Gestures that precede and accompany speech – An analysis of functions and applicability for virtual agents in different activities

Jens Allwood
University of Gothenburg
Gothenburg, Sweden
jens@ling.gu.se

Elisabeth Ahlsén
University of Gothenburg
Gothenburg, Sweden
eliza@ling.gu.se

Abstract

This paper contains an analysis of features of gesture types that are produced before or simultaneously with speech (mainly. nouns and verbs) and in relation to own communication management (choice and change). The types of gestures discussed are arm-hand gestures, head movements and gaze. The analysis is then discussed in relation to two selected social activities, where virtual agents (ECAs) are or can be used. Gesture types and features with different functions are briefly suggested for each of the two activities and also more in general. The analysis is meant provide information about naturally occurring gestures that can serve as a basis for assigning gestural functions to ECAs.

1 Introduction/Background

Human face-to-face interaction is very much characterized by being multimodal. The analysis of spoken interaction and gesture, taking into account also typical interactive features, not traditionally analyzed for written language has been pursued for a couple of decades and has given us more insight into how human-human interactions works on-line. There are, however, still many phenomena related to interactive functions that are not sufficiently studied and understood. Such phenomena include 1) Interactive Communication Management (ICM), i.e. turn taking, feedback and sequences and 2) Own Communication Management (OCM), i.e. choice and change features in speech related to planning and production processes, for example hesitations and self-repeats (Allwood, 2002).

Turning to gestures in spoken interaction and here focusing on arm-hand gestures, head movements and facial expressions, six main types of content of communication have been suggested. A list of what can be conveyed by gestures in face-to-face interaction (Allwood 2007) is the following:

1. Identity: who a communicating person is biologically (e.g. sex and age), psychologically (e.g. character traits such as introvert or extrovert) or socioculturally (e.g. ethnic/cultural background, social class, education, region or role in an activity).

2. Physiological states: e.g., fatigue, illness, fitness etc

3. Emotions and attitudes: expressed continuously with respect to topic, persons etc.

4. Own communication management: gaining time to reflect, plan or concentrate, having difficulties finding a word (cf. Ahlsén, 1985; Ahlsén, 1991) or needing to change what we have said (cf. also Allwood, Nivre & Ahlsén, 1990).

5. Interactive communication management: to regulate turntaking (cf. Duncan & Fiske, 1977; and Sacks, Schegloff & Jefferson, 1975), feedback to show whether we want to continue, whether we have perceived and understood and how we react to the message (cf. Allwood, 1987; and Allwood, Nivre & Ahlsén, 1992).

6. Factual information: especially words and illustrating or emblematic gestures

The role of gestures as enhancing perception and memorization of verbal messages has been demonstrated by Beattie (2005, 2007). Temporally, gestures can be either preceding, simultaneous with or succeeding the corresponding speech. The effects demonstrated by Beattie can, in principle, be achieved by either of this temporal relations. In making communication more efficient, however, gestures preceding speech or accompanying speech are of special interests. A special case is when a gesture replaces a word or a phrase, which is not spoken.

The relation of the content and function of what is spoken and what is conveyed through gestures can be of different types. Analyzing semantic and semiotic features of what speech and gestures convey can shed more light on how speech and gesture co-contribute to the message. The time relation between gesture and speech with respect to the same target message can also provide clues about the unfolding of the production process.

Embodied Communicative Agents (ECAs) are increasingly being introduced in a number of ICT applications, such as front-ends to databases providing various types of information, pedagogical tools and simulation tools for training. The ECAs used in these applications are often very simple with few communicative functions and very limited variation in means of expressions (e.g. three facial expressions), but there are also advanced ECAs designed, at least partially, for purposes of Artificial Intelligence, i.e. in order to simulate and thereby better understand human interaction and/or to make the ECA appear as human-like as possible, by using a number of salient features from human interaction. Some examples of this are experimenting with the generation of eye gaze and smoothness of gestures (Kipp & Gebhard, 2008, Neff et al. 2008), using gesture dictionaries (Poggi et al., 2005), designing production models for iconic gestures (Kopp, Bergmann et al., 2008), providing models for feedback giving (Kopp, Allwood et al, 2008), studying reactions to behaviors increasing intersubjectivity (Cassell and Tartaro, 2007) and to social versus task only interaction style (Bickmore & Cassell, 2005), comparing direction giving by ECAs and humans (Cassell et al., 2007), evaluating culturally dependent features and intercultural communication (Allwood & Ahlsén, forthc). and creating affective behavior (e.g. Strauss & Kipp, 2008).

This paper is an attempt to, by analyzing human-human communication, focus on a number of features of multimodal communication and discuss gestures with different features in relation ECAs in general and for two different activity types. The paper focuses on two of the main categories of what can be conveyed by gestures, factual content (FC) and own communication management OCM).

