AMAB: Automated Measurement and Analysis of Body Motion

Ronald Poppe

University of Twente, NL

Sophie van der Zee

Lancaster University, UK

Dirk K. J. Heylen

University of Twente, NL

Paul J. Taylor

Lancaster University, UK & University of Twente, NL

Address correspondence to: Ronald Poppe, (r.w.poppe@utwente.nl), Department of Human Media Interaction, University of Twente, Enschede, The Netherlands.

Abstract

Although technologies that measure human nonverbal behavior have existed for some time, their use in the analysis of social behavior has been limited. Recent developments in sensor technology enable the automatic recording of full-body movement and psychologists have begun to adopt these tools to address their research questions. However, a standardized methodology to efficiently represent and analyze nonverbal behavior, specifically full-body motion, is absent. In this article, we present Automated Measurement and Analysis Of Body Motion (AMAB), a methodology for examining individual and interpersonal nonverbal behavior from the output of full-body motion tracking systems. We address the recording, screening and normalization of the data. Also, we propose measures to operationalize common research questions in psychological research. Practical examples from several application areas are presented to demonstrate the efficacy of our proposed method to address full-body measurements and comparisons across time, space, body parts and subjects.

AMAB: Automated Measurement and Analysis of Body Motion

Nonverbal behavior is a key ingredient in personal expression (McNeill, 1985) and the regulation of interpersonal exchanges (Ekman, 1967). Its analysis has contributed significantly to our understanding of how human interaction works. It is perhaps not surprising, then, that researchers continue to develop methods for the effective measurement and analysis of such behavior. The most common approach relies on observational coding of behavior, using classification schemes that are developed to serve a particular research question (Dael, Mortillaro, & Scherer, 2012; Lausberg & Sloetjes, 2009; Tiedens & Fragale, 2003). However, while this approach has made important contributions, it is not without its limitations. Its application is both time consuming and open to problems of reliability and validity (Scherer & Ekman, 1982), meaning that there is often an inherent trade-off between the number of actions one codes, the amount of material that is coded, and the reliability of the measurements.

In an effort to circumvent these difficulties, there has been a growing trend toward using technologies to evaluate behavior (Altorfer, Jossen, Wurmle, Kasermann, Foppa, & Zimmermann, 2000; Bente, Senokozilieva, Pennig, Al-Issa, & Fischer, 2008). In particular, researchers have started to undertake automatic measurement of human movement with motion capture devices. To date, such approaches have focused on examining discrete nonverbal behavior, such as head movement or gestures, often in response to particular stimuli (REF). Yet, to explore how nonverbal behavior contributes to the processes of human interaction as observed in more naturalistic settings, there is a need to develop a methodology that allows for the capture of nonverbal behavior over an extended period of time. In this article we introduce a standardized approach to using motion capture methodologies for examining full-body nonverbal behavior. We describe how to process raw data independent of the type of motion capture device and deal with issues such as distortions, normalization and alignment. We exemplify our approach with several case studies of nonverbal behavior measurements to show the efficacy and application of our approach.

Automatic Measurement of Human Body Motion

Recording Equipment

The adaptation of modern computing technology and the development of dedicated technologies has made it easier for researchers to record and analyze human nonverbal behavior (e.g., Dakin, Luu, van den Doel, Inglis, & Blouin, 2010; Krishnan, Juillard, Colbry, & Panchanathan, 2009). Table 1 identifies some of the devices available for recording human motion as a function of distinctions in how they capture and treat movement data. As can be seen from Table 1, the devices differ according to whether they rely on markers or sensors to record movement, and according to whether they offer full-body or single movement capture.

Marker-based approaches employ a set of cameras that are used to detect markers worn on the body. The markers are either passive, such as retro-reflective balls, or active, such as inferred transmitters. The former ensures good visibility but it can lead to confusion across markers, while the latter uses distinct frequencies to avoid confusion but it does require monitoring of the power available to the markers to avoid data loss. For both approaches, in order to obtain a 3D measurement of each marker, it must be visible from at least two cameras, which means that a large number of cameras are needed in order to avoid occlusion, particularly when studying social behavior.

Inertial devices overcome this drawback by measuring movements from the body, typically through sensors worn in a suit or straps. The sensors employ changes in the magnetic field in a gyroscope-like manner to make estimates of their positions. The accuracy of this approach is typically high, though without additional position measurements (such as sonar) the estimated positions might suffer from drift, notably in the presence of metal in the recording environment (e.g., floor, walls and ceiling) and objects therein (e.g., chairs and tables). Moreover, participants wearing such devices might be more conscious of their behavior (REF), which might reduce the ecological validity of any recorded social interaction (REF).

