Nonsymbolic Gestural Interaction for Ambient Intelligence

Rehm, Matthias

Published in:
Human-Centric Interfaces for Ambient Intelligence

Publication date:
2010

Document Version
Early version, also known as pre-print

Non-symbolic gestural interaction for Ambient Intelligence

Matthias Rehm

Multimedia Concepts and Applications, Dept. of Applied Informatics

University of Augsburg, Eichleitnerstr. 30, D-86159 Augsburg, Germany

Abstract

Our gestural habits convey a multitude of information on different levels of granularity that can be exploited for human computer interaction. Gestures can provide additional or redundant information accompanying a verbal utterance, they can have a meaning in themselves, or they can provide the addressee with even more subtle clues for instance about our personality or cultural backgrounds. Thus, gestures are an extremly rich source of communication-specific and contextual information for interactions in ambient intelligent environments. This chapter reviews the different semantic layers of gestural interaction focusing on the layer beyond communicative intent and presents interface techniques to capture and analyse gestural input taking into account non-standard approaches like acceleration analysis or the use of physiological sensors.

Key words: gesture recognition, emotion, personality, culture

Email address: rehm@informatik.uni-augsburg.de (Matthias Rehm).
URL: http://mm-werkstatt.informatik.uni-augsburg.de (Matthias Rehm).

1 Introduction

Imagine a student who has to give an important presentation in front of a university board to apply for a funding of his Ph.D. work. During this presentation, the student exhibits an unusual high amount of hand gestures. An obvious explanation for this excessive show of hand movements is that the speaker is quite nervous due to some unknown reason. This interpretation follows more or less Burgoon's [4] definition of nonverbal communication as "behaviors other than words that form a socially shared coding system (...) [and] have consensually recognizable interpretations" (p. 231). Thus, even if in a specific situation the behavior is not performed intentionally (e.g. excessive use of gestures) it nevertheless conveys meaning relevant for the interaction. Watzlawick, Bavelas and Jackson [43] have put this fact in a short and concise statement by saying that one cannot not communicate. In this chapter, we concentrate on such aspects of gestural interaction that are not directly related to communicative content but convey additional meanings below consciously intended communicative content.

2 Classifying gestural behavior for human-centric Ambient Intelligence

Gestural behavior has mainly been investigated as a co-verbal phenomenon, focusing on the meaning (intentionally) conveyed by the speaker. Gestures in this sense accompany utterances and give sometimes redundant, sometimes additional information about the speaker's message. For instance, a speaker might recount a story where his kid let loose a balloon. He might accom-

pany the utterance "and then the balloon flew up and away" by raising the right hand in a straight line, emphasizing what is also said in the utterance, i.e. that the balloon flew up. He could also accompany this utterance with a hand movement that mirrors the actual ascent of the balloon, for instance while raising the hand, it is going to the right then to the left. This gesture than will give additional information that goes beyond what was explicitly said in the utterance. McNeill established a solid foundation for this perspective on gestures, presenting a taxonomy and coding scheme for conversational gestures. He distinguishes between adaptor, beat, emblem, deictic, iconic, and metaphoric gestures. Adaptors comprise every hand movement to other parts of the body like scratching one's nose. Beats are rhythmic gestures that may emphasize certain propositions made verbally or that may even link different parts of an utterance. Emblems are gestures that are meaningful in themselves, i.e. without an accompanying utterance, but that are highly culture-specific. An example is the American "OK"-emblem, which in Italy is interpreted as an insult. Deictic gestures identify referents in the gesture space. The referents can be real like the addressee or they can be abstract like pointing to the left and the right while uttering the words "the good and the bad". Iconic gestures depict spatial or shape-oriented aspects of a referent, e.g. by using two fingers to indicate someone walking while uttering "he went down the street". Metaphoric gestures at last are more difficult in that they visualize abstract concepts by the use of metaphors, e.g. by using a box gesture to visualize "a story". This is the conduit metaphor that makes use of the idea of a container in this case a container holding information.

Similar taxonomies have been introduced by Kendon [24] and Ekman and Friesen [13], who for instance distiguish between emblems, illustrators, regu-

lators, affect displays, and adaptors, where emblems and adaptors are comparable to McNeill's categories, illustrators summarize McNeill's iconic, deictic, and metaphoric gestures. Affect displays are movements that are triggered by emotional states like Ekman's [12] basic emotions (fear, anger, joy, suprise, sadness, disgust). The relation of affect displays to body movements remains a bit unclear, the face is identified as the main display for emotions. Regulators at last are all movements that do not fall into one of the other categories and that are identified by Ekman and Friesen as necessary to structure the flow of the conversation.

In this chapter, we are looking into a different level of the semantics of gestures. For an ambient intelligence system, more subtle features of gestural activity can provide relevant contextual information for successful interactions. Therefore, the focus in this chapter is not primarily on the symbolic or communicative content of a gesture – whether provided intentionally or not – but rather on the way a gesture is performed, i.e. on qualitative features of movements and their interpretation.

To provide pervasive assistance for complex computing environments it does not suffice to restrict the analysis and interpretation of body movements to the task of finding an appropriate gesture class. It is necessary to focus on the diverse aspects of body movements, which do not only provide emblematic information but have a whole range of communicative functions ([3], [24]) and even allow to identify inherent user characteristics like identity ([29]), personality ([15]), or cultural background ([35]). Such a shift in perspective will allow for infering additional information about the user's patterns of activities and relating them to cognitive or emotional user states. These can be the user's identity, his personality, his current or expected emotional state or mood, his

current or expected state of arousal/concentration, etc.

