Behavior Research Methods (2024) 56:7498–7542
https://doi.org/10.3758/s13428-024-02433-0

ORIGINAL MANUSCRIPT

# EMOKINE: A software package and computational framework for scaling up the creation of highly controlled emotional full-body movement datasets

**Julia F. Christensen[1] · Andrés Fernández[2,3] · Rebecca A. Smith[4] · Georgios Michalareas[1] · Sina H. N. Yazdi[5] · Fahima Farahi[5] · Eva-Madeleine Schmidt[1,6,11] · Nasimeh Bahmanian[7,8] · Gemma Roig[9,10]**

Accepted: 15 April 2024 / Published online: 25 June 2024
© The Author(s) 2024

## Abstract

EMOKINE is a software package and dataset creation suite for emotional full-body movement research in experimental psychology, affective neuroscience, and computer vision. A computational framework, comprehensive instructions, a pilot dataset, observer ratings, and kinematic feature extraction code are provided to facilitate future dataset creations at scale. In addition, the EMOKINE framework outlines how complex sequences of movements may advance emotion research. Traditionally, often emotional-'action'-based stimuli are used in such research, like hand-waving or walking motions. Here instead, a pilot dataset is provided with short dance choreographies, repeated several times by a dancer who expressed different emotional intentions at each repetition: anger, contentment, fear, joy, neutrality, and sadness. The dataset was simultaneously filmed professionally, and recorded using XSENS® motion capture technology (17 sensors, 240 frames/second). Thirty-two statistics from 12 kinematic features were extracted offline, for the first time in one single dataset: speed, acceleration, angular speed, angular acceleration, limb contraction, distance to center of mass, quantity of motion, dimensionless jerk (integral), head angle (with regards to vertical axis and to back), and space (convex hull 2D and 3D). Average, median absolute deviation (MAD), and maximum value were computed as applicable. The EMOKINE software is appliable to other motion-capture systems and is openly available on the Zenodo Repository. Releases on GitHub include: (i) the code to extract the 32 statistics, (ii) a rigging plugin for Python for MVNX file-conversion to Blender format (MVNX=output file XSENS® system), and (iii) a Python-script-powered custom software to assist with blurring faces; latter two under GPLv3 licenses.

**Keywords** Emotion · Motion capture · Computer vision · Affective neuroscience · Aesthetics · Dance · Dataset · Open science

JFC and AF contributed equally to the article (shared first authorship).

✉ Julia F. Christensen
julia.christensen@ae.mpg.de

[1] Department of Cognitive Neuropsychology, Max Planck Institute for Empirical Aesthetics, Frankfurt/M, Germany

[2] Methods of Machine Learning, University of Tübingen, Tübingen, Germany

[3] International Max Planck Research School for Intelligent Systems, Tübingen, Germany

[4] Department of Psychology, University of Glasgow, Glasgow, Scotland

[5] WiseWorld.AI, Porto, Portugal

[6] Max Planck School of Cognition, Leipzig, Germany

[7] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt/M, Germany

[8] Department of Modern Languages, Goethe University, Frankfurt/M, Germany

[9] Computer Science Department, Goethe University, Frankfurt/M, Germany

[10] The Hessian Center for Artificial Intelligence (hessian.AI), Darmstadt, Germany

[11] Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany

## Summary & background

### Summary

EMOKINE is a software and dataset creation framework for highly controlled kinematic datasets of emotionally expressive full-body movements. The primary novelty of this framework is that it provides the research community with a set of 12 readily computed kinematic features (32 statistics in total) that can be used out-of-the-box by researchers in order to model emotional expressivity in full-body movement. For the first time, these 12 features are presented together: speed, acceleration, angular speed, angular acceleration, limb contraction, distance to center of mass, quantity of motion, dimensionless jerk (integral), head angle (with regards to vertical axis and to back), and space (convex hull 2D and 3D). A pilot dataset accompanies this article, of realistic full-body movement stimuli, which have been designed and instantiated with all other parameters controlled so that most of the variability stems from the different expressed emotions. The pilot data for EMOKINE were recorded via the XSENS® system, however, the software is also appliable to data obtained from other motion-capture systems with minimal to no changes.

Based on the creation procedure of this pilot dataset, we describe a process by which such a dataset creation can be scaled up in the future, while ensuring mandatory levels of experimental control that are key for datasets to be used in experimental psychology and affective neuroscience with human participants (Christensen & Calvo-Merino, 2013; Christensen & Jola, 2015), and ensuring the technical detail required for datasets in computer vision and related disciplines. Importantly, instead of following the model of traditional datasets in emotion science that comprise emotional-'action'-based stimuli, like hand-waving or walking motions, the EMOKINE pilot dataset contains complex sequences of movements: 6-s-long dance choreographies. A dancer choreographed a series of short, simple ballet-movement-inspired dance sequences of approximately 6 s in length, which corresponds to eight counts in dance notation. Then she performed the sequences six times each, expressing a different emotional intention at each repetition – namely anger, contentment, fear, joy, neutrality, and sadness. Classically, most datasets include only the 'basic' emotions proposed by Paul Ekman and colleagues (Ekman, 1973/2015; Ekman & Friesen, 1971) – namely, anger, fear, joy, neutrality, and sadness. We extended the spectrum of expressed emotions in the EMOKINE pilot dataset by also adding the emotion 'contentment', which is another positive-valence emotion like joy, but of low arousal; symmetrical to what anger (negative valence, high arousal) is to sadness (negative valence, low arousal).

The EMOKINE suite includes:

(a) a detailed step-by-step procedure guide to create EMOKINE datasets at scale;
(b) a pilot dataset (recorded with the XSENS® system) with four different visual presentations for each video stimulus – namely (i) avatars, (ii) full-light displays (FLDs) with blurred face, (iii) point-light displays (PLDs), and (iv) silhouettes;
(c) the raw kinematic data for each stimulus of the pilot dataset, obtained via the XSENS® motion capture (LINK) system (via 17 sensors distributed on the body, recorded at 240 frames/second);
(d) the code to obtain 32 statistics from the 12 kinematic features;
(e) human observers' emotion recognition rates and value judgments, which confirm the intended emotional categories of the pilot dataset.

The pilot dataset of EMOKINE is available for download on the Zenodo repository and the software on GitHub. Releases on GitHub include:

(a) an extensive repository of code to extract the 32 statistics of the 12 kinematic features – namely, speed, acceleration, angular speed, angular acceleration, limb contraction, distance to center of mass, quantity of motion, dimensionless jerk (integral), head angle (with regards to vertical axis and to back), and space (convex hull 2D and 3D). Average, median absolute deviation (MAD) and maximum value were computed for each;
(b) A MVNX rigging plugin for Python that allows Blender to convert MVNX files to a Blender-friendly format (MVNX = output file of the motion capture XSENS system);
(c) Python-script-powered custom software to assist with the process of blurring faces. The latter two have been released under GPLv3 licenses, and all are available for download on GitHub (see data availability statement, Section 11).

The GitHub readme file includes an explanation on how to apply the EMOKINE software to data obtained from other systems. In particular, points (a) and (c) can be applied to any data including keypoint positions with minimal to no change. Point (b) naturally depends on the MVNX format (which is given by the XSENS system), but it can be ignored for other formats.

### Background

Investigating emotion recognition competence is important due to its significance for psychosocial functioning (Cosmides & Tooby, 2000; Darwin, 1872/2009; Ekman, 1973/2015;

Ekman & Friesen, 1971). In the fields of emotion psychology, affective neuroscience, and computer vision, such research often relies on datasets that comprise stimuli of pictures or videos of facial and bodily expressions of emotions. Yet, compared to the extensive existing research on the recognition of emotions from facial expressions (Byron et al., 2007; Elfenbein & Ambady, 2002; O'Boyle Jr et al., 2011; Rosete & Ciarrochi, 2005; Rubin et al., 2005; Scherer & Scherer, 2011; Walter et al., 2012; Zuskin et al., 2007), the bodily channel of emotional expression has received less empirical attention, despite important calls to extend emotion perception research to this domain (Aviezer et al., 2012; Bellot et al., 2021; de Gelder, 2006, 2009; Keck et al., 2022; McCarty et al., 2017; Vaessen et al., 2018). Less stimulus materials are available, and available datasets suffer from limitations (discussed in, e.g., Christensen & Calvo-Merino, 2013; Christensen & Jola, 2015; Smith & Cross, 2022). Besides, most full-body datasets of emotional expressions show actors performing different emotional actions; for instance, punching a fist in anger, sinking to the floor in sadness, jumping with joy (Atkinson et al., 2004; Crane & Gross, 2013; Dael et al., 2012; Gross et al., 2010). However, this approach likely measures the ability to recognize familiar prototypical *actions* indicative of different emotions, rather than perceptual sensitivity to emotional expressions (Shafir, 2016; Shafir et al., 2013). In order to avoid such confounding effects and investigate the manifestation of different emotions in one same human movement, a same-sequence approach can be a valid alternative, where a set of different movements are repeated several times, with each repetition corresponding to the expression of a different emotion. In other words, the expressor in every repetition performs the exact same reference movements, with the only source of intentional variability in the kinematics being due to the intended expression of a different emotion.

Therefore, walking, pointing, drinking, knocking, or throwing movements have recently been proposed as a valid movement alternative for emotion recognition and emotion perception research (Crane & Gross, 2007; Dekeyser et al., 2002; Heberlein et al., 2004; Krüger et al., 2018; Ma et al., 2006; Pollick et al., 2001; Roether et al., 2009; Vanrie & Verfaillie, 2004). Yet, such movements are rather simple and may be confounded with stereotypical assumptions about these movements. For example, in everyday life, walking patterns are typically associated with how much someone is rushing or with the existence of injuries. With the objective of reducing contextual cues from movement stimulus materials, and movement towards more varied patterns of movements, dance movements have recently been proposed as stimulus materials. Dance is, par excellence, an instance of emotionally expressive full-body movement (Boone and Cunningham, 1998; Boone and Cunningham, 2001; Christensen et al., 2019; Kirsch et al., 2016; Dittrich et al., 1996; Orgs et al., 2016; Van Dyck et al., 2017; Van Dyck et al., 2013; Van Meel et al., 1993). Dance choreographies can be created to be more varied than simple walking or throwing movements. Besides, compared to walking and throwing movements, using choreographies that are novel to participants in emotion perception research reduces possible familiarity effects. Besides, professional dancers are trained to express different emotional states with one same dance gesture and can therefore serve as models for stimulus materials (Christensen et al., 2019; Karin, 2016; Karin et al., 2016). It is relevant to note that we are not proposing that same-sequence stimuli should completely replace emotional-action-based stimuli in emotion research. We propose that these options are alternative stimuli materials and should be chosen depending on the research question. Here we focus on the advantage of same-sequence stimuli materials.

This *same-sequence-different*-emotional expressivity in dance movements is comparable to language. The same sentence can be spoken to sound either angry or happy to a listener, depending on how the expressor modulates their voice with their breathing and the muscles of their vocal tract (Bänziger et al., 2009; Scherer & Scherer, 2011; Scherer et al., 2017). How a dancer performs one same dance movement sequence, at several repetitions, with different emotional intentions, has previously been found to convey these intended emotional states to human observers, even to those without dance experience (Christensen et al., 2017; Christensen et al., 2019; Christensen et al., 2014; Christensen et al., 2023). In classical emotion-recognition tasks with dance movements, average recognition rates are, generally, above chance level, and vary between 18.04% (Christensen et al., major revisions), and 48% (Christensen et al., 2023; Smith & Cross, 2022).

With the advances of filming and motion-capture technologies of the past decades, new horizons have opened up for the digitization and analysis of full-body movement.

By combining these recent technologies with the same-sequence-different-emotion approach, the EMOKINE framework offers a novel route for emotion perception research with full-body movement. The stimuli design is based on previous datasets of dance movements that did, however, not include motion-capture technology (Christensen et al., 2023; Christensen et al., 2017; Christensen et al., major revisions; Christensen et al., 2019; Christensen et al., 2014; Smith & Cross, 2022).

## Objectives

We had four objectives with EMOKINE:

(1) To create a *pilot* dataset of simple dance movement sequences, with each sequence performed by a dancer six times, each with a different intended emotion out of a pool of six possible emotions – namely anger, contentment,

fear, joy, neutrality, and sadness. The novelty of this work is that this same-sequence–different-emotion experimental design was complemented by time-resolved, whole-body kinematics data, measured by motion-capture technology. Please note that this dataset was created to enable the development of the EMOKINE suite and contains portrayals from only one dancer. To ensure generalizability in future research using the EMOKINE software, datasets should include portrayals from more than one dancer.

(2) To render these pilot stimuli in four visual presentations for further study: (i) videos showing an avatar performing the movements, extracted from the XSENS propriety software (avatar videos); (ii) videos showing the dancer in full light, but with blurred face, to avoid emotion recognition from the face (full-light displays; FLDs); (iii) videos showing point-light displays (PLDs), which have been rendered with a Blender-based algorithm (Blender Community, 2018); and (iv) videos showing black-and-white silhouettes, where the movement has been extracted from a greenscreen background to show only a white silhouette moving in front of a black background (silhouette videos).

(3) To compute a total of 32 statistics of 12 kinematic features, for the first time, in one single dataset with same-sequence stimuli materials and make the software freely available.