An increasingly popular alternative to compensate for this validity problem is to use unobtrusive devices that analyze full-body movement using single or multiple cameras. These devices …do what…, possibly aided by projected structured light (as in Microsoft Kinect), to make a volumetric estimation of the scene, in which one or more parametric body models are fitted (Poppe, 2007). The accuracy and robustness are currently lower compared to marker-based and inertial devices and they suffer from the same occlusion problems as the marker-based approaches. However, their unobtrusive nature may make them preferable to some research designs.

AMAB, the method described in this paper, applies to full-body measurements. These can be obtained by using multiple single-sensor devices to simultaneously record the movements of individual body parts. This is a sufficient and economical solution for certain research questions. Without loss of generality, in AMAB it is assumed that all measurements are obtained from a full-body motion capture device.

Movement Representation

Independent of the type of device used to record movement, the representation of these movements in data form is, to some extent, standardized. This is because the human body is most efficiently described in terms of a series of body parts and joints, the former being shapes with a certain length and the latter being single points in space. Together body parts and joints form a tree-like representation of the human body, and movement may be described in terms of the displacement and orientation of the joints. Figure 1 shows a 'kinematic tree' that allows movement calculation to occur in practice. The joint at the top of the tree, usually the pelvis, is termed the root. Of the two joints that are connected to a body part, the one higher in the tree is considered the parent and the other the child. Joints higher in the hierarchy affect those below. For example, movement in the right shoulder affects the right elbow and wrist joints. End-effectors are joints without children (e.g. the right hand and head). Single-sensor measuring devices are typically used to measure the movement of end-effectors. Typically, sensors and markers are not attached at the location of the joints. While one could, in principle, use the sensor locations to analyze human movement, there is no guarantee that these locations are the same between subjects. Small deviations can lead to different values of position and orientation, the main types of data output. Motion capture equipment often employs a calibration phase to determine the joint positions relative to the sensors' placement. In addition, the lengths of the human subject are recorded. These are used to derive body part orientations from joint positions and vice versa. AMAB only considers joint positions, as comparisons between these are more straightforward.

Joint positions are typically expressed as (x, y, z) values, which correspond to distances in space (in meters), factored along each of three axes. Velocity is the derivative (i.e., change) of position over time, measured in meters per second. Velocity has a magnitude and direction along each axis, for example with upward movements being positive. The position of a joint can be described locally, or globally in world space. The former describes positions relative to the parent joint and is suited to analyze the movement of a single joint. Our method considers full-body poses, which can more conveniently be compared using global representations. These require that the origin and x-, y- and z-axes are defined in a calibration step preceding the recording of the body motion. The origin is the base point in space, i.e. the point (0, 0, 0), from which all positions are measured along the axes. Global representations are convenient for displaying the recorded motion, and to directly compare the positions of joints in time and space and across subjects. All devices in Table 1 can output global joint positions.

Output Format

The output of full-body motion capture is numerical and once the movement has been recorded, it can be visualized similar as to recorded videos. Most software packages supplied with the measuring devices allow motion playback from different angles. Different from the coding of videos is the subsequent processing of the body motion. In the former, human annotators typically provide nonverbal behavior class labels for specific time intervals (e.g., raising a hand or turning the head). In contrast, automatic measurement of body motion results in numerical representations of the body's position over time. The raw motion data has to be exported into a format that can be used in numerical analysis software, such as Matlab and Excel. In the remainder of this paper, the k^{th} measurement of body pose is written as a vector (i.e., a row of numbers): $x^k = (x_1^k, \ldots, x_i^k)$. Each component x_i^k ($i \in \{1, \ldots, n\}$) in the vector corresponds to a joint position measurement along an axis. The n position measurements are in fixed order of joints and axes. For each measurement x^k, the recording time t^k is available. There are m measurements in the current session, so $k \in \{1,...,m\}$. When measurements are taken at fixed time intervals, the frame rate in Hertz (Hz) is the number of measurements per second.

Data Screening

Recorded data can be distorted in many ways. Before analyzing the data, it needs to be screened. This process includes removal of data distortions and normalization.

Data Distortions

Data distortions are due to measurement noise and longer-term inconsistencies in the data due to equipment or transmission failure. One has to check, detect and remove all distortions, seen as they can greatly influence the data; with increased chances of biased conclusions. Measurement noise includes incidental values, which are often detected and smoothed by the software that communicates with the device. In the rare occasion that the software does not take care of this, our method applies a moving median filter with a modest window size (e.g. 0.25 seconds) to eliminate incidental off-values while making sure that jitter in the output is suppressed (see Figure 2). Formally, the filtered vector $x'^k = (x'^k_1, ..., x'^k_n)$ is obtained with:

$$x'^k_i = median(\{x^{k-\lambda}_i, ..., x^k_i, ..., x^{k+\lambda}_i\}) \ (k \in \{\lambda + 1, ..., m - \lambda\}) \quad (1)$$

In this equation, λ is the number of measurements before and after the current measurement that is taken into account. The function *median* returns the median value of a list of numbers. For example, with a recording frame rate of 60 Hz and a window size of 0.25 seconds, λ is 7.