Coming back to the example from the beginning, we can speculate a bit more on what this behavior reveals about the speaker. Our first guess was, that he might be nervous because it is an important presentation. On the other hand, the excessive use of hand gestures might also be attributable to the speaker's extrovert personality perhaps strengthened by being nervous. Or the interpretation that his gesture is excessive might just be a misconception by ourselves because we are used to more controlled hand movements in our culture, but in the speaker's culture this might just be a standard behavior routine to underline personal engagement in a topic. These examples make it clear that such non-symbolic, qualitative information can be the source for recognizing a wide range of contextual effects by analyzing a user's gesture usage. At the same time, they underline the complexity of the task as the interpretation of the recognized features is not restricted to a single contextual variable and might be quite ambiguous.. Thus, in the long run, it will be indespensable to come up with an integrated approach for analyzing qualitative features of gesture usage.

As we have seen, a number of contextual factors can relie on gestural activity as an input channel. In the rest of this chapter, contextual influences like the emotional state of the user, the personality of the user, and the cultural background of the user are examined. Beforehand, three attempts are introduced that categorize gestural activity apart from its co-verbal content and that serve as a guideline for analyzing the qualitative features of gestural activity. Table 1 summarizes and groups together the important features from these attempts.

General features	Efron	Gallaher	de Meijer
Location	Plane of gesture		Vertical/Sagittal direction
Distance	Closeness		
Spatial extent	Radius	Constricted vs. expansive	
Speed	Tempo	Tempo	Velocity
Activation		Frequency/Quantity of gesture use	
Fluidity		Jerky vs. smooth	
Power			Force
Body	Body parts Touch	Body Parts Posture	Trunk/Arm Movement

Table 1

General qualitative features of gestural activity for non-symbolic interpretation.

In a study on cultural differences in gesturing (see also Section 5), Efron [11] defined three dimensions for categorizing gestural activity.

(1) Spatio-temporal aspects: This category is based on formal features that allow to describe how a gesture is realized taking into account the radii of

gestures, the plane in which gestures are performed (xy-, xz-, or yz-plane), which body parts are employed, and the speed ("tempo") of gestures.

(2) Interlocutional apects: What Efron calls interlocutional aspects can best be summarized by Hall's [17] notion of "Proxemics", i.e. the way interlocutors use the space available in their face-to-face encounters. This category describes if interlocutors stand close to each other or farther away in such an encounter, if they exhibit frequent body contacts like touching the lower arm of the interlocutor, or if interlocutors gesture while grasping an object, which can be used to emphasize the speakers intention.

(3) Co-verbal aspects: The last category describes the relation of gestures to the content of utterances and is thus in accordance with the above mentioned gestural taxonomies by McNeill or Ekman and Friesen that concentrate on gestures as a co-verbal phenomenon.

A similar set of features can be found in Gallaher's work on personal style (see also Section 4). Gallaher [15] reviews work on expressive movements and shows that intraindividual consistencies exist across a wide range of behavior. For instance, somebody who is walking fast is likely to also do gestures at a high speed, talk faster, and speak louder. Thus, analyzing "tempo", i.e. the speed of movement allows to conclude on more general aspects of a speaker. Gallaher defines a set of expressivity features, which she relates to aspects of personality but which, like Efron's spatio-temporal and interlocutional aspects, are general enough to serve as features for other contextual variables as well. In her analysis, Gallaher focused not only on body movements alone but took other aspects of nonverbal behavior like facial expressions or speech volume into account. A factor analysis revealed four dimensions, which summarize the qualitative

features. The following is an account of (mainly) movement related features from these dimensions.

(1) Expressiveness: This dimension describes which parts of the body are used. Moreover, the frequency of gestures, i.e. how often and how many gestures are used, the speed of gestures and the spatial extent of gestures are features describing the expressiveness of movements.
(2) Expansiveness: To describe how much space a speaker is taking up for doing gestures, the expansiveness dimension is introduced. Features describing this dimension are the spatial extent of gestures and the distance a speaker leaves to his addressee in face-to-face encounters. An example for a non-movement related feature of expansiveness is speech volume.
(3) Coordination: The only movement related feature of the coordination dimension is fluidity, which describes if the movements of a gesture are smooth or jerky.
(4) Animation: The animation dimension is described for instance by postures like slumped vs. erect shoulders or by the speed of gesture or other behaviors like speech.

These dimensions are stable across time and raters. The overlap in features like spatial extent (expressiveness, expansiveness) or speed (expressiveness, animation) shows again, that such qualitative features contribute to different interpretations of behavior. What is evident is that some of the features analyzed by Gallaher are consistent with the features defined by Efron for studying cultural differences in gesture usage.

A third study by de Meijer [10] gives an account on how specific body movements are perceived and what impression they give about the subject's emo-

tional state. To this end, he first defined seven dimensions of body movement that describe in a qualitative way how a specific movement is performed.

(1) Trunk movement: stretching, bowing
(2) Arm movement: opening, closing
(3) Vertical direction: upward, downward
(4) Sagittal direction: forward, backward
(5) Force: strong, light
(6) Velocity: fast, slow
(7) Directness: direct, indirect

These movement qualities are then related to different emotional states of the user (see also Section 3). Again, it is apparent, that there is an overlap in features used to describe body movements with the approaches of Gallaher and Efron. The first two dimensions correspond to some of Efron's spatio-temporal aspects and Gallaher's expressive features. The third and fourth to Efron's location features, the sixth to the tempo or speed feature found again in Efron's spatio-temporal aspects and in Gallaher's expressiveness and animation dimensions. Additonally, de Meijer introduces the force that is used to perform a movement and the feature directness, which unfortunately remains a bit vague and unclear.

Table 1 summarizes the different movement characteristics used by Efron, Gallaher, and de Meijer. Although their analyses focused on very different determinants of behavior – cultural background vs. personality vs. emotional state – there is some overlap in relevant features making these a promising starting point for analyzing gestural activity for ambient intelligent systems on a non-symbolic level.

