Analysing the Modifying Functions of Gesture in Multimodal Utterances

Farina Freigang and *Stefan Kopp*

Social Cognitive Systems Group
Faculty of Technology and Center of Excellence "Cognitive Interaction Technology" (CITEC)
Bielefeld University, Bielefeld, Germany
farina.freigang@uni-bielefeld.de, skopp@techfak.uni-bielefeld.de

Abstract

Gestures may contribute to the meaning of an utterance not only by *adding* information but also by *modifying* the gestural or verbal content uttered in parallel. The phenomenon of modification is more common in natural interaction than it has been given attention. We created a corpus of natural communicative gestures and body movements and conducted a study to examine their modifying functions. Results show that index-finger-pointings are most prominent, which emphasise and affirm an uttered content and, thus, are not only used for referencing but also for modifying. Holds emphasise and colour the utterance by showing a stance towards something. Brushing gestures change the utterance in a discounting or downtoning way. A cluster analysis suggests four distinct categories: a focusing category for emphasising an aspect, an epistemic-attitudinal category to convey one's own stance, an epistemic category for uncertainty, and a category where multiple viewpoints are discussed.

Index Terms: gesture, body movements, modifying function, corpus, empirical study, gesture and speech in dialogue, relevance of gesture unit.

1. Introduction

"Sentences are rarely uttered in a behavioural vacuum. We colour and flavour our speech with a variety of natural vocal, facial and bodily gestures, which indicate our internal state by conveying attitudes to the propositions we express or information about our emotions or feelings." [1, p. 1]

Just as prosody adds modal and affective tones to the semantic propositions carried in speech (e.g., [2]), there can be modifying functions of gesture and body movements. These functions may operate on top of the propositional meaning of either speech or the gesture itself. Thus, a gesture may realise a pragmatic modification of the whole utterance meaning. Uncertainty and miscommunication in human interactions may be minimised if one gets hold of which functions appear in natural communication and how they should be interpreted.

In this paper we present an empirical analysis of these, so far under-researched modifying functions of gesture and body movements. Natural for pragmatic modifications or implications are that they depend on the context to which they are added. We want to investigate in particular the modifying functions that gestures can have in different situations, how gestures and movements can be categorised accordingly, and how those can be possibly combined at the same time (multifunctionality). We assume three general classes of functions that express either positivity or negativity related to importance, opinion/emotion and/or knowledge. Besides that, various other interpretations are possible. One of the main goals of this research is to shed light on how modifying functions influence the overall interpretation of an utterance, hence looking more comprehensively at what pragmatic meaning can be communicated by nonverbal behaviour. In the following, we will discuss related work and present our conceptual approach. We then present a rating study on how human observers perceive and interpret modifying functions carried by natural gestures when their verbal context is present vs. non-present. The analysis of the rating study is twofold: we first present descriptive statistics, followed by a cluster analysis.

2. Background

Within the category of pragmatic functions, "Gestures are said to have modal functions if they seem to operate on a given unit of verbal discourse and show how it is to be interpreted." [3, p. 225] Those "modal functions" may be used to express "an hypothesis or an assertion, and the like" [3, p. 159], they are used as "an implied negative" or an "intensifier for an evaluative statement" [3, p. 225]. One gesture that may carry such a modifying function, is the brushing aside gesture, which "usually serves a modal and discursive function: qualifying something as negative and marking the end of a certain discursive activity" [4, p. 1536]. The term "modal function" will be referred to as "modifying function" in this work.

Another gesture category reported to carry modifying functions are open-palm hand gestures [5], in which a hand flip may express epistemicity or a judgemental modality. Also, [6] investigated functions of hand gestures in two Democratic Party primary debates during the 2004 US presidential campaign and observed the following forms: the extended index finger, the slice gesture, the ring (precision grip) gesture, and the power grip. However, besides analysing gestures with a highlighting function, he only focuses on discourse functions. Additionally, [6]'s analysis is based on politicians, which are assumed to perform practised gestures.