The nature of equipment failure depends on the type of body motion device. When the time of failure is short, the missing measurements can be interpolated from the measurements before and after the failure. The linear interpolation used in AMAB is not the most natural option, but the errors introduced are reasonable. When using higher-order approximations of the measurements, there is the risk of over-estimating the movement.

Normalization

There are a number of common problems that researchers of interaction face, and things are no different for the analysis of automatically recorded body motion. To compare body movements within or between recording sessions, within or between subjects, inter-personal differences in body size and differences in space and time might affect the comparisons. In AMAB, we apply normalization in time and space to cope with these variations.

Normalization in time. There are two relevant types of normalization in time: synchronization and frame rate alignment. When multiple recordings are made simultaneously, synchronization is either handled by the recording software, or needs to be established during data screening. The latter case occurs when recordings have been made on different computers or with different software programs, e.g., motion capture and video recording software. Synchronization ensures that the recordings cover the same time interval. To determine the maximum time interval that is covered by all recordings, the latest start point and earliest end point are determined. For two sequences A and B, with m_A and m_B measurements and recording time vectors t_A and t_B, respectively, the new start (t^S) and end (t^E) points are determined as:

$$t^S = \max(t_A^1, t_B^1),\ t^E = \min(t_A^{m_A}, t_B^{m_B}). \tag{2}$$

For each sequence, the corresponding start and end measurement index are determined, i.e., $k_A^S = \underset{i}{\arg\min}\{t_A^i \geq t^S\}$ is the start index of sequence *A*. This approach can be applied to any number of sequences. See Figure 3 for a schematic representation of the alignment process. Without loss of generality, it is assumed in the remainder of this text that measurements are renumbered, i.e., sequences start at $k = 1$ and end with measurement index $k = m$.

Typically, motion capture devices record at a relatively high fixed frame rate, e.g., 50 or 120 Hz. Such high numbers of measurements per second are computational impractical in some cases. Also, frame rates might differ between sessions. To this end, the measurements are down-sampled in time to a fixed frame rate, with a value depending on the specific research question. Ideally, the factor between the original frame rate and the new frame rate is a natural number *r*. In this case, down-sampling is achieved by selecting every r^{th} measurement, i.e. $k = \{1 + ri\}$ with $i < \frac{m}{r}$. When *r* is not a natural number, *k* can be a fraction and the weighted average of the measurement before (k^-) and after (k^+) target frame *k* is calculated as:

$$x'^k = \frac{(k - k^-)x^{k^-} + (k^+ - k)x^{k^+}}{r}. \tag{3}$$

Normalization in space. Global joint positions depend on the location of the root and the locations of the joints relative to the root. When comparing the body pose of a single subject at different time instances, or when comparing the body postures of multiple subjects, the global position within the recording space is an important factor. Without normalization, the difference in global position is likely to largely influence any pose comparison. Even if a subject performs the exact same movement but at different locations in a room, there will be differences in the pairwise comparisons of the positions of each joint. For those research questions where this effect is undesirable, poses are normalized for position by expressing all position measurements relative to the root of the body. When multiple subjects have been recorded simultaneously, this step can be applied for each of them individually. Typically, the pelvis is used as the root joint. Consequently, the location of the pelvis in the recording space is translated to (0, 0, 0). This translation is also applied to the positions of all other joints, i.e. the pelvis joint position (x_{Px}, x_{Py}, x_{Pz}) is subtracted from the position of each joint j individually:

$$(x'_{jx}, x'_{jy}, x'_{jz}) = (x_{jx} - x_{Px}, x_{jy} - x_{Py}, x_{jz} - x_{Pz}) \tag{4}$$

Figure 4(I) shows a schematic example of position normalization in the case of two subject seated at opposite sides of a table.

The global body orientation, i.e., facing direction of a subject is not important for all research questions. For example, the orientations of two subjects seated oppositely at both sides of a table are approximately 180 degrees apart. When looking at the similarity of their poses, it makes sense to rotate the pose of one of the subjects 180 degrees around a vertical axis, see Figure 4(II).To apply heading normalization, it is assumed that poses are normalized for position and the x- and z-axis are parallel with the ground plane and the y-axis is pointing upward. All joints are rotated around the y-axis in such a manner that the subject faces the positive x-axis. To this end, the left hip is placed above the positive z-axis, i.e., with an x-position of zero. The angle of rotation θ is determined as:

$$\theta = -\arctan(x_{LHx} / x_{LHz}) \tag{5}$$

with $(x_{LHx}, x_{LHy}, x_{LHz})$ denoting the position of the left hip. Next, all joints are rotated around the y-axis with angle *θ*. For joint *I* with position (x_{Ix}, x_{Iy}, x_{Iz}), the rotated x- and z-positions are determined by:

$$\left[x'_{Ix}, x'_{Iz}\right] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_{Ix} \\ x_{Iz} \end{bmatrix} \tag{6}$$

The y-position remains unchanged. An example of the result of position and orientation normalization is shown in Figure 5. The graph shows the sum of all pairwise joint distances (see equation 7) between two subjects seated at opposite sides of a table (see Figure 4). The graph clearly shows the effect of normalization on the pose distance.