3 Emotions

Since Picard's seminal book [33], affective interactions have increasingly become the center of interest for Human Computer Interaction due to the fact that emotions – either our own or those attributed to others – play a fundamental role on different levels of our communicative and decision making behavior as has convincingly been shown by Damasio [9]. Especially in situations where the user experiences negative emotions like frustration and/or anger, the interaction might greatly benefit from the system's ability to take the user's state into account in its next move, either to prevent the user from breaking up the interaction altogether or in the ideal case to change the user's emotional state and his attitude towards the system in order to provide for a more positive interaction experience.

Whereas it is undeniable that our faces often reveal our emotional state in face-to-face encounters (see e.g. [14]), the mapping between body movements and emotional states is still subject of discussion. For instance, de Meijer [10] gives an account of a controlled study on how specific body movements are perceived and what impression they give about the subject's emotional state. To this end, de Meijer defines twelve emotion categories follwing Ekman's basic emotions [12] (joy, grief, anger, fear, surprise, disgust) and adding additional categories by Izard [20] (interest, shame, contempt) and some so-called emotional attitudes taken from Machotka [30] (sympathy, antipathy, admiration). Relating given movements to one of the twelve emotion categories allowed de Meijer to identify the qualitative movement features and combinations of them that play a central role in the perception of this emotion. The dimensions are related to the parameters identified in the previous section

Emotion	Activity	Spatial Extent	Power
hot anger	high	high	high
elated joy	high	high	high
happiness	-	-	low
disgust	-	low	-
contempt	low	-	-
sadness	-	-	low
despair	-	high	-
boredom	low	low	low

Table 2

Correlations between emotion categories and movement profiles accoding to Walbott.

and take into account body parts, location, speed and power of movements. A number of correlations between these features and emotions were found. Especially the difference between positive and negative emotions was reliably distinguishable. As a general result of this study, de Meijer was able to define movement profiles for emotions. Thus, single qualitative features were not realiable enough to distinguish emotions, but more complex combinations of movement features had high predictive value.

A similar study was conducted by Walbott [42], who tried to correlate specific movements with specific emotional states. Wallbott is more cautious in his account and states that the quality of body movements cannot directly be

mapped to emotional states. Rather, it is indicative of an emotion's quantity, i.e. it signifies the intensity of an emotion. But also that vice versa, differences in body movement are sometimes explained by the intensity of a given emotion instead of a difference between emotional states. Nevertheless, his results show distinctive patterns of movement and postural behavior for some of the studied emotions. In this study he used 14 emotional categories: elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride, and boredom. Twelve actors (six male, six female) had to act these 14 emotions in two scenarios uttering nonsense sentences to prevent emotional priming by the content of the utterace. 1344 samples were recorded under these conditions. For the analysis, 224 takes were selected from this database. The coding system introduced by Walbott is a combination of a categorical approach similar to Ekman and Friesen for emotions, expressive parameters (activity, spatial extension, power/dynamics) for qualitative movement features, and posture coding following Bull's ideas [3]. Table 2 gives an account for expressive movement profiles for some of the studied emotions. What becomes evident for hot anger and elated joy is the influence of the emotions' intensity on the expressive profile. On the other hand, this data shows that it is feasible to distinguish between low intensity emotions like disgust and contempt based on the expressive features.

Crane and Gross [8] show not only that emotions can be recognized in the body movements of others but also that body movements are affected by felt emotions. Four emotions plus a neutral state were elicited (angry, sad, content, joy and no emotion), subjects were then asked to walk across the room. This movement was recorded by video and motion capture. Afterwards, subjects gave a self-report on the felt emotion. Additionally, recordings were rated by

observers, who could choose out of ten different emotions. Although emotions were recognized beyond chance (62% of anger trials, 76% of sad trials, 74% of content trials, 67% of joy trials, 83% of neutral trials), observers' ratings do not necessarily correspond to the self-reports of the subjects making evident a fundamental problem with these kinds of studies. Actors or laypersons are instructed to display emotions or emotions are elicited by specific means, subjects then rate these expressions. Because this happens in a laboratory setting, the displayed emotions might be not felt but simply acted. Thus, although humans are able to interpret body movements as having affective content, it cannot be guaranteed that a person exhibiting such movements really feels the emotion that is attributed to him. It remains to be seen if these results scale up to natural situations. Crane and Gross analyzed movement taking qualitative movement features into account. Results show that apart from speed and velocity of the walking movement, posture and limb motions were affected that also play a crucial role in hand gestures. Especially sadness seems to influence movement qualities of the arms and hands. The spatial extent – measured in this case by shoulder and elbow ranges – is significantly less compared to all the other emotions, i.e. anger, content, and joy. Categorizing the elicited emotions according to valence and activation dimensions gives another insight. Emotions in the high activation group (anger, joy) show a higher spatial extent in elbow flexions.

Some words are in place on emotional models used in the context of the studies and applications presented here. Most of these relie either on categorical approaches like Ekman's [12] basic emotions or on dimensional approaches, which date back as far as Wundt [45]. Categorical approaches define distinct emotion categories that are often claimed to be universal and that can be mapped to

specific behavior routines like facial displays or – as we have seen above – to expressive movement features. Dimensional approaches on the other hand define emotions as a continuous phenomenon, taking up to three dimensions into account: (i) arousal denotes the intensity of a felt emotion, (ii) valence denotes if this emotion is positive or negative, and (iii) dominance denotes if the emotion is more outgoing like anger or more self-directed like fear. Crane and Gross combine both types of model for their analysis to capture the effects of the intensity of an emotion. As was also shown by Walbott, intensity of emotions is a crucial feature that influences the gestural activity.

Kapur and colleagues [21] present a system that was trained to detect four basic emotions based on movement patterns and performed with a recognition rate similar to a human observer. The emotions were sadness, joy, anger, and fear. To create the necessary database, motion capture data was collected for five subjects who were told to represent the emotional states by moving around. 500 samples were collected, i.e. every subject performed every emotion 25 times. To capture the dynamics of the movements, the velocity of the movement, the acceleration, as well as the position of body parts were used as features. No further movement analysis was conducted, i.e. movements were taken into account as whole samples. As their system was able to perform similar to a human observer, the employed features seem to represent a promising starting point for the recognition task.