In the present work, we are interested in investigating `modifying functions` (MF) in more depth. We concentrate on naturally produced human gestures and body movements that occur in (dyad) interactions, which may carry MF and were accompanied by speech of the same person. The following body movements (BM) are considered: head and shoulder movements, hand and arm gestures and upper BM; and, additionally, coarse facial expressions. We define MF as follows:

If P is the propositional meaning of an utterance (verbal and/or nonverbal), BM or gesture may additionally signal MF which act as an operator F such that F(P) is the combined meaning of the entire multimodal utterance with:

$$F(P) \neq P.$$

Freigang, F. & Kopp, S. (2015). Analysing the Modifying Functions of Gesture in Multimodal Utterances. In Proceedings of Gesture and Speech in Interaction (GESPIN 4), pp. 107-112.

Our approach is based on the assumption that, when accompanying speech, BM and gesture may *not only* carry propositional meaning(s) or aspects referring to some content, instance, object, referent or situation of interest in the real word, but that they carry meaning *beyond* any of those propositions. We are interested in exactly all of these BM and gestures.

The BM and gestures meant here belong to the category of 'pragmatic gestures', meaning gestures that take up a pragmatic function [3, p. 158]. However, our definition goes a little further: We disregard any so termed 'pragmatic gestures' that influence the *structure* (e.g., marking the beginning of discourse ("attention-refocusing" [7]), feedback), or the *timing* (e.g., turn taking) of an utterance, or *refer* to a person or an issue under discussion with a little point or nod [8]. We solely consider MF in pragmatic gestures. More specifically, MF that accompany a propositional message (which is either verbal or nonverbal, i.e., expressed by speech, BM, or gestures) *change* or *add* something to the meaning of the overall message, so that the resulting overall message is different from the 'purely propositional' message. Thus, MF in BM and gestures frame the overall meaning of the utterance, namely, they indicate what a person intentionally and non-intentionally communicates and which BM and gestures are used in this process. MF in BM are comparable to modifying words or prosody in speech, which may modify the propositional meaning of a message (e.g., certain acoustic cues are used to convey irony [9]). Although we are interested in the functions of these gestures, we will be using form categories in order to describe how these MF in pragmatic gestures may manifest themselves.

3. Study Design

We conducted a study to investigate and unravel the MF that humans see in BM and gestures. The study was carried out in two parts: first, the corpus creation part included recording and annotating utterances of the candidates and, second, in the rating study part data was presented and rated by naïve participants.

3.1. NIC - Natural Interaction Corpus

The Bielefeld `Natural Interaction Corpus` (NIC) of nonverbal behaviour is comprised of eight dyads (4 female-female, 4 male-male interactions); each dyad consisting of two iterations, one after each stimulus video (the two stimulus videos were shown in alternating order). This results in 16 interactions with three audio recordings (over the participant's headset and a room microphone) and three video recordings from different angles (each participant recorded from a slight side-angle and both participants from a bird's-eye view perspective). The total length of all videos is 1 hour and 45 minutes, with an average length of 6 minutes and 30 seconds per interaction, and the recording took place at a CITEC laboratory in September 2014. The stimulus videos consisted of technical instructions with varying complexity and relevance: one was about how to operate a mobile working platform (from which to cut trees, among others) and another video was about how to grout the joints of tiles using silicone. The participants[1] were university students and university staff, all were German native speakers (self-reported), with an average age of 25 (in a range of approx. 20 to 40) years and were paid for 50 minutes of participation in the study. After watching one stimulus video, the two participants talked about the video (with no other person being present in the study room). The participants were informed that we are interested in natural human behaviour in spontaneous dialogues in order to shed light on facets of human communication. In no situation it was referred to BM or gesture. However, we seated the participants on three-legged stools that are a little higher to make it easier for them to use their arms and hands (the rest position was usually the thighs) and which were placed in contiguity and facing one another. The participants performed many natural BM and gestures (although as expected, this depended on the extroversion of the person), among which we also found MF in BM and gesture, mainly pointings, holds and brushings.

In the post-processing of the data, we created manual annotations in ELAN[2] [10] of BM and gestures that we speculated may carry a MF. We define MF to have a *focusing*, an *attitudinal* and an *epistemic* component.

