DHQ: Digital Humanities Quarterly
2017
Volume 11 Number 2

Automated Pattern Analysis in Gesture Research: Similarity Measuring in 3D Motion Capture Models of Communicative Action

Daniel Schüller <schueller_at_humtec_dot_rwth-aachen_dot_de>, Natural Media Lab, Human Technology Centre, RWTH Aachen University
Christian Beecks <christian_dot_beecks_at_uni-muenster_dot_de>, University of Münster
Marwan Hassani <m_dot_hassani_at_tue_dot_nl>, Data Management and Exploration Group, RWTH Aachen University
Jennifer Hinnell <hinnell_at_ualberta_dot_ca>, Department of Linguistics, University of Alberta
Bela Brenger <brenger_at_rwth-aachen_dot_de>, Natural Media Lab, Human Technology Centre, RWTH Aachen University
Thomas Seidl <seidl_at_dbs_dot_ifi_dot_lmu_dot_de>, Ludwig Maximilian University of Munich
Irene Mittelberg <mittelberg_at_humtec_dot_rwth-aachen_dot_de>, Natural Media Lab, Human Technology Centre, RWTH Aachen University

Abstract

The question of how to model similarity between gestures plays an important role in current studies in the domain of human communication. Most research into recurrent patterns in co-verbal gestures – manual communicative movements emerging spontaneously during conversation – is driven by qualitative analyses relying on observational comparisons between gestures. Due to the fact that these kinds of gestures are not bound to well-formedness conditions, however, we propose a quantitative approach consisting of a distance-based similarity model for gestures recorded and represented in motion capture data streams. To this end, we model gestures by flexible feature representations, namely gesture signatures, which are then compared via signature-based distance functions such as the Earth Mover's Distance and the Signature Quadratic Form Distance. Experiments on real conversational motion capture data evidence the appropriateness of the proposed approaches in terms of their accuracy and efficiency. Our contribution to gesture similarity research and gesture data analysis allows for new quantitative methods of identifying patterns of gestural movements in human face-to-face interaction, i.e., in complex multimodal data sets.

Introduction

Given the central place of the *embodied mind* in experientialist approaches to language, co-verbal gestures have become a valuable data source in cognitive, functional and anthropological linguistics (e.g. [Sweetser 2007]). While there exist various views on embodiment, the core idea is that human higher cognitive abilities are shaped by the morphology of our bodies and the way we interact with the material, spatial and social environment (e.g. [Gibbs 2006]; [Johnson 1987]). Drawing on these premises, some gesture scholars stress that gestures are conditioned by the forms and affordances of their material habitat as well as the speakers' interactive and collaborative practices (e.g. [Enfield 2009]; [Streeck 2011]). Pioneering work done by Kendon (e.g. [Kendon 1972], [Kendon 2004]), McNeill (e.g. [McNeill 1985], [McNeill 1992], [McNeill 2005]) and Müller [Müller 1998] has shown that manual gestures are an integral part of utterance formation and communicative interaction. The state-of-the-art of research in the growing interdisciplinary field of *gesture studies* has recently been presented in the *International Handbook on Multimodality in Human Interaction* ([Müller 2013], Vol. 1 2013, [Müller 2014] Vol. 2 2014). One quintessence to be drawn from this large body of work is that language, whether spoken or signed, is embodied, dynamic, multimodal and intersubjective (see also [Duncan 2007]; [Gibbs 2006]; [Jäger 2004]; [Mittelberg 2013]; [Müller 2008]). 1

Indeed, human communication typically involves multiple modalities such as vocalizations, spoken or signed discourse, manual gestures, eye gaze, body posture and facial expressions. In face-to-face communication, manual gestures play 2

an important role by conveying meaningful information and guiding the interlocutors' attention to objects and persons talked about. Gestures here are understood as spontaneously emerging, dynamic configurations and movements of the speakers' hands and arms that contribute to the communicative content and partake in the interactive organization of a spoken dialogue situation (e.g. [Bavelas 1992]; [Kendon 2004]; [McNeill 1992]; [Müller 1998]; [Mittelberg 2016]). The contribution of gestures to multimodal interaction may consist in, e.g., deictic reference to locations, ideas, persons or things (both abstract and concrete); they may also fulfill metalinguistic functions, e.g., in referring to citations of other speakers or outlining the structure of their argumentation. Gestures may also provide schematic, iconic portrayals of actions, things or spatial constellations [Mitteberg 2014]; [Rieser 2012]. As we use the term here, gestures are not to be confused with emblems (e.g. the victory sign), which have a culturally defined form-meaning correlation in the same sense that words may have a fixed meaning [McNeill 1992]. So gestures are always to be investigated in view of the co-occurring speech with which they jointly create (new) semiotic material and support speakers in organizing their thoughts or drawing connections to their social and physical environment [Streeck 2011]; [Mittelberg 2016].

3 Drawing on this large body of gesture research across various fields of the humanities and social sciences, the interdisciplinary approach presented here aims at identifying and visualizing patterns of gestural behavior with the help of custom-tailored computational tools and methods. Although co-speech gestures tend to be regarded as highly idiosyncratic in respect to their spontaneous individual articulation by speakers in spoken dialogue situations, it is safe to assume that there are recurring forms of dynamic hand configurations and movement patterns which are performed by speakers sharing the same cultural background. From this assumption follows the hypothesis that, on the one hand, a general degree of similarity between gestural forms may be presumed – trivially – due to the shared morphology of the human body (e.g. [Tomasello 1999]). On the other hand, the specific cultural context plays an important role in both language acquisition and the adoption of culture-specific behavioral patterns (see, e.g. [Bourdieu 1987] on Habitus and Hexis). Moreover, gesture has also been ascribed a constitutive role regarding the human capacity for language both ontogenetically and phylogenetically (e.g. [Goldin-Meadow 2003]; [McNeill 2012]). Previous empirical gesture research indeed shows that co-verbal gestures exhibit recurrent form features and movement patterns, as well as recurring form-meaning pairings: see, for instance, Bressem [Bressem 2013] on form features; Fricke [Fricke 2010] on "kinaesthemes"; Kendon [Kendon 2004] on "locution clusters", McNeill [McNeill 1992], [McNeill 2000] on cross-linguistic path and manner imagery portraying motion events; McNeill [McNeill 2005] on "catchments"; Ladewig [Ladewig 2011] and Müller [Müller 2010] on "recurring gestures", and Cienki [Cienki 2005] and Mittelberg [Mittelberg 2010] on image-schematic patterns. From this perspective, a question central to gesture research concerns the factors that may motivate and lend a certain systematicity to forms and functions of human communicative behaviors not constituting an independent sign system in the Saussurian sense as spoken and signed languages do. Manual gesture is a semiotically versatile medium, for it may, depending on a given context, assume more or less language-like functions: from accompanying spoken discourse in the form of pointing, accentuating beats, schematic iconicity or managing social interaction to carrying the full load of language (e.g. [Goldin-Meadow 2007]).