(4) To obtain emotion-recognition judgments and beauty ratings from human participants for all the created stimuli mentioned above. The emotion-recognition judgments were obtained to validate the pilot dataset in terms of the intended emotional expression of the dancer. Beauty judgments were obtained to encourage the use of aesthetic judgment as an implicit emotion recognition task with future datasets. Previous research has shown that while emotion-recognition rates for a stimulus set may be low, significant differences can usually be found between stimuli intended to express different emotions (Christensen et al., 2023; Christensen et al., 2019), making aesthetic judgment (e.g., beauty, liking, etc.) an interesting implicit emotion-recognition task for the future.

## Method

Ethical approval for the experiment was in place through the Umbrella Ethics approved by the Ethics Council of the Max Planck Society (Nr. 2017_12). For the observer experiment (performed online via the platform Prolific®; Peer et al., 2021; Stanton et al., 2022), informed consent was obtained from all participants, and was given online through a tick-box system. All methods were performed in accordance with the relevant guidelines and regulations.

### Open science statement

In accordance with the framework for open and reproducible science (Munafò et al., 2017), all measures that we collected in the study are reported here.

### Participants

#### Participant pilot dataset creation (one dancer)

One female former professional dancer with 20 years of ballet dance experience choreographed and performed the dance movement sequences.

#### Participants' online experiment

In total, 172 participants (57 male, one other) participated in the human emotion-recognition task (mean age = 35.89 years, *SD* = 11.93, range, 18–65). From the original sample, 22 participants were excluded due to technical issues (the video stimuli did not play), or for not passing attention checks (on two of the emotion-recognition trials, cartoon videos were shown with very obvious emotional expressions (Sponge Bob crying a river of tears; correct response: sad; and Mikey Mouse's head turning red and exploding; correct response: angry). The experiment took approximately 50 min and participants were paid via the Prolific platform (£8/h). Participants had an average of 1.5 years of hobby dance experience (*SD* = 5.05, range, 0–40). We had set the Prolific® filter to return only participants whose first language was English, to ensure complete comprehension of study instructions.

To determine the sample size, we used G*Power 3.1. (Faul et al., 2007). Because the stimuli of the EMOKINE dataset were presented to participants in four different types of visual presentations (avatars, full-light displays (FLDs), point-light displays (PLDs), silhouettes), there was a total of 216 stimuli. Rating this many stimuli could have led to participant fatigue. To avoid this, we opted for dividing stimuli randomly into four sets and determined the sample size for each group of participants for these four sets of stimuli. Subsequently, we confirmed that the percentage of correct responses given by participants in these four different groups to the stimuli was equivalent. We choose a threshold for a large effect size of $d = .80$ (Cohen, 1988) because large effect sizes indicate that the research finding has practical significance. We had initially planned to compare the percentage of correct responses between the four groups with an independent $t$ test. As a result, the suggested sample size calculation for independent samples $t$ test (effect size = .80; alpha = .05; power = .90) was 28 per group. However, we tested at least 30 participants on each of the four sets to ensure full randomization (30 can be divided by six emotions, 28 cannot). Due to technical difficulties, the final

number of participants in each group was; group 1: $N = 36$; group 2: $N = 32$; group 3: $N = 33$; group 4: $N = 31$.

## Materials

### Hardware

Motion capture was performed by means of the MVN Link system (XSENS®, 2020, 2023). Motion capture in this context is the act of recording the motion through time of a set of landmarks (also called keypoints) that are representative of a full-body human pose. This technology has matured mainly via two different approaches: optical solutions, in which markers on the body allow to locate the keypoints, and inertial/magnetic solutions, in which a set of sensors is placed on the body. Both approaches have advantages and disadvantages, but it is generally understood that while optical systems provide very high positional precision, inertial systems are more robust, versatile, and provide more stable acceleration readings (Lim et al., 2016; Skogstad et al., 2011). We are using the latter method in the current research.

The XSENS® system combines inertial and magnetic sensors with advanced algorithms and biomechanical models to provide highly reliable and accurate readings with high spatio-temporal resolution (Schepers & Giuberti, 2018). This allows an optimal assessment of kinematic parameters for complex movement such as dance.

In its full-body configuration, the MVN Link system provides kinematic information from 23 keypoints via a set of 17 wireless sensors embedded on different parts of a spandex suit that fits the dancer's body. This setup is designed to allow for highly free and complex motion.

The keypoints are called "segments" in the XSENS Manual (see Sections 7.2.5 and 15.4 for more details; XSENS Manual, 2020). For an overview of the 17 sensors and how they translate into information about the 23 keypoints, see Table 1 and Fig. 1 (reproduced from the XSENS manual, Section 15.4. (XSENS Manual, 2020), and the XSENS fact sheet about the biomechanical model; XSENS, 2023).

**Table 1** The 23 body keypoints provided by the XSENS® software

| Keypoint number | Segment Label (Bone) | Description | Joint |
|---|---|---|---|
| 1 | Pelvis | Segment between both hip joints and joint L5S1 | jL5S1 |
| 2 | L5 | Segment between joints jL5S1 and jL4L3 | jL4L3 |
| 3 | L3 | Segment between joints jL4L3 and jL1T12 | jL1T12 |
| 4 | T12 | Segment between joints jL1T12 and jT9T8 | jT9T8 |
| 5 | T8 | Segment between joints jT9T8 and jT1C7 | jT1C7 |
| 6 | Neck | Segment between joints jT1C7 and jC1Head | jC1Head |
| 7 | Head | End segment above joint jC1Head | jRightC7Shoulder |
| 8 | Right shoulder | Segment between joints jRightC7Shoulder and jRightUpperArm GH | jRightShoulder |
| 9 | Right upper arm | Segment between joints jRightUpperArm GH and jRightElbow | jRightElbow |
| 10 | Right forearm | Segment between joints jRightElbow and jRightWrist | jRightWrist |
| 11 | Right hand | End segment after joint jRightWrist | jLeftC7Shoulder |
| 12 | Left shoulder | Segment between joints jLeftC7Shoulder and jLeftUpperArm GH | jLeftShoulder |
| 13 | Left upper arm | Segment between joints jLeftUpperArm GH and jLeftElbow | jLeftElbow |
| 14 | Left forearm | Segment between joints jLeftElbow and jLeftWrist | jLeftWrist |
| 15 | Left hand | End segment after joint jLeftWrist. | jRightHip |
| 16 | Right upper leg | Segment between joints jRightHip and jRightKnee | jRightKnee |
| 17 | Right lower leg | Segment between joints jRightKnee and jRightAnkle | jRightAnkle |
| 18 | Right foot | Segment between joints jRightAnkle and jRightToe | jRightBallFoot |
| 19 | Right toe | End segment after joint jRightToe. | jLeftHip |
| 20 | Left upper leg | Segment between joints jLeftHip and jLeftKnee | jLeftKnee |
| 21 | Left lower leg | Segment between joints jLeftKnee and jLeftAnkle | jLeftAnkle |
| 22 | Left foot | Segment between joints jLeftAnkle and jLeftToe | jLeftBallFoot |
| 23 | Left toe | End segment after joint jLeftToe. | |

Overview of the keypoints that were provided by the XSENS system and were used in subsequent analyses. This table has been reproduced from the XSENS manual, Section 15.4. In its "full-body" configuration, the MVN Link system from XSENS provides information for 23 keypoints (called "segments" in the XSENS Manual, see Sections 7.2.5 and 15.4; XSENS Manual, 2020), based on the computations via the 17 sensors, placed throughout the body. Reproduced with permission from the XSENS manual, Section 15.4. (XSENS Manual, 2020), and the XSENS fact sheet about the biomechanical model (XSENS, 2023)

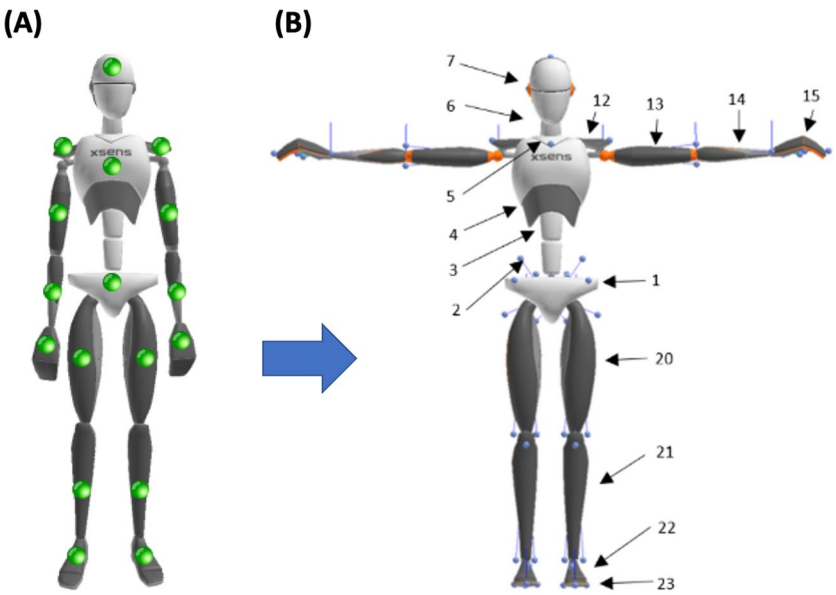



**Fig. 1** Sensors and segments (keypoints) of the XSENS MVN biomechanical model. *Note*: **(A)** The *large green spheres* on the model illustrate the location of the 17 sensors embedded in the spandex suit of the XSENS Link System. **(B)** The *small blue spheres* on the model illustrate the joints that are used to compute the 23 segment (keypoint) values, shown with *arrows* and *numbers*, via the biomechanical model of the XSENS software. Reproduced with permission from the XSENS fact sheet about the biomechanical model; XSENS, 2023)

The XSENS® recording and filming took place in the ArtLab foyer of the Max Planck Institute for Empirical Aesthetics in Frankfurt am Main, Germany, in front of a standard 6 × 3m chroma-key greenscreen background (LTT Junior Truss system with Premium green Buehnenmolton). This allowed for the creation of additional visual preparations of the stimuli, such as silhouette videos. For this, dedo-stage lights (AX3 light drop; 15W RGBW CREE LED) were used to illuminate the entire greenscreen and to minimize shadows.

To produce additional visual presentations of the dataset (FLDs and silhouettes), the dancer was also filmed using a camera Canon EOS 5D Mark IC camera with a Canon EF 24–105 mm f/4 L IS USM lens (settings: e.g., framerate (raw) at 50 fps and framerate (output) at 25 fps. White balance: 5000k, shutter speed: 1/100 sec, and ISO: 400. Video format: H.264, aspect ratio: 16:9, and resolution: 1920 × 1080).

Postproduction of the video footage was done on a 15-inch MAC Book Pro (2017; Processor: 2.9-GHz Quad-Core Intel Core i7; Memory: 16 GB 2133-MHz LPDDR3; Graphics: Radeon Pro 560 4 GB; Intel HD Graphics 630 1536 MB.

### Software

The XSENS® company provides proprietary software that allows calibrating and monitoring the setup during recordings, as well as reading the sensors with a framerate of up to 240 Hz, and also the ability to edit and export the recorded data in various formats including video, and to MVNX (a form of XML; see the XSENS Manual (2020), chapters 6 to 10, for more details). The MVNX provides raw sensor data, as well as refined readings for positions, accelerations, and angles of the full-body keypoints (for more details, see the XSENS Manual, 2020, chapter 15). This data format contains all essential information and is open, so it can be further processed without any proprietary restrictions.

The point-light display (PLD) stimuli were rendered using Blender (Blender Community, 2018), an open-source 3D rendering engine that allows flexible creation and editing of scenes, including positioning and configuration of camera viewpoints, and recording of sequential data into various formats (including video). Specifically, we developed a Python plugin that allows Blender to process the MVNX sequences and convert them into a Blender-friendly format, based on hierarchical relationships between different movements in a human skeleton (a process called "rigging"). We have released our MVNX rigging plugin to GitHub under GPLv3 license (see data availability statement, Section 11).

We also developed custom software to assist with the process of blurring faces, in the form of a series of Python scripts that make use of third-party, open-source deep learning models to first detect the dancer's head (Wu et al., 2019), and then identify the pixels that correspond to the face (more details are provided in the following sections). We have also released these scripts to GitHub under GPLv3 license, with the hope that they can be useful to the community (see data availability statement, Section 11).

Finally, the software *Adobe After Effects* 2019 and *Adobe Premiere Pro* 2019 were used for rendering the

video clips in postproduction. The online survey tool Limesurvey® was used for the observer experiment, and the experiment was launched via the Prolific® online platform.

### Overall procedure

The recording of the pilot stimuli was carried out by a team of five researchers and three filmmakers. The recording procedure followed the recommended standard practice by the XSENS Company (see the XSENS Manual, 2020); chapters 7 and 8), including body measurements and a calibration routine, before the start of each recording session. At the end of each calibration, the dancer was instructed which sequence to perform and which emotion to express, based on an a priori established list of choreographies and emotion orders. The dancer proceeded to say the name of the sequence and the emotion out loudly. Then, clapped twice to signal the beginning of the sequence and to secure the alignment of XSENS®, video and audio recordings. As specified in more detail in section "Stimuli (of the pilot dataset)", 12 separate choreographies were created for the EMOKINE pilot dataset (i.e., 12 different sequences of movements). Each of these 12 choreographies was then performed by the dancer six times, to express a different emotion at each repetition. Hence, each emotion was expressed once for each of the 12 choreographies. The order of the 'emotion takes' for each sequence was always: neutrality, joy, contentment, sadness, fear, and anger. If the dancer was not satisfied with the performance (e.g., made a mistake in the choreography, or felt that the emotion was not expressed), a second (or third) 'emotion take' was performed of the same movement until the dancer agreed with the performance. If the dancer made a mistake in the choreography, the stimulus was discarded without further analysis. If the dancer felt the emotion was not expressed, the "best" sequence of these duplicate takes was chosen by the dancer.