- A A **focusing** function highlights or brings into or out of focus an aspect of communication that was communicated by the utterance giver. The utterance giver wants to ensure that the interaction partner perceives the piece in or out of focus.
- B An **attitudinal** or an emotional function expresses an utterance giver's stance, opinion or feeling regarding an aspect of communication. The utterance giver wants to communicate a personal viewpoint and maybe even convince the interaction partner of it.
- C An **epistemic** function refers to knowledge or lack of knowledge of an utterance giver regarding an aspect of communication. The utterance giver may want to communicate an assessment or rating of a knowledge content of the same or a different utterance.

BM and gestures that seemed to carry any of these MF were annotated according to three categories: (1) salient movements, those which obviously have a MF and were executed quite clearly, (2) relevant movements which belong to the mainly chosen category, and (3) borderline movements, which showed only very fast, short, small, not easy to recognise MF in movements. BM and gestures were annotated if all of the following criteria could be satisfied: the BM fits the definition of MF (A-C), the BM carries a MF which operates 'on top' of a propositional meaning of BM, the BM shapes at least a referent *and* a MF and not a referent alone, the BM is integrated in a person's utterance and does not stand alone, the BM does not involve any of the following: turn-taking, feedback, word finding, questions, self-adaptors. Additionally, we annotated which of the following body parts were involved in a movement: right/left/both hand(s) (also referring to fingers, e.g., pointing with an index finger), right/left/both arm(s), and (right/left) shoulder(s). We plan to extend the annotation scheme similar to the one created for interpreting the clusters of the cluster-analysis (cf. section 5.2).

3.2. Rating Study

The judgement of uninformed participants in a subsequent rating study had to prove whether other persons also see the MF in BM and gestures. The participants rated the utterances in terms of 14 adjectives that we assumed, first, to be intuitively understandable and, second, correspond to the range of possible combined meanings that can be related back to specific MF.

This study being a proof of concept, we chose not to include fillers and use mostly BM and gestures of the first category

[1] In the following, only pictures of those participants are depicted that agreed to it in a consent form.

[2] EUDICO Linguistic Annotator developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(salient movements) and only in order to balance for various aspects (see below) we added six gestures of the second category (relevant movements). On the basis of the annotations, we automatically extracted video snippets, adding 500 milliseconds before and 150 milliseconds after an annotation. The snippets were between 1194 and 1784 milliseconds long, leaving almost no space for contextual information. Although we are looking at pragmatic functions here, it seemed important to investigate these functions in a minimal time frame in order to extract context independent features.

The experiment consists of two conditions: the first one is the speech-and-gesture (S+G) condition, in which the videos are shown as described above (with speech and with head) and in the second condition, the gesture-only (Gonly) condition, the videos are muted and cropped so that the snippets show only the region between the neck down to the upper legs of the participants (without speech and without head), putting movements of hands, arms, and shoulders into focus (cf. Figure 1). The Gonly-condition provides an isolated view on BM and gestures carrying MF in order to see how much meaning is left. The same 36 video snippets were provided for both conditions. Participant group A watched 18 snippets in the S+G-condition and the other half in the Gonly-condition and vice versa for participant group B, i.e., the videos that a participant saw were all different. We made sure the groups were balanced according to coarse gesture groups (e.g., pointing, hold, brushing), the different participants of NIC and the gender of the participants. Every participant watched the video snippets in the S+G-condition before the Gonly-condition, to mask that this study concentrates on BM and gestural behaviour. For each participant and each condition, the videos were shown in a random order.

The procedure of the rating study was as follows: The participants started with the S+G-condition and every video snippet was played to them three times in a row on the left side of the screen. After these automatic displays, a button could be pressed as often as desired to replay the video. The right side of the screen displayed the heading 'The utterance of the person is ...' with a 7-point Likert scale (excluding forced decision making; 'matches exactly' (1) to 'does not match at all' (7)) and then listing 14 adjectives (displayed in random order with every new video): 'discounting/downtoning', 'revaluing', 'affirmative', 'emphasising', 'classifying', 'emotionally coloured', 'focused', 'critically', 'opinionative', 'negative', 'positive', 'relevant', 'humorous', 'uncertain'.[3] The participants had to rate how much each adjective fitted the expressive behaviour of the person in the video and an answer for every adjective was necessary in order to move on to the next video. No definitions were given for the adjectives, leaving it up to the participants to decide what they mean to them. An optional text field was provided at the bottom of each screen asking for adjectives that would be more characteristic for the utterance. After 18 videos in the S+G-condition had been answered, all participants rated the other set of videos in the Gonly-condition. The final part of the study consisted of definitions for MF and a rating of how the 14 adjectives fit each definition (evaluated by a 7-point Likert scale). The study has been implemented in the Python programming language as a guided user interface, extracting and saving the answers of the participants automatically.

