Multimodal Communicative Feedback in Swedish

Gustaf Lindblad
SCCIIL (SSKKII)
University of Gothenburg
Gothenburg, Sweden
gustaf.lindblad@gu.se

Jens Allwood
SCCIIL (SSKKII)
University of Gothenburg
Gothenburg, Sweden
jens.allwood@gu.se

Abstract

This study investigates multimodal communicative feedback among speakers of Swedish. We find that the most common way of providing feedback in Swedish is by a multimodal combination of a gestural verbal and a vocal-verbal basic feedback unit, or by just a feedback word or a verbal head gesture on its own. The most common verbal head gestures are nods, and the most common vocal-verbal feedback is just one of four short words. We also find that while nods are primarily used for giving feedback, all other head gestures are more typically used for non-feedback purposes.

1 Introduction

In this paper, it is our intention to describe multimodal communicative feedback in Swedish. Our aim is to present a fairly general overview of Swedish multimodal feedback. The present paper thus continues the work presented in Cerrato (2002) and (2007). The present examination is based on a different data set than previous work. On the basis of this, a second aim is to substantiate or call in to question previous results. The paper will focus on the most common types of communicative feedback, trying to see the typical and broader patterns. Because of the combinatorial properties of multimodal communication, an in depth description would be too extensive, if it were to handle all possible combinations and aspects.

Based on Allwood et al. (1992) and Allwood, Kopp et al. (2007), we define communicative feedback as unobtrusive vocal and gestural communicative contributions that "inform an interlocutor about the ability and willingness to (i) continue the interaction, to (ii) perceive, and (iii) understand what is communicated, and (iv) in other ways attitudinally and emotionally react to this" (Boholm and Lindblad, 2011).

We define gestures as all non-vocal bodily movements that are used for communication. This includes non-voluntary movements that are nevertheless interpreted by the second party as giving information about the message or states of the first party. This inclusive definition is motivated by the fact that it is difficult to draw a definitive line between volitional and non-volitional communicative behavior.

2 Method

The data consists of ANVIL annotations (Kipp, 2001) of six dyadic first acquaintance interactions of Swedish people. In total 11 different persons participate in the interactions (one person participates in two), four of which are female-male interaction, one female-female and one male-male. Each interaction lasts approximately eight minutes (the total length of the six interactions is 48 minutes, 5 seconds), and was filmed using three different camera angles (see Figure 1).

The annotations were transcribed using the Gothenburg Transcription Standard, GTS, (Nivre, 2004), transcriptions imported into ANVIL using Praat (Boersma, 2001), and annotated using the MUMIN coding scheme (Allwood, Cerrato et al., 2007). Regrettably it is not possible to present inter-coder reliability, as the data set used in this article has not been double coded. However, transcriptions and ANVIL annotations alike were checked by at least one person other than the annotator to make sure that they complied with the specifications. We therefore have a fairly strong confidence in the

reliability of this data. It should also be noted, that inter-coder reliability is a somewhat blunt measure of data usefulness, as it does not measure the most valuable characteristic, which is validity.

Figure 1. Examples of what the three camera angels captured during one conversation.

The MUMIN coding scheme provides guidelines for classification of bodily behavior into discrete units in our annotations. We will not describe all the different possibilities here, but because head movements are the most commonly used gestures to provide feedback in conversation, a short description of these varieties is called for.

In accordance with the MUMIN coding scheme we differentiate four different types of nods based on two dimensions of expression: the direction of the initial movement of the nod, and whether it is a single or a repeated nod. This yields the four basic types: down-nod single (ds), down-nod repeated (dr), up-nod single (us) and up-nod repeated (ur). Previous research (e.g. Boholm and Allwood, 2010; Boholm and Lindblad, 2011) has supported this classification, as these different types show different patterns of production. Apart from nods we classify head movements into seven further categories: shake, side turn, tilt, waggle, head forward, head backward and other. The 'shake' refers to the repeated turning of the head from side to side around the longitudinal axis common in most European cultures, 'side turn' refers to just turning the head non-repeatedly. 'Tilt' is a sideways (left or right) slanting of the head away from the longitudinal axis of the body, 'waggle' refers to a rapidly repeated 'tilt'. 'Head forward' and 'head backward' are somewhat similar to nods, but features a rapid initial movement and subsequent slower normalization of the head position, whereas nods are characterized by a more oscillating movement. The 'other' category is used for all other conceivable movements of the head that are not captured by the specified categories.