Subsequently, the recordings were rendered in the four different visual presentations. This coarse parameterization was devised in order to enable research into how emotion recognition and beauty ratings are affected by these four different levels of information detail in the representation of emotional kinematics.

#### Procedure for human observer experiment

Four separate experiments were set up to allow the 216 stimuli to be rated by four separate groups of participants. Stimuli were divided randomly, but equally between the experiments, to include the same number of stimuli of each type of visual presentation. To ensure that ratings would be equivalent across all four experiments, one sequence (i.e., 6 stimuli × 4 visual presentations = 24 stimuli) was selected and presented in all four experiments, for an interrater reliability check. For the order of stimuli presentation during the experiment, visual presentation was blocked, but stimuli and blocks were randomized across participants.

Participants watched the stimuli one by one, and rated, first, what emotion they recognized in the movement (forced choice task; anger, contentment, fear, joy, neutrality, sadness), and then, how beautiful they found the movement (Likert scale; 0 = not beautiful; 100 = very beautiful). Participants could only watch each stimulus once and were then asked to provide their rating. As this was an online experiment operated via Prolific®, viewing angle and distance were not controlled. However, filters on Prolific were set so that the experiment could only be performed on a computer desktop (not on a tablet or mobile phone). Figure 2 sets out one trial of the observer experiment.

After the four blocks (with breaks in between), participants were asked to fill in three questionnaires: the Aesthetic Responsiveness Assessment (AReA) (Schlotz et al., 2020), the Interpersonal Reactivity Index (IRI) (Davis, 1980), and demographics questions. The questionnaire data are not presented here.

## Pilot dataset specifications

### Stimuli (of the pilot dataset)

Originally, 12 dance sequences were created by the dancer. However, three of these were deemed not good enough by the dancer and therefore discarded before any further analysis, yielding a total of nine sequences included in the subsequent computations presented here. Each sequence was performed six times to express a different emotional intention at each repetition, namely, anger, contentment, fear, joy, neutrality, sadness; i.e., 9 sequences × 6 emotions = 54 emotional dance movement stimuli. In addition, for each sequence, the dancer did a seventh repetition of the sequence, during which she explained the movements while doing them, like an instruction video for a dance class; yielding nine explanation videos (used elsewhere; Schmidt et al., 2023), yet the videos are provided here as part of the full EMOKINE dataset). Therefore, the total number of stimuli in the EMOKINE dataset is 63 (see Fig. 3).

Seven of the nine sequences that had been choreographed involved only arm movements, while the lower body was held relatively still (only one step to the side). The two remaining sequences involved some side steps ("second-position" in the ballet syllabus). However, as a first step towards quantifying the kinematics of emotional expressivity in dance, the dancer kept most of the emotional expressivity in the EMOKINE dataset to the arms, following previous dataset creations (that did not use motion capture),

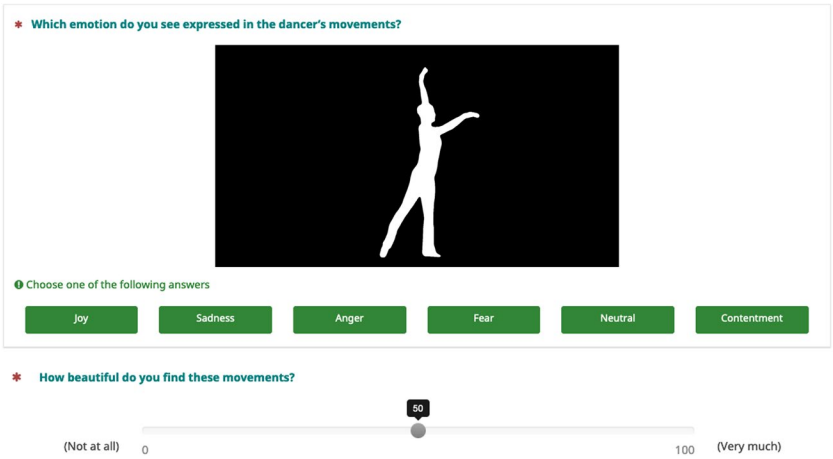



**Fig. 2** One trial of emotion recognition experiment with human observers. *Note*: Sample trial from the human observer experiment. Stimuli were shown one by one (here, a silhouette video). Participants were instructed to guess the emotion the dancer was expressing (forced choice task; anger, contentment, fear, joy, neutrality, sadness), and to rate how beautiful they found the movement (slider question; 0 = not beautiful; 100 = very beautiful)

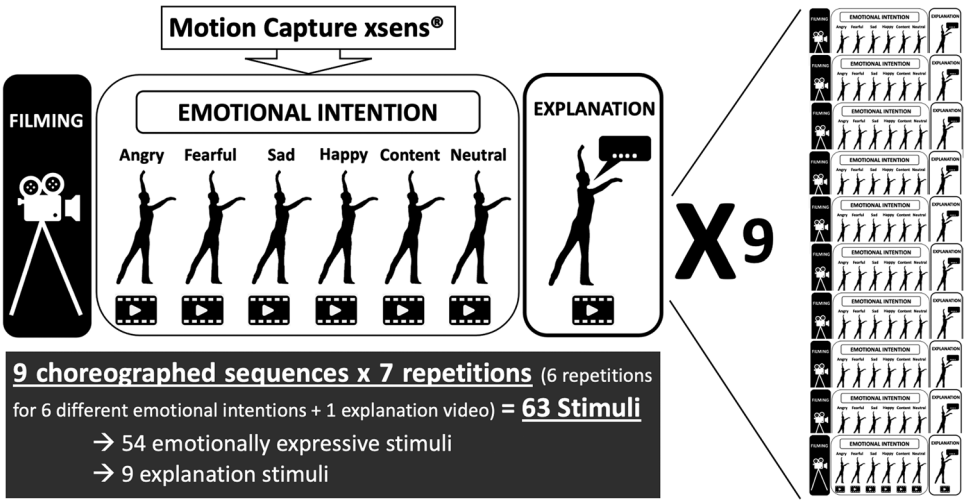



**Fig. 3** Pilot dataset creation. *Note*: The dancer wore an XSENS® Motion Capture Suit LINK while performing nine different sequences of choreographed movements (see Table 2). She repeated these choreographies seven times each. Six times she performed the movements with different emotional intentions at each repetition (anger, contentment, fear, joy, neutrality, sadness). The seventh time she explained the movements, as for a dance class instruction. The latter stimuli were included in a different experiment (Schmidt et al., 2023), but are also included in the present dataset for sake of completeness

which has focused on the arms only (Sawada et al., 2003) (see Table 2 for an overview of the choreographies of each sequence's movements).

In post-production, all 63 pilot stimuli (the 54 emotional stimuli plus the nine explanation videos) were rendered in four different visual presentations (avatars, full-light displays (FLDs), point-light displays (PLDs), silhouettes). None of the stimuli contains facial information, there is no costume, color, or music in the clips. Each clip was faded in and out and contains one full dance phrase (eight counts in dance theory) (see Fig. 4 for the four visual presentations).

The EMOKINE pilot dataset is available for download online, and a selection of stimuli has been used in published work (Schmidt et al., 2023). However, please note that this dataset was created to enable the development of the EMOKINE software and contains portrayals from only one dancer. To ensure generalizability of results from future research using the EMOKINE software, new datasets should be created that include portrayals from more than one dancer. We hope that the details about the stimuli creation procedure set out above may help this endeavor.

**Table 2** Movement sequences of the nine EMOKINE pilot dataset choreographies

| Sequence no. | Sequence movements | Repetitions |
|---|---|---|
| Seq1 | Arms: bras-bas position move up to both arms in 2nd position, palms down, elbows lead. Then, back to bras-bas. Port-de-bras to 1st position, and back to bras-bas. Legs: remain feet parallel, first position. | seq1_angry<br>seq1_content<br>seq1_fearful<br>seq1_joy<br>seq1_neutral<br>seq1_sad<br>seq1_explanation |
| Seq2 | Arms: Move from bras-bas via port-de-bras to first position, open to second, then allongé arms slightly over 180 degrees, palms turn downward, then back to bras-bas position. Legs: remain feet parallel, first position. | seq2_angry<br>seq2_content<br>seq2_fearful<br>seq2_joy<br>seq2_neutral<br>seq2_sad<br>seq2_explanation |
| Seq3 | Arms: bras-bas position move up to both arms in 2nd position, palms down, elbows lead. Then rotate to open 4th allongé, leading with the left arm front, right arm back, then rotate the right arm front, left back. Legs: remain feet parallel, first position. | seq3_angry<br>seq3_content<br>seq3_fearful<br>seq3_joy<br>seq3_neutral<br>seq3_sad<br>seq3_explanation |
| Seq4 | Arms: from bras-bas position move right arm though 1st position, up to 5th. Then left arm follows the same path. Both arms open simultaneously to 2nd position (held elbows in 2nd), subtle allongé and close arms in bras-bas. Make a final allongé movement from bras-bas to 5th position, palms turned out, elbows supple ("swan arms"). Legs: remain feet parallel, first position. | seq4_angry<br>seq4_content<br>seq4_fearful<br>seq4_joy<br>seq4_neutral<br>seq4_sad<br>seq4_explanation |
| Seq5 | Arms: from bras-bas position move right arm though 1st position, up to 5th. Then left arm follows the same path. Both arms lower back down simultaneously to 1st position, then open in second position, allongé, close in bras-bas. Legs: remain feet parallel, first position. | seq5_angry<br>seq5_content<br>seq5_fearful<br>seq5_joy<br>seq5_neutral<br>seq5_sad<br>seq5_explanation |
| Seq6 | Arms: from bras-bas position move both arms to 1st position (elbows held), then allongé arms to open 4th position croisé to the right, pass through 1st position, then to open 4th position croisé to the left. Back to 1st position and move arms back down to bras-bas with elbows held. Legs: remain feet parallel, first position. | seq6_angry<br>seq6_content<br>seq6_fearful<br>seq6_joy<br>seq6_neutral<br>seq6_sad<br>seq6_explanation |
| Seq7 | From parallel 0-position, arms bras-bas: arms and legs together: Slide through plié to the right, to an open wide second position (legs), while arms go from bras-bas to 1st position. Rotate to arms croisé open allongé 4th position, legs, right leg tendu derriere. Pass back through 2nd position, arms 1st. Rotate to left, again, arms croisé open allongé 4th position, legs, right leg tendu derriere. Back to 2nd position legs, plié, arms open to second position, stretch arms allongé as legs stretch second position. Close in from second position turned out, right leg slides back in, to parallel 0-position. | seq7_angry<br>seq7_content<br>seq7_fearful<br>seq7_joy<br>seq7_neutral<br>seq7_sad<br>seq7_explanation |
| Seq8 | From parallel 0-position, arms bras-bas: arms and legs together: step with the left leg to an open efface 4th position with the legs, while arms go up, through 1st position to 5th position. Open arms to 2nd position, allongé arms to an open third (left arm back, right arm croisé front), then pass back through 2nd position with the arms and make an open third again, to the other side (right arm back, left arm effacé front). Arms back to second position. Then, at the same time, close arms to bras-bas and step left leg back to parallel 0-position. | seq8_angry<br>seq8_content<br>seq8_fearful<br>seq8_joy<br>seq8_neutral<br>seq8_sad<br>seq8_explanation |
| Seq9 | Arms and legs together: step out with left foot to an open second, left arm makes a sweeping movement, port-de-bras through 1st position to second, but upper body tilted, so left hand almost touches floor. Again, simultaneously, left arm closes up to 5th position, and does cambré to the right while left leg stretches, so a diagonal line between left toe and left arm is formed. Right leg is 2nd position plié, head looks to the left. Left arm opens to 2nd position, while legs plié in 2nd position. Then arm makes a fast port-de-bras through bras-bas then 1st and up to 5th, while legs do a balancé (a round shape is created in the air). During the movement, the right arm and hand are fairly still, right hand resting lightly on the lower right hip. When left arm reaches second position again, freeze and legs close back to 0-position and arms back to bras-bas. | seq9_angry<br>seq9_content<br>seq9_fearful<br>seq9_joy<br>seq9_neutral<br>seq9_sad<br>seq9_explanation |

Choreographies of the nine movement sequences included in the EMOKINE dataset. The movements contained in each sequence are set out in this table using Western classical ballet vocabulary

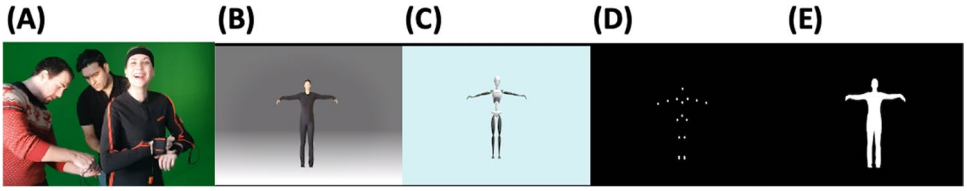



**Fig. 4** The four visual presentations of the EMOKINE pilot dataset. *Note*. **(A)** One dancer performed nine dance sequences of eight counts six times with six different emotional intentions, while wearing an XSENS® Motion Capture Suit in front of a greenscreen. Filming was done with a Panasonic camera. **(B)** Full-light display (FLD) dancer with blurred face: Background grey set with Adobe After Effects Keylight effect. Face blurred with automated computational methods (Cheng et al., 2019; Nirkin et al., 2017). **(C)** XSENS® Avatar dancer: Extracted from XSENS® system propriety software (XSENS Manual, 2020). **(D)** Point-light display (PLD) dancer: XSENS® output data with skeleton information (MNVX) fed into a customized Blender module and white spheres attached to skeleton (Blender Community, 2018; Schepers & Giuberti, 2018). **(E)** Silhouette dancer: Adobe After Effects Keylight effect used to remove background (greenscreen). Adobe After Effects Opacity function used to color foreground (dancer) white

#### Full-light displays (FLDs) with blurred face

For the full-light displays (FLDs) with blurred face, videos were rendered by importing them into Adobe Premiere Pro. The videos were trimmed to the start and end points of the movements with the help of a dancer (academic dance sequences have specific start and end points that are only detectable for the expert). Each clip was rendered into a separate file in an uncompressed format and the title was added, as specified verbally by the dancer during the recording. In this saving procedure, the sound track (ambient noise) of the clips was removed. Then, all rendered files were imported to Adobe After Effects. The "Keylight" effect was used to set the background to be a shade of grey.