The rating study took place in a seminar room of the university building in March and April 2015. The task was described to the participants as classifying natural utterances of humans nonverbal behaviour; BM and gestures were not mentioned at all. A total of 27 participants took part in the study (13 female, 13 male, 1 other gender). The participants were university students and university staff, all were German native speakers (self-reported), had an average age of 29 (in a range of 21 to 55 years) and were paid for participating in the study (taking from 20 to 50 minutes, depending on answering speed).

Figure 1: Video snippets of the S+G-condition (left) and the Gonly-condition with muted videos (right). The gesture depicted here is a space holding gesture: the participant performs a circle while saying "I know a lot about this topics".

4. First Rating Study Results

In the following, preliminary results of the video ratings will be presented. Given all ratings for all adjectives, we got a rather normal distribution of all votes with a small tendency towards adjectives that 'do not match' a BM or gesture in a given video snippet. This tendency is a little bigger in the Gonly-condition, when only BM and gestures are observed without sound. In the present analysis, we concentrate on what raters do see in the videos, namely, which adjectives fit the utterance of the person in the video. Tables of the results of the rating study can be downloaded online.[4] In section 5, we will present a cluster analysis based on the same data.

4.1. Adjectives Describing MF of BM and Gesture

The 14 adjectives (as mentioned in section 3.2) were the items of the rating study, which were used with varying frequency to describe the BM and gestures in the videos. Those adjectives that were rated as 'matching a video positively' (*adjectives that describe well what the utterance of the person in the video* does *express*) a lot of times, were 'affirmative' and 'emphasising' and also quite frequent were 'focused' and 'opinionative'. Predicative for 'matching a video negatively' (*adjectives that describe well what the utterance of the person in the video* does not *express*) were 'discounting/downtoning', 'revaluing', 'affirmative', 'humorous', 'uncertain', and also 'critically', 'negative' and 'positive'. In fact, 'humorous' was the most often and clearly rated adjective for describing well what the utterance of the person in the video *does not* express: which on average was the case for every fourth adjective in the S+G-condition and every sixth adjective in the Gonly-condition.

[3]The exact German words were: 'Die Äußerung der Person ist ...', 'abtuend', 'aufwertend', 'bestimmt', 'betonend', 'einordnend', 'emotional gefärbt', 'fokussiert', 'kritisch', 'meinungstragend', 'negativ', 'positiv', 'relevant', 'scherzhaft', 'unsicher'.

[4]Tables of the results of the rating study grouped according to the clusters of the cluster analysis: `http://pub.uni-bielefeld.de/luur/download?func=downloadFile&recordOId=2763501&fileOId=2763503`

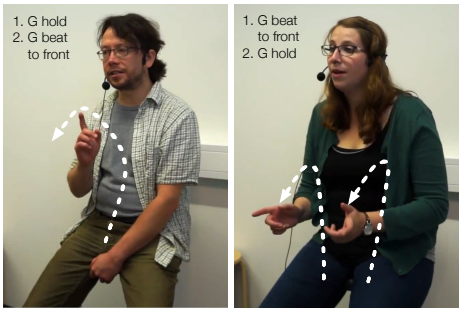

Figure 2: Index-finger-pointings: The 'baseline' pointing gesture (left) and a pointing with two hands (right).

4.2. Stable Modifying Functions

As a first analysis, we only considered the video snippets that were rated with maximum clarity and agreement, i.e., with a low standard deviation, namely $\sigma \leq 1.0$, and showing a clear tendency towards one of the poles: $\mu \leq 2.5$ for 'does match' and $\mu \geq 5.5$ for 'does not match'. For now, we will only consider the positive or 'matching' cases, that is, the videos with BM and gestures that have been rated with one or more adjectives *matching* what the utterance of the person expresses.