Every distinct bodily gesture was coded as its own feature (element) in ANVIL, and coded as either feedback or non-feedback. In some cases it is not immediately clear where one gesture ends and the next one begins, but as a general rule we would separate a continuous bodily movement into two or more elements if the movement had salient different parts described by the MUMIN coding scheme. This was primarily an issue with regards to hand gestures, whereas facial expression, head movements and other bodily movements generally had a more pronounced beginning and end.

Vocal verbal contributions were annotated as their own units according to the GTS, with one exception, which is contributions beginning with feedback and then continuing with non-feedback. In these cases, the feedback part and the rest of the contribution was coded as separate units.

3 Results

Out of 4993 annotated features (elements) in our data set, 1486 were coded as providing communicative feedback. Of these, 1406 included either vocal-verbal or verbal head gestures. This means that there were only 80 feedback features using facial, hand or other bodily gestures. Because there are so few of each kind, these are excluded from the further analysis in the present paper.

Gesture category	n.	Multimodal
Body posture	15	14
Facial expression	53	50
Hand gesture	12	10

Table 1. Non-vocal, non-head gesture feedback.

Of the 1406 remaining feedback features 912 are annotated as being multimodal (456 vocal-verbal, 456 verbal head gestures), which means that there are 950 feedback units (1406 - 456 = 950) in the data set. This means that, on average, there is feedback every 3 seconds in these recordings (((48 * 60 + 5) seconds) / (950 feedback units) = 3.04 seconds/feedback unit), illuminating the ubiquity of this phenomenon in conversation.

3.1 Multimodal and unimodal overview

The most common way to give feedback is by means of a multimodal combination of vocal-verbal plus verbal head movement, 456 out of 950 instances (48%). Second most common is a unimodal vocal-verbal feedback, 331 of 950 (35%), and third a unimodal verbal head movement, 163 of 950 (17%). Overall, we see that multimodal and unimodal feedback are equally common, but from the perspective of the respective modalities you can also say that both vocal-verbal feedback and gestural verbal feedback is more often produced as a multimodal unit than as a unimodal unit, with 456 out of 787 (58%) of vocal-verbal feedback and 456 out of 619 (74%) of feedback head movements being produced in a multimodal unit. The ratios are close to identical with what Boholm and Lindblad (2011) found in a different but comparable data set, indicating that these patterns are stable in this kind of casual conversation.

	This study		**Boholm & Lindblad (2011)**	
	n.	**%**	**n.**	**%**
Multimodal	456	48,0%	413	48,9%
Unimodal vocal-verbal	331	34,8%	290	34,4%
Unimodal head movement	163	17,2%	141	16,7%
Total	950	100,0%	844	100,0%

Table 2. Comparison of overall multimodal and unimodal feedback in this study to a study by Boholm and Lindblad (2011).

3.2 Head gestures

There were 1297 head gestures annotated in our data set, of which 621 were annotated as feedback and 676 as non-feedback head gestures. Since there were only two instances of the 'waggle' head gesture used for feedback, this type has been left out from further analysis as a feedback gesture in this paper. Table 3 presents all occurrences of all head gesture types.

Head gesture	**dr**	**ds**	**ur**	**us**	**back**	**forward**	**shake**	**side turn**	**tilt**	**waggle**	**other**
Total	242	127	103	135	89	109	48	179	167	31	67
Non-feedback	68	44	17	26	50	76	33	163	129	29	41
Feedback	174	83	86	109	39	33	15	16	38	2	26
% feedback	72%	65%	83%	81%	44%	30%	31%	9%	23%	6%	39%
Multimodal fb	116	63	63	97	36	10	12	11	32	0	16
Unimodal fb	58	20	23	12	3	23	3	5	6	2	10
% multimodal	67%	76%	73%	89%	92%	30%	80%	69%	84%	0%	62%

Table 3. Occurrences of the different types of head gestures.
(dr = down repeated, ds = down single, ur = up repeated, us = up single)

Something that immediately stands out is that all types of nods are much more frequently used for giving feedback, whereas all other head gestures are more frequently used for non-feedback gestures. This is shown more clearly in Figure 2. We also note that this is most pronounced for up-nods, that seem to be used predominantly for giving feedback, as well as for 'side turn', 'tilt' and 'waggle' which are mainly used for non-feedback gesturing.