Blurring the face required locating the pixels that correspond to the face, which can be a very time-consuming task if done manually for video datasets. To speed up the process, we developed rigging software for a semi-automated pipeline. Each video was split in consecutive, deinterlaced image frames that were processed separately. For each image, the Detectron2 deep learning model for human keypoint estimation was used (Wu et al., 2019). Since we only had a single person on static background, averaging any detected keypoints for the nose, eyes and ears provided a very robust estimation of the head position, and given that the dancer was always more or less centered and at the same distance from the camera, a fixed-length frame of 140 × 140 pixels around the estimated head position was used, in order to extract a patch containing the head. This allowed the face segmentation model by Nirkin and colleagues (2017) to produce a binary mask that accurately matched the actual face at pixel level. This binary mask was then translated from the head-patch back to the main image. The resulting masks for the whole dataset were then grouped and paired with the corresponding videos in order to blur the faces at the regions where the masks were active.

#### Silhouettes

To render the footage into silhouette dancer videos, all footage was imported into Adobe Premiere Pro as before. Here, the "Keylight" effect was added, and settings adapted to remove the background from each clip, and the "Level" effect (setting: output black = 255) was added to each clip to color the extracted foregrounds white (the visible dancer silhouette). "Opacity" keyframes were then added to the beginning and the end of each clip to allow for a fade-in and fade-out of each clip (eight frames). Finally, each clip was rendered as a separate file in H264 format (see Fig. 2).

#### Point-light displays (PLDs)

The point-light display (PLD) videos were created using the XSENS output data, MNVX (Blender Community, 2018; XSENS Manual, 2020). The MNVX file contains information about the skeleton (bone geometry and connections), and each "frame" (240 frames per second) contains kinematic information about the position and angle of the bones. We wrote a custom Blender plugin (Blender Community, 2018) that read each MVNX file and created a skeleton with the corresponding geometry and connections. Then, the module read the frame information and created an animation. Based on previous models for marker placement, namely the frontal view of the Plug-In Gait Model (Kainz et al., 2017; Piwek et al., 2016), we identified a series of key landmarks on the skeleton, and attached a white sphere to each landmark, in order to create the "light points" that convey the information about the movements. For the video rendering, we positioned a virtual camera "in front of" the PLDs with an angle, position and focal length that closely resembled the data obtained with the video camera (for an example of the result, see Fig. 3). Then, the skeleton was made transparent (making it black on a black background) and the spheres bright white (increasing the contract),

allowing to extract the pixel-position of each point in the rendering. Videos were faded in and out.

#### XSENS avatar dancer

The XSENS® avatar dancers were extracted from the propriety software of the XSENS® system (XSENS Manual, 2020).

### Data formats (of the pilot dataset)

Beyond the already-discussed four modalities for the stimuli, we include with the EMOKINE pilot dataset, several modalities of data records that we recommend to use in the future, as they may help with downstream tasks. The pilot files are available on Zenodo (see Section 11; Data Availability Statement), and consist of the following file formats MVNX, comma-separated values, and camara position (CamPos). Extensive details are provided in the readme files along with the data and software on Zenodo and GitHub.

#### MVNX files

We include the raw MVNX motion capture recordings, as produced by the XSENS® software for the pilot dataset.

#### Comma-separated values (CSV) files

To facilitate easy integration with other data analysis tools, we recommend converting a subset of the MVNX files into comma-separated value (CSV) files. For each sequence and emotion, we extract per-keypoint time series for position, orientation, velocity, angular velocity, acceleration, angular acceleration, center of mass and foot contacts. In the EMOKINE software package, we provide the script and instructions to perform this conversion.

#### Camera position (CamPos) files

While the positional data in the MVNX files is provided in a global three-dimensional frame of reference, the stimuli are rendered from a specific camera perspective. We use Blender to extract the positions of the bones and PLD spheres relative to the camera, as *x*/*y*/*depth* coordinates, where $x$ goes from 0 (leftmost pixel) to 1 (rightmost pixel), $y$ from 0 (bottom pixel) to 1 (top pixel) and *depth* is provided in meters. The result, dubbed here CamPos (for camera positions), is provided as JSON files containing the time series in frames at 60 Hz, where each frame contains the camera-relative positions. This can be useful for example in analyzing kinematic features from the perspective of the observer. In the EMOKINE software package, we also provide the script and instructions to produce these files.

#### Using the EMOKINE software with data obtained from other motion-capture systems

Although the EMOKINE pilot dataset was recorded via the XSENS system, most of the EMOKINE software provided here can be directly applied to data obtained from other motion capture systems with little to no modification. Specifically, only the files "1a_mvsn_to_csv.py" and "1b_mvnx_blender.py" are relevant to the MVNX formatted data. These data are then converted to tabular format (see GitHub repository for examples), which is then consumed by the remaining scripts (together with plain video data whenever needed, e.g., for the file "2b_face_blur.py"). Researchers intending to use this software with motion capture data from other systems simply need to ensure that their data follows the same tabular format, and that they have video data available whenever needed (e.g., for face blur or silhouette extraction).

## Kinematic features

Making use of the Silhouette, MVNX, and CamPos data modalities, we compute a series of kinematic features. We extracted 32 statistics from 12 kinematic features. We group the extracted kinematic features in the following categories: Speed and Acceleration (speed, acceleration, angular speed and angular acceleration; Section "Speed and acceleration"), Expansion/Contraction (limb contraction, distance to center of mass; Section "Expansion/contraction"), Movement Activity (quantity of motion, QoM, ratio; Section "Movement activity"), Fluidity/Smoothness (dimensionless jerk (integral); Section "Fluidity/smoothness"), Body Tilt (head angle, with regards to vertical axis and with regards to back; Section "Body tilt"), and Space (convex hull 2d and 3D; Section "Space"). For each of these, we computed the per-sequence *average*, *median absolute deviation* (MAD) and *maximum value* as described above.

The resulting computed features for each sequence and emotion are provided in the EMOKINE dataset, and the script and instructions to compute them from the raw data are included in the EMOKINE software package. In Section "Validation of kinematic features and results of observer experiments", we demonstrate the usefulness of these features and the meaningfulness of the EMOKINE data through a series of quantitative and qualitative analyses.

Before we outline each kinematic feature in detail, we give an overview of the math behind the kinematic features. More formally, for each sequence $s(t) \in \mathcal{S}$ (from the 63 total in the EMOKINE dataset $\mathcal{S}$), we have $\{K_i^{(s)}(t)\}_{i=1}^{12}$ nonnegative scalar features, where $t \in \{0, \ldots, T_s\}$ indicates discrete time with a duration of $T_s$ frames. Thus, $K_i^{(s)}(t) : \mathcal{S} \mapsto \mathbb{R}_{>0}^{T_s}$

**Table 3** Formulas for kinematic feature extraction

| Name | Source | Formula | Unit |
|---|---|---|---|
| Speed | MVNX | $v_j(t) := \| \nabla_t p_j(t) \|_2$ | $\frac{m}{s}$ |
| Acceleration | MVNX | $a_j(t) := \| \nabla_t^2 p_j(t) \|_2$ | $\frac{m}{s^2}$ |
| Angular speed | MVNX | $\dot{\omega}_\xi(t) := \| \nabla_t \omega_j(t) \|_2$ | $\frac{rad}{s}$ |
| Angular acceleration | MVNX | $\ddot{\omega}_\xi(t) := \| \nabla_t^2 \omega_j(t) \|_2$ | $\frac{rad}{s^2}$ |
| Limb contraction | MVNX | $l_c(t) := \frac{1}{4}(\| p_h(t) - p_a(t) \|_2 + \cdots)$ | $m$ |
| CoM distance | MVNX | $m_j(t) := \| p_j(t) - \mu(t) \|_2$ | $m$ |
| Quantity of motion | Silhouette | $q(t) := \frac{\lvert Q_\delta(t) \rvert_1}{\lvert f(t) \rvert_1}$ | ∅ |
| Dimensionless jerk | MVNX | $\lambda_j(t) := \frac{(T_s)^5}{\Delta_p^2} \int_{t=0}^{T_s} \dot{a}_j(t)^2$ | ∅ |
| Head angle wrt. back | MVNX | $\alpha(t) := cos^{-1}\left(u_{ab}(t)^\top u_{bc}(t)\right)$ | *rad* |
| Head tilt wrt. vertical | MVNX | $\beta(t) := \cos^{-1}\left(u_\uparrow(t)^\top u_{bc}(t)\right)$ | *rad* |
| Convex hull 3D | MVNX | $volume\left(\mathcal{C}_{3D}(t)\right)$ | $m^3$ |
| Convex hull 2D | CamPos | $area\left(\mathcal{C}_{2D}(t)\right)$ | ∅ |

Summary of the computations used to extract the kinematic features. Some were extracted directly from the MVNX files, as provided by the XSENS® software (see Section 3.3. for more details), while others were extracted from the CamPos data (see Section 3.3.3.) and the silhouette stimuli videos (see Section "Silhouettes"). We refer the reader to Sections "Speed and acceleration" for the comprehensive definitions of the quantities

is the kinematic feature of sequence $s$ out of a total of 12 kinematic features, represented by a nonnegative vector of dimension $T_s$.

Some of the kinematic features were extracted directly from the MVNX files as provided by the XSENS software (see Section "Pilot dataset specifications" for more details), while others were extracted from the CamPos data (see Section 3.3.3.) and the silhouette stimuli videos (see Section "Silhouettes"). In the following sections, we describe in detail how kinematic features were extracted and/or computed. This information is summarized in Tables 3 and 4.

We aggregate each kinematic feature across time $t$ to obtain a single scalar statistic that summarizes the kinematics of each sequence $s$. The following aggregation techniques are used in multiple features:

- Average: $\overline{K}_i^{(s)} := \frac{1}{T_s}\sum_{t=1}^{T_s} K_i^{(s)}(t)$
- Median absolute deviation (MAD): $\widetilde{K}_i^{(s)} := \underset{t}{\text{median}}\left( \left| \underset{t}{median}\left(K_i^{(s)}(t)\right) - K_i^{(s)}(t) \right| \right)$
- Maximum: $\widehat{K}_i^{(s)} := \max_t\left(K_i^{(s)}(t)\right)$

Qualitatively, the $K_i^{(s)}(t)$ features tell us "how much" of a given feature is given at each timepoint. Then, the main difference between these three aggregations is their sensitivity to outliers: $\widehat{K}_i^{(s)}$ is the most sensitive, and $\widetilde{K}_i^{(s)}$ is the least sensitive. The average: $\overline{K}_i^{(s)}$ lies inbetween. Varying sensitivity to outliers is important if a sequence relies on punctual strong phenomena to convey crucial information (e.g., a short burst in velocity in a generally slow sequence will still have a large maximum), or conversely, to recover the underlying information in cases, where the sequence is exposed to outliers (e.g., if a movement is supposed to be smooth, but is slightly shaky, or contaminated with noise).

## Speed and acceleration

Speed is one of the most frequently explored kinematic parameters, and research suggests that it plays a substantial role in an observer's ability to distinguish between specific kinds of emotional expressivity. The majority of this work suggests that slow movements are associated with sadness, and in some cases with expressions of neutrality and fear, while fast movements are typically associated with happiness (or joy) and anger (Bernhardt & Robinson, 2007; Crane & Gross, 2007; Crane & Gross, 2013; Gross et al., 2010; Halovic & Kroos, 2018; Masuda et al., 2010; Montepare et al., 1999; Roether et al., 2009; Smith & Pollick, 2022).

Acceleration is less studied in relation to observer judgements of emotional expressivity in the kinematics literature. But in a study conducted by Sawada and colleagues, a similar pattern emerged across these movement features. Namely, they found that high acceleration in arm movements was associated with anger, and low acceleration was associated with sadness (Sawada et al., 2003). We here provide a series of speed and acceleration related features for to enrich emotional kinematics research in the future; speed (Section "Speed"), acceleration (Section "Acceleration"), angular speed (Section "Angular speed") and angular acceleration (Section "Angular acceleration").