4 In this paper, we will focus on certain kinds of co-verbal gestures, i.e. specific image-schematic gestalts, e.g. spirals, circles, and straight paths [Cienki 2013], [Mittelberg 2010]. Figure 1 shows visualizations of several different gestural traces belonging to these movement types. In what follows, we will present a novel method designed to identify, search and cluster in an automatized fashion gestural movement patterns throughout our data set, and potentially also any other motion capture data set, e.g. to recognize and group those traces that are similar regarding their formal features and the way in which they unfold in gesture space.

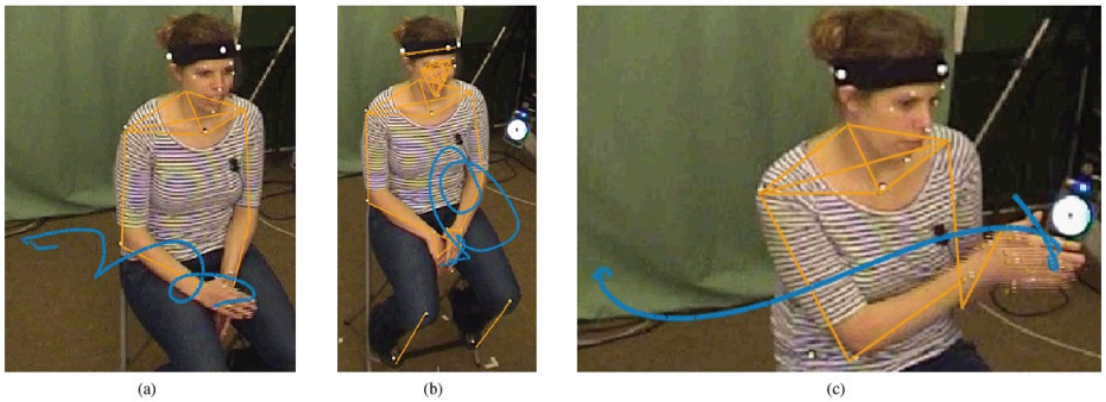

Fig. 1. Three example gestures of different movement types: (a) gesture of type spiral, (b) gesture of type circle, and (c) gesture of type straight. Blue trajectories indicate the main movement of the relevant markers.

Figure 1.

Research Objective

Whereas the gesture research discussed above mostly relies on observational methods and qualitative video analyses, our aim is to add to the catalogue of methods for empirical linguistics and gesture studies by outlining a computational, quantitative and comparative 3D-model driven approach in gesture research. While there is a trend to combine qualitative with quantitative as well as experimental methods in multimodal communication research [Gonzalez-Marquez 2007]; [McNeill 2001]; [Müller 2013], standardized and widely applicable tools and methods still need to be developed. In order to derive statistically significant patterns from aligned linguistic and behavioral data, some recent research initiatives have started to compile and work with larger-scale corpora, e.g. drawing on technological advances in data management and automated analyses (e.g. the Little Red Hen Project, [Steen 2013]). While using audio and video technology to record co-speech gestures remains the dominant way to construct multimodal corpora, some research teams have begun to employ 3D motion capture technology to overcome the limits of 2D video, e.g. Lu and Huenerfauth [Lu 2010] for signed language, and Beskow et al. [Beskow 2011] and Pfeiffer et al. [Pfeiffer 2013a] for spoken discourse (see also [Pfeiffer 2013] for an overview). Our contribution is to obtain some kind of numerical instrument for graduating gestural similarity for measuring gesture similarity in sets of recorded behavioral data. While this instrument, in its current state, in no way addresses the problem of the meaning of certain gestural forms, it is the first step towards a model of measuring similarity between recurring dynamic gesture form patterns. Our goal is to first establish a robust, flexible and automated methodology which allows us to determine

5

1. whether there are shared or common reoccurring gestural movement patterns in a given set of 3D recorded, behavioral communication data,
2. exactly which forms there are, and
3. the extent to which they occur,

and then to apply this methodology to the recorded 3D numerical MoCap data of a group of participants.

Both the alignment of gestures with the co-occurring speech, and the semantic comparison of the established (formally) sufficiently similar gesture-speech constructions, still have to be done manually by human gesture researchers, through semiotic analyses of the multimodal, speech and behavioral data corpora. The primary aim of developing an automated indicator of gesture similarity is to identify recurrent movement patterns of interest from the recorded 3D corpus data computationally, and thus to enable human gesture researchers to handle these data sets in a more efficient manner. In order to make gesture similarity automatically accessible, we propose a distance-based similarity model for gestures arising in three-dimensional motion capture data streams. In comparison to two-dimensional video capture technology, working with numerical three-dimensional motion capture technology has the advantage of measuring and visualizing the temporal and spatial dynamics of otherwise invisible movement traces with the highest possible accuracy. We aim at maintaining this accuracy by aggregating movement traces, also called trajectories, into a *gesture signature* [Beecks 2015]. This gesture signature has the ability of weighting trajectories according to their relevance. Based on this

6

lossless feature representation, we propose to measure similarity by means of distance-based approaches [Beecks 2013], [Beecks 2010]. We particularly investigate the *Earth Mover's Distance* [Rubner 2000] and the *Signature Quadratic Form Distance* [Beecks 2010] for the comparison of two gesture signatures on real conversational motion capture data.

Properties of the 3D Data Model

From a philosophy of science point of view, before being able to apply computational algorithms to naturalistic real-world gestures, there must be a translation from the real-world dialogue situations, involving people speaking and gesturing, from which data are captured, to a computable set of data. For this purpose, a marker-based *Vicon* Motion Capture system was used in this study. Participants wear a series of markers attached to predetermined body parts of interest (fingers, wrists, elbows, neck, head, etc.). The *Vicon* system automatically generates a chart of numerical 4-tuples of Euclidean space-time coordinates for each marker attached to these points on the participants' bodies. The movement of the markers is tracked by 14 *Vicon* infrared cameras, and the physical trajectories of the markers are represented in a chart of space-time coordinates. These space-time charts form the data sets that are investigated algorithmically, relieving the gesture analyst of the difficult, and subjective, task of manually examining highly ephemeral real-world dialogue situations. But what are the crucial features that such a numerical representation must have in order to enable researchers to not only investigate a model but also to finally derive statements and theories about a modeled real-world situation? We address the following research questions: Which logical features of the model are essential if one wants to investigate the real world by investigating a model? And secondly, what are the epistemic benefits of investigating models instead of real-world situations? 7

The most important feature is that the model *represents* its representandum. The representandum in question is a set of relevant features and relations of a given part of reality, namely the change in space-time location of certain body parts caused by the participants' kinetic movement. The representing model, on the other hand, is the virtual, computational 3D recording of the real-world kinetic movement, mapped onto Euclidean space. Representation itself is the relation which holds between the model and its representandum. Representation, as we understand the term here, is a non-reflexive and non-symmetric relation, which simply means that an *a* does not necessarily represent *a*, and that if *a* represents *b*, then *b* does not necessarily represent *a*. We further assume that representation depends on transitive relations, such as identity between some complex relational features of the entities to be represented in a model and the entities that represent/model them. In short, this means that if there are entities *x, y, z,* there must be at least one mutual complex relational feature *Fr(x,y,z),* which these entities have in common. The transitive relation *R*, then, is the "identity" relation, in that two entities have an identical relation to a third entity – a frame of reference, or a *tertium comparationis*. 8

Definition: Transitivity 9

For a binary relation *R* over a set *A* consisting of *x, y, z,* the following statement holds true: 10

$$\forall x,y,z \in A: xRy \,\&\, yRz \rightarrow xRz$$

Example 1.

