions, further emotional, attitudinal information (E) may be expressed concurrent to the word *yes*. For example, the word may be articulated with enthusiastic prosody and a friendly smile, which would give the interlocutor further information about the recipient's emotional state. Similarly, the willingness to continue (facilitating communication) might be expressed by posture mirroring.

Types of reception
As explained above, feedback behavior is a more or less aware and controlled expression of reactions and responses based on appraisal and evaluation of information contributed by another communicator. We think of these reactions and responses as produced in two main stages: First, an unconscious appraisal is tied to the occurrence of perception, emotions and other primary bodily reactions. If perception and emotion is connected to further processing involving meaningful connections to memory, then understanding, empathy and other cognitive attitudes, like surprise or hope, might occur. Secondly, this stage can lead to more aware appraisal, or evaluation concerning the evocative functions (C, P, U) of the preceding contribution and especially its main evocative function (A), which can be accepted, rejected or possibly met with some form of intermediary reaction, (often expressed by modal words like *perhaps*, *maybe* etc). We distinguish between these two types of reception and use the term "reactive" when the behavior is more automatic and linked to earlier stages in receptive processing, and the term "response" when the behavior is more aware and linked to later stages. For example, vocal feedback words like *yes*, *no* and *m* as well as head gestures are typically responses associated with evaluation, while posture adjustment and facial gestures are more reactive and linked more directly to appraisal and perception.

Types of appraisal and evaluation
Responses and reactions with a certain feedback function occur as a result of continuous appraisal and evaluation on the part of the communicators.

We suggest that the notion of "appraisal" be used for processes that are connected to low levels of awareness and control, while "evaluation" is used when higher levels are involved. The functions C, P, U all pose requirements that can be evaluated as to whether they are met or not (positive or negative). Positive feedback in this sense can be explicitly given by the words *yes* and *m* or head nod (or implicitly by making a next contribution), and negatively by words like *no* or head shakes. The attitudinal and emotional function (E) of feedback is more complex and rests upon both appraisal, i.e. processes with a lower degree of awareness and control, as well as evaluation processes. What dimensions are relevant here is not clear. One possibility is the dimensions suggested by Scherer (1999), where it is suggested that the appraisal dimensions most relevant are (i) novelty (news value of stimulus), (ii) coping (ability to cope with a stimulus), (iii) power (how powerful does the recipient feel in relation to the stimulus), (iv) normative system (how much does the stimulus complies with norms the recipient conforms to), (v) value (to what extent does the stimulus conform to values of the recipient). The effect of appraisal that runs sequentially along these dimensions is a row of emotional reactions, which may include a certain prosody or other behavioral reactions, primarily through prosody and facial display. Additionally, there will be a cognitive evaluation of whether or not the recipient is able and/or willing to comply with the main evocative function of the preceding contribution (A), e.g., can the statements made be believed, the questions answered or the requests complied with.

Types of communicative intentionality
Like any other information communicated by verbal or bodily means, feedback information concerning the basic functions (C, P, U, A, E) can be given on many levels of awareness and intentionality. Although such levels almost certainly are a matter of degree, we, in order to simplify matters somewhat, here distinguish three levels from the point of view of the sender (cf. Allwood 1976): (i) Indicated information is information that the sender is not aware of, or intending to convey. This information is mostly communicated by virtue of the recipient's seeing it as an indexical (i.e., causal) sign. (ii) Displayed information is intended by the sender to be "showed" to the recipient. The recipient does not, however, have to recognize this intention. (iii) Signaled information is intended by the sender to "show" the recipient that he is displaying and, thus, intends the recipient to recognize it as displayed. Display and signaling of information can be achieved through any of the three main semiotic types of signs (indices, icons and symbols, cf. Peirce 1955/1931). In particular, we will regard

ordinary linguistic expressions (verbal symbols) as being signals by convention. Thus, a linguistic expression like *It's raining*, when used conventionally, is intended to evoke the receiver's recognition not merely that "it's raining" but that he/she is "being shown that it's raining".

Degree of continuity (i.e. analog vs. digital)
Feedback information can be expressed in analog ways, such as prosodic patterns in speech, continuous body movements and facial expressions, which evolve over stretches of interaction. It may also be more digital and discrete, such as feedback words, word repetitions or head nods and shakes. Normally, analog and digital expressions are used in combination.

Type of semiotic information carrying relation
Following Peirce's semiotic taxonomy, where indices are based on contiguity, icons on similarity and symbols on conventional, arbitrary relations between the sign and the signified, we can find different types of semiotic information expressed by feedback.

Falsification
A relevant question to ask in relation to all theories is the question of how the theory could be falsified. Since the aspect of the theory that have been presented in this paper mainly consist of a taxonomy of the theoretical dimensions of the theory, falsification in this case consists in showing that the taxonomy is ill founded, i.e. that it is not homogeneous, that the categories are not mutually exclusive, not perspicuous, not economical or not fruitful. Since the question of whether this is so or not can be meaningfully asked, we conclude that theory has empirical content.