Normalization for different subjects. Subjects differ in their body sizes, e.g., arm, leg and upper body lengths. These differences cause different subjects with similar joint rotations to have different joint positions and vice versa. Consequently, these differences affect the comparison of their joint positions. Typically, these effects are rather small when looking at full-body motion. To reduce them even further, the change in position, i.e., velocity, can be used to encode the direction and magnitude of the movement.

Besides structural differences in body sizes, subjects usually display different nonverbal behavior, for example in the amount and type of gesturing. When manually labeling videos, such measurable differences might give rise to normalization for occurrence or the introduction of subject-specific baselines for the amount or types of behavior. A similar treatment can also be applied to the numerical analysis of body movement, for example when subjects differ in the amount of movement they exhibit. Care should be taken when evening out these differences, as they are often meaningful and the normalization might not be straightforward. Normalization for differences in body size is typically not required to answer most research questions. Due to the specific nature of the normalization, it is not part of AMAB.

Interpretation and Operationalization

In this section, it is demonstrated how the AMAB approach can be applied to address various research questions that involve full-body motion measurements. First, common dependent variables are described, followed by more extensive discussions focused on time, space, body parts and multiple subjects.

Dependent Variables

For most of the research that employs full-body measurements, the operationalization of the research questions involves calculating differences between poses, movement velocity or a quantization of the amount of body movement. Their calculation from screened data is discussed subsequently.

Pose difference. When comparing two poses *A* and *B*, their difference can be expressed as the sum of the distances between each of the joints *j* in the set *J*:

$$\delta_{A,B} = \sum_{j \in J} \sqrt{(x_{j_A x} - x_{j_B x})^2 + (x_{j_A y} - x_{j_B y})^2 + (x_{j_A z} - x_{j_B z})^2} \,. \quad (7)$$

The distance for each joint individually is calculated using Pythagoras theorem. Pose differences can only be compared when the sets of joints *J* are equal and both poses have been equally normalized.

Movement velocity. The change in the pose of one subject is the difference between two subsequent poses and is calculated with equation 7. Changes in pose are most conveniently expressed as velocities, with meter per second (m/s) as the unit. To this end, the pose distance needs to be calculated per second, which depends on the frame rate of the recording or the down-sampling applied in data screening. For a frame rate f and pose difference δ, the velocity v is calculated as:

$$v = f\delta \; . \tag{8}$$

Amount of movement. The total amount of movement for a single joint or all joints can be calculated by summing all pairwise distances between subsequent measurements over an interval. For a sequence of length m and all joints in J, the amount of movement α is:

$$\alpha = \sum_{i=1}^{m-1} \delta_{x^i, x^{i+1}} \; . \tag{9}$$

When comparing the amount of movement between two sequences, these should cover time intervals of equal length and with equal frame rates.

Comparisons across Time

Motion capture devices measure full-body poses over time. While the pose, i.e. the spatial configuration of the body, is informative, one is sometimes particularly interested in the temporal aspect of human movement. Changes in pose or movement over time can be used to measure consistency of movement or response time. For example, fatigue can be measured by analyzing the decrease in total amount of movement over time. The analysis of response times to certain stimuli is particularly important in sports research, e.g. the swing of a bat in response to a baseball pitch or the start of a 100 meter sprint. The latter will be used here as an example to demonstrate how the AMAB method can be applied in practice.

Currently, response times in the start of sprinting are typically measured using pressure-sensors in the starting block (Bezodis, Salo, & Trewartha, 2010). A threshold on this pressure is set to exclude false alarms due to small changes in body pose. Analysis of these response times allows for a more accurate measurement of false starts, but also for the development of more effective start poses. While pressure measurements are accurate, foot pressure is the result of all movements of body parts higher in the kinematic chain (see Figure 1). As such, it is an indirect measurement. Ideally, one wants to analyze not only the final pressure of the foot but also the evolvement of the movement of all parts in the body. The use of motion capture devices enables one to perform such analysis. For example, Slawinski, Dumas, Cheze, Ontanon, Miller and Mazure–Bonnefoy (2013) employ full-body motion capture devices to study pose change during the start of a sprint.