The transitivity of *R* is so important here because, in terms of modeling, it is the crucial feature that R must have. The different relata involved are: movements of body parts *(x)*, movements of markers *(y)*, and computational trajectories *(z)*. The relation R which holds between these relata is the identity of their curve with respect to space – either to a given virtual Euclidian space, or the physical space (which functions as a reference frame). The identity of the movement curve of body parts and the movement curve of markers simply stems from their physical attachment/ conjunction in physical space. The identity of marker movements and the computational trajectories is a result of metering the markers' light reflections, by means of the 14 Vicon infrared cameras, and numerically mapping the outcome onto Euclidean space coordinates. But if one differentiates between physical and Euclidean space, there is hardly any identity one could honestly speak of in this case. In what sense could physical and virtual movement curves ever be identical? Only in the sense that we *identify* physical and Euclidean space by conventions of scientific modeling and metering practices; the term *curve* is itself an indication of how familiar that convention is. Since Euclidean space is a 11

conventionalized and well-accepted geometrical model of physical space – we are well accustomed to talking about physical movement in terms of, distances, trajectories, vectors, miles, kilometres and change in space/time coordinates etc. – one can say that Euclidean space is our familiar standard model for describing our perception of movement in physical space, both in everyday conversation and scientific discussion. Thus, it is justified to speak of the *identity* of the spatial curves of trajectories in Euclidean space and those of moving physical objects such as Motion Capture markers, only if we accept this convention. Regarding representation, what does this imply? If representation depends on the transitivity of relation R holding between the entities *x, y, z,* and the identity of that relation depends on conventions of metering and modeling, representation additionally depends on following conventions. Given a 4-tuple of coordinates of a MoCap marker, the movement of this marker is modeled by a vector which points from tuple-1 to tuple-2 to tuple-n. The crucial feature of this kind of modeling is that the movement of a single marker *a* is represented as a dynamic space-time trajectory which aggregates the consecutively changing coordinates of *a* over a given time frame.

Regarding the above-mentioned definition of transitivity, let our variables take the following values: 12

x = movement of body part from position a to b; 13
y = movement of marker M from position a to b;
z = trajectory of marker M

Given these values, we outline the transitivity relation as follows: 14

> ∀x,y,z ∈ A: xRy & yRz → xRz: x [movement of body part from position a to b] **R** y [movement of marker M from position a to b] & y [movement of marker M from position a to b] **R** z [trajectory of marker M] → x [movement of body part from position a to b] **R** z [trajectory of marker M].

Example 2.

This means that if *x, y, z* obtain a transitive relation *R* in the above sense, then *x, y,* and *z* are to be regarded as homomorphous abstract *concepts* that all denote the same event of spatiotemporal movement, and *R* is an equivalence relation. So the *extensional* equivalence of these concepts is a necessary and sufficient condition that allows us to investigate reality by investigating the model: If a language and its translation are equivalent, it should be equally valid to investigate one or the other. However, since the translation of the concept "*movement* of marker" into "*aggregated* marker *coordinates*" fails to be an *intensionally* adequate translation, i.e. the concepts do not *mean* the same (event vs. aggregated states of affairs), it could at first seem odd to describe real movement in terms of trajectories. But, since the concepts are at least phenomenologically and extensionally equivalent, this basically remains a question of the interpreter's ontology (see [Quine 1980]) and how the final research result is to be formulated. If we decide to treat "event" and "aggregated states of affairs" as being synonymous, the problem completely disappears. Otherwise, we have to re-translate the problematic concept into one which suits our needs. In terms of epistemic benefits, one major advantage of the proposed distance-based gesture-similarity model (see the following section), i.e. the combination of gesture signatures with signature-based distance functions, is its applicability to any type of gestural pattern and to data sets of any size. In fact, distance-based similarity models can be utilized in order to model similarity between gestural patterns whose movement types are well known and between gestural patterns whose inherent structures are completely unknown. In this way, they provide an unsupervised way of modeling gesture similarity. This flexibility is attributable to the fact that the proposed approaches are model independent, i.e. no complex gesture model has to be learned in a comprehensive training phase prior to indexing and query processing. Another advantage of the proposed distance-based gesture-similarity model is the possibility of efficient query processing. Although calculating the distance between two gesture signatures is a computationally expensive task, which results in at least a quadratic computation time complexity with respect to the number of relevant trajectories, many approaches such as the independent minimization lower bound of the Earth Mover's Distance on feature signatures [Uysal 2014] and metric indexing [Beecks 2011], as well as the Ptolemaic indexing [Hetland 2013] of the Signature Quadratic Form Distance, are available for efficient query processing and, thus, for assessing gesture similarity in a larger quantitative way. 15

Modeling Gesture Similarity

In this section, we present a distance-based similarity model for the comparison of gestures within three-dimensional motion capture data streams. To this end, we first introduce *gesture signatures* as a formal model of gestures arising in motion capture data streams. Since gesture signatures comprise multiple three-dimensional trajectories, we continue with outlining distance functions for trajectories before we investigate distance functions applicable to gesture signatures. 16

Gesture Signatures

Motion capture data streams can be thought of as sequences of points in a three-dimensional Euclidean space. In the scope of this work, these points arise from several reflective markers which are attached to the body and in particular to the hands of a participant. The motion of the markers is triangulated via multiple cameras and finally recorded every 10 milliseconds. In this way, each marker defines a finite trajectory of points in a three-dimensional space. The formal definition of a trajectory is given below. 17

Definition: Trajectory 18

Given a three-dimensional feature space $\mathbb{R}^3$, a trajectory $t:\{1,\ldots,n\}\rightarrow \mathbb{R}^3$ is defined for all $1\leq i\leq n$ as: 19

$$t(i)=(x_i,y_i,z_i)$$

Example 3.

A trajectory describes the motion of a single marker in a three-dimensional space. It is worth noting that the time information is abstracted to integral numbers in order to model trajectories arising from different time intervals. Since a gesture typically arises from multiple markers within a certain period of time, we aggregate several trajectories including their individual relevance by means of a gesture signature. For this purpose, we denote the set of all finite trajectories as trajectory space $\mathbb{T}=\bigcup_{k\in\mathbb{N}}\{t|\ t:\{1,\ldots,k\}\rightarrow \mathbb{R}^3\}$, which is time-invariant, and define a gesture signature as a function from the trajectory space $\mathbb{T}$ into the real numbers $\mathbb{R}$. The formal definition of a gesture signature is given below. 20

Definition: Gesture Signature 21

Let $\mathbb{T}$ be a trajectory space. A *gesture signature* $S\in\mathbb{R}^{\mathbb{T}}$ is defined as: 22

$$S:\mathbb{T}\rightarrow \mathbb{R} \text{ subject to } |S^{-1}(\mathbb{R}\setminus\{0\})|<\infty$$

Example 4.