Five videos fulfil these criteria and all of them show pointings with the extended index finger, labelled with the four positive adjectives (cf. section 4.1), but particularly 'affirmative' and 'emphasising'. Three of these video examples are extremely prominent as they have been rated clearly with $\mu \approx 1.6$. Besides the index-finger-pointing, these gestures include an obvious hold of the index finger, a beat before the hold and in one case the index-finger-pointing was done with two hands in parallel (cf. Figure 2). The gestures in the other two videos show only a short index-finger-pointing, in one case the participant snaps his fingers during pointing.

Analysing the data further, and allowing for more uncertainty, the results are not as clear but we can observe a few tendencies. As stated before, the coarse gesture groups are pointings, holds and brushings (example gestures are depicted in Figures 1 to 3). In the group of holdings, some gestures are rated as 'emphasising' and 'opinionative', but also 'affirmative', 'classifying' and 'focused'. Brushing gestures are often seen as 'discounting/downtoning' but also 'emotionally coloured'.

4.3. MF in the Gonly-Condition

The three prominent pointing gestures of the S+G-condition are also most prominent in the Gonly-condition with $\mu \approx 1.6$. In one of the three videos, an index-finger-pointing is rated 'affirmative' with even $\mu \approx 1.4$. The other adjectives associated with pointing gestures are 'emphasising', 'focused' and 'opinionative'. This suggests a certain amount of communicative 'self-containment' of the gestures, even without verbal context. Within the set criteria ($\sigma \leq 1.0$, $\mu \leq 2.5$), another less prominent and quickly performed pointing gesture emerges, rated as 'emphasising'. This comprises four positive examples in the Gonly-condition with index-finger-pointings. Further analyses of the Gonly-condition shows even weaker but the same tendencies towards 'emphasising' and 'opinionative' for holding gestures and 'discounting/downtoning' for brushings.

Figure 3: From left to right: First, two holding gestures (one hold open and one hold with one hand) and, second, two brushing gestures (one brushing over the own hand and the other one is brushing/shovelling something away with both hands).

4.4. Interaction Between Modalities

When allocating solely a positive or negative meaning to a BM or gesture and a verbal utterance separately, we observed incongruences between the two modalities. These incongruences are often overwritten by one modality, unless this modality is missing (as in the Gonly-condition). For instance, while performing a shovelling away gesture with two hands (negative), the person has an outstanding positive attitude reflected in the voice and her facial expressions (cf. Figure 3, picture on the right). One prominent rating (according to the criteria above) for this example is that this utterance is 'not negative' ($\mu \approx 6.1$), assigning less weight to the gesture while interpreting the mismatching cues in her utterance. In the Gonly-condition, the gesture is interpreted as neutral ($\mu \approx 4{,}5$). In a similar example, the brushing away gesture and the manner of performance (fast, hitting, with a final flap at the end) is purely negative (cf. Figure 3, third picture) just as rated by the participants in the Gonly-condition: 'not revaluing' ($\mu \approx 6.0$) and 'not positive' ($\mu \approx 5.8$). However, the voice in the video and the facial expressions are rather positive and, consequently, the ratings in the S+G-condition where these features were observed are less negative ($\mu \approx 5.0$ for 'not revaluing' and $\mu \approx 4.4$ for 'not positive').

5. A Cluster Analysis

In the following, we will present results of a cluster analysis on the rating study data with Ward's method.

5.1. Method of Cluster Analysis

Ward's method or *minimum variance method* is a criterion applied in hierarchical cluster analysis. Other clustering methods are used to fuse cluster pairs with the smallest distance (or greatest similarity) in each step. With the Ward criterion, however, clusters and objects are merged step by step in order to reach the smallest increase of heterogeneity within a cluster. Heterogeneity is measured by the sum of variances within the clusters.