Bernhardt and Robinson [2] go a step further and present a machine learning approach that takes the inner structure of movements into account to allow for a more context-dependent classification of emotions based on movement patterns. To this end, they build on work from Bull [3] that shows that affective states can be recognized from body movements. To this end, they

develop a recognition framework by defining motion primitives that are used to recognize affective states. Such primitives are created by clustering motion samples that are found in specific contexts. To exemplify their approach they consider a very small context, which is "knocking at a door". They make use of a database containing around 1200 knocking motions recorded by motion capturing and done in affective ways to realize neutral, happy, angry, and sad knocking. Their clustering approach is based on some apriori knowledge that allows for segmenting the knocking movement into four phases: (i) lift arm, repeatedly (ii) knock and (iii) retract, (iv) lower arm. To recognize the affective states, features are calculated on the motion primitives that are similar to those described in Section 2 (general names given in brackets): maximum distance of hand from body (body parts), average hand speed (speed), average hand acceleration (power), average hand jerk (fluidity). Additionally the same features were calculated for the elbow. Their recognition algorithm first segments a motion into motion primitives for each of the four phases, then calculates the expressive features to classify the affective content of the motion. Results show that this approach is very promising with recognition rates far above chance, i.e. up to 92% for the four class problem.

Castellano, Villalba and Camurri [6] compare the applicability of a time-series classification approach (Dynamic Time Warping) with feature-based approaches (Nearest Neighbour, Bayesian Network, Decision Trees) for recognizing emotions based on nonpropositional gestural qualities. Movements are described by power (amplitude), speed, fluidity, activation, and velocity. To train and test the approach, ten subjects were asked to provide gestures for eight emotional states (anger, despair, interest, pleasure, sadness, irritation, joy and pride). These were chosen because they are equally distributed in the

two dimensional valence and arousal space. Each subject repeated each gesture three times resulting in 240 gestures. The approach then focuses only on four emotions (joy, anger, pleasure, sadness), which represent the four quadrants of the valence-arousal space. Consequently, the approach is based on a very small sample size of 30 samples for each emotion and it remains to be shown if the results scale up. Apart from the movement features mentionend above, Castellano and colleagues calculate some second order statistical features like initial and final slope, initial and final slope of the main peak, maximum, mean, etc. on these motion cues. It remains unclear why this is necessary and how recognition rates benefit from the inclusion of these features. Results show that expressive motion cues allow to discriminate between high and low arousal emotions and between positive and negative emotions. This is in line with Walbott's results (see above), who has shown that such motion cues are a good predictor for the intensity of emotions.

Shan, Gong and McOwan's [38] work on emotion recognition is in line with Efron's analysis. They focus on spatio-temporal aspects for modeling body gestures that allow for recognizing emotional states. Instead of defining specific spatio-temporal features like Efron has done, they analyze video sequences without investing further knowledge into the definition of specific features. Instead they use spatial and temporal filters to identify regions and time-series that show strong spatial or temporal activity. There work is based on the general assumptions that although strong variance can be seen in doing a gestures, spatio-temporal features related to emotions are stable over subjects. Features are directly calculated on the video image as points of interest in the space-time by employing spatial (Gaussian) and temporal (Gabor) filters on the video image to derive these interest points. To classify emtions, a

clustering approach is used to identify movement prototypes based on these interest points. Recognition rates using support vector machines range between 59% and 83% for a seven class problem (anger, anxiety, boredom, disgust, joy, puzzle, surprise). To train their recognition system they make use of a database containing around 1900 videos. Additionally they showed that fusing information from gestural activity and facial expressions can result in higher recognition rates.

To sum up, a number of studies show that there is a correlation between qualitative features of gestural activity as described in Section 2 and emotional states but also that this correlation is not unambiguous and sometimes only allows to derive the intensity of an emotion or its valence but not the distinct emotion itself. Some first approaches to automatically recognize emotions based on such correlations have been presented that are very promising but at the moment lack comparability due to different sets of emotions and quite different databases that were employed for training and testing the recognition techniques.

4 Personality

Whereas the analysis of emotional states have become very popular in recent years, other contextual factors influencing interactions like personality or cultural heuristics for behavior have not been in the central focus of attention, although for instance Gallaher's expressive parameters have been defined to capture the relation between body movements and personality.

Ball and Breese [1] present a first model of integrating personality as a factor

influencing gestural behavior. To this end, they define a Bayesian network that models the causal relations between gestural activity as well as posture and personality traits. Their model is based on studies that show that people are able to reliably interpret personality traits based on movement features. Their approach is primarily concerned with conveying the personality of an embodied agent by characteristic movements but because they model this relation with a Bayesian network, the same approach can be employed to recognize the user's personality based on his movement characteristics, which have already been modeled in the network. Apart from defining specific postures and gestures that are most likely to occur in correlation with a given personality, qualitative characteristics like frequency, speed, and timing of a gesture have been integrated to convincingly convey information about personality.

To integrate personality as a contextual factor influencing the movements of an embodied agent, Pelachaud [32] drew from Gallaher's analysis of personal style to define expressive features that serve as control parameters for the animation (gestures and face) of the virtual character. The aim of this work was to create individual behaviors for an agent instead of generic one's, in this case trying to integrate some kind of personal style for the agent. To this end, she defined a set of six parameters, which are based on Gallaher's dimensions: spatial extent, speed, fluidity, power, repetivity, quantity. Perception studies were conducted, showing that combinations of these parameters establish consistent behavior patterns like sluggishness or vigorous movements. Moreover it was shown that participants are able to recognize the differences in some of these parameters with good results for spatial extent and speed, and less good results for fluidity and power.