A gesture signature formalizes a gesture by assigning a finite number of trajectories non-zero weights reflecting their importance. Negative weights are immaterial in practice but ensure the gesture space $\mathbb{S}=\{S\in\mathbb{R}^{\mathbb{T}}\wedge|S^{-1}(\mathbb{R}\setminus\{0\})|<\infty\}$ forms a vector space. While a weight of zero indicates insignificance of a trajectory, a positive weight is utilized to indicate contribution to the corresponding gesture. In this way, a gesture signature allows us to focus on the trajectories arising from those markers which actually form a gesture. For example, if a gesture is expressed by the participant's hands, only the corresponding hand markers and thus trajectories have to be weighted positively. 23

A gesture signature defines a generic mathematical model but omits a concrete functional implementation. In fact, given a subset of relevant trajectories $\mathcal{T}^+\subset\mathbb{T}$, the most naive way of defining a gesture signature S consists in assigning relevant trajectories a weight of one and irrelevant trajectories a weight of zero, i.e. by defining S for all $t\in\mathbb{T}$ as follows: 24

$$S(t) = \begin{cases} 1, & if\ t \in \mathcal{T}^+ \\ 0, & otherwise \end{cases}$$

Figure 2.

The isotropic behavior of this approach, however, completely ignores the inherent characteristics of the relevant trajectories. We therefore weight each relevant trajectory according to its inherent properties of *motion distance* and *motion variance*. These properties are defined below.

Definition: Motion Distance and Motion Variance

Let T be a trajectory space and t:{1,…,n}→ R^3 be a trajectory. The *motion distance* m_δ:T→R of trajectory t is defined as:

$$m_\delta(t) = \sum_{i=1}^{n-1} \|t(i) - t(i+1)\|_2$$

Figure 3.

The motion variance m_{σ^2}:T→R of trajectory t is defined with mean

$$m_{\sigma^2}(t) = \frac{1}{n} \cdot \sum_{i=1}^{n} \|t(i) - \mu(t)\|_2^2$$

Figure 4.

as:

$$S_{m_\delta}(t) = \begin{cases} m_\delta(t), & if\ t \in \mathcal{T}^+ \\ 0, & otherwise \end{cases}$$

Figure 5.

The intuition behind motion distance and motion variance is to take into account the overall movement and vividness of a trajectory. The higher these qualities, the more information the trajectory may contain and vice versa. Their utilization with respect to a set of relevant trajectories finally leads to the definitions of a *motion distance gesture signature* and a *motion variance gesture signature*, as shown below.

Definition: Motion Distance Gesture Signature and Motion Variance Gesture Signature

Let T be a trajectory space and $\mathcal{T}^+ \subset T$ be a subset of relevant trajectories. A motion distance gesture signature $S_{m_\delta} \in R^T$ is defined for all $t \in T$ as:

$$S_{m_{\sigma^2}}(t) = \begin{cases} m_{\sigma^2}(t), & if\ t \in \mathcal{T}^+ \\ 0, & otherwise \end{cases}$$

Figure 6.

A motion variance gesture signature $S_{m_{\sigma^2}} \in R^T$ is defined for all $t \in T$ as:

$$DTW_\delta(t_n, t_m) = \delta(t_n(n), t_m(m)) + \min \begin{cases} DTW_\delta(t_{n-1}, t_{m-1}) \\ DTW_\delta(t_n, t_{m-1}) \\ DTW_\delta(t_{n-1}, t_m) \end{cases}$$

Figure 7.

Motion distance and motion variance gesture signatures are able to reflect the characteristics of the expressed gestures with respect to the corresponding relevant trajectories by adapting the number and weighting of relevant trajectories. As a consequence, the computation of a (dis)similarity value between gesture signatures is frequently based on the (dis)similarity values among the involved trajectories in the trajectory space. We thus outline applicable trajectory distance functions in the following section.

Trajectory Distance Functions

Due to the nature of trajectories whose inherent properties are rarely expressible in a single figure, trajectories are frequently compared by aligning their coincident similar points with each other. A prominent example is the *Dynamic Time Warping Distance*, which was first introduced in the field of speech recognition by Itakura [Itakura 1975] and Sakoe and Chiba [Sakoe 1978] and later brought to the domain of pattern detection in databases by Berndt and Clifford [Berndt 1994]. The idea of this distance is to locally replicate points of the trajectories in order to fit the trajectories to each other. The point-wise distances finally yield the Dynamic Time Warping Distance, whose formal definition is given below.

Definition: Dynamic Time Warping Distance

Let $t_n: \{1,\ldots,n\} \rightarrow R^3$ and $t_m: \{1,\ldots,m\} \rightarrow R^3$ be two trajectories from T and $\delta: R^3 \times R^3 \rightarrow R$ be a distance function. The *Dynamic Time Warping Distance* $DTW_\delta: T \times T \rightarrow R$ between t_n and t_m is recursively defined as:

$$DTW_\delta(t_0, t_0) = 0$$
$$DTW_\delta(t_i, t_0) = \infty \quad \forall 1 \leq i \leq n$$
$$DTW_\delta(t_0, t_j) = \infty \quad \forall 1 \leq j \leq m$$

Figure 8.

with

$$EMD_\delta(S_1, S_2) = \min_F \left\{ \frac{\sum_{t \in \mathbb{T}} \sum_{t' \in \mathbb{T}} \delta(t, t') \cdot f(t, t')}{\min\{\sum_{t \in \mathbb{T}} S_1(t), \sum_{t' \in \mathbb{T}} S_2(t')\}} \right\}$$

Figure 9.

As can be seen in the definition above, the Dynamic Time Warping Distance is defined recursively by minimizing the distances *δ* between replicated elements of the trajectories. In this way, the distance *δ* assesses the spatial proximity of two points while the Dynamic Time Warping Distance preserves their temporal order within the trajectories. By utilizing Dynamic Programming, the computation time complexity of the Dynamic Time Warping Distance lies in $\mathcal{O}$(n·m).

Although there exist further approaches for the comparison of trajectories, such as *Edit Distance on Real Sequences* [Chen 2005], *Minimal Variance Matching* [Latecki 2005], and *Mutual Nearest Point Distance* [Fang 2009], we have decided to utilize the Dynamic Time Warping Distance for the following reasons: (i) The distance value is based on all points of the trajectories with respect to their temporal order and is not attributed to partial characteristics of the trajectories, (ii) it provides the ability of exact indexing by lower bounding [Keogh 2002], and (iii) it indicates superior quality in terms of accuracy within preliminary investigations.

Given a ground distance in the trajectory space T, we will show in the following section how to lift this ground distance to the gesture space $S \subset R^T$ in order to compare gesture signatures with each other.