EMPIRICAL BASIS
To test our theoretical framework for its adequacy and usability in analyzing multimodal corpora, we have started to gather and analyze data on 30 video-recorded dyadic interactions with two subjects in standing position. The dyads were systematically varied with respect to sex and mutual acquaintance. The subjects were university students and their task was to find out as much as possible about each other within 3 minutes. Extractions of one minute from the video-recordings were transcribed and coded, according to an abbreviated version of the MUMIN coding scheme for feedback (Allwood et al. 2005). The coding schema identifies the feedback units (either verbal or non-verbal), which are coded for function type (giving, eliciting) and attitudes (continued contact, perception, under-standing; acceptance of main evocative function; emotional attitudes). It further captures the following non-vocal behaviors: posture shifts, facial expressions, gaze, and head movements. In addition, intensity and pitch of the (single) audio track were computed using the PRAAT software; movement analysis was applied to measure how the interlocutors' movements vary and coordinate over time. Finally, subjects were asked to fill in a questionnaire about their socio-cognitive perception of the other (e.g. rapport). Fig. 1 shows a snapshot of the annotation board during a data coding session.

CONCLUSIONS
We have presented a theory for communicative feedback, describing the different dimensions involved. This theory is supposed to provide the basis of a framework for analyzing embodied feedback behavior in natural interactions. We have started to design a coding scheme and a data analysis method suited to capture those features that are decisive in this account (such as type of expression, relevant function, or time scale). Currently, we are investigating how the resultant multimodal corpus can be analyzed for patterns and rules as required for a predictive model of embodied feedback. Ultimately, such a model should afford its simulation and testing in a state-of-the-art embodied conversational character.
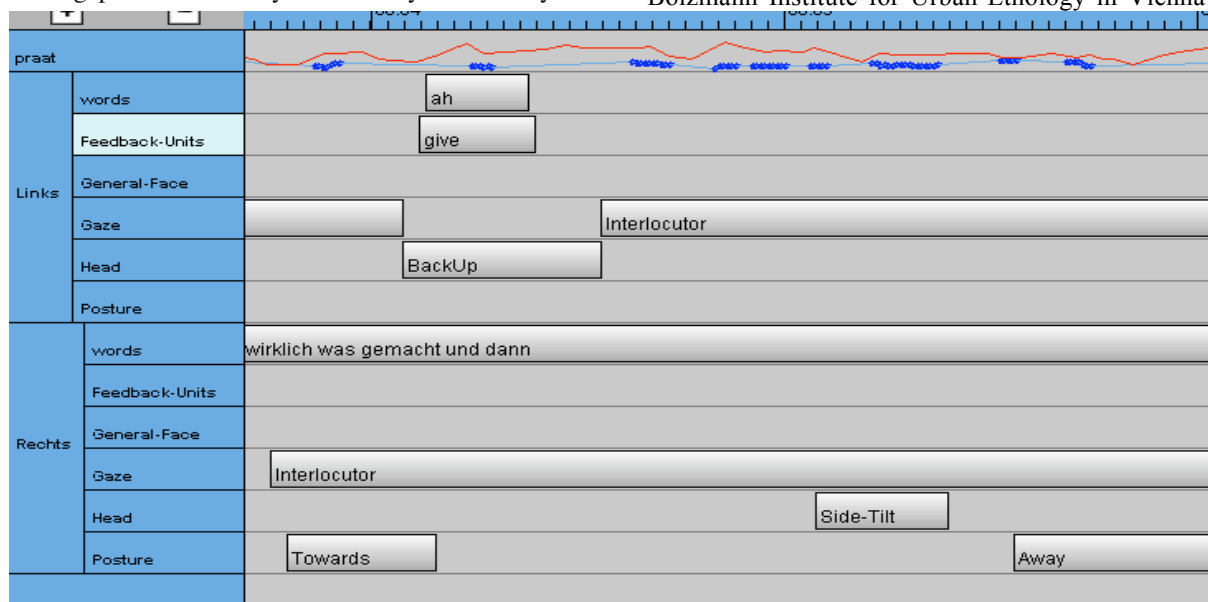
for help with data collection and transcription.

REFERENCES

1. Allwood, J. (1976). Linguistic Communication as Action and Cooperation. *Gothenburg Monographs in Linguistics* 2. Göteborg University, Department of Linguistics.

2. Allwood, J. (2000). Structure of Dialog. In Taylor, M., Bouwhuis, D. & Neel, F. (eds.) *The Structure of Multimodal Dialogue* II, Amsterdam, Benjamins.pp. 3 - 24.

3. Allwood, J., Cerrato, L., Dybjær, L., Jokinen, K., Navaretta, C. & Paggio, P. (2005). The MUMIN Multimodal Coding Scheme. *NorFA Yearbook* 2005.

4. Allwood, J, NivreJ, & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback, *Journal of Semantics*, vol. 9, no. 1, 1-26.

5. Peirce, C. S. (1931). *Collected Papers of Charles Sanders Peirce*, 1931-1958, 8 vols. Edited by Charles Hartshorne, Paul Weiss, and Arthur Burks. Cambridge, Mass., Harvard University Press.

6. Scherer, K. T- (1999). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.) *Handbook of Emoiton and Cognition* (pp.637-663). Chichester: New York.

7. Wiener, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. BIT Press.