In: Proceedings of the International Conference on
Affective Computing and Intelligent Interaction (ACII-09),
IEEE Press, 2009

Gesture and Emotion: Can basic gestural form features discriminate emotions?

Michael Kipp
DFKI, Saarbrücken, Germany
kipp@dfki.de

Jean-Claude Martin
LIMSI-CNRS, Paris, France
martin@limsi.fr

Abstract

The question how exactly gesture and emotion are interrelated is still sparsely covered in research, yet highly relevant for building affective artificial agents. In our study, we investigate how basic gestural form features (handedness, hand shape, palm orientation and motion direction) are related to components of emotion. We argue that material produced by actors in filmed theater stagings are particularly well suited for such analyses. Our results indicate that there may be a universal association of gesture handedness with the emotional dimensions of pleasure and arousal. We discuss this and more specific findings, and conclude with possible implications and applications of our study.

1. Introduction

An emotion can be seen as an *episode of interrelated, synchronized changes in five components in response to an event of major significance to the organism* [43]. These five components are: the cognitive processing, the subjective feeling, the action tendencies, the physiological changes, and the motor expression. While facial expressions have been studied extensively as a main vehicle for the motor expression component [16], studies about gestural expressions of emotions are sparse and less conclusive.

An early investigation on gesture and affect partly resulted in a link between intensity of felt emotion and quality or activity of gesture [17]. More recently, several studies observed how the whole body expresses emotion. Discriminative features of some emotions were found in the whole body posture and its movement quality (e.g. [45]). Such studies do not provide precise descriptions of hand gestures. Moreover, movement quality is usually described with continuous parameters which makes coding more more error-prone than discrete features of hand gestures: handedness, hand shape, palm orientation etc. These properties have been used in the past to study discourse function [34], meaning [6] and recurrent patterns of gestural form [4]. However, to the best of our knowledge there is no work investigating the relationship between *hand* gestures (more specifically the components of handedness, hand shape and motion direction) and emotion.

Our main motivation lies in creating an empirical basis for the design of believable virtual characters, also called *embodied conversational agents* [8]. For the automatic production of nonverbal behavior, our approach relies on a lexicon of gestures where gesture entries, so-called gesture *lexemes*, are described using basic components (cf. [29]). This is motivated by studies indicating that conversational gestures originate from a shared set of recurring patterns [4, 6, 46]. Each single pattern can be described as a set of constraints on gestural components like handedness, hand shape, palm orientation and motion direction. These components bear similarity to morphemes in spoken language, being the atomic building blocks of gestural movement. Such a view facilitates operationalization in the field of automatic gesture synthesis for virtual characters, to procedurally generate a stream of connected gestures for a given piece of text/speech [30, 37]. Therefore, our more specific target is to exploit our findings to build an emotion-to-gesture-component morphological lexicon to complement the meaning-form lexicon in our gesture synthesis approach. These lexicons consists both of a general part and a character-specific one.

Surprisingly, staged theater plays have not been explored much as a potential multimodal corpus for the emprically based design of virtual characters or for experimental studies on multimodal perception. We found videos of filmed theater plays a promising source for our studies mainly because they contain a wide range of strong emotional displays in relatively short time (since a theater play must condense real life situations) which would be difficult if not impossible to elicit under lab conditions. Moreover, the featured behaviors are expressive because they are rehearsed (i.e. evaluated by the director as sufficiently expressive and communicatively ”correct”), there are numerous spatial multiparty behaviors situated in an environment (as opposed to laboratory surroundings), and there is a rich context in terms of character history and interpersonal relations (available from stage directions in the script). Note that this data is different from recent efforts to collect acted data un-

978-1-4244-4799-2/09/$25.00

der lab conditions [2, 5] since it encompasses a whole play instead of short episodes, thus including a rich context in terms of emotional development and interpersonal relations.

This work has the following research contributions:

- A method for studying multimodal correlations, from coding PAD-based emotions gesture components to the analysis of correlations
- Introducing filmed theater plays as a promising source data for the study of gesture and affect and other multimodal relations
- Several correlation results, both character-specific and (possibly) universal, showing that emotion is highly correlated with simple gesture components like handedness

In the following we will first review related work and then present our corpus, the emotion and gesture coding schemes. We will present a detailed correlation analysis and discuss the results. We conclude with possible applications and future work.