We chose a hierarchical clustering method – in contrast to a partitional clustering method like *K-means*[5] – in order to avoid defining the number of clusters before running an algorithm, which we could not know. Then, our goal was to find homogeneous groups, which was perfectly given by Ward's method. Data outliers cannot be identified but we hypothesise that all pragmatic gestures and BM have a particular meaning or func-

[5]K-means also calculates the sum of variances within the clusters and is therefore quite similar to Ward's method; although it proceeds differently.

tion and no outliers exist. Several sources confirm that Ward's method is the preferred hierarchical clustering method, citing the comparison of [11].[6] We analysed the rating study (cf. section 3.2) using this method in SPSS.[7]

After applying Ward's method, we used the *Elbow method* to decide on the number of clusters. Here, the percentage of variance is plotted, showing where the slope between data points increases notably and the one data point that forms this 'elbow' indicates the number of clusters. However, the best number of clusters is subjective and not clear-cut. For the S+G-condition an obvious cut exists after four clusters (first 'elbow') but if we are very precise and allow for smaller clusters, we could also agree on seven clusters (second 'elbow'). In order to provide the full picture of the data, we will describe the S+G-condition with four clusters and the according subclusters. Exactly five clusters appeared for the Gonly-condition. For the cluster results compare the tables online.[4]

5.2. Cluster Descriptions and Form Annotations

In a first step, we identified the adjective with highest passing in each cluster. Secondly, we analysed to which extent our categories of MF (focusing, attitudinal, epistemic) were depicted by the clusters; the categories have been annotated in the first post-processing step of the corpus. In order to describe these clusters in more detail, we subsequently annotated the video clips according to form features of nonverbal behaviour. The annotations were carried out without reference to a certain cluster, thus, no similar annotations were made due to neighbouring videos. The annotations included the following categories: 'gesture class', 'hand', 'hand shape', 'hand description', 'palm orientation', 'back of hand orientation', 'arm', 'taken space of movement', 'point in space of movement', 'direction of movement', 'duration of movement', 'additions', 'BM', 'face', 'perspective of movement'. This category system was formed in parts on the basis of two annotation schemes [12] [13] and extended relevant aspects, e.g., shoulders, head and upper BM, and facial expressions.

5.3. Clusters in the S+G-Condition

The four main clusters of the S+G-condition are the following: one cluster with videos depicting primarily a *focusing* MF (`1.A`), one with *epistemic* and *attitudinal* MF (`1.B`), one with negative *epistemic* features (`1.C`) and one cluster with a mixture of *various MF* (`1.D`). For an overview of the clusters of the S+G-condition, cf. Figure 4.

The main gesture class of `cluster 1.A` is 'deictic' including a lot of pointings, carried out primarily by the right arm. The direction of the movement is rather frontward and has an accentuated ending. It may include additions like a beat or a hold or BM like a head nod. Many positive adjectives dominate this cluster such as 'affirmative', 'emphasising', 'classifying', 'focused', 'opinionative' and 'relevant'. Given all of these criteria and considering the previous annotations of MF, we consider this group as representing **(positive**[8]**) *focusing* MF**, since relevant aspects are marked. When looking closer, two

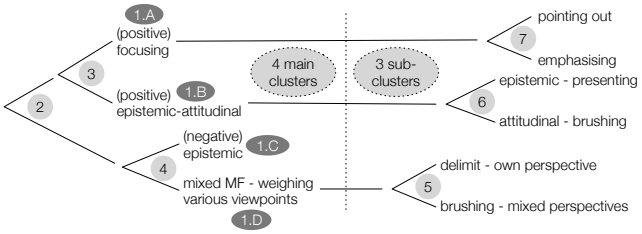

Figure 4: This dendrogram illustrates the cluster partitions and the four main clusters (`1.A – 1.D`) of the S+G-condition. Note that the distances between the cluster partitions are not accurate.

subclusters appear. One resembles the 'deictic' *focusing* category, in that something is 'pointed out', including the nods and a solely frontward and accurate movement with the right arm. The second subcluster forms the 'emphasising' *focusing* category, which contains a lot of hold gestures in addition to the deictic gestures and the gestures are carried out with both hands, in a frontward and upward direction and are realised a little sloppy in a few cases.

`Cluster 1.B` accumulates primarily gesture holds and also brushes; both hands are used and negative facial expressions appear. The hands seem to hold up and push down facts, and with these upward and downward movements it is a bidirectional cluster, which is more inherently consistent when looking at the subclusters. Adjectives that *do not* describe this cluster well are 'discounting/downtoning', 'revaluing', 'negative', 'positive', 'humorous' and 'uncertain'. We conclude this to be the cluster of **(positive) *epistemic* and *attitudinal* MF**, since the participants are self-confident about their utterance and have an attitude towards the topic, overall presenting their point of view and statement. The subclusters form two groups. One depicts the rather *epistemic* MF with gestures that present facts on hands, with a neutral face and hold additions. The other group rather accumulates *attitudinal* MF with brushing away gestures, (negative) facial expressions (pinched face, little angry, raised eye brows) and beat additions in some cases.