AMAB can be used for the numerical analysis of response times, in order to detect false starts. First, the full-body movement needs to be filtered (equation 1) and aligned in time relative to the starting stimulus (equation 2), e.g. a gunshot. After normalization in space (equations 4-6), frame-to-frame differences between poses can be calculated using equation 7. Subsequently, a threshold can be set on the movement velocity (equation 8) in order detect the movement offset and to prevent false alarms.

Comparisons across Space

While normalization of position is typically carried out to compare different poses, the position itself can also be used as an independent measurement. In proxemics research, one is interested especially in the location of a person in space, for example the distribution of groups of subjects in a room.

The physical distance between individuals is termed interpersonal distance and is influenced by factors such as age (Hayduk, 1983) and culture (Hall & Whyte, 1963). Hall and colleagues distinguished four different zones of interpersonal distance: intimate (<

1.5 feet), personal (1.5-4 feet), social (4-12 feet) and public (12-25 feet). Researchers have studied interpersonal distance on all four levels, usually using video recordings to estimate distances (Hayduk, 1983). According to Hayduk, especially the measurements and methodological strategies to study interpersonal distance needed refinement, which can be achieved through the use of new technologies. One line of research has focused on the use of head-mounted displays, which track the subject's head position and orientation, and adapt the view on a virtual environment accordingly in the display (Bailenson, Blascovich, Beall, & Loomis, 2003). Another line of research studies interpersonal distance with the use of virtual characters, which allows for the control of stimuli but reduces the ecological validity, especially when the interactions involve dialogue (Dotsch & Wigboldus, 2008). The use of full-body motion capture devices enables the accurate recording of interactions between human subjects without hindering other means of expression (e.g., facial movement and speech). In addition, more informative measures can be employed, such as those looking at open or closed postures and body sway, which allows one to study interpersonal space accurately on all four levels.

The proposed method can be used to determine interpersonal distances automatically. One could record the motion of two subjects A and B using motion capture devices, and filter the measurements over time by applying a median filter over each sequence (equation 1). Next, the sequences are aligned and converted to the same frame rate (equations 2 and 3). The distance in position can be obtained by calculating the pose difference between head positions of A and B, observed at the same time. To this end, equation 7 is applied with only the head joint H in set J. As the position in world space is required, no position and orientation normalization is applied. Note that equation

7 also takes into account the height of the subjects' heads. Alternatively, this height can be ignored by calculating the inter-personal distance $\delta_{A,B}$ between head positions x_{H_A} and x_{H_B} in only the x- and z-plane as: $\delta_{A,B} = \sqrt{(x_{H_Ax} - x_{H_Bx})^2 + (x_{H_Az} - x_{H_Bz})^2}$. The average and standard deviation of the interpersonal distance can be calculated over all frames in a sequence. In addition, one can look at the orientation of both subjects towards each other by applying equation 5 and comparing the two values of θ. If these are not 180 degrees apart, subjects are not facing each other directly.

Comparisons across Body Parts

Instead of treating the body as a whole and calculating differences between full-body poses, joint positions can be compared pairwise when body movements have been recorded. In rehabilitation studies, the movement of affected and unaffected limbs are often compared. Similarly, research into gestures often differentiates between both hands.

One prominent line of research is concerned with the relation between body pose and perceived affect. Typically, the positions or orientations of body parts are varied systematically, e.g. a head rotation of 20 degrees left and a right elbow bend of 45 degrees. Subjects are shown stimuli of manipulated body poses and asked to assign affective labels. These ratings are then used to determine patterns in the relation between the position or orientation of individual body parts and the perceived affective state. The stimuli can be photographs of posed mannequins (James, 1932) or computer-generated characters (e.g. Coulson, 2004). While these stimuli can be varied systematically, their ecological validity is often lower as the body parts are arranged in an unnatural manner. To this end, researchers have employed motion capture devices to record full-body poses while eliciting emotions and thereby capturing body poses that correlate with genuine emotions (e.g., Kleinsmith, Bianchi-Berthouze, and Steed (2011)).

With AMAB, one could record full-body poses while eliciting known emotions. After normalization for position (equation 4) and orientation (equations 5 and 6), one can analyze the relative position of end-effectors, e.g. the positions of the hand relative to the elbows and shoulders, feet relative to the knees and hip. In line with Coulson (2004), one could expect raised hands to be more correlated to happiness than to sadness. The use of motion capture devices also allows for the analysis of dynamic aspects of affect, e.g. by measuring the velocity or amount of movement for individual body parts using equations 8 and 9.

Comparisons across People

Often, one is interested in comparing body movements of multiple people. For example, subjects performing the same task at different moments in time, such as performing gestures. Alternatively, one can look at body movement of multiple interacting subjects at the same moment in time, e.g. in studies on pedestrian avoidance in crowded places and on turn-taking in interactions. The example that will be explained in more detail here is the occurrence of behavioral mimicry in interactions.