2. Related work

Basic gestural components like *handedness* have been analyzed by semiotician Calbris [7] who showed in a study of a French politician how the use of left hand and right hand corresponds with whether he is talking about his left-wing party (left hand) or the opposing party (right hand). This dichotomy is overlaid with other functions that left/right fulfill but the approach shows that the "often non-conscious choice of the hand used to express oneself is semantically motivated". In our study, we are interested whether the choice of hands can also be motivated by emotional state, especially in a play where emotion is something the actors wants to communicate, so the emotion in this context may even be considered "semantic".

Most experimental studies about the expression of emotion have focused on speech [44] or facial expressions [16]. Only a few studies considered gestural and bodily expression of emotions. Wallbott [45] collected video takes in which actors portrayed the emotions of elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride, and boredom via a scenario approach. He observed discriminative features of emotions both in static descriptors of the body posture and in the body movement quality. De Meijer [11] looked at gross body movements (trunk movement, arm movement, vertical direction, sagittal direction, force, velocity and directness) and correlated these with emotion ratings. Emotion categories differed as to the amount, type, and weights of predicting movement features. Three factors were extracted from the original ratings and interpreted as rejection-acceptance, withdrawal-approach, and preparation-defeatedness. Boone [3] asked adults to discriminate the emotions of happiness, sadness, anger, and fear from dance performances. Six specific cues were used to distinguish among the four categories of emotion: changes in tempo (anger), directional changes in face and torso (anger), frequency of arms up (happiness), duration of arms away from torso (happiness), muscle tension (fear), duration of time leaning forward (sadness).

Digital corpora have recently been collected with high resolution in 3D (motion capture) in conjunction with controlled episodes of emotion expression. They also use acting-based designs to collect in-lab motion capture data: use of markers on body to recognize four acted basic emotions [27], use of motion capture of static postures during acting of two nuances of four basic emotions (e.g. sad and depressed) [12], use of video processing of facial expressions and upper body gestures during six acted emotional behaviors [23]. Most of these studies consider the whole body expression but do not describe hand gestures form features such as hand shape or handedness. Furthermore, these experimental studies use acting in a very restricted sense, when compared with actual drama and staged plays. They consist of recording, in a laboratory environment, amateur or professional actors who are expressing emotions specified using single labels of emotions or short scenario scripts.

Enos and Hirschberg [18] suggest two more sophisticated approaches for eliciting acted emotional speech. The scenario approach uses a description of the character, the situation, the actor's goals and the goals of other characters. The script approach makes direct use of the actor's training and skills via the acting of original scenes taken from theater plays. However, they consider the speech modality only. Busso and Narayan [5] also highlight the importance of contextualization to collect genuine databases of expressions of emotions. They collected the Interactive Emotional Dyadic Motion Capture (IEMOCAP) database which uses plays (scripted sessions), and improvisation based on hypothetical scenarios (spontaneous sessions). The IEMOCAP database includes speech and motion-captured facial expressions but not hand gestures.

A few multimodal emotional corpora include gestures. The GEMEP corpus uses scripts with nonsense words and professional actors video-taped in lab [2]. The Belfast Naturalistic and the EmoTV Databases [14] aim at collecting multimodal expressions of in real-life settings (e.g. TV interviews, reality games). People are usually recorded for a few seconds or minutes. Most of the suggested approaches for collecting the spontaneous expression of emotion have limitations such as ethical issues, restricted domain, or lack of control (i.e., type of sensors, modalities, quality of recording, noise background, lexical and emotional content) [5]. Naturalistic behaviors can also be replayed by actors to get high-resolution spoken behaviors: the genuineness and validity of such re-enacted recordings was evaluated by naive subjects in order to study the differences in speech between naturalistic versus acted conditions [21]. Clips from movies have also been used as behavioral data in some studies about emotional expression with a focus on speech [9].

Most emotional corpora focus on spoken or facial expressions of emotions, or both. When including hand gesture expression of emotions, current corpora are limited with respect to the emotional context (physical environment, dyadic social interaction with long-term relations, consistent behavior of single actor across, duration of behaviors). We argue that the social setting and the average duration of the dialogues should be long enough to contextualize the signs and flow of emotions [13]. As suggested by Busso and Narayan, the semantic context of the material should be congruent with the intended emotion, to avoid adding extra difficulties to the actors. Therefore, we suggest filmed theater plays as a basis for the analysis of multimodal behaviors.