Karpouzis and colleagues [22] present a gesture recognition system that takes

the same parameters into account to extract quantitative information related to gestural expressivity from the user's hand movements: spatial extent, speed, fluidity, power, repetitions. Similar to Pelachaud's work, expressivity is not restricted to hand movements but takes head movements and facial expressions into account, too. Spatial extent for instance describes for hand and head movement if this movement is wider or narrower movement, for facial expressions it describes increased vs. decreased muscular contraction.

Caridakis and colleagues [5] then combine both approaches to realize a system that allows to mimic the behavior of a human by a virtual agent based on the recognized expressive features and corresponding profiles of the agent. The general idea is that the agent is not directly mirroring the user's behavior but instead by extracting the expressive parameters, the agent's individual behavior is modified to fit the user's expressive behavior profile. Thus, the same gesture is realized by the agent qualitatively different depending on the set of parameters. For instance, the user might show an expression of sadness accompanied by slow and narrow movements. To mirror this behavior in the agent, the agent's behavior profile for this emotion is combined with the user's expressive parameters to result in a display of the same emotion with a similar profile that nevertheless is idiosyncratic for this agent. This example application represents a first step in analyzing the user's gestural activity as a basis for deriving information about his personality profile.

5 Culture

Labarre [28] reviews a large body of evidence on the cultural differences in using and interpreting body movements including gesture repertoires that have

specific meanings in a given culture (emblems). Most embarrasing situations might occur if someone uses such emblematic gestures unconsciously in interactions with people from other cultures. The best known example might be the American "OK"-sign formed by thumb and index finger which in Italy is a severe insult. Another example taken from Labarre is a gesture, where the open right hand is raised to the face, with the thumb on the bridge of the nose. This is used by the Toda in South India to express respect, the almost identical gesture is used in Germany as a mocking gesture, i.e. as a sign of disrespect. Thus, the recognition of specific gestures may either give interesting insights into the cultural background of the user or it might cause severe problems in interpreting the semantic content of the gesture if the cultural background is not known. Again, the quality of the movement can serve as necessary evidence for a successful disambiguation.

The kinesthetic features defined by Efron (see Section 2) derived from his study of cultural differences in gesturing, but so far there are only very few approaches that take this information into account in an interactive ambient intelligent system. In his study, Efron [11] examined differences in gesturing between Italian and Jewish immigrants as well as assimilated subjects from the same two cultural groups. Based on his large amount of data (around 2500 subjects), he could show significant differences in all the categories he analysed, i.e. apatio-temporal aspects, interlocutionary aspects, and co-verbal aspects (see Section 2). With his sample of assimilated subjects, i.e. subjects already living for a long time in the US, he was also able to show that differences vanished, giving clear evidence that the differences in gestural activity are a learned cultural heuristics. An example of the differences he found is the following: Whereas Italian subjects used their whole arm for gesturing, Jewish

subjects kept their upper arms close to the body resulting in movements from the elbow downwards, i.e. in narrower movements.

This empirical evidence of cultural differences in the way gestures are realized on the spatio-temporal level is accompanied by a number of anecdotal references found in the literature. Hall [16] for instance gives a number of such references to culture-specific differencs in gesture usage. Similar information can be found in Ting-Toomey [40], claiming for instance that Germans use more gestures than Japanese or that Southern Europeans gesture more frequently then Northern Europeans. As we have seen in Section 2, Efron's spatio-temporal and interlocutionary aspects are very similar or identical to Gallaher's expressive dimensions [15], which she uses to distinguish different personal styles of gesturing. This implies again that these dimensions might also be useful for describing cultural differences in gesture use.

Rehm and colleagues [36] present a corpus study designed to shed light on specific differences in gesture usage in individualistic and collectivistic cultures with the aim of deriving expressive profiles for these cultures to adapt the behavior of virtual agents to the user's cultural background. To this end, they recorded around 20 hours of material of interactions in Germany (21 pairs) and Japan (26 pairs). Their analysis focused on nonverbal behavior like gesture use and postures. Gestural expressivity was analysed focusing on parameters, which have been proven to be successful for animating a virtual agent [32]: spatial extent, speed, overall activation, fluidity, power. Results from this corpus analysis show significant differences in the expressive profiles of participants from the two cultures. The frequency of gesture use is consistent with information from the literature [40] in that a significant difference could be seen in the number of gestures that were used in the German and

	Hierarchy	Identity	Gender	Uncert.	Orient.
Germany	35	67	66	65	31
Japan	54	46	95	92	80
Sweden	31	71	5	29	33
US	40	91	62	46	29

Table 3

Hofstede's ratings on a scale from 1 to 100 for some selected countries.

the Japanese samples. German participants used more than three times more gestures than Japanese participants on average. Other significant differences were found for the two expressive parameters spatial extent and speed of a gesture.

Rehm, Bee, and André [35] give an example how this information can be used to infer the cultural background of the user based on his gestural expressivity. They present a Bayesian network model of cultural influences on expressivity that is employed to analyse the user's expressive behavior and derive his cultural background. Culture in their approach is defined as a dimensional model following Hofstede's suggestions [18]. A given culture is thus a point in a five-dimensional space where dimensions describe dichotomies like individualistic vs. collectivistic or high power vs. low power distance. Table 3 gives cultural profiles for some exemplary countries.

(1) Hierarchy: This dimension describes the extent to which different distribution of power is accepted by the less powerful members. According to Hofstede more coercive and referent power (based on personal charisma and identification with the powerful) is used in high-H societies and more reward, legitimate, and expert power in low-H societies.

(2) Identity: Here, the degree to which individuals are integrated into a group is defined. On the individualist side ties between individuals are loose, and everybody is expected to take care for himself. On the collectivist side, people are integrated into strong, cohesive in-groups.