Gesture Signature Distance Functions

Gesture signatures can differ in size and length, i.e., in the number of relevant trajectories and in the lengths of those trajectories. In order to quantify the distance between differently structured gesture signatures, we apply signature-based distance functions [Beecks 2013], [Beecks 2010]. In this paper, we focus on those signature-based distance functions that consider the entire structure of two gesture signatures in order not to favor partial similarity between short and long gesture signatures. For this reason, we investigate the transformation-based *Earth Mover's Distance* [Rubner 2000] and the correlation-based *Signature Quadratic Form Distance* [Beecks 2010] in the remainder of this section.

The Earth Mover's Distance, whose name was inspired by Stolfi and his vivid description of the transportation problem, which he likened to finding the minimal cost to move a total amount of earth from earth hills into holes [Rubner 2000], has been originated in the computer vision domain. It defines the distance between two gesture signatures by measuring the cost of transforming one gesture signature into another one. The formal definition of the Earth Mover's Distance is given below.

Definition: Earth Mover's Distance

Let $S_1, S_2 \in S$ be two gesture signatures and $\delta: T \times T \rightarrow R$ be a trajectory distance function. The *Earth Mover's Distance* $EMD_\delta: S \times S \rightarrow R$ between S_1 and S_2 is defined as a minimum cost flow of all possible flows $F = \{f \mid f: T \times T \rightarrow R\}$ as:

- $\forall t, t' \in \mathbb{T}; f(t, t') \geq 0$
- $\forall t \in \mathbb{T}; \sum_{t' \in \mathbb{T}} f(t, t') \leq S_1(t)$
- $\forall t' \in \mathbb{T}; \sum_{t \in \mathbb{T}} f(t, t') \leq S_2(t')$
- $\sum_{t \in \mathbb{T}} \sum_{t' \in \mathbb{T}} f(t, t') = \min\{\sum_{t \in \mathbb{T}} S_1(t), \sum_{t' \in \mathbb{T}} S_2(t')\}$

Figure 10.

subject to the constraints:

$$\langle S_1, S_2 \rangle_s = \sum_{t \in \mathbb{T}} \sum_{t' \in \mathbb{T}} S_1(t) \cdot S_2(t') \cdot s(t, t')$$

Figure 11.

As can be seen in the definition above, the Earth Mover's Distance between two gesture signatures is defined as a linear optimization problem subject to non-negative flows which do not exceed the corresponding limitations given by the weights of the trajectories of both gesture signatures. The computation of the Earth Mover's Distance can be restricted to the relevant trajectories of both gesture signatures and follows a specific variant of the simplex algorithm [Hillier 1990].

The idea of the Signature Quadratic Form Distance consists in adapting the generic concept of *correlation* to gesture signatures. In general, correlation is the most basic measure of bivariate relationship between two variables [Rodgers 1988] and can be interpreted as the amount of variance these variables share [Rovine 1997]. In order to apply the concept of correlation to gesture signatures, all trajectories and corresponding weights are related with each other based on a trajectory similarity function $s: T \times T \rightarrow R$. The resulting *similarity correlation* between two gesture signatures $S_1, S_2 \in S$ is then defined as:

$$SQFD_s(S_1, S_2) = \sqrt{\langle S_1, S_1 \rangle_s - 2 \cdot \langle S_1, S_2 \rangle_s + \langle S_2, S_2 \rangle_s}$$

Figure 12.

The similarity correlation between two gesture signatures finally leads to the definition of the Signature Quadratic Form Distance, as shown below.

Definition: Signature Quadratic Form Distance

Let $S_1, S_2 \in S$ be two gesture signatures and $s: T \times T \rightarrow R$ be a trajectory similarity function. The *Signature Quadratic Form*

Distance $SQFD_s$:S×S→R between S_1 and S_2 is defined as:

$$SQFD_s(S_1, S_2) = \sqrt{\langle S_1, S_1 \rangle_s - 2 \cdot \langle S_1, S_2 \rangle_s + \langle S_2, S_2 \rangle_s}$$

Figure 13.

The Signature Quadratic Form Distance is defined by adding the intra-similarity correlations $<S_1,S_1>_s$ and $<S_2,S_2>_s$ of the gesture signatures S_1 and S_2 and subtracting their inter-similarity correlation $<S_1,S_2>_s$. The smaller the differences among the intra-similarity and inter-similarity correlations the lower the resulting Signature Quadratic Form Distance, and vice versa. The computation of the Signature Quadratic Form Distance can be restricted to the relevant trajectories of both gesture signatures and has a quadratic computation time complexity with respect to the number of relevant trajectories. 52

More details regarding the Earth Mover's Distance and the Signature Quadratic Form Distance as well as possible similarity functions can be found for instance in the PhD thesis of Beecks [Beecks 2013b]. Among other approaches, such as the matching-based Signature Matching Distance [Beecks 2013], the aforementioned signature-based distance functions have been shown to balance the trade-off between retrieval accuracy and query processing efficiency. Before we investigate their performance in the context of gesture signatures, we devote the next section to a discussion about the properties of the proposed distance-based gesture similarity model. 53

Experimental Evaluation

Evaluating the performance of distance-based similarity models is a highly empirical discipline. It is nearly unforeseeable which approach will provide the best retrieval performance in terms of accuracy. To this end, we qualitatively evaluated the proposed distance-based approaches to gesture similarity by using a natural media corpus of motion capture data collected for this project. This dataset comprises three-dimensional motion capture data streams arising from eight participants during a guided conversation. The participants were equipped with a multitude of reflective markers which were attached to the body and in particular to the hands. The motion of the markers was tracked optically via cameras at a frequency of 100 Hz. In the scope of this work, we used the right wrist marker and two markers attached to the right thumb and right index finger each. The gestures arising within the conversation were classified by domain experts according to the following types of movement: spiral, circle, and straight. Example gestures of these movement types are sketched in Figure 1. A total of 20 gesture signatures containing five trajectories each was obtained from the motion capture data streams. The trajectories of the gesture signatures have been normalized to the interval $[0,1]^3 \in R^3$ in order to maintain translation invariance. 54

The resulting distance matrices between all gesture signatures with respect to the Earth Mover's Distance and the Signature Quadratic Form Distance are shown in Figure 2 and Figure 3, respectively. As described in the previous Section, we utilized the Dynamic Time Warping Distance based on Euclidean Distance as trajectory distance for the Earth Mover's Distance and converted this trajectory distance by means of the power kernel [Schölkopf 2001] with parameter α=1 into a trajectory similarity function for the Signature Quadratic Form Distance. Since weighting of relevant trajectories by motion distance and motion variance, approximately shows a similar behavior, we include the results regarding motion variance gesture signatures only. We depict small and large distance values by bluish and reddish colors in order to visually indicate the performance of our proposal: gesture signatures from the same movement type should result in bluish colors while gesture signatures from different movement types should result in reddish colors. 55

As can be seen in Figure 2 and Figure 3, both Earth Mover's Distance and Signature Quadratic Form Distance show the same tendency in terms of gestural dissimilarity. Although distance values computed through the aforementioned 56

distance functions have different orders of magnitude, both gesture signature distance functions are generally able to distinguish gesture signatures from different movement types. On average, gesture signatures belonging to the same movement type are less dissimilar to each other than gesture signatures from different movement types. We further observed that the distinction between gesture signatures from the movement types spiral and straight are most challenging. This is caused by a similar sequence of movement of these two gestural types. While gesture signatures belonging to the movement type straight follow a certain direction, e.g., movement on the horizontal axis, gesture signatures from the movement type spiral additionally oscillate with respect to a certain direction. Since this oscillation can be dominated by the movement direction, the underlying trajectory distance functions are often unable to distinguish oscillating from non-oscillating trajectories and thus gesture signature of movement type spiral from those of movement type straight.