Non-conscious behavioral mimicry is the automatic tendency to imitate the behaviors of other people, including poses, gestures, mannerisms, speech rates and facial expressions (Chartrand & Bargh, 1999; Stel, van Dijk, & Olivier, 2009) at the same time or within a time window of 3-5 seconds (Chartrand & Lakin, 2012). Increased levels of mimicry facilitate smooth interactions and foster liking (Chartrand & Bargh, 1999) and recent research has focused on the moderators and consequences of behavioral mimicry

(Chartrand & Lakin, 2012). So far, in most studies, manually coded events from video recordings are used to measure behavioral mimicry. Besides issues with the subjective and time-consuming nature of the task (Scherer & Ekman, 1982), these comparisons are usually only made between isolated behaviors (e.g., face touching and gesturing). This excludes more detailed analysis of the form, magnitude and direction of the behavior. When using full-body measurements, these factors can be taken into account. For example, an event binary coded as a posture shift can instead be quantitatively analyzed in terms of directed joint motion, which provides more objective and detailed information about the movement.

The AMAB methodology can be used to numerically analyze the amount of mimicry between two subjects A and B, by looking at their poses or their motion. After simultaneously recording the body movements, noisy measurements can be removed (equation 1) and the measurements can be aligned in time (equation 2). When using manually coded videos, the occurrence of individual nonverbal behaviors is typically rather low, which requires the use of fairly large time intervals. In contrast, the frame rate of the body motion recordings is typically high which enables analysis of mimicry at a much more finer time scale. To make sure only the pose, not global position in space, is taken into account, the poses of both subjects are normalized using equation 4. As interacting subjects typically face each other, poses are also normalized for rotation using equations 5 and 6. This ensures that both subjects have similar positions and facing directions (cf. Figure 4). Additionally, one might left-right mirror the pose of one of the subjects, to make the direction of the movements of both subjects similar. As both subjects are rotated to face the positive x-axis, the z-values of all joints j in J of one of the subjects can be negated as follows: $x_{jz} = -x_{jz}$.

Mimicry can also be operationalized using body motion by analyzing the velocity of the body. This representation mitigates the effect of inter-personal differences in body sizes to some extent. To this end, the values for each joint along each of the axes are converted to z-scores as follows: $x'^{k}_{i} = \frac{(x'^{k}_{i} - \mu)}{\sigma}$ with frame $k \in \{1 \dots m\}$ and m the total number of measurements, and μ and δ the mean and standard deviation over the sequence of measurements, respectively. Once these steps have been taken, the screened data of both subjects can be compared using windowed cross-correlation (e.g., Boker, Xu, Rotondo, and King (2002)) or based on spectral methods (e.g., Oullier, de Guzman, Jantzen, Lagarde, and Kelso (2007)) where distances between poses can be calculated using equation 7.

Conclusions and Future Research

We have introduced Automated Measurement and Analysis of Body motion (AMAB), a methodology for the study into nonverbal behavior measured with full-body motion tracking devices. The increasingly wider availability and applicability of these devices provide opportunities for psychologists working on nonverbal behavior. AMAB is the first standardized methodology that addresses the automatic measurement and analysis of full-body motion for a broad range of applications and research questions.

In this paper, we discussed the recording and representation of full-body motion and provided a sequence of steps to screen the data and eliminate nuisance factors. Common dependent variables have been introduced and it is explained how to derive these from the motion representations. Finally, we discussed four case studies to exemplify how the AMAB methodology can boost nonverbal behavior research, by both improving existing research and opening up new research opportunities. These demonstrated the method’s efficacy to analyze human full-body motion across time, space, body parts and people. We have shown that this methodology can be applied in a wide range of research areas, including rehabilitation, sports performance and the nonverbal components of interpersonal interaction research.

While this paper focused on hypothesis-driven research questions, AMAB can also be used in explorative, data-driven research. The automatic measurement and subsequent screening of the data provide an excellent starting point for the construction of features to be used for statistical analysis, including pattern recognition approaches (see e.g. Kleinsmith et al. (2011); Krishnan et al. (2009)).

There are several extensions possible in the development of the AMAB methodology. First, we plan to use AMAB as a basis for the representation and analysis of the body motion of groups of people. These analyses go beyond the pairwise comparisons presented in the current paper and will provide opportunities to objectively study group behavior, e.g. in recognizing threatening situations in public spaces. Second, inter-personal differences in body movement are currently not explicitly addressed. Dealing with different body sizes and variations in the amount and type of movements could lead to a notion of baseline behavior, which is instrumental in many research questions, especially in one-on-one interviewing situations. We intend to explore this avenue of research, which further increases the applicability of AMAB.