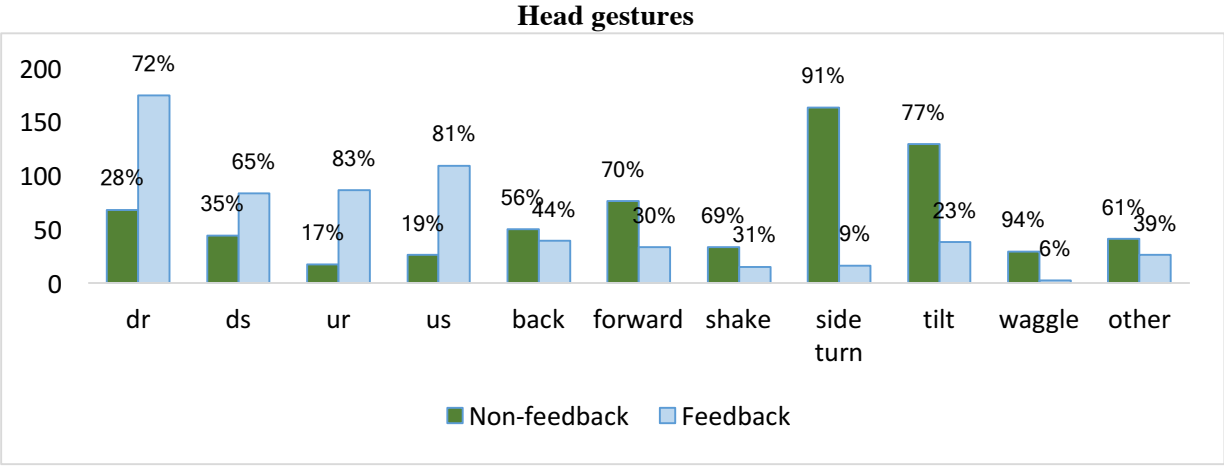

Figure 2. Comparison of non-feedback to feedback head gestures.

Nods are the most common head gestures used in Swedish to express feedback, with 452 out of 619 instances of verbal feedback head gestures in our data (73%) being nods. By contrast, headshakes are the least common of our basic types of head movements, with only 15 instances (2%). This is especially interesting considering that nods and shakes are often regarded as basic head gestures expressing 'yes' and 'no' respectively. For comparison, different basic varieties of 'yes' ('ja') account for 292 out of 787 instances (37%) of vocal-verbal feedback, and basic varieties of 'no' ('nej') for 56 (7%). Considering the multimodal combinations, we find only one instance of a headshake coupled with a vocal-verbal 'yes', whereas seven are coupled with a single 'no', three are coupled with a short phrase beginning with the word 'no', three are unimodal, and one is coupled with a feedback cluster containing the word 'no'.

Most feedback head gestures are multimodal (74%), but broken down into the different types, we find that there are differences. The single up-nod and the head backwards gestures are the most likely gestures to be produced multimodally (around 90% of the time), which is interesting as these gestures are quite similar in their initial phase with an upward-backward movement of the head. The head forward gesture is the only gesture that is produced unimodally most of the time.