(3) Gender: The gender dimension describes the distribution of roles between the genders. In feminine cultures the roles differ less than in masculine cultures, where competition is rather accepted and status symbols are of importance.

(4) Uncertainty: The tolerance for uncertainty and ambiguity is defined in this dimension. It indicates to what extent the members of a culture feel uncomfortable in unstructured situations which are novel, unknown, surprising, or different from usual. Whereas uncertainty avoiding cultures have rules to avoid unknown situations, uncertainty accepting cultures are more tolerant of opinions different from what they are used to and they try to have as few rules as possible.

(5) Orientation: This dimension distinguishes long and short term orientation. Values associated with long term orientation are thrift and perseverance whereas values associated with short term orientation are respect for tradition, fulfilling social obligations, and saving one's face.

According to Hofstede [18], nonverbal behavior is strongly affected by cultural affordances. The identity dimension e.g. is tightly related to the expression of emotions and the acceptable emotional displays in a culture in that for instance individualistic cultures tolerate the expression of individual anger more easily than do collectivistic cultures. Hofstede, Pedersen, and Hofstede [19] explicitely examine the differences that arise in the use of sound and space for the five dimensions. By relating the results from their corpus study

to Hofstede's dimensional model, Rehm and colleagues show how the user's expressive gestural behavior can be recognized with high accuracy and can then be used to infer the user's position on Hofestede's cultural dimensions. With this information at hand it becomes possible to modify the behavior of an interactive system according to this contextual information.

6 Recognizing gestural behavior for human-centric Ambient Intelligence

In the preceding chapters of Part I, vision-based techniques for gesture recognition have already been presented in depth. Here the focus is on input techniques that make use of sensoric equipment that allows for more private interactions. Although vision-based techniques present the most unobstrusive method for movement analysis and have proven to be very successful for recognizing gestural activity (perhaps apart from some minor occlusion problems), they may present a severe threat to privacy in ambient intelligent environments if the user is unaware of the devices and does not know which information is processed, e.g. his affective state, his personality traits, or his cultural background. Thus, more obstrusive input methods might be more appropriate for such sensitive personal information as they give the control about which information is transmitted to the environment into the hands of the user.

In the remainder of this chapter we therefore present input techniques that make use of acceleration or physiological sensors like EMG. Both techniques rely on sensors that are meanwhile small enough to be worn by the user either as handheld devices or attached to his body. It is not unreasonable to assume that by and large such sensors will become integrated in everyday objects like

rings or items of clothing removing this annoyance altogether.

6.1 Acceleration-based gesture recognition

With the advent of Nintendo's new game console, acceleration-based interactions have become very popular. Although most commercial games seem to rely on relatively primitive information like the raw acceleration, more sophisticated gesture recognition is possible with such a device. Schlömer and colleagues [37] make use of HMMs to analyze the acceleration data. They evaluate their approach with an arbitrary set of five gestures and present user-dependent recognition rates up to 93% for this five class problem. Rehm and colleagues [35] make use of acceleration-based recognition to capture gestural activity that can relate to the cultural background of the user and exemplify this approach with the Wiimote. In their approach, features are calculated on the raw signal. Different classification approaches like Naïve Bayes, Nearest Neighbour and Multilayer Perceptron are compared for different gesture sets like expressivity parameters or German emblems. Results show that recognition rates are user-dependent and that this approach is feasible with recognition rates for a seven class problem of German emblems up to 94% making use of a standard Nearest Neighbour classfier.

In an earlier study, Kela and colleagues [23] present a similar approach tailored to gestures for controlling a video recorder making use of a cubelike handheld device, which was equipped with three acceleration sensors quite similar to Nintendo's controller. To come up with a realistic gesture set, they conducted a participatory design study, which resulted in eight suitable gestures. Gesture analysis was based on HMMs taking the filtered data into account.

User-dependent recognition rates are up to 99% depending on the number of training samples provided to estimate the model parameters.

Urban and colleagues [41] examined the feasibility of using acceleration sensors for a marshalling task designed to control unmanned aircrafts on a flight desk. The general idea was to allow the marshaller to make use of the same gesture signals that are employed with manned vehicles. Two main tasks had to be solved for this 20-class recognition problem. On the one hand, they evaluated the best placement of the acceleration sensors on the upper and lower arm for robust gesture recognition. On the other hand they showed that time-series classifiers like Dynamic Time Warping can be an efficient technique for acceleration-based gesture recognition.

Strachan and colleagues [39] faced the problem of reconstructing the 3D-movement of the hand from acceleration data. This is no trivial task due to inherent drift of the sensors making the prediction of the exact trajectory difficult. By decomposing gestures in linearly combined motion primitives they were able to build personalized models of gestures that a user is going to use in an application. Thus, they integrated subjective idiosyncracies of gestural activity into their recognition system. Whereas this is only a byproduct of their approach, the work by Lester, Hannaford, and Borriello [29] is directly tailored to this challenge.

Wheras most approaches so far focus on the recogntion of discrete gesture classes, Lester and his colleagues exploit the applicability of accerlation-based techniques to identify users by their subjective idiosyncrasies in handling devices. In an ambient intelligence environment, the user will be carrying a number of devices, which have to be coordinated to a certain degree and have

to interact with one another, with the environment and of course with the user. By enabling the device to identify who is currently carrying it might rid the user of some management load. Lester and colleagues make use of information about the user's specific movement qualities to solve this problem. To this end, they employ a complex coherence function measuring to which extent two signals are correlated at given frequencies.

The approaches presented here show that acceleration-based gesture recognition is feasible and that not only gestures as such can be recognized but also more subtle aspects of gestural activity like expressivity or other idiosynchratic features allowing for instance to identify the user.

6.2 Gesture recognition based on physiological input

Another currently not very well explored way of gesture recognition is the use of physiological sensors. Such sensors have increasingly been used over the last years to recognize emotional states or at least a user's state of arousal (e.g. [25], [34]). Some sensors like EMG measure muscle activity and can thus be adapted to capture certain aspects of gestural movements that might not easily be recognizable by vision- or acceleration-based techniques.