2 Method

The analysis was based on a sample of 100 occurrences of gestures preceding or accompanying words, mainly nouns and verbs, in videorecorded spoken face-to-face interaction dyads. 60 of the gestures were primarily identified as illustrating factual content of nouns and verbs, whereas 40 gestures were primarily identified as occurring with own communication management (OCM), i.e. choice and change behavior.

The gestures were coded according to the following features:

- Time: beginning and end

- Target: target word, target word category (mainly for the 60 factual information gestures)

- Contributions: preceding contribution, speech, gesture

- Timing of gesture stroke in relation to spoken contribution/target word: before, same, after (for FC gestures in relation to target word, for OCM gestures in relation to vocal-verbal OCM and target word where a target word can be identified)

- Representational features:

  Description of preparation, prestroke hold, stroke, poststroke hold, retraction

  Gesture form: body part, direction of movement, hand shape

  Complexity: two hands, finger movements, change of hand shape other than fist or open hand shape

- Semantic features of gesture: shape, location in relation to body, functional arm and finger movement, functional hand shape, movement of an object, illustrating action

- Information of gesture in relation to speech: same, added (earlier, more content)

- Gaze direction

- Choice or change function (for OCM gestures)

The features applicable to each of the gestures were coded and used as a basis for the analysis, together with the video.

Two activity types typical for ECA:s were then selected for discussion of types and features of gestures:

- Front-end to database
- Education-training

3 Results

The target word types for factual information gestures are presented in table 1.

	Target word type	
	Noun	**Verb**
Factual information gestures	42%	58%

Table 1. Target word type

There are a more factual information gestures accompanying verbs than nouns in natural spoken interaction.

	Timing	
	Preceding word	**Simultaneous with word**
Factual information gestures	30%	70%

Table 2. Timing of gestures (temporal relation between gesture and target word)

Most of the factual information gestures are produced simultaneously with the target word, as can be seen in table 2, but still a substantial part of them start and have the peak of their strokes before the target word is produced. This is of special interest with respect to the planning and production process as well as for the perception and comprehension process in the interlocutor.

Table 3 presents how much different body parts are used in gestures in the data.

	Body parts		
	Hands		**Head**
	1	2	
Factual information gestures	50%	48%	2%
OCM gestures	0%	88%	12%

Table 3:Body parts used in gestures

Comparing gestures that are mainly arm-hand-finger movements of one hand, two hands or movement of the head, differences are found in distribution between gestures used mainly with factual information and gestures used mainly for own communication management. Gestures used for factual information are fairly evenly distributed between the use of both hands and the use of only one hand, with only very few head movements, Gestures for own communication management, on the other hand are almost always made with one hand only, practically never with two hands, but not infrequently with head movement or gaze.

This indicates that the gestures used with own communication management are most often of a different type than illustrating gestures used with factual information. Iconic gestures which use only one hand are sometimes considered as less complex than if two hands are used. In the case of OCM gestures, this can be one interpretation, while they might also be more fundamentally different, in the typical cases. There is, however, also a considerable overlap and possibly a continuous scale between more representational gestures occurring with nouns and verbs and OCM gestures occurring with communication management. Verbal-vocal OCM as well as OCM gestures often also occur in the context of verb and noun production, when searching for and trying to produce the right noun or verb.

In table 4, a further clue to the planning and production process, i.e. the gaze direction of the speaker during the production of gesture is shown.

	Gaze	
	At interlocutor	**Up/Down/Out**
Factual information gestures	90%	10%
Own Comm.. Management gestures	50%	50%

Table 4. Gaze direction during gesture

Also for gaze direction during gesture, there is a substantial difference between the two types of gestures. During factual information gestures, gaze is almost always directed towards the interlocutor, whereas with own communication management gestures, there is an even distribution of gaze between looking at the interlocutor and ”looking away” (up, down, in front of you or at an object or one’s own hands).

This also points to illustrating gestures used with nouns and verbs perhaps being used more deliberately in order to enhance the listener’s comprehension, by showing/specifying form, size, action, location etc. This does not seem to take so much effort in planning that the gaze has to be averted. With OCM gestures, on the other hand, there is generally a problem of choice or change of verbal-vocal production, which calls for effort and more often requires gaze aversion. In this case, both gaze aversion and gesture indicate planning problems.

For OCM gestures, choice and change functions are distributed as follows (table 5).

	OCM function	
	Choice	**Change**
Own Comm.. Management gestures	82%	18%

Table 5. Own communication management: choice vs. change function

Choice OCM is much more common, both in speech and gesture, than change OCM. This could, however, vary with both individual speaker type and activity type. It is also the case that about 40% of all speech based choice related OCM involves gestures, whereas only 15% of speech based change related OCM is accompanied by gestures (Allwood et al., 2002).