`Cluster 1.C` consists of gesture holds carried out in vicinity of the initial (or resting) position. The movements contain hedging elements, may be sloppy and without tension and are carried out in an upward direction. Shrugs and head tilts are included and the facial expressions are neutral. Most adjectives have very negative ratings and those which represent the cluster a little are 'discounting/downtoning' and 'uncertain'. We interpret this cluster to show **(negative) *epistemic* MF**, since the persons in the videos seem to be uncertain about what they are uttering. There is no further partitioning.

`Cluster 1.D` consists of very mixed features. It is the only cluster that shows delimiting of something with a movement, e.g., the participants block out space with their hands and arms. However, hold and brushing gestures are similarly prominent. Another interesting aspect is that various perspectives are taken (concluded from the overall utterance of the person): 'my point of view', 'someone else's point of view', 'our point of view'. The movements have a medium to large extend in space and are usually carried out without tension. In some cases circles or wriggles are included. Head shakes and smiles appear with the utterance. There are only minimal trends of adjectives that represent this cluster: 'discounting/downtoning', 'emotionally coloured', 'positive' and 'humorous'; and more often oc-

[6] Bashfield states that Ward's method "clearly obtained the most accurate solutions [...] the minimum variance method is generally preferable" [11, p. 385]

[7] IBM Corp. Released 2013. IBM SPSS Statistics for Macintosh, Version 22.0. Armonk, NY: IBM Corp.

[8] 'Positive' and 'negative' indicate the direction of a MF. An example for a negative *focusing* MF is a brushing away gesture that is used to 'brush' an aspect out of focus.

curring adjectives which *do not*: 'revaluing', 'critically', 'negative' and 'uncertain'. This cluster frames mainly a mixture of ***various (positive)*** **MF**, in particular *attitudinal and epistemic*, but rather no *focusing* MF. We call it the cluster of 'weighing different viewpoints' since it is very diverse and different point of views are taken. Two subclusters exists which divide delimiting and brushing gestures. Delimiting gestures are connected to one's own perspective, beat and hold additions and a very positive attitude. Brushing gestures present rather the mixed perspectives, carrying no additions and are accompanied by a little less positive attitude (in comparison to the other subcluster).

5.4. Clusters in the Gonly-Condition

The five clusters in the Gonly-condition will shortly be described in the following. `Cluster 2.A` consists of similar form annotations and adjectives like `1.A` and represents movements carrying (positive) *focusing* MF. Then, `cluster 2.B` has quite similar adjective ratings and form annotations as `1.C` and, therefore, accumulates (negative) *epistemic* MF. The following three clusters are different from those of the S+G-condition. `Cluster 2.C` is characterised by adjectives like 'affirmative' and 'emphasising' and *not* by 'humorous' and 'uncertain' and by brushing movements and beats. We interpret it as the cluster of (positive) *attitudinal* MF, since a person indicates her stance towards an aspect (only in parts similar to `1.B`). Then, `clusters 2.D` and `2.E` consist of hold gestures with different implications: `Cluster 2.D` carries parts of (negative) *epistemic* MF ('don't know') with a mix of various perspectives. `Cluster 2.E` consists of a mix of *various* MF and various perspectives are discussed. Some videos group similarly in the two conditions, although differences exist already due to the fact that one more cluster emerged in this condition.

6. Discussion and Conclusion

Although the ratings of the first analysis are not clear-cut, they indicate that MF, pinpointed here in terms of a set of adjectives, exist in BM and gesture. In tendency, pointing-like gestures are 'affirmative' and 'emphasising', hold gestures are rather 'emphasising' and brushing gestures are rather 'discounting/downtoning'. So, one important result is that pointing gestures are not solely used to refer to entities in the world, but also have a function of marking an utterance as, e.g., important or meaningful. It is also noteworthy that the most prominent gestures included a beat, which supports the viewpoint that a beat can also have a modal rather than a parsing function [14]. Additionally, the ratings make sense when looking within a gesture: in the prominent cases, if a gesture is rated 'affirmative', it is also rated 'not uncertain'.