The speaking face and body of the characters, the explicit and implicit paralanguage, kinesics and silences in theater have been approached by [41]. The use of gesture in Shakespeare plays has been studied by [1]. In real theater, actors have to manage several types of emotions during the acting process such as *task emotions* (acting in front of an audience) and *character emotions* (inferred from the script) [32]. Johnstone [26] introduced the concept of *status* as a tool to elicit more interesting improvisations, claiming that the audience relished drastic changes in status (e.g. the servant teasing the master). His concepts proved so clear that they could be operationalized in a computer system with virtual actors [24]. While these concepts were hand-coded, one of our long-terms goals is to model such interaction patterns by analyzing our annotated corpus.

3. Theater Corpus

Our corpus consists of sections from two movie versions of the play *Death of a Salesman* by Arthur Miller (see Fig. 1). We focus on the main protagonist, Willy Loman. The first version, henceforth **DS-1**, was released in 1966 and featured Lee J. Cobb as Willy. The second version, **DS-2**, was released in 1985 and featured Dustin Hoffman as Willy. The total duration of the clips and the number of annotated entities is listed in Table 1. Note that the emotion annotation was performed on the speech segments. The **DS-1** material is larger in total duration because the gesture events were sparser than in **DS-2** so we had to add data.

We chose this particular play for three reasons. First, the main protagonist Willy displays a wide range of emotions (sad, exuberant, hostile, etc.). This is caused by his different relationship to the other protagonists (problematic relationship with his son Biff, dominant relationship with his wife Linda) but also because of time shifts, where in the past the general mood is depicted as cheerful and optimistic and in the present is quite depressing. Moreover, in the past the interpersonal relations are much more positive. The second reason is that the play is naturalistic, i.e. both language and set are designed to mimic real life. Therefore, also the nonverbal behavior has a natural look, whereas e.g. classical verse plays (e.g. Shakespeare) or more contemporary absurd plays (e.g. Beckett) may cause stylized behaviors by stylized language. The third reason is that this play has at least three movie versions of which we selected two. By having selected equivalent scenes for DS-1 and DS-2 we have a somewhat similar distribution of emotional states when comparing behaviors.

4. Emotion Coding

4.1. Previous Work

Two main approaches are used to represent and code emotions: categorical approaches and dimensional approaches. A categorical approach consists of selecting a finite set of discrete labels. Several lists have been proposed such as Ekman's six primary emotions [15] or the set of 24 emotions in the OCC model [38].

Dimensional approaches represent an affective state by its location along one or more continuous axes. Russell and Mehrabian described two studies showing that three independent and bipolar dimensions adequately define affective states [42]. These three dimensions are: **P**leasure (positive versus negative affective state), **A**rousal (level of physical activation and/or mental alertness), and **D**ominance (feelings of control and influence over others and situations, versus feeling controlled and influenced by external circumstances). This so-called PAD space was proposed for representing temperament [36] but also more transitory affective states such as emotions. The locations of 151 emotional terms in these three dimensions were experimentally found [42]. As opposed to studies that solely rely on pleasure and arousal, the inclusion of dominance as a third dimension can separate different states like anger and fear by putting them in different parts of the 3D space. PAD is used both for experimental studies in psychology and in computational models of emotions for virtual characters [10, 22].

movie	*duration*	*speech seg.*	*gestures*
DS-1	16:56	104	141
DS-2	8:48	69	117

Table 1. Annotated clips of the two versions of *Death of a Salesman*.

4.2. Emotion Coding Scheme

Both the categorical and the dimensional approach to emotion coding can be relevant to studying gestural expression of affect. We examined gestural features that are discriminative of specific affect categories (e.g. anger) or specific affect dimensions (e.g. valence). A key issue in annotating emotion is reliability, especially in varied emotional situations with a rich social context. That is why in our coding scheme, when selecting an emotional category, the coder was able to see corresponding dimensional values as a means for checking that his annotation was consistent. Since few studies have considered the relations between gestural features and emotions, we used a set of well-documented categories. A set of 6 basic emotion categories would have been too limited with respect to the rich video data that we collected in the various theater scenes. The OCC set of 24 emotions categories might have been too large to ensure consistent coding. Furthermore, the OCC model does not provide correspondence between categories and dimensions. We decided to test the applicability of the PAD space. It proposes a main set of 8 affective states that correspond to the eight corners of PAD space [36]:

- Bored, sad, fatigued (-P-A-D)
- Disdainful, indifferent, unconcerned (-P-A+D)
- Anxious, aghast, distressed, insecure (-P+A-D)
- Hostile, angry, nasty (-P+A+D)
- Docile (+P-A-D)
- Relaxed, satisfied, comfortable (+P-A+D)
- Dependent, amazed, grateful, respectful (+P+A-D)
- Exuberant, admired, bold, excited (+P+A+D)

Coders were able to check during annotation the PAD correspondence between PAD dimensions and PAD labels. The coder also separately coded the three single dimensions. The dimensions are specified as pleasure (from negative, e.g. pain, to positive, e.g. happiness), arousal (from negative, e.g. drowsiness, to positive, e.g. excitement), dominance (from negative, e.g. total lack of control over the emotional situation, to positive, e.g. extreme feeling of control). Intensity of the emotion category was encoded from low to high, independently of the level of arousal (e.g. low arousal resulting from a very intense sadness). The PAD dimensions and the intensity were coded with three values: low, neutral, high.

Emotion annotation was done on the utterance level. When necessary, the coder was allowed to decompose a single utterance into segments in order to annotate multiple emotional states perceived in the utterance. Coding was performed in the Anvil tool [28].

5. Gesture Coding

5.1. Previous Work

A number of different approaches to coding gestures have been developed in the past, driven by the particular needs of the individual research aims. One can distinguish between purely descriptive schemes and schemes that involve interpretation by the human coder. One of the earliest purely descriptive scheme was developed by Frey et al. in Bern [19, 20]. For the transcription, the coder would look at static frames of the video (15 frames per second) and estimate the angles of various body parts at this particular instance. The result is a large matrix of body part positions over time. In a more recent approach called FORM [33], the coder would also specify a position of a body part, however only when the position changes. One of the pioneers of modern gesture research, McNeill, transcribed gestures using a segmentation-and-classification approach [34, 35]. Segmentation was done using the concept of movement phases (cf. [31]), and classification included objective features like handedness, palm orientation, motion direction, etc. While the descriptive part is similar to our own scheme, their descriptive features are either more detailed (e.g. using the hand shape repertoire of American Sign Language) or less detailed (4 categories for motion direction). More recently, Bressem [4] used a reduced set of descriptive features (hand shape, hand orientation, movement pattern, position in space) for a cluster analysis of recurrent patterns; her coding set is comparable in size and categories to ours.

Interpretative schemes consider the *function* of a gesture. Examples are Poggi [39] who describes the function of gesture in terms of a semiotic goal and belief model, or Webb [46] who categorizes metaphoric gestures partly on the basis of recurrent form but also on the basis of meaning.

5.2. Gesture Coding Scheme

For the design of our coding scheme we selected categories that would be easy to objectively encode. The encoding of temporal events like gestures always involves two principal steps, segmentation and categorization. For segmentation the challenge is to clearly identify beginning and end of a gesture which involves the decision whether a gesture ends or is simply *on hold*. In previous work, we encoded both the movement phases (preparation, stroke, hold, retraction etc.) and the whole gesture which consists of a number of phases (typically preparation-stroke-retraction). This adds the challenge of identifying phase boundaries and introduces an interpretative process where the codes decides what part of the movement is the most "meaningful" one, then to be declared the *stroke*. So for our coding we decided to simply encode the motion part of what the coder deemed to be a gesture, which corresponds to the *expressive phase*, as defined by Kita et al. [31]. Also note, that

Hand	Shape	Palm	Dir
LH	open	up	up
RH	claw	down	down
2H	fist	side-forward	forw
	index finger	side-up	sideways-out
	thumb out	out-up	sideways-in
	finger ring	out-side	to body
	purse	in-up	wrist only
		in-side	
		in-down	

Table 2. Categories for gesture encoding.

DS-1
high pleasure ∼ forward motion
low pleasure ∼ sideways-out motion
high arousal ∼ index finger hand shape

Table 3. Correlations for DS-1.

DS-2
high pleasure ∼ wrist-only motion
low pleasure ∼ downward motion
low dominance ∼ open hand shape
high dominance ∼ index finger hand shape

Table 4. Correlations for DS-2.

for our analysis the precise boundaries of the gesture do not matter as long as a temporal overlap with the corresponding emotion coding is guaranteed.