References

Altorfer, A., Jossen, S., Wurmle, O., Kasermann, M-L., Foppa, K., & Zimmermann, H. (2000). Measurement and meaning of head movements in everyday face-to-face communicative interaction. *Behavior Research Methods, Instruments, & Computers, 32,* 17-32.

Bailenson, J.N., Blascovich, J., Beall, A.C., & Loomis, J.M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin, 29,* 819-833.

Bente, G., Senokozlieva, M., Pennig, S., Al-Issa, A., & Fischer, O. (2008). Deciphering the secret code: A new methodology for the cross-cultural analysis of nonverbal behavior. *Behavior Research Methods, 40,* 269-277.

Bezodis, N.E., Salo, A.I.T., & Trewartha, G. (2010). Choice of sprint start performance measure affects the performance-based ranking within a group of sprinters: Which is the most appropriate measure? *Sports Biomechanics, 9*, 258-269.

Boker, S.M., Xu, M., Rotondo, J.L., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series, *Psychological Methods, 7*, 338-355.

Chartrand, T.L., & Bargh, J.A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893-910.

Chartrand, T.L., & Lakin, J.L. (2013). The antecedents and consequences of human behavioral mimicry, *Annual Review of Psychology, 64*, 285-308.

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior, 28*, 117-139.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). The body action and posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior, 26,* 97-121.

Dakin, C.J., Luu, B.L., van den Doel, K., Inglis, J.T. & Blouin, J. 2010). Frequency-specific modulation of vestibular-evoked sway responses in humans. *Journal of Neurophysiology, 103*, 1048-1056.

Dotsch, R., & Wigboldus, D.H.J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology, 44*, 1194-1198.

Ekman, P. (1965). Communication through nonverbal behavior: A source of information about an interpersonal relationship. In S.S. Tomkins & C. Izard (Eds.), Affect, cognition, and personality (pp. 390-442). New York: Springer.

Hall, E.T., & Whyte, W.F. (1963). Intercultural communication: A guide to men of action. *Practical Anthropology, 10,* 216-299.

Hayduk, L.A. (1978). Personal space: An evaluative and orienting overview. *Psychological Bulletin, 85*, 117-134.

Hayduk, L. A. (1983). Personal space: Where we now stand. *Psychological Bulletin, 94,* 293-335.

James, W.T. (1932). A study of the expression of bodily posture. *Journal of General Psychology, 7*, 405-437.

Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A. (2011). Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B, 41*, 1027-1038.

Krishnan, N.C., Juillard, C., Colbry, D., & Panchanathan, S. (2009). Recognition of hand movements using wearable accelerometers. *Journal of Ambient Intelligence and Smart Environments, 1*, 143-155.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 41,* 841-849.

Mead, R., Atrash, A., & Mataric, M.J. (2013). Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction, *International Journal of Social Robotics*, to appear.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350-371.

Oullier, O., de Guzman, G.C., Jantzen, K.J., Lagarde, J., & Kelso, J.A.S. (2007). Social coordination dynamics: Measuring human bonding, *Social Neuroscience, 3*, 178-192.

Poppe, R. (2007). Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding, 108*, 4-18.

Scherer, K.R., & Ekman, P. (1982). Methodological issues in studying nonverbal behavior. In K.R. Scherer & P. Ekman (Eds.), Handbook of methods in nonverbal behavior research (pp. 45-135). New York: Cambridge University Press.

Slawinski, J., Dumas, R., Cheze, L., Ontanon, G., Miller, C., & Mazure–Bonnefoy, A. (2013). Effect of postural changes on 3D joint angular velocity during starting block phase. *Journal of Sports Sciences*, to appear.

Stel, M., van Dijk, E., & Olivier, E. (2009). You want to know the truth? Then don't mimic! *Psychological Science, 20*, 693-699.

Tiedens, L.Z., & Fragale, A.R. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology, 84,* 558-568.

Table 1

Overview of body motion measuring devices

Device Characteristics		Example Devices	Example Studies
Full-Body	Type		
Yes	Marker-based	Vicon MX, MotionAnalysis Raptor, Advanced Realtime Tracking ARTTRACK, Optitrack Arena, PhaseSpace Impuls X2, Phoenix Technologies Inc. Visualeyez, Qualisys Oqus	Slawinski et al., 2013
Yes	Inertial-based	Animazoo IGS, Ascension MotionStar, Xsens MVN	Kleinsmith et al., 2011; Krishnan etal., 2009
Yes	Vision-based	Microsoft Kinect, Ipi Soft, Organic Motion Openstage	Mead et al., 2013
No	Inertial-based	Ascension TrakSTAR, Polhemus Liberty Latus, Sparkfun Electronics Witilt	Bailenson et al., 2003; Dotsch & Wigboldus, 2008

Figure 1 Human body representation (left) and kinematic tree (right). Best viewed in color.