Naik and colleagues [31] first separate the muscle activity from different muscles with a four channel EMG sensor before attempting to classify specific movements. Making use of independent component analysis and a neural network model they are able to distinguish accurately between three different types of motion: wrist flexion, finger flexion, and wrist and finger flexion. Depending on the recognition task, this information can be crucial to distinguish

between different gesture classes, for instance in sign language, where finger movements play a crucial role.

Kim, Mastnik and André [26] allow a user to radiocontrol a toy car by different hand gestures, which are recognized from an EMG signal. Four gestures were identified as suitable for this task. Sensors are placed on the lower arm below the wrist. Gesture classification makes use of a combination of Naïve Bayes and Nearest Neighbour classifiers. The system was evaluated with 30 subjects to find the optimal combination of classifiers. User-independent recognition rates for this small set of four gestures vary between 87% and 98% and exemplify convincingly that gesture recognition based on such physiological information is possible.

Whereas Naik and and colleagues are independent from the sensor placement, this is not true for more specific gesture recognition tasks. Wheeler [44] uses EMG sensors to emulate a joystick and a keyboard and depending on the device, i.e. on the movements necessary for the device, number and placement of electrodes is different. In the joystick trial, users had to perform four gestures (up, down, left, right), which were recognized making use of four HMMs, one for each gesture class. Recognition results are accurate for all but the gesture "left", which was only recognized in 30% of the cases and otherwise confused with "up". In the more complex keyboard trial, users had to perform 11 gestures (0 to 9, enter). Again, one HMM was trained for each gesture class. Recognition rates vary between 70% and 100% depending on the gesture class.

All approaches show that gesture recognition with EMG is possible but that it is not easy to get robust recognition rates especially due to problems in placing the sensors. Recognition results are dependent on the muscles that the sensors

are placed to and on the specific gestures that are realized in an application making it difficult to come to a general conclusion. A promising solution seem to be the combination of acceleration-based and EMG-based recognition as was recently shown by Chen and colleagues [7] for the recognition of Chinese and by Kim and colleagues [27] for the recognition of German sign language.

7 Conclusion

This chapter provided insights into how qualitative aspects of gestural activity can be exploited as an input channel for a variety of contextual variables like the emotional state of the user, his personality, or his cultural background. It was shown by evidence from studies on these different aspects that a general set of qualitative movement features can be defined and how these features can then further the recognition of emotion, personality or cultural background from the user's gestures.

Although all of the presented approaches are very stimulating and relevant, it remains to be shown how such social-psychological context variables can be integrated for human-centric ambient intelligence because for instance the speed and spatial extent of a gesture might give hints on the personality profile of the user but these features might also allow for infering the cultural background of the user. The fact that the same set of features (or at least subsets of this general set) are applicable for all of the variables presented in this chapter emphasizes the fact that such an integrated account is feasible and also necessary.

References

[1] Gene Ball and Jack Breese. Relating personality and behavior: Posture and gestures. In A. M. Paiva, editor, *Affective Interaction*, pages 196–203. Springer, Berlin, Heidelberg, 2000.

[2] Daniel Bernhardt and Peter Robinson. Detecting affect from non-stylised body motions. In A. Paiva, R. Prada, and R. W. Picard, editors, *ACII 2007*, pages 59–70. Springer, Berlin, Heidelberg, 2007.

[3] P. E. Bull. *Posture and Gesture*. Pergamon Press, 1987.

[4] Judee K. Burgoon. Nonverbal signals. In Mark L. Knapp and Gerald R. Miller, editors, *Handbook of Interpersonal Communication*, pages 229–285. SAGE Publications, Thousand Oaks, 1994.

[5] George Caridakis, Amaryllis Raouzaiou, Elisabetta Bevacqua, Maurizio Mancini, Kostas Karpouzis, Lori Malatesta, and Catherine Pelachaud. Virtual agent multimodal mimicry of humans. *Language Resources and Evaluation*, 41:367–388, 2007.

[6] Ginevra Castellano, Santiago D. Villalba, and Antonio Camurri. Recognising human emotions from body movement and gesture dynamics. In A. Paiva, R. Prada, and R. W. Picard, editors, *ACII 2007*, pages 71–82. Springer, Berlin, Heidelberg, 2007.

[7] X. Chen, X. Zhang, Z. Zhao, J. Yang, V. Lantz, and K. Wang. Hand Gesture Recognition Research Based on Surface EMG Sensors and 2D-accelerometers. In *IEEE International Symposium on Wearable computers*, pages 11–14, 2007.

[8] Elizabeth Crane and Melissa Gross. Motion Capture and Emotion: Affect Detection in Whole Body Movement. In *Affective Computing and Intelligent Interaction*, pages 95–101. Springer, Berlin, Heidelberg, 2007.

[9] Antonio R. Damasio. *Descartes Irrtum*. dtv, 1995.

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

[11] David Efron. *Gesture, Race and Culture*. Mouton and Co, 1972.

[12] Paul Ekman. Basic emotions. In Tim Dalgleish and Mick Power, editors, *Handbook of Cognition and Emotion*, chapter 3, pages 45–60. John Wiley and Sons Ltd., Chichester, 1999.

[13] Paul Ekman and Wallace Friesen. The repertoire of nonverbal behavior: categories, origins, usage and coding. *Semiotica*, 1:49–98, 1969.

[14] Paul Ekman and Erika Rosenberg, editors. *What the Face Reveals: Basic & Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, 1998.

[15] Peggy E. Gallaher. Individual Differences in Nonverbal Behavior: Dimensions of Style. *Journal of Personality and Social Psychology*, 63(1):133–145, 1992.