### Speed

Velocity is a vector provided by the MVNX system that points in a specific 3D direction, and speed is the "length" of the vector. This length tells us how fast is a given keypoint moving in that direction, in meters per second. More formally, if the position in meters of a given joint $j$ in the 3D space at time $t$ is:

$$p_j(t) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Where $x$ is aligned (and pointing to) the magnetic north, $y$ is aligned (and pointing to) the west, and $z$ is pointing up (for more details, see the XSENS Manual (2020; section 23.8).

Then, the velocity $v$ is the derivative of the position with respect to time ($\nabla_t$), and our *speed* feature is the Euclidean norm of the velocity, i.e. $\|v_j(t)\|_2$:

$$v_j(t) = \nabla_t p_j(t) = \begin{pmatrix} \nabla_t x \\ \nabla_t y \\ \nabla_t z \end{pmatrix}$$

Note that, in discrete time, this quantity could be approximated by computing $\frac{p(t+\Delta_t)-p(t)}{\Delta_t}$, where $\Delta_t$ is a small amount of discrete time (e.g., one frame), but in this case it is not necessary since it is provided directly by the MVNX file, and estimated by the XSENS system using a proprietary algorithm; see XSENS Manual (2020; section 23.8). We provide the average, MAD and maximum velocity for each joint $j$ and sequence $s$ in EMOKINE.

#### Acceleration

For each joint and timepoint we define the acceleration $a_j(t)$ as a three-dimensional vector, encoding the rate of change in the speed with respect to time. Our *acceleration* feature is then the Euclidean norm of that vector i.e., $\|a_j(t)\|_2$ with:

$$a_j(t) = \nabla_t^2 p_j(t) = \nabla_t v_j(t) = \begin{pmatrix} \nabla_t^2 x \\ \nabla_t^2 y \\ \nabla_t^2 z \end{pmatrix}$$

The acceleration vectors $a_j(t)$ are also estimated through the XSENS proprietary algorithm (XSENS Manual, 2020; section 23.8) and provided directly through the MVNX files. The acceleration is conceptually associated to the "force" applied to a joint. As with the *speed* feature, the Euclidean norm does not convey information about the directionality. We provide the average, MAD and maximum acceleration for each joint $j$ and sequence $s$ in EMOKINE.

#### Angular speed

Joints not only have positions, but they also have orientations. A joint can change position without changing orientation (e.g., walking with a straight head), and vice versa (rotating the neck while standing still). The angular speed focuses on the orientation: It measures the change of "angle" as a function of time, so instead of meters per second, we have radians per second. If a dancer is rotating a full circle per second, then the angular speed of their body would be $2\pi$ radians per second ($2\pi$ radians = 360 degrees).

In the XSENS system, each keypoint is considered the beginning of a segment (can be thought of as a "bone") with its own, local three-dimensional coordinate system. When the subject stands in T-pose, all local coordinate systems are aligned with the global system (see Fig. 92 in section 23.5 of the XSENS Manual, 2020). Then, the segment rotations follow the $Z$ (flexion/extension), $X$ (abduction/adduction), $Y$ (internal/external) convention (see XSENS Manual, 2020; section 23.6, for exhaustive details). More formally, for each segment $\xi$, the orientation in radians $\omega_\xi(t)$ is given as a three-dimensional vector in Euler representation, which varies as a function of time:

$$\omega_\xi(t) = \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

Then, the *angular speed* feature is the Euclidean norm of the derivative of $\omega_\xi(t)$ with respect to time, i.e., $\dot{w}_\xi(t) = \|\nabla_t \omega_\xi(t)\|_2$, given in $\frac{rad}{s}$. Like the rest of quantities presented so far, this quantity is estimated by the XSENS proprietary algorithm, and directly provided via the MVNX file. Analogously to the linear velocity previously discussed, the Euclidean norm retains the information about the amount of angular speed, but does not contain information about the specific directions. We provide the average, MAD and maximum angular velocity for each joint $j$ and sequence $s$ in EMOKINE.

#### Angular acceleration

Analogously to the case of linear acceleration, angular acceleration is the second derivative of angle with respect to time, i.e., $\ddot{w}_\xi(t) = \|\nabla^2{}_t \omega_\xi(t)\|_2$. Like the rest of quantities presented so far, this quantity is estimated by the XSENS proprietary algorithm, and directly provided via the MVNX file. The Euclidean norm retains the information about the amount of angular acceleration but does not contain information about the specific directions. We provide the average, MAD, and maximum angular velocity for each joint $j$ and sequence $s$ in EMOKINE.

### Expansion/contraction

Body expansion and contraction is another kinematic feature of movement commonly explored in emotion perception research. However, unlike with speed, the results in this area do not present such a clear pattern of associations, likely because many datasets used in this area focus on emotional actions (instead of the same-sequence approach proposed in EMOKINE).

Of the available literature, most research in this area seems to agree that expansion is associated with happiness or joy, and some suggest that it also leads to the perception of anger (Camurri et al., 2003; Gross et al., 2010; Gross et al., 2012;

Masuda et al., 2010; Montepare et al., 1999; Shafir, 2016; Shikanai et al., 2013; Wallbott, 1998). Castellano and colleagues' study is a notable exception in that they found anger to be associated with contraction instead (Castellano et al., 2007). However, contraction is more commonly noted to align with the perception of fear and sadness (Camurri et al., 2003; Masuda et al., 2010; Shafir et al., 2016; Shikanai et al., 2013; Wallbott, 1998), and in some cases with neutral expressivity (Montepare et al., 1999). For body expansion/contraction, the EMOKINE framework includes limb contraction (Section "Limb contraction") and distance to center of mass (Section "Distance to center of mass (CoM)").

### Limb contraction

Limb contraction regards the positions of five keypoints: head, right hand, left hand, right toe and left toe (Poyo Solanas et al., 2020). Respectively: $\left(p_h(t), p_a(t), p_b(t), p_c(t), p_d(t)\right)$. Then, at each timepoint , this metric consists in the mean Euclidean distance between each of the four extremity endpoints and the head, i.e.,

$$\mathrm{l_c(t)} = \frac{1}{4}\left(\|p_h(t) - p_a(t)\|_2 + \|p_h(t) - p_b(t)\|_2 + \|p_h(t) - p_c(t)\|_2 + \|p_h(t) - p_d(t)\|_2\right)$$

This metric is a proxy for body contraction, with the idea that contracted bodies tend to have shorter distances between the limb endpoints and the head, while expanded poses tend to have longer distances. We provide the average and MAD limb contraction for each joint $j$ and sequence $s$ in EMOKINE.

### Distance to center of mass (CoM)

At each timepoint $t$, and together with the joint positions $p_j(t)$, the XSENS system also estimates and retrieves the position of the person's *center of mass* $\mu(t)$, in meters, which is a "weighted average" among all the points in the body, thus representing the idea of its "central point". For each sequence in EMOKINE and for each keypoint $j$, we compute the CoM distance between each keypoint position $p_j(t)$ and the CoM as follows:

$$m_j(t) = \| p_j(t) - \mu(t)\|_2$$

We then compute and retrieve the average and MAD.

## Movement activity

Movement activity has been examined in a number of ways in the kinematics literature depending on the particular movement stimuli used. For example, in studies exploring the kinematics of walking motions it is typically measured via step frequency (e.g., Crane & Gross, 2007). More often, however, it is quantity of motion that is assessed. Research in this area, as with speed, suggests that high movement activity is typically associated with the portrayal of joy and anger and low movement activity is associated with sadness, fear, and sometimes neutral expressivity (Bernhardt & Robinson, 2007; Camurri et al., 2003; Crane & Gross, 2007; Crane & Gross, 2013; Gross et al., 2010; Halovic & Kroos, 2018; Masuda et al., 2010; Montepare et al., 1999; Roether et al., 2009; Shikanai et al., 2013). Wallbott (1998) is an exception to this, in that they found variations in movement activity to distinguish between happiness with different levels of intensity; high activity was associated with "elated joy", but the more general "happiness" was associated with low movement activity. Based on this research, the EMOKINE framework includes Quantity of Motion (QoM) as a measure of movement activity (Section "Quantity of motion (QoM)").

### Quantity of motion (QoM)

Unlike the other quantities presented so far, the QoM is not extracted from the MVNX sequential data. Instead, the input are silhouette videos at 25 fps, where each frame is a Boolean matrix $\mathrm{f(t)} \in \{0,1\}^{1080\times1920}$, with pixel values of 0 corresponding to the background, and values of 1 to the dancer. The QoM is a time-dependent feature that can be intuitively understood as a ratio between how much has the silhouette moved in the recent past, in proportion to how big the silhouette is right now (Castellano et al., 2007). More formally, given a time span of $\delta$ frames, we make use of the boolean operations of pixel-wise union ($\vee$), intersection ($\wedge$), negation ($\neg$) and sum ($|\cdot|_1$) to define the QoM $q(t)$ as follows:

$$Q_\delta(t) = \left(\bigvee_{i=1}^{\delta} f(t-i)\right) \wedge \neg f(t)$$

$$q(t) = \frac{|Q_\delta(t)|_1}{|f(t)|_1}$$

The Boolean array $Q_\delta(t) \in \{0,1\}^{1080\times1920}$, also called the *silhouette motion mask*, is active for the pixels that were active immediately before the current time and are not currently active. Thus, it encodes the "recent activity": it is all zeros if there is no movement, and it contains more active pixels as movement increases.

Then, the QoM is the ratio between the sum of active pixels in $Q_\delta(t)$ and the sum of currently active pixels. Note that the QoM is a full-body quantity and does not depend on a given joint. Therefore, for each sequence, we provide only one scalar average, MAD, and integral QoM. While the average QoM is trivially the integral divided by the number

of frames, we included both here for convenience, since the integral depends on the sequence length and is a quantity of interest in the literature.

## Fluidity/smoothness

Comparatively, there is less research examining the role of movement fluidity in the perception of emotional expressivity, but there does exist some evidence to suggest that movement fluidity is associated with happiness or joy, while stiff or low-fluidity motion is associated with anger (Montepare et al., 1999) and other negative valence emotions like grief and fear (Camurri et al., 2003). For the EMOKINE framework, we computed dimensionless jerk (integral), as a measure of movement fluidity/smoothness (Section "Dimensionless jerk (integral)").

### Dimensionless jerk (integral)

Based on Hogan & Sternad (2009), this feature proposes a variation of the jerk, which is the time-derivative of acceleration (Hogan & Sternad, 2009):

$$\dot{a}_j(t) = \nabla_t a_j(t) = \nabla_t^3 p_j(t) = \begin{pmatrix} \nabla_t^3 x \\ \nabla_t^3 y \\ \nabla_t^3 z \end{pmatrix}$$

In order to quantify smoothness in trajectories from $t_1$ to $t_2$ the integral of the square jerk is typically used in the literature, i.e., $\int_{t_1}^{t_2} \dot{a}_j(t)^2$. One major problem with this, as pointed out by Hogan & Sternad (2009), is that this yields a high-order polynomial unit of $\frac{length^2}{time^5}$ being sensitive to noise and changes in scale. They propose a dimensionless variant, where the integral is multiplied by $\frac{(t_2-t_1)^5}{\Delta_p^2}$, where $\mathbf{\Delta_p}$ is the extent of the length achieved between $t_1$ and $t_2$ (i.e., if in a sequence the individual moves more, $\mathbf{\Delta_p}$ will be larger for that sequence). The result yields a notion of movement smoothness that is normalized against duration and size, and (by design) it is also void of any units, hence the dimensionless characterization. For each sequence $s$, we provide the dimensionless jerk between $t_1 = 0$ and $t_2 = T_s$, i.e.:

$$\lambda_j(t) = \frac{(T_s)^5}{\Delta_p^2} \int_{t=0}^{T_s} \dot{a}_j(t)^2$$

Note that this feature is full-body and does not depend on specific joints.