Apart from the quality of accuracy, efficiency is another important aspect when evaluating the performance of gesture similarity models. For this purpose, we measured the computation times needed to perform single distance computations on a single-core 3.4 GHz machine. We implemented the proposed distance-based approaches in Java 1.7. The Earth Mover's Distance, which needs on average 148.6 milliseconds for a single distance computation, is approximately three times faster than the Signature Quadratic Form Distance, which needs on average 479.8 milliseconds for a single distance computation. In spite of the theoretically exponential and empirically super-cubic computation time complexity of the Earth Mover's Distance [Uysal 2014], this distance is able to outperform the Signature Quadratic Form Distance. The reason for this is the high number of computationally expensive trajectory distance computations. While the computation of the Earth Mover's Distance is carried out on the trajectory distances between the two gesture signatures, the computation of the Signature Quadratic Form Distance additionally takes into account the trajectory distances within both gesture signatures. Therefore the number of trajectory distance computations is significantly higher for the Signature Quadratic Form Distance than for the Earth Mover's Distance.

To sum up, the experimental evaluation reveals that the proposed distance-based approaches are able to model gesture similarity in a flexible and model-independent way. Without the need for a preceding training phase, the Earth Mover's Distance and the Signature Quadratic Form Distance are able to provide similarity models for searching similar gestures which are formalized through gesture signatures.

		spirals					circles								straights					
spirals	0	49	50	35	60	50	103	93	92	69	48	51	72	51	57	35	46	22	53	60
	49	0	18	29	43	39	105	84	79	72	56	75	49	41	41	52	37	38	40	32
	50	18	0	22	37	27	102	78	54	46	41	55	37	27	28	60	30	37	29	35
	35	29	22	0	22	24	126	94	94	51	37	40	46	17	15	41	13	29	19	31
	60	43	37	22	0	47	168	123	153	98	64	77	71	36	30	45	23	42	30	37
	50	39	27	24	47	0	97	73	47	23	22	31	26	19	28	75	27	39	26	57
circles	103	105	102	126	168	97	0	13	77	105	72	103	74	122	134	163	135	66	136	153
	93	84	78	94	123	73	13	0	71	96	63	98	50	95	102	155	101	60	99	133
	92	79	54	94	153	47	77	71	0	47	32	72	23	73	94	136	102	71	109	135
	69	72	46	51	98	23	105	96	47	0	25	44	30	36	49	101	52	66	58	96
	48	56	41	37	64	22	72	63	32	25	0	26	27	32	41	80	39	36	42	72
	51	75	55	40	77	31	103	98	72	44	26	0	51	39	50	77	44	45	37	76
	72	49	37	46	71	26	74	50	23	30	27	51	0	40	48	107	47	48	52	87
	51	41	27	17	36	19	122	95	73	36	32	39	40	0	14	69	16	39	22	51
straights	57	41	28	15	30	28	134	102	94	49	41	50	48	14	0	65	12	37	23	53
	35	52	60	41	45	75	163	155	136	101	80	77	107	69	65	0	49	38	69	59
	46	37	30	13	23	27	135	101	102	52	39	44	47	16	12	49	0	32	18	40
	22	38	37	29	42	39	66	60	71	66	36	45	48	39	37	38	32	0	39	56
	53	40	29	19	30	26	136	99	109	58	42	37	52	22	23	69	18	39	0	32
	60	32	35	31	37	57	153	133	135	96	72	76	87	51	53	59	40	56	32	0

Fig. 2. Distance matrix for the Earth Mover's Distance based on motion variance gesture signatures with respect to different movement types. Bluish and reddish colors indicate small and large distance values, respectively.

Figure 14.

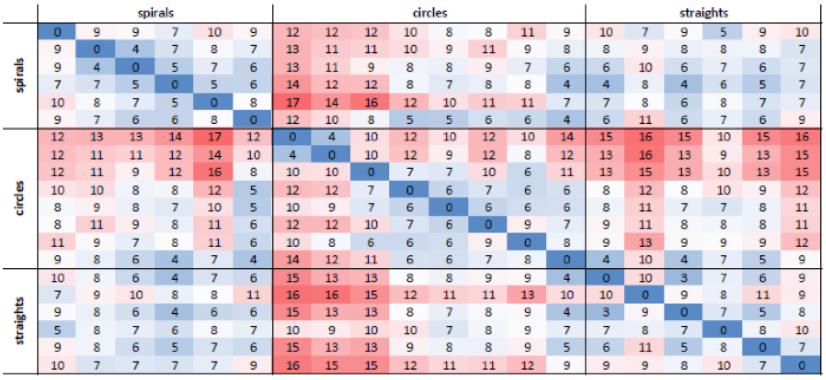

Fig. 3. Distance matrix for the Signature Quadratic Form Distance based on motion variance gesture signatures with respect to different movement types. Bluish and reddish colors indicate small and large distance values, respectively.

Figure 15.

Conclusions and Future Work

In this paper, we have investigated distance-based approaches to measure similarity between gestures arising in three-dimensional motion capture data streams. To this end, we have explicated gesture signatures as a way of aggregating the inherent characteristics of spontaneously produced co-speech gestures and signature-based distance functions such as the Earth Mover's Distance and the Signature Quadratic Form Distance in order to quantify dissimilarity between gesture signatures. The experiments conducted on real data are evidence of the appropriateness in terms of accuracy and efficiency of the proposal.

In future work, we intend to extend our research on gesture similarity towards indexing and efficient query processing. While the focus of the present paper lies on dissimilarity between pairs of gestures, we further plan to quantitatively analyze motion capture data streams in a query-driven way in order to support the domain experts' qualitative analyses of gestural patterns within multi-media contexts. The overall goal of this research is to contribute to the advancement of automated methods of pattern recognition in gesture research by enhancing qualitative analyses of complex multimodal data in the humanities and social sciences. While this paper focuses on formal features of the gestural movements, further steps will entail examining the semantic and pragmatic dimensions of these patterns in light of the cultural contexts and embodied semiotic practices they emerge from.

Acknowledgment

This work is partially funded by the Excellence Initiative of the German federal and state governments and DFG grant SE 1039/7-1. This work extends [Beecks 2015].

Works Cited

Bavelas 1992 Bavelas, J. B., Chovil, N., Lawrie, D. A. and Wade, A. “Interactive Gestures”, *Discourse Processes*, 15 (1992): 469-489.

Beecks 2010 Beecks, C., M.S. Uysal, and Seidl, T. “A Comparative Study of Similarity Measures for Content-Based Multimedia Retrieval”. *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 1552-1557 (2010).