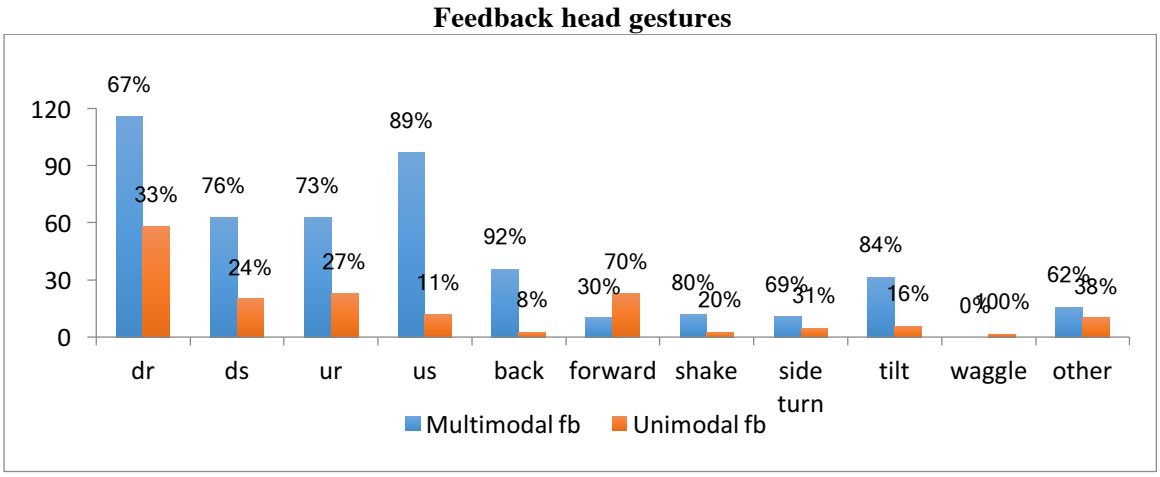

Figure 3. Occurences of feedback head gestures.

3.3 Vocal-verbal

There were 1570 vocal-verbal contributions (or utterances) in our annotated material, with 787 annotated as containing communicative feedback, which leaves 783 as non-feedback. This means that half of all utterances in our data are feedback, which does not mean that half of what is being said is feedback, as the average duration of feedback utterances is 0.49 seconds (st.dev. 0.37) and the average

duration of non-feedback utterances is 2.97 seconds (st.dev. 3.67). It should be noted that in cases where a feedback expression heads up a longer contribution, only the initial feedback part is being used for our calculations. The possible different forms of vocal-verbal feedback are many, but in reality the majority of the feedback utterances fall into a more limited set of categories. In our data 458 of the 787 feedback utterances are one of four basic Swedish feedback words: 'ja', 'm', 'okej' and 'nej'. These words can be produced in some different varieties, for instance with reduction of the 'j' phoneme in 'ja', 'okej' and 'nej'. For the sake of brevity we will disregard these differences and focus only on the basic word types in this paper, though we acknowledge that these differences can be of significance.

There are also 119 cases of what we call feedback clusters or feedback phrases, which are two or more of the basic feedback words produced together in rapid succession. It is very common to repeat the same word (e.g. 'ja ja ja'), but also combinations of two or more different words occurs (e.g. 'ja okej'). In total, this means that 577 out of 787 feedback utterances (73%) consist of one or more of the four most common feedback words in Swedish.

There are 20 cases of what we call 'other repetition', which is when a person gives feedback by repeating a word or utterance that the interlocutor has just said (e.g. A: "I will come tomorrow" B: "Tomorrow", where B's utterance would count as other repetition feedback). Basic feedback words are excluded from this category as not to be counted twice. But it should be noted that also these words can be other-repeated, which reinforces their feedback function.

Of the remaining 190 feedback contributions, no one type has an occurrence of 20 times or more, and most only occur once. Many of them consist of a basic feedback word and a few other words, e.g. 'ja det är det' ('yes it is'), 'ja visst' ('yes sure') or 'nä jag förstår' ('no I see').

Feedback type	ja	m	okej	nej	cluster	other repetition	all others	TOTAL
Total	238	139	38	43	119	20	190	787
Unimodal	106	65	14	22	27	7	90	331
Multimodal	132	74	24	21	92	13	100	456
% Multimodal	55%	53%	63%	49%	77%	65%	53%	58%

Table 4. The most common types of vocal-verbal feedback and their multimodal frequencies.