[16] Edward T. Hall. *The Silent Language*. Doubleday, 1959.

[17] Edward T. Hall. *The Hidden Dimension*. Doubleday, 1966.

[18] Geert Hofstede. *Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage Publications, Thousand Oaks, London, 2001.

[19] Gert J. Hofstede, Paul B. Pedersen, and Geert Hofstede. *Exploring Culture: Exercises, Stories, and Synthetic Cultures*. Intercultural Press, Yarmouth, 2002.

[20] C. E. Izard. *Human Emotions*. Plenum Press, 1977.

[21] Asha Kapur, Ajay Kapur, Naznin Virji-Babul, George Tzanetakis, and Peter F. Driessen. Gesture-based affective computing on motion capture data. In

J. Tao, T. Tan, and R. W. Picard, editors, *Affective Computing and Intelligent Interaction (ACII)*, pages 1–7. Springer, Berlin, Heidelberg, 2005.

[22] Kostas Karpouzis, George Caridakis, Loic Kessous, Noam Amir, Amaryllis Raouzaiou, Lori Malatesta, and Stefanos Kollias. Modeling naturalistic affective states via facial, vocal, and bodily expressions recognition. In *Human Computing*, pages 91–112, Berlin, Heidelberg, 2007. Springer.

[23] Juha Kela, Panu Korpipaää, Jani Mäntyjärvi, Sanna Kallio, Giuseppe Savino, Luca Jozzo, and Sergio Di Marca. Accelerometer-based gesture control for a design environment. *Pers. Ubiquitous Computing*, 10:285–299, 2006.

[24] Adam Kendon. *Gesture — Visible Action as Utterance*. Cambridge University Press, 2004.

[25] Jonghwa Kim, Elisabeth André, Matthias Rehm, Thurid Vogt, and Johannes Wagner. Integrating Information from Speech and Physiological Signals to Achieve Emotional Sensitivity. In *Proceedings of Interspeech/Eurospeech*, 2005.

[26] Jonghwa Kim, Stephan Mastnik, and Elisabeth André. Emg-based hand gesture recognition for realtime biosignal interfacing. In *Proceedings of Intelligent User Interfaces*, pages 30–39, 2008.

[27] Jonghwa Kim, Johannes Wagner, Matthias Rehm, and Elisabeth André. Bi-channel Sensor Fusion for Automatic Sign Language Recognition. In *Automatic Face and Gesture Recognition*, 2008.

[28] Weston Labarre. The cultural basis of emotions and gestures. In *NACHSCHAUEN*. NACHSCHAUEN, 49–68, 199X.

[29] Jonathan Lester, Blake Hannaford, and Gaetano Borriello. Are you with me? — using accelerometers to determine if two devices are carried by the same person. In A. Ferscha and F. Mattern, editors, *PERVASIVE 2004*, pages 33–50, Berlin, Heidelberg, 2004. Springer.

[30] P. Machotka. Body movements as communication. *Dialogues: Behavioral Science Research*, 2:33–65, 1965.

[31] Ganesh R. Naik, Dinesh Kant Kumar, Vijay Pal Singh, and Marimuthu Palaniswami. Hand gestures for hci using ica of emg. In *HCSNet Workshop on the Use of Vision in HCI (VisHCI)*, pages 67–72, 2006.

[32] Catherine Pelachaud. Multimodal expressive embodied conversational agents. In *Proceedings of ACM Multimedia*, pages 683–689, 2005.

[33] Rosalind Picard. *Affective Computing*. MIT Press, Cambridge, 1997.

[34] H. Prendinger, H. Dohi, H. Wang, S. Mayer, and M. Ishizuka. Empathic embodied interfaces: Addressing users' affective state. In E. André et al., editor, *Affective Dialogue Systems (ADS-04)*, pages 53–64, Berlin, Heidelberg, 2004. Springer.

[35] Matthias Rehm, Nikolaus Bee, and Elisabeth André. Wave like an Egyptian — Acceleration based gesture recognition for culture-specific interactions. In *Proceedings of HCI 2008 Culture, Creativity, Interaction*, pages 13–22, 2008.

[36] Matthias Rehm, Yukiko Nakano, Elisabeth André, and Toyoaki Nishida. Culture-specific first meeting encounters between virtual agents. In Helmut Prendinger et al., editors, *Intelligent Virtual Agents*. Springer, 2008.

[37] Thomas Schlömer, Benjamin Poppinga, Niels Henze, and Susanne Boll. Gesture recognition with a wii controller. In *Proceedings of Tangible and Embedded Interaction (TEI)*, 2008.

[38] Caifeng Shan, Shaogang Gong, and Peter W. McOwan. Beyond facial expressions: Learning human emotion from body gestures. In *Proceedings of British Machine Vision Conference (BMVC)*. 2007.

[39] Steven Strachan, Roderick Murray-Smith, Ian Oakley, and Jussi Ängeslevä.

Dynamic primitives for gestural interaction. In S. Brewster and M. Dunlop, editors, *MobileHCI 2004*, pages 325–330, Berlin, Heidelberg, 2004. Springer.

[40] Stella Ting-Toomey. *Communicating Across Cultures*. The Guilford Press, New York, 1999.

[41] Martin Urban, Peter Bajcsy, Rob Kooper, and Jean-Christophe Lementec. Recognition of arm gestures using multiple orientation sensors: Repeatability assessment. In *IEEE Intelligent Transportation Systems Conference*, pages 553–558, 2004.

[42] Harald G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[43] Paul Watzlawick, Janet H. Beavin Bavelas, and Don D. Jackson. *Menschliche Kommunikation*. Huber, Bern, 1969.

[44] Kevin R. Wheeler. Device control using gestures sensed from emg. In *IEEE International Workshop on Soft Computing in Industrial Applications*, 2003.

[45] Wilhelm Wundt. *Grundriss der Psychologie*. Engelmann, Leipzig, 1896.