## Body tilt

Most of the research exploring tilt or angularity within body motion or positioning has focused on the head. There is compelling evidence to suggest that a downward or forward-tilted orientation of the head is uniquely associated with the portrayal of sadness (Crane & Gross, 2007; Masuda et al., 2010; Shafir et al., 2016; Wallbott, 1998), and these works suggest other patterns that may emerge. Wallbott (1998) suggests that having the head oriented backwards is associated with elated joy, raised shoulders are associated with both elated joy and hot anger, and raised arms are associated with each of the

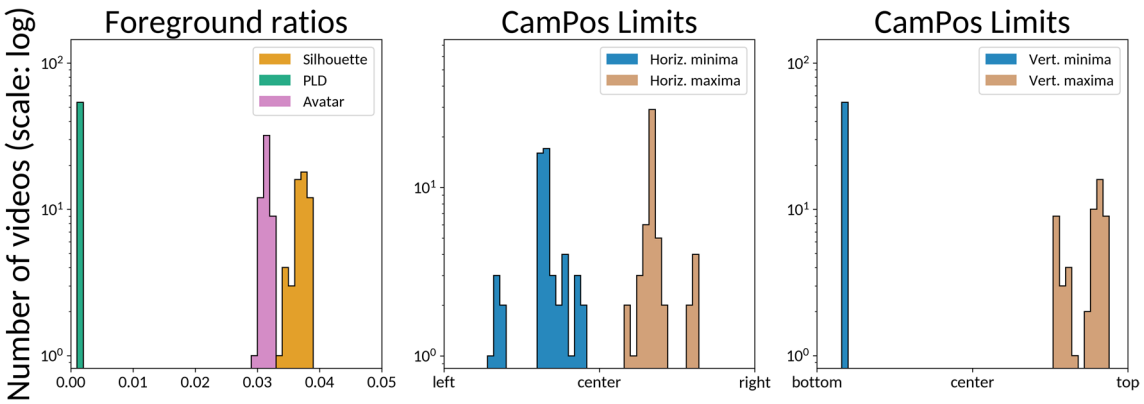



**Fig. 5 Foreground statistics**. *Note*: Illustrations of the foreground statistics of all videos together (each unit on the *y*-axis is one video). We see a quite homogenous distribution of the dancer (= the foreground) across videos. **Foreground ratios**. This statistic indicates the ratio of the foreground with regards to the background. We see a concentrated distribution which indicates that there is not a big variation across stimuli of this set. **Camera positions (CamPos) for horizontal minima and maxima**. This figure answers the question of whether the dancer moves 'out' of the frame of the video horizontally, i.e., to the sides (*x*-axis). We see a quite symmetrical distribution around the center of the videos, and to the sides, with a very slight bias to the left. **Camera positions (CamPos) for vertical minima and maxima**. We see that the dancer occupies the vertical space quite homogeneously (*x*-axis). The high values are due to the arm movements up over the head, apart from that all camera positions are roughly at the same height

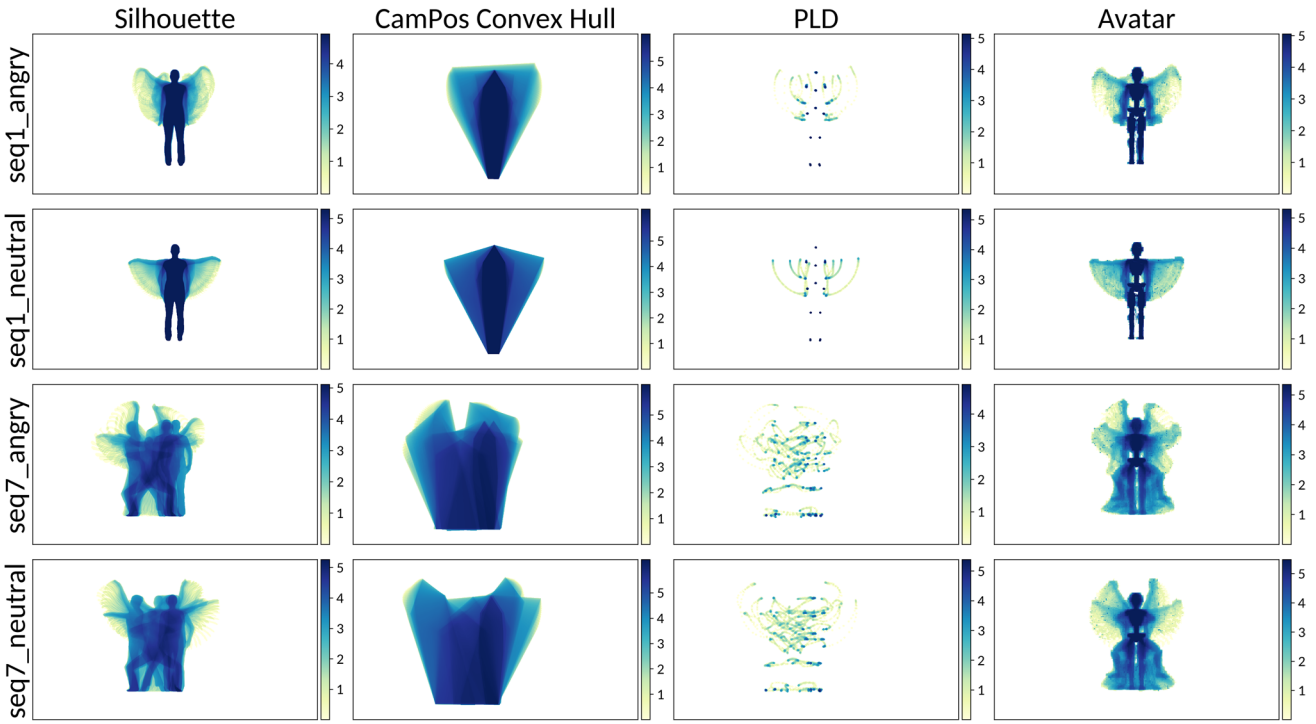



**Fig. 6 Histograms of kinematics that depend on silhouette images**. *Note*: Frequencies of frames occupied by the dancer in space. Figure shows two illustrative stimuli as examples, sequence 1 (perfect alignment) and sequence 7 (good alignment except for the avatar stimulus). The illustrations for all 54 stimuli are in the appendix of the paper. The computation of the frequencies of frames occupied for the stimuli **silhouettes**, **convex hull**, **point-light displays (PLDs)**, and **avatar** stimuli depend on silhouette images. Silhouettes of the stimuli and CamPos (camera position) of the convex hull are calculated with the coordinates of the PLD (in Blender). The column with the avatar is from the XSENS® software. Across the columns, we see that the distribution of a single video is very equal among different modalities of videos, which indicates a very good alignment. It makes sense, since PLDs, convex hull and silhouettes are all from the same source, and they clearly show the same distribution. The silhouette occupies most space, then convex hull, avatar, PLDs. However, the illustration also shows a defect of the avatar (extracted from XSENS® software – the software is automatically adjusting the camera position. In videos 7–9 the dancer turns the upper body, which is corrected by the software so the camera position remains frontal throughout. This results in the very symmetric histograms in that column, while the other columns show movement also to the sides. As long as no turns are in the movement, the four stimuli modalities are aligned in terms of the frequencies of frames occupied by the dancer in space. There are no units on the axes because this is the horizontal and vertical space that the dancer occupies. The scale on the right side of the histograms shows the log of the frequency (color = log frequency of frames; 4 = frequency is 100 (np.log(freq)) = 4.61). The log scale allows seeing more differentiation the distribution of frames occupied by the dancer in space

following: elated joy, cold anger, hot anger, and terror. Masuda et al. (2010) found that reclined posture is associated with pleasure, and that a mixture of reclined and straight posture is associated with relaxation. Shafir and colleagues (2016), on the other hand, found that a reclined tilt to the upper body is associated with fear. For the EMOKINE framework, we computed the head tilt with regards to the back (Section "Head tilt with respect to back") and with regards to the vertical axis (Section "Head tilt with respect to vertical").

### Head tilt with respect to back

For the computation of this kinematic feature, we consider the three-dimensional positions for three keypoints: T8 vertebra, neck and head, dubbed here $p_a(t)$, $p_b(t)$, and $p_c(t)$, respectively. Then, we define the unit vectors going from T8 to the neck, and from neck to head, as:

$$u_{ab}(t) = \frac{p_b(t) - p_a(t)}{\| p_b(t) - p_a(t) \|_2}$$

$$u_{bc}(t) = \frac{p_c(t) - p_b(t)}{\| p_c(t) - p_b(t) \|_2}$$

Then, the head tilt with respect to the back $\alpha(t)$ is the angle between $u_{ab}(t)$ and $u_{bc}(t)$, which can be computed as: $\alpha(t) = cos^{-1}\left(u_{ab}(t)^{\top} u_{bc}(t)\right)$, in radians, since the dot product between 2 unit vectors yields their cosine. This cosine is always non-negative, since we do not expect an angle larger than 90 degrees. We provide the average and MAD of $\alpha(t)$ across time.

### Head tilt with respect to vertical

This feature is similar to the head tilt with respect to the back, but instead we measure the angle between $u_{bc}(t)$ and the global vertical:

$$u_{\uparrow}(t) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This yields our desired feature, in radians: $\beta(t) = cos^{-1}\left(u_{\uparrow}(t)^{\top} u_{bc}(t)\right)$. We also provide the average and MAD as aggregation statistics.

## Space

The kinematic features in relation to space, in the context of this paper, refer to the physical area used by the dancer when performing the movement sequences. Previous research exploring space in this way as a kinematic parameter suggests that large travel distances (i.e., using a high proportion of the movement space) is associated with joy, and low travel distances (i.e., using a low proportion of the space) is associated with sadness (Sawada et al., 2003). It has also been suggested that movement in a variety of directions (i.e., a wider range in the movement space used) is associated with anger (Masuda et al., 2010). As measures of space within the EMOKINE framework, we provide the convex hull 2D (Section "Convex hull 3D") and 3D (Section "Convex hull 2D").

### Convex hull 3D

Given the $p_j(t)$ locations for all body keypoints $\mathcal{J}$ at a given time $t$, the convex hull is the smallest convex envelope that contains all points. For example, if a person is extending their arms and legs in a t-pose and we take a frontal image of their keypoints, the corresponding 2D convex hull would be a convex polygon going from the head, to the hands, then from each hand to its respective foot, and then connecting the feet. In 3D, it follows the same principle but it also takes depth into account, yielding a convex polytope, where the vertices are the $p_j(t)$ keypoint positions. The convex hull can be used as a proxy for how much space is the person effectively occupying.

Formally, given the set of all keypoints $\mathcal{J}$, the 3D convex hull can be defined as the set (Boyd & Vandenberghe, 2004):

$$\mathcal{C}_{3D}(t) = \left\{ \sum_j \theta_j p_j(t) | j \in \mathcal{J}, \theta \geq 0, \sum_j \theta_j = 1 \right\}$$

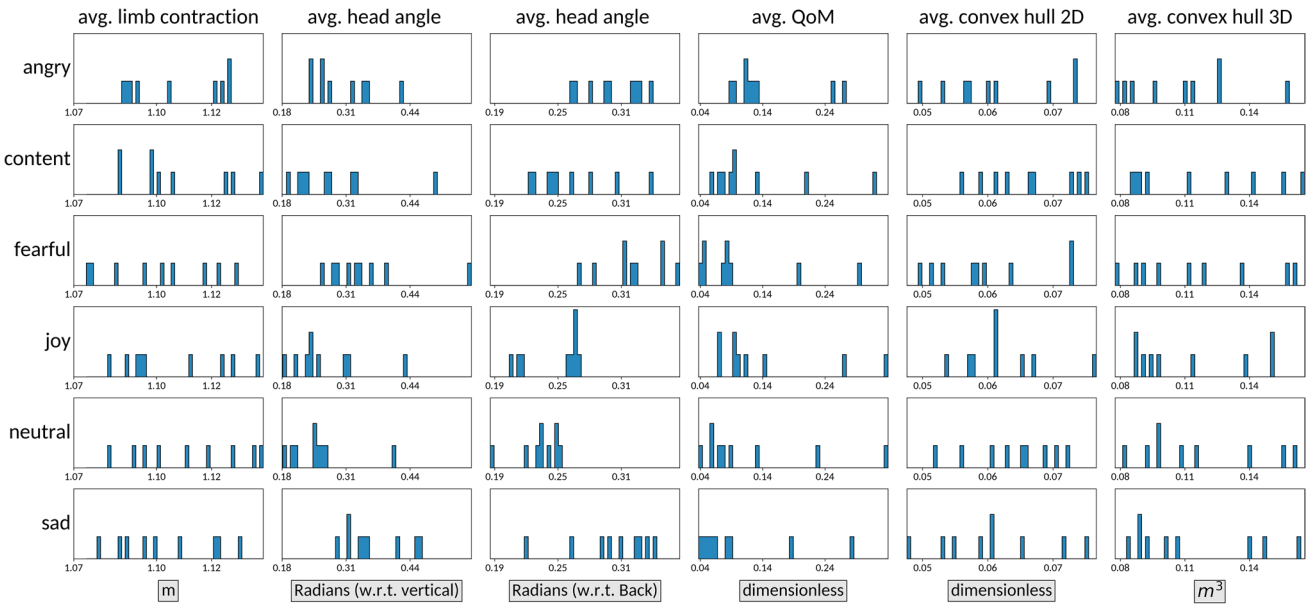
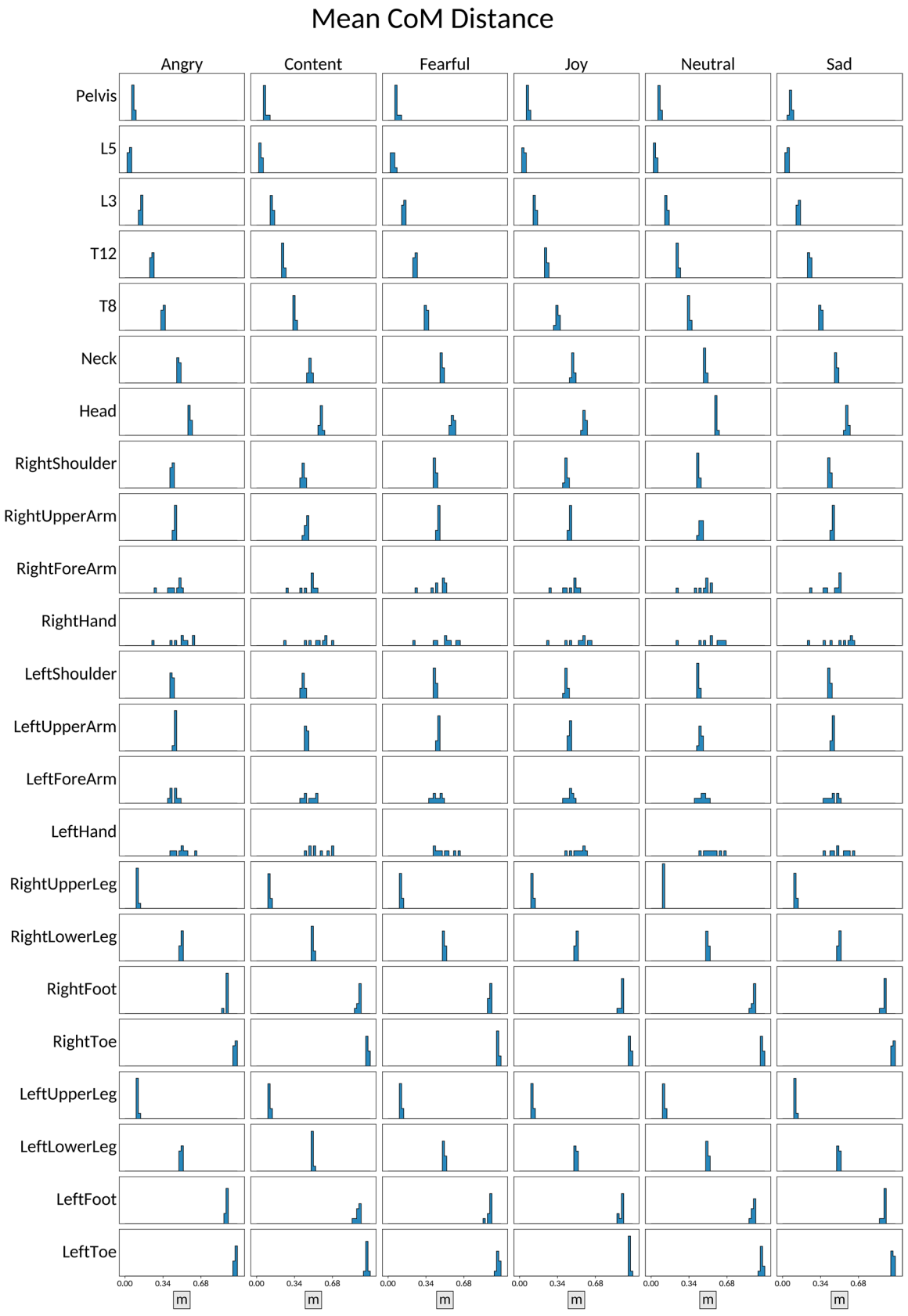