Each segment would then be annotated with categorical information that describes the gesture is very simple terms. We encoded the following 4 aspects: handedness (Hand), hand shape (Shape), palm orientation (Palm) and motion direction (Dir). Table 2 shows all labels per category. The labels for palm orientation mainly refer to the normal vector orthogonal to the palm, pointing away from it. In cases of ambiguity (e.g. pointing to the side), the orientation of the vector from wrist to index finger knuckle is specified also (e.g. side-forward means palm oriented sideways, index finger knuckle is pointing away from body). We did not propose labels for all anatomically possible palm directions but only for those occurring a minimal number of times.

Figure 1. Sample frames from the corpus.

Fig. 1 shows sample frames from the corpus to exemplify our categories. (A) shows a RH gesture with hand shape *open*, palm *up* and direction *down*. (B) shows a RH gesture with hand shape *open*, palm *side-forward* and direction *forw*. (C) shows a RH gesture with hand shape *index finger*, palm *down* and direction *forw*. Finally, (D) shows a LH gesture with hand shape *open*, palm *down* and direction *wrist-only*.

6. Analysis

Our analysis was targeted at possible correlations between emotion and gesture. On the emotion side, we used the *emotion category*, and, for more detail, the single emotion dimensions of pleasure (P), arousal (A) and dominance (D). For gestures, we used the features of handedness, hand shape, palm orientation, and motion direction. We computed chi square (χ^2) to find correlated dimension pairs (e.g. pleasure and handedness). We then looked at the deviations between expected occurrences and actual occurrences to find out the magnitude and direction (positive/negative) of correlations between concrete values. To compare occurrences of emotion episodes with gestural events in our data, we counted co-occurrences between encoded emotion segments and encoded gestures, looking at all pairs where the emotion interval temporally *contained* the gesture event.

6.1. Results: Emotion category vs. gesture features

In the pairwise analysis we found highly significant correlations between *emotion category* and *handedness* for both DS-1 ($\chi^2 = 40.14; p < .001$) and DS-2 ($\chi^2 = 35.37; p < .001$). We found one more correlation, for DS-2 only, for *emotion category* and *palm orientation* ($\chi^2 = 42.50; p < .05$). All other pairs did not reach significance.

To get a more detailed impression of the effect one can look at the deviations from expected values for each value combination, resulting in a difference matrix which relates to emotion vs. handedness. Looking at the most accentuated deviations we found almost identical results for the emotion-handedness relationships for DS-1 and DS-2. Both actors consistently use their LH in a *relaxed* state and their RH in a *hostile* state. For DS-2, there is an additional effect linking 2H to an *exuberant* state.

6.2. Results: Emotion dimensions vs. gesture features

While emotion category gives us an intuitive categorization of the PAD space, it is also of interest whether and

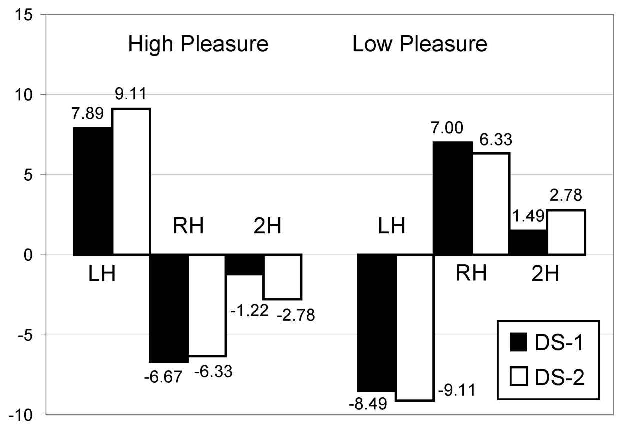

Figure 2. Deviation from expected values (pleasure – handedness).