Beecks 2010a Beecks, C., M.S. Uysal, and Seidl, T. “ Signature Quadratic Form Distance”. *Proceedings of the 9th International Conference on Image and Video Retrieval*, pp. 438-445 (2010).

Beecks 2011 Beecks, C., Lokoc, J., Seidl, T., and Skopal, T. “Indexing the signature quadratic form distance for efficient content-based multimedia retrieval”. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, 2011, pp. 24:1–8.

Beecks 2013 Beecks, C., S. Kirchhoff, and Seidl, T. “On Stability of Signature-Based Similarity Measures for Content-Based Image Retrieval”. *Multimedia Tools and Applications*, pp. 1-14.

Beecks 2013a Beecks, C., S. Kirchhoff, and Seidl, T. “Signature matching distance for content-based image retrieval”. *Proceedings of the ACM International Conference on Multimedia Retrieval*, pp. 41-48 (2013)

Beecks 2013b Beecks, C. *Distance-Based Similarity Models for Content-Based Multimedia Retrieval*. Ph.D. thesis, RWTH Aachen University (2013). Available online: http://darwin.bth.rwth-aachen.de/opus3/volltexte/2013/4807/

Beecks 2015 Beecks, C., Hassani, M., Hinnell, J., Schüller, D., Brenger, B., Mittelberg, I. and Seidl, T. “Spatiotemporal Similarity Search in 3D Motion Capture Gesture Streams”. *Proceedings of the 14th International Symposium on Spatial and Temporal Databases*, pp. 355-372 (2015).

Beecks 2015a Beecks, C., J. Lokoc, T. Seidl, and Skopal, T. “Indexing the Signature Quadratic Form Distance for Efficient Content-Based Multimedia Retrieval”. *Proceedings of the ACM International Conference on Multimedia Retrieval*, pp. 24:1-8 (2015).

Berndt 1994 Berndt, D. and Clifford, J. “Using dynamic time warping to find patterns in time series”. *AAAI94 Workshop on Knowledge Discovery in Databases*, pp. 359-370 (1994)

Beskow 2011 Beskow, J., Alex, S., Al Moubayed, S., Edlund, J. and House, D. “Kinetic Data for Large-Scale Analysis and Modeling of Face-to-Face Conversation”. In *Proceedings of the International Conference on Audio-Visual Speech Processing*, Stockholm, pp. 103–106 (2011).

Bourdieu 1987 Bourdieu, P. Sozialer Sinn. *Kritik der theoretischen Vernunft*. Suhrkamp, Frankfurt am Main (1987).

Bressem 2013 Bressem, J. “A Linguistic Perspective on the Notation of Form Features in Gestures”. In Müller, C. et al. (eds): *Body – Language – Communication. An International Handbook of Multimodality in Human Interaction*. (HSK 38.1). De Gruyter Mouton, Berlin/Boston,pp. 1079–1098 (2013).

Chen 2005 Chen, L., Özsu, M.T. and Oria, V. “Robust and Fast Similarity Search for Moving Object Trajectories”. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 491-502 (2005).

Cienki 2005 Cienki, A. “Image Schemas and Gesture”. In Hampe, B. (ed). From *Perception to Meaning: Image Schemas in Cognitive Linguistics*. Mouton de Gruyter, Berlin/New York (2005).

Cienki 2013 Cienki, A. “Cognitive Linguistics: Spoken Language and Gesture as Expressions of Conceptualization”. In Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D. and Teßendorf, S. (eds), *Body– Language–Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1. Mouton de Gruyter, Berlin/New York, pp. 182–201 (2013).

Duncan 2007 Duncan, S., Cassell, J. and Levy E.T. (eds), *Gesture and the Dynamic Dimension of Language*. John Benjamins, Amsterdam/ Philadelphia, pp. 269–283 (2007).

Enfield 2009 Enfield, N. *The Anatomy of Meaning. Speech, Gestures, and Composite Utterances*. Cambridge University Press, Cambridge (2009).

Fang 2009 Fang, S. and Chan, H. “Human Identification by Quantifying Similarity and Dissimilarity in Electrocardiogram Phase Space”. *Pattern Recognition*, vol. 42, 9 (2009): 1824-1831.

Fricke 2010 Fricke, E. “Phonaestheme, Kinaestheme und Multimodale Grammatik”. *Sprache und Literatur*, 41 (2010): 69–88.

Gibbs 2006 Gibbs, R. W., Jr. *Embodiment and Cognitive Science*. Cambridge University Press, Cambridge (2006).

Goldin-Meadow 2003 Goldin-Meadow, S. *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press, Cambridge, MA (2003).

Goldin-Meadow 2007 Goldin-Meadow, S. “Gesture with Speech and Without It”. In Duncan, S. (ed), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 31–49 (2007).

Gonzalez-Marquez 2007 Gonzalez-Marquez, M., Mittelberg, I., Coulson, S. and Spivey, M.J. (eds). *Methods in Cognitive Linguistics*. John Benjamins, Amsterdam/Philadelphia (2007).

Hetland 2013 Hetland, M.L., Skopal, T., Lokoc, J. and Beecks, C. “Ptolemaic Access Methods: Challenging the Reign of the Metric Space Model”. *Information Systems*, vol.38, no.7 (2013): 989-1006.

Hillier 1990 Hillier, F. and Lieberman, G.. *Introduction to Linear Programming*. McGraw-Hill.(1990)

Itakura 1975 Itakura, F. “Minimum Prediction Residual Principle Applied to Speech Recognition”. *IEEE Transactions on*

Acoustics, Speech and Signal Processing, vol.23, no.1 (1975): 67-72.

Johnson 1987 Johnson, M. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press, Chicago (1987).

Jäger 2004 Jäger, L. & Linz, E. (eds) *Medialität und Mentalität. Theoretische und empirische Studien zum Verhältnis von Sprache, Subjektivität und Kognition*. Fink Verlag, München (2004).

Kendon 1972 Kendon, A. “Some Relationships between Body Motion and Speech. An analysis of an example”. In Siegman, A. and Pope, B. (eds), *Studies in Dyadic Communication*. Pergamon Press, Elmsford, NY, pp. 177–210 (1972).

Kendon 2004 Kendon, A. *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge (2004).

Keogh 2002 Keogh, E. “Exact Indexing of Dynamic Time Warping”. *Proceedings of 28th International Conference on Very Large Data Bases*, 406-417 (2002).

Ladewig 2011 Ladewig, S. “Putting the cyclic gesture on a cognitive basis”. *CogniTextes*, 6 (2011).

Latecki 2005 Latecki, L.J., Megalooikonomou, V., Wang, Q., Lakämper, R., Ratanamahatana, C.A. and Keogh, E.J. “Elastic Partial Matching of Time Series.Knowledge Discovery in Databases”, *9th European Conference on Principles and Practice of Knowledge Discovery in Databases, Lecture Notes in Computer Science*, vol. 3721 (2005), Springer, pp. 577-584.

Lu 2010 Lu, P. and Huenerfauth, M. “Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research”. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*. Los Angeles, pp. 89–97 (2010).