**Fig. 7 Kinematics that depend on silhouette images as a function of emotion**. Note: Qualitative histograms illustrating the distributions of the silhouette-dependent kinematics **average limb contraction, average head angle with respect to vertical, average head angle with respect to back, average quantity of motion (QoM), mean convex hull 2D, and mean convex hull 3D**. The kinematics are plotted separately for each emotion (angry, content, fearful, happy, neutral, sad) across all videos. Comparing the distribution of each kinematic feature across the different emotions (column wise, top to bottom), we observe differences in the distributions, as a function of intended emotion. For example, the distribution of the mean head angle for angry videos (top graph) is left skewed, while it is right skewed for sad videos (bottom graph). The *y*-axis shows number of videos. The *x*-axis is the scale of the respective kinematic (see label on the bottom of each column)


Mean CoM Distance
Angry
Content
Fearful
Joy
Neutral
Sad
Pelvis
L5
L3
T12
T8
Neck
Head
RightShoulder
RightUpperArm
RightForeArm
RightHand
LeftShoulder
LeftUpperArm
LeftForeArm
LeftHand
RightUpperLeg
RightLowerLeg
RightFoot
RightToe
LeftUpperLeg
LeftLowerLeg
LeftFoot
LeftToe
0.00
0.34
0.68
m

◂**Fig. 8** Distance to the center of mass. *Note*: Distance from center of mass (CoM). Histograms for each of the 23 keypoints given by the XSENS® system, across the six emotions (**angry, content, fearful, joy, neutral, sad**). The further away from the center of mass the keypoint is (pelvis = very close; hands = very far), the distribution changes. We also see that the legs remain relatively stable throughout, which is in accordance with the choreographies set out in Table 2 above: the movements were mostly confined to the arms, with little leg movement. We also observe that the distributions vary across emotions, especially for the arms, which again is in accordance with the intention of the dancer during stimuli creation, where the intention was to confine the expressivity mostly to the arms

And the $c_{3D}(t)$ feature we provide is the volume of $\mathcal{C}_{3D}(t)$, in cubic meters $m^3$ (since the MVNX input positions $p_j(t)$ are given in meters). We used the SciPy Python library to compute this feature. Apart from the average and MAD aggregated statistics, we provide the following 2 aggregations:

- Global convex hull $c_{3D}(1, \ldots, T_s)$: This is the convex hull obtained from all points in a given sequence (as opposed to a specific timepoint), i.e., it covers all the locations where any keypoints has been at any time.
- Union of convex hulls $\bigcup_{t=1}^{T_s} c_{3D}(t)$: The main difference with the global convex hull, is that the union of convex hulls is a subset of the global convex hull, and is not necessarily convex: if the dancer jumps from the bottom left corner of the screen to the bottom right corner, the bottom center of the screen is part of the global convex hull, but not of the union of convex hulls. The reason is that there is no single timepoint where the bottom center is being covered, but if we consider all timepoints at once, we need to connect the left and right corners through the bottom center, thus making it part of the global convex hull.

#### Convex hull 2D

We compute this feature analogously to the 3D convex hull described before. The difference for this version is that we use the CamPos two-dimensional coordinates as source for this feature, i.e., two-dimensional vectors $\rho_j(t) \in [0, 1]^2$ given in coordinates relative to the camera, going from $(0, 0)$ for the bottom-left corner to $(1, 1)$ for the top-right corner (i.e., it is a dimensionless ratio). The resulting convex hull is then a convex polygon (Boyd & Vandenberghe, 2004):

$$\mathcal{C}_{2D}(t) = \left\{ \sum_j \theta_j \rho_j(t) | j \in \mathcal{J}, \theta \geq 0, \sum_j \theta_j = 1 \right\}$$

And the computed feature $c_{2D}(t)$ is the surface of this polygon. Since the horizontal and vertical CamPos coordinates go from 0 to 1, the result is itself a (dimensionless) ratio between 0 and 1, telling how much of the total screen is covered by the convex hull of the dancer.

As with $c_{2D}(t)$, we provide four aggregations: average, MAD, global convex hull and union of convex hulls. We compute the surface of the 2D convex hull using the Python Shapely library.

## Validation of kinematic features and results of observer experiments

A series of validations of our kinematic and observer data are provided in this section. First, we provide a series of illustrations of the distributions of our kinematic features across the stimuli of the dataset. This is to compare the alignment of the stimuli between each other, to ensure that they are equivalent (e.g., between different visual presentations), and that differences emerge, where differences were expected (e.g., between different intended emotional expressions) (Section "Computational tests"). Second, we provide statistical tests of the observer ratings. In particular, we test that emotion recognition is above chance level (i.e., that the stimuli transmit the emotions that they were intended to transmit by the dancer), for all visual presentations, and we confirm that the beauty ratings vary as a function of the intended emotions by the dancer (Section "Results of observer experiments").

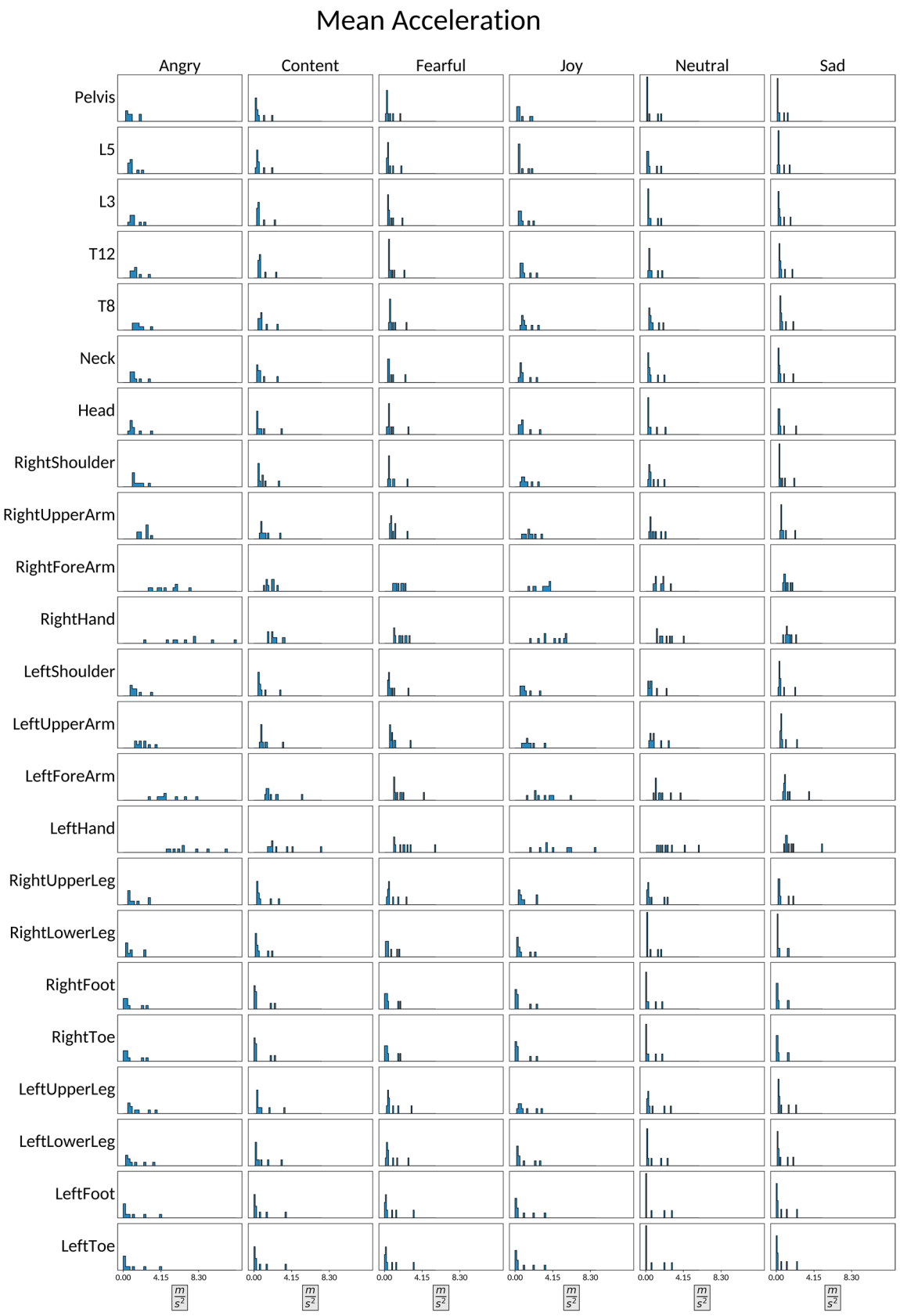
### Computational tests

We here provide a series of illustrations of computational validations, using either silhouette-dependent images or the keypoint-dependent data (that we retrieved from the XSENS® sensors) from our stimulus set. We use:

- foreground statistics to show that stimuli are balanced and within frames (Section "Foreground statistics").
- qualitative histograms (silhouettes) to show that stimuli are aligned with each other (Section "Qualitative histograms (silhouette-dependent kinematics)").
- kinematic histograms to show that the stimuli and features yield meaningful signal, not random noise (Section "Qualitative histograms (keypoint-dependent kinematics)").

#### Foreground statistics

The foreground statistics describe the distribution of space occupied by the dancer (= the foreground) across videos. Using the silhouette images of the stimuli, we see a very homogenous distribution of the foreground throughout all videos for our three metrics (foreground ratios, and camera position limits for horizontal and vertical distributions). These results indicate that the videos of the stimuli set are equivalent in terms of foreground


Mean Acceleration
Angry
Content
Fearful
Joy
Neutral
Sad
Pelvis
L5
L3
T12
T8
Neck
Head
RightShoulder
RightUpperArm
RightForeArm
RightHand
LeftShoulder
LeftUpperArm
LeftForeArm
LeftHand
RightUpperLeg
RightLowerLeg
RightFoot
RightToe
LeftUpperLeg
LeftLowerLeg
LeftFoot
LeftToe
0.00
4.15
8.30
m/s²

◂**Fig. 9** Mean acceleration. *Note*: Mean acceleration. Histograms for each of the 23 keypoints given by the XSENS® system, across the six emotions (**angry, content, fearful, joy, neutral, sad**). Acceleration changes are evident across emotions especially for arms and pelvis and the head. The legs remain relatively stable throughout, which is in accordance with the choreographies set out in Table 2: the movements were mostly confined to the arms, with little leg movement

distribution (see Fig. 5 for illustration and a short description of the findings).

### Qualitative histograms (silhouette-dependent kinematics)

We use qualitative histograms (silhouette images of the videos) to show that the stimuli are aligned with each other. We computed the frequencies of frames occupied for the silhouettes, convex hull, point-light displays (PLDs), and (D) the avatar stimuli, which all depend on silhouette images. The Silhouette occupies most space, then convex hull, avatar, PLDs. The histograms show a nice alignment across all four modalities. However, they also show a defect of the avatar (extracted from XSENS® software) – the software is automatically adjusting the camera position. In videos 7–9, the dancer turns the upper body, which is corrected by the software so the camera position remains frontal throughout. This results in the very symmetric histograms in that column, while the other columns show movement also to the sides. This means that as long as no turns are in the movement, the four modalities are aligned in terms of the frequencies of frames occupied by the dancer in space (see Fig. 6 for illustration and a short description of the findings for two of the nine sequences; sequence 1 and 7). The illustrations for the remaining sequences are in the appendix of the paper (see Figs. 15, 16, 17, 18, 19).