how the single PAD dimensions relate to gestural features. Therefore we conducted $3 \times 4 \times 2 = 24$ χ^2 tests for each PAD dimension, each gesture feature in each movie version. We found 8 significant relationships. Most applied equally for DS-1 and DS-2: pleasure-handedness (DS-1: $\chi^2 = 12.36, p < .01$; DS-2: $\chi^2 = 13.48, p < .01$), arousal-handedness (DS-1: $\chi^2 = 16.57, p < .01$; DS-2: $\chi^2 = 16.44, p < .01$), and pleasure-motion (DS-1: $\chi^2 = 28.29, p < .05$; DS-2: $\chi^2 = 14.92, p < .05$). Two more significant effect applied to DS-1 only, arousal-shape ($\chi^2 = 24.39, p < .01$), and DS-2 only, dominance-shape ($\chi^2 = 5.99, p < .05$). To get an impression how concrete value relationships can be visualized with the help of deviation from expected values, Fig. 2 shows these deviations for *pleasure* vs. *handedness*. One can see that the left hand is used when in a positive mood, while the right hand seems to be reserved for negative moods. Tables 3 and 4 summarize further correlations from the other significant correlations.

7. Discussion

7.1. Emotion, handedness, motion direction

Our analysis showed that handedness is closely correlated with the emotion categories in the sense that *relaxed* correlates with LH and *hostile* with RH. Looking at the emotion dimensions we found positive correlations between high pleasure and LH, and low pleasure and RH. This is not surprising because in exploratory studies as well as in related work it has been shown that speakers seem to reserve left/right hands for specific concepts [7] or for connecting discourse entities (cf. McNeill's concept of *catchments* [34, 35]). So it seems plausible that for the portrayal of a certain emotional state, handedness may be used as well. What is surprising is the fact that the same correlation holds for both actors. From an evolutionary perspective it can be argued that, given that both actors are right-handed, the right hand is used for fighting and may therefore be more active during hostile moods. By way of contrast, one could argue that then the left hand can be more active during non-dangerous situations, i.e. in relaxed moods. Posner formally demonstrated how concrete acts (like punching) can evolve to semiotic signs [40]. This would close the loop in the evolutionary argument from fighting to gesture. Of course, we are aware of the fact that two sample subjects cannot prove a universal validity. Extending the study in this direction may be a promising avenue for the future.

The second significant result of our study showed that the pleasure (or valence) dimension of emotion is correlated with motion direction (for both DS-1 and DS-2). This finding can be explained with certain (idiosyncratic) gestures occurring in certain moods. For instance, when aggressive, the DS-1 actor performed a particular gesture that wipes outward, therefore we find a positive correlation of low pleasure and the sideward-outward motion direction. The DS-2 actor preferred a downward motion, usually with his raised index finger. Thus, the correlation analysis gives us individualized *components* of behavior that could be used to drive synthesis methods.

7.2. Implications and Applications

The obtained results are preliminary but have a number of perspectives. The first perspective is empirical research in the question: *Is handedness a general predictor for affect?* This question could be analyzed in a larger database with various side conditions like cultural group, acted vs. spontaneous, conversational setup (negotiation, small talk, narration etc.) and others. The second perspective is the application of such data for the animation of believable virtual humans [8]. A number of projects are trying to put their virtual humans on an empirical basis to drive their nonverbal behavior [30, 37]. If gesture synthesis is based on a *morphological* approach where certain aspects of a gesture can be preferred or suppressed, our findings can inform both the general part of the algorithm but also contribute to individualized profiles that strive to *imitate* behavioral aspects of one (DS-1) or the other (DS-2) speaker [37]. Such an approach necessitates the use of a character animation engine with a high level of control (handedness, hand shape, trajectory etc.), a number of which are currently under development [25]. Finally, since technology is advancing toward automatic recognition of emotion and gesture, our finding could, for instance, predispose face-based emotion recognizer if the active hand is known from motion capture devices.

8. Conclusion

We presented a study analyzing the relation between emotion and gestural features on a corpus of theater movies.

We argued that theater data is well suited for this kind of analysis because of the wide range of emotions displayed and the fact that emotion is inherent part of the actors' message. We presented a coding scheme for emotions, based on the PAD dimensions, and for gestural features (handedness, hand shape, palm orientation, motion direction). Analyzing comparable material of two different actors acting in the same role (Willy Loman) in two film versions of *Death of a Salesman* we found significant correlations between handedness and emotion. For both actors, the right hand was used in conjunction with negative and aggressive feelings, whereas the left hand was used more when in a relaxed and positive mood. Other significant correlations differed for the two actors.