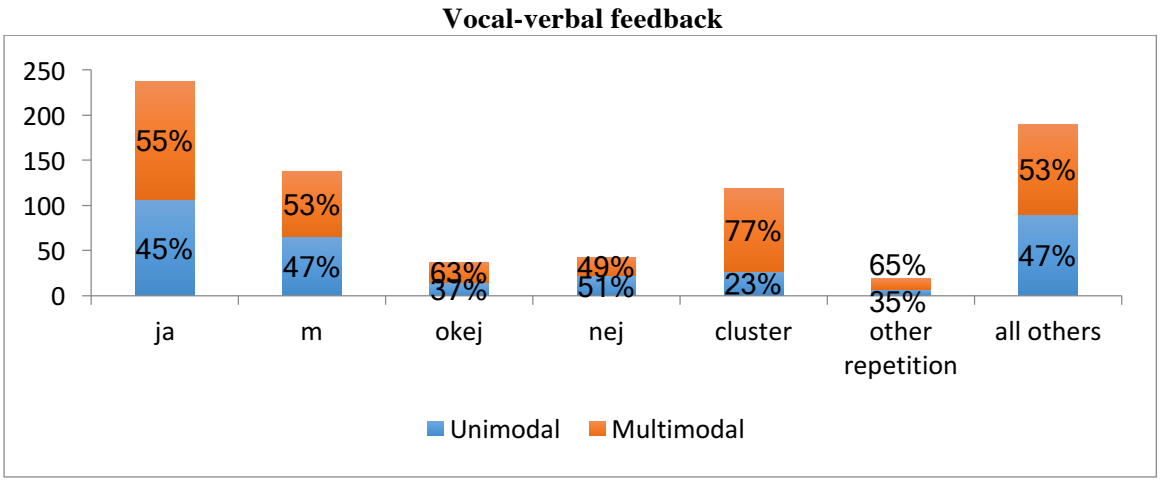

Figure 4. The most common types of vocal-verbal feedback and their multimodal distribution.

It is clear that the most common feedback word in Swedish is 'ja' followed by 'm', whereas 'okej' and 'nej' are much less common although still fairly frequent. This pattern has previously been shown in several studies (e.g. Allwood (2000), Boholm and Lindblad, 2011; Navarretta et al., 2012), and seems to be fairly stable. We also note that the basic feedback words are produced multimodally with head gestures about 50% of the time, with the exception of 'okej' that has a tendency to be coproduced with head gestures more often. Other repetition is also more likely to be co-produced with a head gesture, and feedback clusters even more so.

3.4 Multimodal: vocal-verbal and head gesture

When we consider the combinations of vocal-verbal and head gesture feedback, we find that some combinations seem to be more common than others. It is difficult to make a table that reflects all the interplay between the types, as their frequencies are so varied. Some trends are more easily discernable though. In table 5 we have shaded the cells darker for higher numbers, comparing on the horizontal axis, from the perspective of vocal-verbal feedback. Table 6 is shaded vertically, from the perspective of the verbal head gestures. Each perspective tells a somewhat different story, but we also see several cells where there seems to be some agreement between the perspectives.

Feedback type	dr	ds	ur	us	backward	forward	shake	side turn	tilt	other
ja	40	28	8	27	6	2	1	0	15	5
m	28	11	21	10	3	0	0	1	0	0
okej	1	1	4	12	4	1	0	0	1	0
nej	0	3	2	1	3	0	7	2	2	1
feedback cluster	26	5	18	18	8	1	1	3	6	6
other repetition	3	1	3	2	1	1	0	1	0	1
all others	18	14	7	27	11	5	3	4	8	3

Table 5. Multimodal combinations of vocal-verbal and head gesture feedback, shaded horizontally.

Feedback type	dr	ds	ur	us	backward	forward	shake	side turn	tilt	other
ja	40	28	8	27	6	2	1	0	15	5
m	28	11	21	10	3	0	0	1	0	0
okej	1	1	4	12	4	1	0	0	1	0
nej	0	3	2	1	3	0	7	2	2	1
feedback cluster	26	5	18	18	8	1	1	3	6	6
other repetition	3	1	3	2	1	1	0	1	0	1
all others	18	14	7	27	11	5	3	4	8	3

Table 6. Multimodal combinations of vocal-verbal and head gesture feedback, shaded vertically.