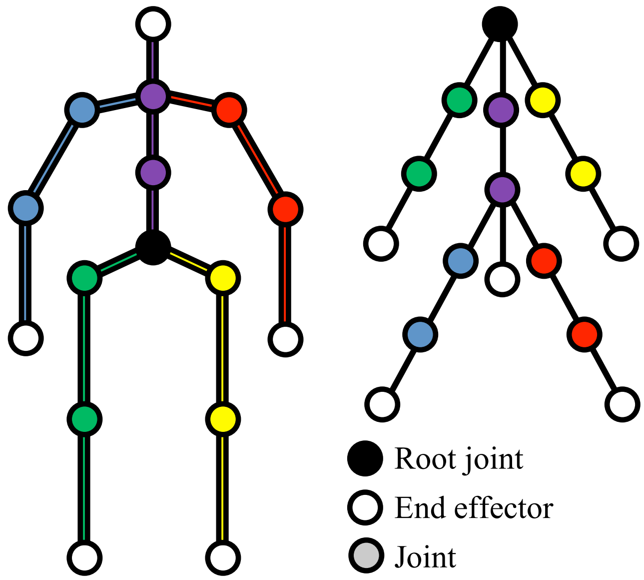

Figure 2 Example of median filtering.

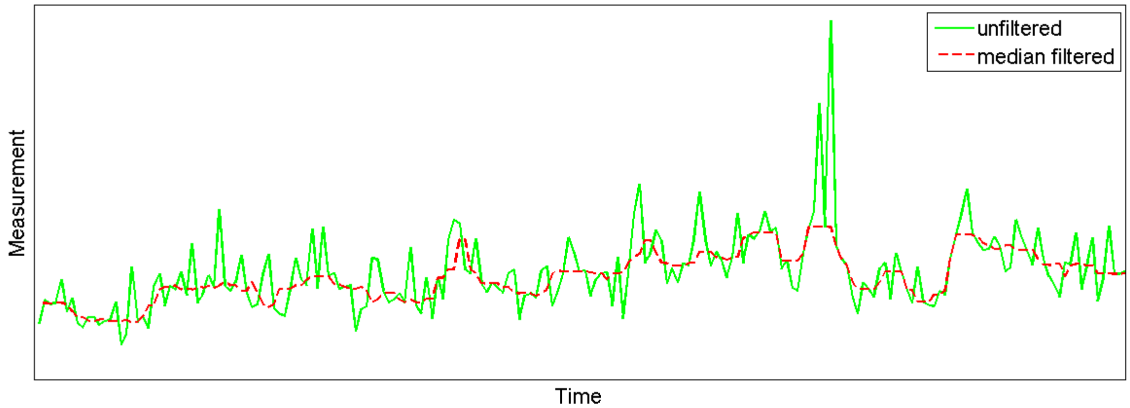

Figure 3 Example of the temporal alignment of two sequences.

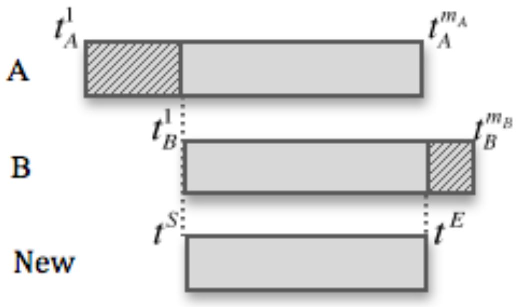

Figure 4 Example of position normalization (I) and orientation normalization (II).

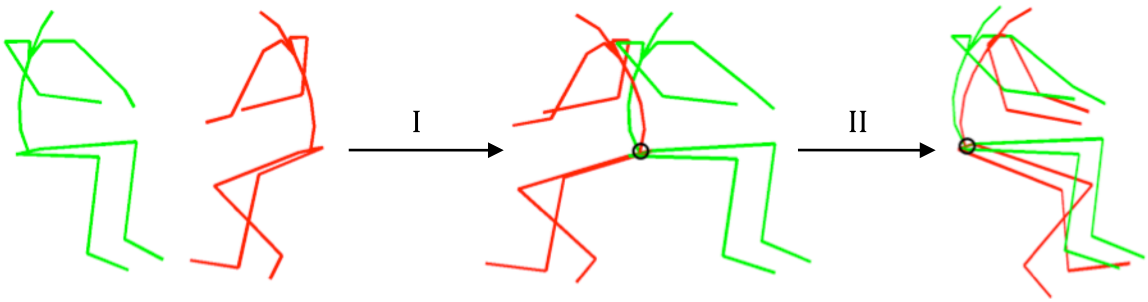

Figure 5 Example of position and orientation normalization. The sums of pairwise joint distances (in meter) are shown over time.