For the future, we plan to utilize our findings for guiding the production of synthetic gestures. Further studies will also include inter-coder reliability tests to validate the robustness of our coding scheme. For increasing the limited scope of this study it would be interesting to include more actors but also to compare our findings with naturally occuring dialogue. Note, however, that finding naturally evoked emotions of similar intensities and categories in a social context is very difficult and that inducing such emotions for research purposes may raise ethical issues.

To conclude, we argue that our findings can be used both for general and character-specific gesture synthesis for affective virtual characters. Follow-up work should also investigate the question whether a significant preference for the leading hand in aggressive moods is a universal phenomenon.

Acknowledgements

This research was partially funded by the German Academic Exchange Service (DAAD) and the French Égide partenariat Hubert Curien within PROCOPE 2008 No. 178 63 RL, and conducted within the framework of the Excellence Cluster *Multimodal Computing and Interaction* (MMCI), sponsored by the German Research Foundation (DFG).

References

[1] J. H. Astington. Actors and the body: Meta-theatrical rhetoric in shakespeare. *Gesture*, 6(2):241–259, 2006.

[2] T. Bänziger, H. Pirker, and K. Scherer. Gemep - geneva multimodal emotion portrayals: A corpus for the study of multimodal emotional expressions. In *Workshop "Corpora for research on emotion and affect" at the 5th International Conference on Language Resources and Evaluation (LREC'2006)*, pages 15–19, 2006.

[3] R. T. Boone and J. G. Cunningham. Children's understanding of emotional meaning in expressive body movement. In *Biennial Meeting of the Society for Research in Child Development*, 1996.

[4] J. Bressem. Recurrent form features in coverbal gestures. Talk at the Third International Conf. of the International Society for Gesture Studies (ISGS), June 2007.

[5] C. Busso and S. Narayanan. Scripted dialogs versus improvisation: Lessons learned about emotional elicitation techniques from the iemocap database. In *Interspeech*, Brisbane, Australia, 2008.

[6] G. Calbris. *Semiotics of French Gesture*. Indiana University Press, Bloomington, Indiana, 1990.

[7] G. Calbris. From left to right...: Coverbal gestures and their symbolic use of space. In *Metaphor and Gesture*, pages 27–53. John Benjamins, 2008.

[8] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill. *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[9] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette. The safe corpus: illustrating extreme emotions in dynamic situation. In *Workshop "Corpora for research on emotion and affect". 5th International Conference on Language Resources and Evaluation (LREC'2006)*, pages 76–79, Genova, Italy, 2006.

[10] M. Courgeon, J.-C. Martin, and C. Jacquemin. User's gestural exploration of different virtual agents' expressive profiles. In *7th Int. Conf. on Autonomous Agents and Multiagent Systems*. ACM, 12-16 May 2008.

[11] M. de Meijer. The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13(4):247–268, 1989.

[12] P. De Silva, A. Kleinsmith, and N. Bianchi-Berthouze. Towards unsupervised detection of affective body posture nuances. In J. Tao, T. Tieniu, and R. Picard, editors, *1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, LNCS 3783, pages 32–40, Beijing, China, 2005. Springer.

[13] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach. Emotional speech; towards a new generation of databases. *Speech Communication*, 40:33–60, 2003.

[14] E. Douglas-Cowie, R. Cowie, Sneddon, C. Cox, Lowry, McRorie, J.-C. Martin, L. Devillers, and A. Batliner. The humaine database: addressing the needs of the affective computing community. In A. Paiva, R. Prada, and R. Picard, editors, *2nd International Conference on Affective Computing and Intelligent Interaction (ACII'2007), LNCS, vol. 4738*, pages 488–500, Lisbon, Portugal, 2007.

[15] P. Ekman. Facial Expression of Emotion: New Findings, New Questions. *Psychological Science*, 3(1):34–38, January 1992.

[16] P. Ekman. Basic emotions. In T. Dalgleish and M. J. Power, editors, *Handbook of Cognition and Emotion*, pages 301–320. John Wiley, New York, 1999.

[17] P. Ekman and W. V. Friesen. Hand and Body Cues in the Judgment of Emotion. *Perceptual and Motor Skills*, 24:711–724, 1967.

[18] F. Enos and J. Hirschberg. A framework for eliciting emotional speech: Capitalizing on the actor's process. In *Workshop "Corpora for research on emotion and affect". 5th International Conference on Language Resources and Evaluation (LREC'2006)*, pages 6–10, Genova, Italy, 2006.