What is, then, the content and function of the factual information gestures? In table 6, the semantic features are presented ranked according to frequency of occurrence.

Semantic features of factual information gestures	
Illustrating action	53%
Illustrating shape	33%
Illustrating location	10%
Functional hand movement	3%
Functional hand shape	1%

Table 6. Semantic features of gestures used for factual information

In the 60 gestures used for factual information with nouns and verbs, 72 semantic features were coded and among these 72 features the distribution was, as shown in table 6, that illustrating an action was the most common feature, followed by shape and location (on the body or in the room). This is consistent with more gestures occurring with verbs than with nouns, although the difference between gestures illustrating actions and gestures illustrating shape is somewhat greater than that between gestures with verbs and gestures with nouns. Gestures illustrating action are also sometimes used to illustrate the meaning of nouns and gestures illustrating shape can be used also with verbs.

Turning to OCM gestures, about 20-25%, according to Allwood & Ahlsén (2002), are illustrating content in a similar way to that of factual content gestures. The rest have more general functions having to do also with self-activation and interaction regulation, especially turn keeping.

4 Discussion

A summary of findings is that

- factual information gestures are used more with verbs than with nouns
- they are most often simultaneous with the noun or verb, but in 30% of the cases precede the target word
- there is an even distribution between use of one and two hands in factual information gestures, but only use of one hand gestures and to some extent head movements (10%) in OCM gestures
- gaze is practically always directed at the interlocutor when factual information gestures are produced, but evenly distributed between gaze at the interlocutor and gaze directed elsewhere with

OCM gestures (gaze aversion could in itself also be considered an OCM gesture).
- more than 80% of the OCM gestures have choice function, rather than change function
- the most frequent semantic feature of factual information gestures is illustration of action, followed by illustrating of shape and location.

This overview is based on a limited sample of data, but can be compared also to earlier studies (e.g. Allwood, Ahlsén et al. 2002) and it gives a general idea of how the two types of gestures are used.

Turning to web-ECAs, the use of both types of gesture (FC and OCM) can be useful in many contexts and the findings reported in this study can be applied more or less directly in the design of ECAs. It is, however, no trivial task to implement gestures of these types in an ECA, so that they will (i) occur with the right context and timing, (ii) be chosen correctly, and (iii) be produced in a way that looks more natural than disturbing. Especially the factual content gestures produced with verbs and nouns are most often quite specific and require either (i) that the ECA has a fixed repertoire of verbal-vocal and gestural output for a specific task, modeled on what a human produces in the same task, i.e. a form of copying of sequences and combinations in context or (ii) that extensive dictionaries of gesture-word correspondences are available and are culturally adapted and can be linked to specific words in production. Considering the OCM gestures, the same is true to some extent, since they often occur when a person has problems producing the right words and then also in a fairly specific way illustrate the content of the intended word or phrase. There are, however, a number of other typical features of OCM gestures, such as discrete pointing to oneself when referring to oneself, to the interlocutor when referring to the interlocutor, pointing to one's head or mouth when referring to own memory or production problems, moving hand in a certain direction in relation to movement verbs and also more metaphorically in relation to more abstract words (forward for words indicating traveling, walking, running, biking etc as well as progress and reference to future; to the side for throwing away, canceling etc; to the back for past time, leaving behind etc.). These types of gestures contain factual content and can also be found with nouns and verbs.

Looking at our two exemplifying activity contexts for an ECA, both types of gesture can be useful in both activity types.

1. The more general OCM gesture types can be used when there is unclarity, "hesitation", time needed for processing, need for change and problems of understanding.
2. The more specific types of gestures illustrating the content of verbs and nouns (most often actions and shapes or locations) can be exploited in enhancing the salience and clarity of spoken (or written) output.

In both cases, timing is essential, as well as gestures that appear fairly natural in the context.

For an ECA as front end to a database, gestures illustrating the content of frequently used words could be included linked to the words, e.g. illustrating the shape of a paper, form, ticket etc., index finger tracing line for reading, writing movement for writing, typing movements for entering data via the computer, hand-to-ear movement for phoning, driving movement (holding driving wheel) for driving, stop sign for stopping etc. Pointing to clock for opening hours, gestures illustrating packing, sending, picking up etc. as well as many kinds of directive pointing can also be useful.

An ECA in an education interface should have specific gestures adapted to what type of education it is used for. Here, pre-prepared sequences of actions for specific procedures can be used, that are specifically designed and related to the words (e.g. nouns and verbs) included, e.g. for learning to make something (practical-procedural education). For more theoretical education, pointing, showing and giving directions by gesture in combination with reference to pictures could be used, also here possibly with gesture-word links from a dictionary.