However, the roles of the verbal and gestural utterance and their influence on each other are still not clear. It seems that a MF in BM or gesture is not as prominent when being accompanied by speech and facial expressions, as when being perceived on its own, namely in the Gonly-condition. Here, it seem to be interpreted as more negative which could be a result of the increased uncertainty in this unimodal condition.

The results of the cluster analysis suggest four distinct groups to which our MF relate in a plausible way: focusing/emphasising an aspect of the own utterance, conveying an epistemic-attitudinal statement, expressing epistemic uncertainty and discussing and weighing multiple viewpoints. In order to investigate these groups in more detail and to show each function group with its according form features in all its facets, more data is required (36 video snippets were used in this work).

The approach whether to use adjectives to measure MF in BM and gesture is up for discussion. From our point of view this was a viable first step as a proof of concept; the direct matching of the video snippets to the definitions of MF were difficult to realise due to the complexity of the definitions. By performing further analyses, we hope to find more answers regarding the possible modifications of utterance meaning. It would be interesting to observe this change concentrating on differences in modalities and cases when one modality is omitted.

7. Acknowledgements

This work was supported by a scholarship of the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG). The authors would like to thank Cord Wiljes for dubbing one of the stimulus videos and Sören Riechers for his support in the rating study setup. Additional thanks goes to the anonymous reviewers for their helpful comments.

8. References

[1] Wharton, T., *Pragmatics and non-verbal communication*, Cambridge University Press, 2009.

[2] Lu, Y., Aubergé, V. and Rilliard, A., "Do You Hear My Attitude? Prosodic Perception of Social Affects in Mandarin", Int. Conf. on Speech Prosody Proc., 685–688, 2012.

[3] Kendon, A., *Gesture: Visible Action as Utterance*, Cambridge University Press, 2004.

[4] Payrató, L., Teßendorf, S., "Pragmatic Gestures", In: C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and S. Teßendorf (eds.), Body Language Communication: An International Handbook on Multimodality in Human Interaction. Handbooks of Linguistics and Communication Science 38(1):1531–1539, 2013.

[5] Ferré, G., "Functions of Three Open-palm Hand Gestures", *Multimodal Communication*, 1(1):5–20, 2011.

[6] Streeck, J., "Gesture in Political Communication: A Case Study of the Democratic Presidential Candidates During the 2004 Primary Campaign", *Research on Language and Social Interaction*, 41(2):154–186, 2008.

[7] Norris, S., "Three Hierarchical Positions of Deictic Gesture in Relation to Spoken Language: A Multimodal Interaction Analysis", *Visual Communication*, 10(2):129–147, 2011.

[8] Enfield, N. J., Kita, S. and de Ruiter, J. P., "Primary and Secondary Pragmatic Functions of Pointing Gestures", *Journal of Pragmatics*, 39(10):1722–1741, 2007.

[9] Bryant, G. A. and Fox Tree, J. E., "Recognizing Verbal Irony in Spontaneous Speech", *Metaphor and Symbol*, 17(2):99–119, 2002.

[10] Sloetjes, H. and Wittenburg, P., "Annotation by category: ELAN and ISO DCR", Int. Conf. on Language Resources and Evaluation Proc., 2008.

[11] Blashfield, R. K., "Mixture model tests of cluster analysis: Accuracy off our agglomerative hierarchical methods", Psychological Bulletin, 83(3):377–388, 1976.

[12] Lücking, A., Bergman, K., Hahn, F., Kopp, S., Rieser, H., "Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications", Journal on Multimodal User Interfaces 7(1–2):5-18, 2013.

[13] Bressem, J., Ladewig, S.H., Müller, C., "Linguistic Annotation System for Gestures (LASG)", In: C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and S. Teßendorf (eds.), Body Language Communication: An International Handbook on Multimodality in Human Interaction. Handbooks of Linguistics and Communication Science 38(1):1098–1124, 2013.

[14] Kendon, A., "Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation", *Journal of Pragmatics*, 23(3):247–279, 1995.