[19] S. Frey, H. P. Hirsbrunner, A. Florin, W. Daw, and R. Crawford. A unified approach to the investigation of nonverbal and verbal behavior in communication research. In W. Doise and S. Moscovici, editors, *Current Issues in European Social Psychology*, pages 143–199. Cambridge University Press, Cambridge, 1983.

[20] S. Frey and J. Pool. A new approach to the analysis of visible behavior. Technical Report 1976-2, University of Berlin, 1976.

[21] S. Frigo. The relationship between acted and naturalistic emotional corpora. In *Workshop "Corpora for research on emotion and affect". 5th International Conference on Language Resources and Evaluation (LREC'2006)*, pages 34–36, Genova, Italy, 2006.

[22] P. Gebhard. Alma a layered model of affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 2005.

[23] H. Gunes and M. Piccardi. Fusing face and body display for bi-modal emotion recognition: Single frame analysis and multi-frame post integration. In J. Tao, T. Tieniu, and R. Picard, editors, *1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, LNCS 3783, pages 102–110, Beijing, China, 2005. Springer.

[24] B. Hayes-Roth, R. van Gent, and D. Huber. Acting in character. In R. Trappl and P. Petta, editors, *Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents*, pages 92–112. Springer, New York, 1997.

[25] A. Heloir and M. Kipp. Embr - a realtime animation engine for interactive embodied agents. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA-09)*, 2009.

[26] K. Johnstone. *Impro. Improvisation and the Theatre*. Routledge/Theatre Arts Books, New York, 1979. (Corrected reprint 1981).

[27] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. Driessen. Gesture-based affective computing on motion capture data. In J. Tao, T. Tieniu, and R. Picard, editors, *1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, LNCS 3783, pages 1–8, Beijing, China, 2005. Springer.

[28] M. Kipp. Anvil – a Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of Eurospeech*, pages 1367–1370, 2001.

[29] M. Kipp. *Gesture Generation by Imitation: From Human Behavior to Computer Character Animation*. Dissertation.com, Boca Raton, Florida, 2004.

[30] M. Kipp, M. Neff, K. H. Kipp, and I. Albrecht. Toward Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach. In *Proc. of the 7th International Conference on Intelligent Virtual Agents (IVA-07)*, pages 15–28. Springer, 2007.

[31] S. Kita, I. van Gijn, and H. van der Hulst. Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth and M. Fröhlich, editors, *Gesture and Sign Language in Human-Computer Interaction*, pages 23–35, Berlin, 1998. Springer.

[32] E. Konijn. *Acting emotions: Shaping emotions on stage.* Amsterdam: Amsterdam University Press, 2000.

[33] C. H. Martell. Form: An extensible, kinematically based gesture annotation scheme. In *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*. Kluwer Academic Press, 2004.

[34] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, 1992.

[35] D. McNeill. *Gesture and Thought*. University of Chicago Press, Chicago, 2005.

[36] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14:261–292, 1996.

[37] M. Neff, M. Kipp, I. Albrecht, and H.-P. Seidel. Gesture Modeling and Animation Based on a Probabilistic Recreation of Speaker Style. *ACM Transactions on Graphics*, 27(1):1–24, March 2008.

[38] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge University Press, MA, 1988.

[39] I. Poggi. *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. Weidler Buchverlag, Berlin, 2007.

[40] R. Posner. Believing, causing, intending: The basis for a hierarchy of sign concepts in the reconstruction of communication. In R. J. Jorna, B. van Heusden, and R. Posner, editors, *Signs, Search, and Communication: Semiotic Aspects of Artificial Intelligence*, pages 215–270. deGruyter, Berlin, New York, 1993.

[41] F. Poyatos. *Nonverbal communication across disciplines*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2002.

[42] J. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11:273–294, 1977.

[43] K. R. Scherer. Emotion. In M. Stroebe and W. Hewstone, editors, *Introduction to Social Psychology: A European perspective*, pages 151–191. Oxford: Blackwell, 2000.

[44] M. Schröder. Experimental study of affect burst. *Speech Communication. Special Issue following the ISCA Workshop on Speech and Emotion*, 40(1-2):99–116, 2003.

[45] H. G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896, 1998.

[46] R. Webb. *Linguistic Properties of Metaphoric Gestures*. UMI, New York, 1997.