McNeill 1985 McNeill, D. “So You Think Gestures are Nonverbal?”. *Psychological Review*, 92,3 (1985): 350–371.

McNeill 1992 McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*. Chicago University Press, Chicago (1992).

McNeill 2000 McNeill, D. (ed). *Language and Gesture*. Cambridge University Press, Cambridge (2000).

McNeill 2001 McNeill, D., Quek, F., McCullough, K.-E., Duncan, S., Furuyama, N., Bryll, R., Ma, X.-F. and Ansari, R. 2001. “Catchments, Prosody and Discourse”, *Gesture* 1, 1 (2001): 9–33.

McNeill 2005 McNeill, D. Gesture and Thought. Chicago University Press, Chicago (2005).

McNeill 2012 McNeill, D. *How Language Began: Gesture and Speech in Human Evolution*. Cambridge University Press, Cambridge (2012).

Mitteberg 2014 Mittelberg, I. and Waugh, L.R. “Gestures and Metonymy”. In Müller, C., Cienki, A., Fricke, E., Ladewig, S.H., McNeill, D. and Bressem, J. (eds), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. Vol. 2. Handbooks of Linguistics and Communcation Science. De Gruyter Mouton, Berlin/Boston, pp. 1747–1766 (2014).

Mittelberg 2010 Mittelberg, I. “Geometric and Image-Schematic Patterns in Gesture Space”. In Evans, V. and Chilton, P. (eds), *Language, Cognition, and Space: The State of the Art and New Directions*. Equinox, London, pp., 351–385 (2010).

Mittelberg 2010a Mittelberg, I. “Interne und externe Metonymie: Jakobsonsche Kontiguitätsbeziehungen in redebegleitenden Gesten”. *Sprache und Literatur* 41,1 (2010): 112–143.

Mittelberg 2013 Mittelberg, I. “Balancing Acts: Image Schemas and Force Dynamics as Experiential Essence in Pictures by Paul Klee and their Gestural Enactments”. In B. Dancygier, M. Bokrent and J. Hinnell (eds), *Language and the Creative Mind*. Stanford: Center for the Study of Language and Information, pp. 325–346.

Mittelberg 2013a Mittelberg, I. “The Exbodied Mind: Cognitive-Semiotic Principles as Motivating Forces in Gesture”. In Müller, C., Cienki, A., Fricke, E., Ladewig, S.H., McNeill, D. and Teßendorf, S. (eds). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*. Handbooks of Linguistics and Communication Science (38.1). Mouton de Gruyter,Berlin/New York, pp. 750-779 (2013).

Mittelberg 2016 Mittelberg, I., Schüller, D. “Kulturwissenschaftliche Orientierung in der Gestenforschung”. In Jäger, L., Holly, W., Krapp, P., Weber, S. (eds.). *Language – Culture – Communication. An International Handbook of Linguistics as Cultural Study*. Mouton de Gruyter, Berlin/New York, pp. 879-890 (2016).

Müller 1998 Müller, C. Redebegleitende Gesten. *Kulturgeschichte - Theorie - Sprachvergleich*. Berlin Verlag A. Spitz, Berlin (1998).

Müller 2008 Müller, C. *Metaphors Dead and Alive, Sleeping and Waking. A Dynamic View*. University of Chicago Press, Chicago (2008).

Müller 2010 Müller, C. “Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive”, *Sprache und Literatur* 41 (2010): 37–68.

Müller 2013 Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D. and Teßendorf, S. (eds). *Body– Language–Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1. Mouton de Gruyter, Berlin/New York (2013).

Müller 2014 Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D. and Bressem, J.(eds). *Body– Language–Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2. Mouton de Gruyter, Berlin/New York (2014).

Pfeiffer 2013 Pfeiffer, T. “Documentation with Motion Capture”. In Müller, C., Cienki, A., Fricke, E., Ladewig, S.H., McNeill, D. and Teßendorf, S. (eds). *Body- Language-Communication: An International Hand- book on Multimodality in Human Interaction*, Hand- books of Linguistics and Communication Science. Mouton de Gruyter, Berlin, New York (2013).

Pfeiffer 2013a Pfeiffer, T., Hofmann, F., Hahn, F., Rieser, H., and Röpke, I. “Gesture Semantics Reconstruction Based on Motion Capturing and Complex Event Processing: A Circular Shape Example”. In M. Eskenazi, M. Strube, B. Di Eugenio, and J. D. Williams (eds). *Proceedings of the Special Interest Group on Discourse and Dialog (SIGDIAL) 2013 Conference*, pp. 270–279 (2013).

Quine 1980 Quine, W.V.O. *Wort und Gegenstand*. Reclam, Stuttgart (1980).

Rieser 2012 Rieser H., Bergmann K. and Kopp, S. “How do Iconic Gestures Convey Visuo-Spatial Information? Bringing Together Empirical, Theoretical, and Simulation Studies.” In Efthimiou, E. and Kouroupetroglou, G. (eds). *Gestures in Embodied Communication and Human-Computer Interaction*. Springer, Berlin/Heidelberg (2012).

Rodgers 1988 Rodgers, J. and Nicewander, W. *Thirteen Ways to Look at the Correlation Coefficient*. American Statistician, pp. 59-66 (1988).

Rovine 1997 Rovine, M. and Von Eye, A. “A 14th Way to Look at a Correlation Coefficient: Correlation as the Poportion of Matches”. *American Statistician*, pp. 42-46 (1997)

Rubner 2000 Rubner, Y., C. Tomasi, and Guibas, L.J. “The Earth Mover's Distance as a Metric for Image Retrieval”, *International Journal of Computer Vision*, vol. 40,2 (2000): 99-121.

Sakoe 1978 Sakoe, H. and Chiba, S. “Dynamic Programming Algorithm Optimization for Spoken Word Recognition”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, 1(1978): 43-49.

Schölkopf 2001 Schölkopf, B. “The Kernel Trick for Distances”. *Advances in Neural Information Processing Systems*, 301-307.

Steen 2013 Steen, L. and Turner, M. “Multimodal Construction Grammar”. In Dancygier, B., Bokrent, M. and Hinnell, J. (eds). *Language and the Creative Mind, Stanford: Center for the Study of Language and Information* (2013).

Streeck 2011 Streeck, J., Goodwin, C. and LeBaron, C.D. *Embodied Interaction: Language and Body in the Material World: Learning in doing: Social, Cognitive and Computational Perspectives*. Cambridge University Press, New York (2011).

Sweetser 2007 Sweetser, E. “Looking at Space to Study Mental Spaces: Co-speech Gesture as a Crucial Data Source in Cognitive Linguistics”. In Gonzalez-Marquez, M., Mittelberg, I., Coulson, S. and Spivey, M. (eds). *Methods in Cognitive Linguistics*. John Benjamins, Amsterdam/Philadelphia, pp. 201¬224 (2007).

Tomasello 1999 Tomasello, M. *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, MA (1999).

Uysal 2014 Uysal, M.S., Beecks, C., Schmücking, J., and Seidl, T. “Efficient Filter Approximation using the Earth Mover's Distance in Very Large Multimedia Databases with Feature Signatures”. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 979-988 (2014).