There seems to be a strong coupling of nods and all positive feedback words. Repeated down-nods are most strongly connected with ‘ja’ and repeated up-nods that are mostly coupled with ‘m’ and feedback clusters. Similarly to what Boholm and Lindblad (2011) found, we see that ‘m’ has a correlation with repeated nods. Boholm and Allwood (2010) found a correlation between ‘okej’ and single up-nods, a result that is repeated here. Head shakes and ‘no’ have a strong coupling, as discussed earlier. We also notice that feedback clusters seem to favor repeated head nods somewhat, and it would be interesting to see whether this is correlated to word repetition within these clusters. In the previously cited study by Boholm and Allwood (2010), no such relation was found, but since that study relied on a fairly small data sat, further investigation would still be interesting. Repeated up-nods show the interesting pattern of being somewhat disassociated from ‘ja’ but closely associated with ‘m’ and clusters, raising the question of whether these clusters have ‘m’ in them, or if there is something else going on.

4 Discussion

Even if many of the subtleties of the use of feedback are still unknown, there are some patterns in Swedish communicative feedback that we have noticed re-emerging (e.g. Boholm and Allwood, 2010; Boholm and Lindblad, 2011; Navarretta et al., 2012). Nods are the most common head gestures for feedback, and among them the repeated down nod is the most common, with the single up-nod being the second most common in Swedish feedback. These two nod types show an interesting dissimilarity, in that single up-nods are almost always multimodal, whereas repeated down-nods are the type of nod

most often produced unimodally. One reason for this, we hypothesise, could be that the single up-nod is more often used for emphasis or uptake, while the repeated down-nod is more typically used for giving silent agreement. Single up-nods are sometimes used to signal that the information is new or surprising. It is likely that other aspects of the head gestures, such as intensity, are important for their functions in this regard. In order to investigate these kinds of issues, more in-depth qualitative analysis is needed.

Feedback clusters need to be broken down into their components to see if they show any patterns depending on their parts, such as if repeated nods are correlated to repetition of words, if there are ordering effects or dominant words. We also need to look closer at the big lump of 'others' and we acknowledge that more statistical analysis is needed to substantiate our findings. A very interesting challenge is to look into individual variation in this regard.

It is our intention to increase our sample size, as it is somewhat small. However, we are encouraged by the fact that many of our findings replicate what has been found in other comparable studies. We suspect that there might be more order in this chaos than first meets the eye, and this warrants further investigation.

References

Allwood J., Cerrato L., Jokinen, K., Navarretta C. and Paggio P. (2007). The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing. In Martin et al. (eds) *Multimodal Corpora for Modelling Human Multimodal Behaviour*.

Allwood J., Kopp S., Grammer K., Ahlsén E., Oberzaucher E. and Koppensteiner M. (2007). The analysis of embodied communicative feedback in multimodal corpora: a prerequisite for behavior simulation. *Language Resources and Evaluation*, *41*(3-4), 255-272.

Allwood J., Nivre J., Ahlsén E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics, 9*(1), 1-26.

Allwood J. (ed.) (2000) Talspråksfrekvenser (Spoken Language frequencies). *Gothenburg Papers in Theoretical Linguistics S21*. University of Gothenburg. ISSN 0281-2847.

Boersma P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5*:9/10, 41-345.

Boholm M. and Allwood J. (2010). Repeated head movements, their function and relation to speech. In Kipp et al. (eds.) *Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. LREC 2010.

Boholm M. and Lindblad G. (2011). Head movements and prosody in multimodal feedback. *NEALT Proceedings Series: 3rd Nordic Symposium on Multimodal Communication*, pp. 25-32.

Cerrato, L., (2002), Some characteristics of feedback expressions in Swedish, Proceedings of Fonetik, TMH-QPSR, vol. 44, n.1, 2002, pages: 101-104.

Cerrato L., (2007), Investigating Communicative Feedback Phenomena across Languages and Modalities. University dissertation from Stockholm : KTH, TRITA-CSC-A 2007:3 ISSN-1653-5723 ISRN-KTH/CSC/A--07/03—SE ISBN 978-91-7178-632-6 Format (including language).

Kipp M. (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 1367-1370.

Navarretta C., Ahlsén E., Allwood J., Jokinen K., and Paggio P. (2012). Feedback in Nordic First-Encounters: a Comparative Study. *Proceedings of the Language Resources and Evaluation Conference 2012*, 494-2499.

Nivre J. (2004). *Göteborg Transcription Standard. (GTS) V. 6.4*. Department of Linguistics, University of Gothenburg.