These are just exemplifications of how the types of gestures in this study can be used in ECAs. IN general, it can be concluded that most of the gestures are fairly specifically linked to specific content words and fairly hard to implement in a natural way, except for pre-prepared and human-based "scenarios" or "sequences" of an ECA. The use of gesture dictionaries is cumbersome and still needs considerable work. For the more general types of OCM gestures, it should, however be much easier to implement them in ECAs in general and they could potentially add to naturalness in the appearance and interaction of ECAs, especially in problematic sequences.

Since gestures are known to enhance perception and memory processes by adding redundancy but also by specifying semantic features, multimodal presentation is a worthwhile enterprise, even though it is fairly complex, as for most of the gestures of this study. The study has presented some of the features to be considered for gestures used for factual content and own communication management.

References

Ahlsén, E. 1985. Discourse Patterns in Aphasia. *Gothenburg Monographs in Linguistics*, 5. University of Gothenburg, Department of Linguistics.

Ahlsén, E. 1991. Body Communication and Speech in a Wemicke's Aphasic - A Longitudinal Study. *Journal of Communication Disorders*, 24:1-12.

Allwood, J. 2002. Bodily communication – dimensions of expression and content. B. Granström, D. House & I. Karlsson (eds) *Multimodality in Language and Speech Systems*. Kluwer, Dordrecht.

Allwood, J. & Ahlsén, E. Multimodal intercultural Information and Communication Technology – A conceptual framework for designing and evaluating Multimodal Intercultural Communicators. (Forthcoming in M. Kipp, J.-C. Martin, P. Paggio & D. Heylen (eds) *Multimodal Corpora*. Springer Verlag.)

Allwood, J., Ahlsén, E. ; Lund, J. et al. 2007. Multimodality in own communication management.. *Current Trends in Research on Spoken Language in the Nordic Countries*. II, 10-19.

Allwood, J., Nivre, J. & Ahlsén, E. 1990. Speech Management: On the Non-Written Life of Speech. *Nordic Journal of Linguistics*, 13:3-48.

Allwood, J., Nivre, J. & Ahlsén, E. 1992. On the Semantics and Pragmatics of Linguistic Feedback. *The Journal of Semantics,* 9.1.

Beattie, G. 2005. Why the spontaneous images created by the hands during talk can help making TV advertismements more effective. *British Journal of Psychology*, 97:21-37.

Beattie, G. 2007. The role of iconic gestures in semantic communication and its theoretical and practical applications. In Duncan, S., Cassell, J. &Levy, E. (eds.) *Gesture and the Dynamic Dimensions of Language*, pp. 221-241.

Bickmore, T., Cassell, J. .2005. Social Dialogue with Embodied Conversational Agents. In van Kuppevelt, J., Dybkjaer, L. & Bernsen, N. (eds.), *Advances in Natural, Multimodal Dialogue Systems.* New York: Kluwer Academic.

Cassell, J., Kopp, S., Tepper, P., Ferriman, K. & Striegnitz, K . 2007. Trading Spaces: How Humans and Humanoids use Speech and Gesture to Give Directions. In Nishida, T. (ed) *Conversational Informatics.* New York: John Wiley & Sons, pp. 133-160

Cassell, J., Thórisson, K.: (1999). The power of a nod and a glance. Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence,* 13:519-538.

Duncan, S. and Fiske, D. 1977. *Face-to-Face Interaction.* Lawrence Erlbaum, Hillsdale. N.J.

Kipp, M. & Gebhard, P. 2008. IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA-08), LNAI 5208*, Springer, pp. 191-199.

Kopp, S., Allwood, J., Ahlsén, E. et al. 2008. Modeling Embodied Feedback with Virtual Humans. In: Springer series Lecture Notes in Computer Science (LNBCS) subseries Lecture Notes in Artificial Intelligence (LNAI). I. Wachsmuth & G. Knoblich (Eds.) *Modeling Communication with Robots and Virtual Humans*, LNAI 4930. p. 18-37, Springer, Berlin.

Kopp,S., Bergmann, K. & Wachsmuth, I. 2008. Multimodal communication from multimodal thinking - Towards an integrated model of speech and gesture production. *Int. Journal Semantic Computing* 2(1):115-136.

Neff, M., Kipp, M., Albrecht, I. and Seidel, H.-P. 2008. Gesture Modeling and Animation Based on a Probabilistic Recreation of Speaker Style. In *ACM Transactions on Graphics* 27 (1), ACM Press, pp. 1-24.

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., De Carolis, N.: 2005. GRETA. A Believable Embodied Conversational Agent. In Stock, O., Zancarano, M. (eds.) *Multimodal Intelligent Information Presentation* Kluwer, Dordrecht.

Sacks, H-, Schegloff, E. & Jeffersson, G. 1974. A Simplest Systematics for the Organization of Turntaking in Conversation. *Language,* 50:696-735.

Strauss, M. & Kipp, M. 2008. ERIC: A Generic Rule-based Framework for an Affective Embodied Commentary Agent. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*