We use another series of qualitative histograms to explore visually the kinematics of average limb contraction, mean head angle (with respect to vertical and back), average quantity of motion (QoM), mean convex hull 2D and mean convex hull 3D. We plotted these separately for each emotion (angry, content, fearful, happy, neutral, sad), across all videos (see Fig. 7 for illustration and a short description of the findings).

### Qualitative histograms (keypoint-dependent kinematics)

For the keypoint-dependent kinematics (retrieved from the XSENS® sensors), we provide illustrations for the distance to the center of mass (Fig. 8) and average acceleration (Fig. 9). The figures show histograms for each of the 23 keypoints across the six emotions (angry, content, fearful, joy, neutral, sad). The distance to center of mass figure shows that the further away from the center of mass the keypoint is (pelvis = very close; hands = very far), the distribution changes. We also see that the legs remain relatively stable throughout, which is in accordance with the choreographies set out in Table 2 above: the movements were mostly confined to the arms, with little leg movement. Hence, these distributions confirm the choreographies. We also observe that the distributions vary across emotions especially for the arms, which again is in accordance with the intention of the

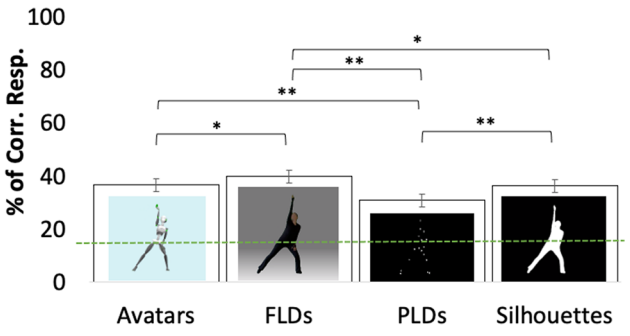



**Fig. 10** Emotion recognition rate: Main effect of visual presentation. *Note*: Observer emotion recognition scores as a function of visual presentation was above chance level (16.67%) for all four visual presentations (*dotted line* illustrates chance level). Observer emotion recognition was highest for FLDs than for all other visual presentations, while PLDs had the lowest observer emotion recognition rates. Observer recognition rates for avatars and silhouettes were higher than recognition rates of PLDs. Observer recognition rates for avatars and silhouettes did not differ between each other. ** $p > .001$. * $p > .02$. *Bars* show averages; *error bars* represent SE. *Dotted line* represents emotion recognition chance level at 16.5%. FLD = full-light displays; PLDs = point-light displays

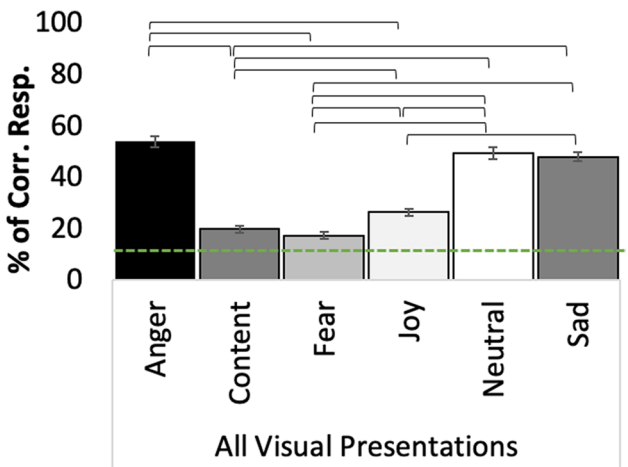



**Fig. 11** Emotion recognition rate: Main effect of emotion. *Note*: Observer emotion recognition scores as a function of emotion (regardless of visual presentation) was above chance level (16.67%) for five of the emotions intended for the dancer (anger, content, joy, neutral, sad), but not for fear (*dotted line* illustrates chance level). Observer emotion recognition rates were highest for angry videos > neutral videos > sad videos > joyful video > content videos > fearful videos. There were significant differences between all pairs of emotions (all *p*s > .01), except between the pairs anger-neutral, content-fear, and sad-neutral. *Brackets* indicate significant differences, all $p > .001$, except the pair anger-sad, which was $p = .015$

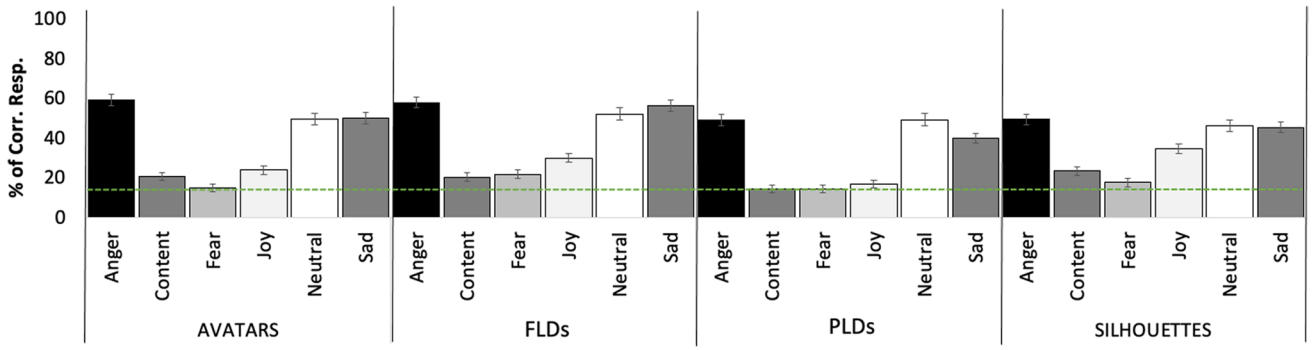



**Fig. 12** Visual presentation x emotion for percentage of correct responses. *Note*: Illustrating observer emotion recognition rates for the different emotions between the different visual presentations, for all emotion categories. For the avatars, all emotions were recognized above chance (all *p*s > .050), except for fear ($p = .363$). For the FLDs, all emotions were recognized above chance (all *p*s > .016), except for content ($p = .071$). For the PLDs, emotions were recognized above chance (all *p*s > .001), except for content ($p = .226$), fear ($p = .315$) and joy ($p = .898$). For the silhouettes, emotions were recognized above chance (all *p*s > .050), except for fear ($p = .643$). *Bars* show emotion averages; *dotted line* illustrates chance level; *error bars* represent SE. *FLD* full-light displays; *PLDs* point-light displays

dancer during stimuli creation, where the intention was to confine the expressivity mostly to the arms (see Figs. 8 and 9 for illustrations and a short description of the findings).

The remaining keypoint-dependent kinematics are provided in the appendix; for dimensionless jerk, see Fig. 20, for angular acceleration, see Fig. 21, and for angular velocity, see Fig. 22.

## Results of observer experiments

As described in Sections "Participants' online experiment" and "Procedure for Human Observer Experiment", human observers performed an emotion recognition task and an aesthetic judgment task on all pilot stimuli. We here provide a technical test to ensure that:

- Intended emotional expression was recognized above chance level (Sections "Chance level analysis: Visual presentation", "Chance Level Analysis: Emotion Category", and "Chance Level Analysis: Visual Presentation x Emotion Category").
- Beauty ratings depended on the intended emotional expression of the dancer (Section "Observer experiment beauty ratings").

### Chance level analysis: Visual presentation

Chi-square *t* tests were used to determine whether observer recognition rates were above chance for the four visual presentations (four levels; avatars, full-light displays (FLDs), point-light displays (PLDs), silhouettes; chance level: 100%/6 emotion categories = 16.67%). Results showed that the stimuli of all four visual presentations had been recognized above chance level (all *p*s < .001).

To explore whether there was a difference between the four visual presentations in terms of emotion recognition accuracy, we performed a 1 x 4 RM ANOVA with the factor *visual presentation* (four levels; avatars, FLDs, PLDs, silhouettes), and the dependent variable *percent of correct responses* ('correct responses' = when observers guessed the emotion that the dancer was intending while performing the movements). There was a main effect of *visual presentation* ($F(3,393) = 21.352$, $p < .001$, partial $\eta^2 = .140$). Estimated marginal means showed that FLD videos were recognized best (EMM = 39.82%; SE = 1.12), followed by avatar videos (EMM = 36.45%; SE = .98), then silhouette videos (EMM = 36.20%; SE = 1.10), and finally PLDs (EMM = 30.77%; SE = .98). Bonferroni corrected pair-wise comparisons showed

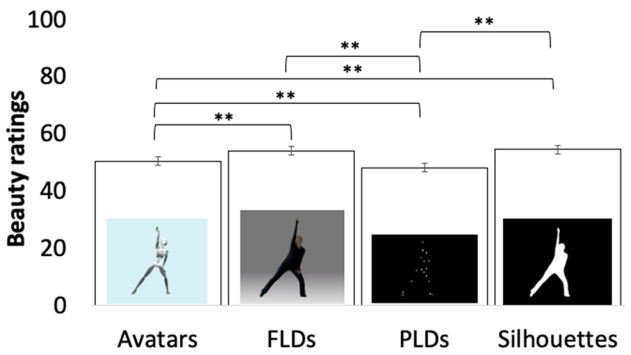



**Fig. 13** Beauty ratings: Main effect of visual presentation. *Note*: Observer beauty ratings as a function of visual presentation from 0 (not beautiful) to 100 (very beautiful). Observer beauty ratings were highest for silhouette stimuli followed by FLDs that did not differ between each other ($p = .280$), followed by avatar stimuli, and, finally, PLDs which all differed significantly between each other (all *p*s > .001). *Bars* show averages; *error bars* represent SE. *FLD* full-light displays, *PLDs* point-light displays

**Table 4** Confusion matrix: EMOKINE avatar stimuli

| Clip name | Intended emotion by dancer | Decoded emotion by human observers (%) | | | | | | Responses (%) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Angry | Content | Fear | Joy | Neutral | Sad | Incorrect | Correct |
| seq1_angry | Angry | 66.67 | 6.67 | 6.67 | 6.67 | 13.33 | 0.00 | 33.33 | 66.67 |
| seq2_angry | Angry | 59.85 | 5.30 | 3.79 | 20.45 | 8.33 | 2.27 | 40.15 | 59.85 |
| seq3_angry | Angry | 88.89 | 0.00 | 2.78 | 0.00 | 5.56 | 2.78 | 11.11 | 88.89 |
| seq4_angry | Angry | 86.11 | 0.00 | 2.78 | 0.00 | 8.33 | 2.78 | 13.89 | 86.11 |
| seq5_angry | Angry | 50.00 | 3.13 | 15.63 | 12.50 | 18.75 | 0.00 | 50.00 | 50.00 |
| seq6_angry | Angry | 28.13 | 9.38 | 9.38 | 12.50 | 34.38 | 6.25 | 71.88 | 28.13 |
| seq7_angry | Angry | 42.42 | 12.12 | 3.03 | 15.15 | 21.21 | 6.06 | 57.58 | 42.42 |
| Seq8_angry | Angry | 51.52 | 6.06 | 9.09 | 15.15 | 18.18 | 0.00 | 48.48 | 51.52 |
| seq9_angry | Angry | 51.61 | 6.45 | 6.45 | 19.35 | 16.13 | 0.00 | 48.39 | 51.61 |
| seq1_content | Content | 0.00 | 19.35 | 16.13 | 9.68 | 0.00 | 54.84 | 80.65 | 19.35 |
| seq2_content | Content | 20.45 | 12.12 | 3.03 | 36.36 | 22.73 | 5.30 | 87.88 | 12.12 |
| seq3_content | Content | 2.78 | 30.56 | 11.11 | 5.56 | 22.22 | 27.78 | 69.44 | 30.56 |
| seq4_content | Content | 2.78 | 19.44 | 0.00 | 13.89 | 44.44 | 19.44 | 80.56 | 19.44 |
| seq5_content | Content | 0.00 | 28.13 | 6.25 | 21.88 | 21.88 | 21.88 | 71.88 | 28.13 |
| seq6_content | Content | 0.00 | 21.88 | 3.13 | 6.25 | 31.25 | 37.50 | 78.13 | 21.88 |
| seq7_content | Content | 3.03 | 30.30 | 0.00 | 15.15 | 36.36 | 15.15 | 69.70 | 30.30 |
| seq8_content | Content | 3.03 | 36.36 | 3.03 | 0.00 | 30.30 | 27.27 | 63.64 | 36.36 |
| seq9_content | Content | 0.00 | 12.90 | 3.23 | 48.39 | 19.35 | 16.13 | 87.10 | 12.90 |
| seq1_fearful | Fear | 3.23 | 29.03 | 19.35 | 6.45 | 19.35 | 22.58 | 77.42 | 22.58 |
| seq2_fearful | Fear | 6.82 | 16.67 | 12.12 | 18.94 | 29.55 | 15.91 | 84.09 | 15.91 |
| seq3_fearful | Fear | 2.78 | 13.89 | 16.67 | 0.00 | 33.33 | 33.33 | 66.67 | 33.33 |
| seq4_fearful | Fear | 2.78 | 19.44 | 11.11 | 0.00 | 33.33 | 33.33 | 66.67 | 33.33 |
| seq5_fearful | Fear | 0.00 | 12.50 | 28.13 | 0.00 | 21.88 | 37.50 | 62.50 | 37.50 |
| seq6_fearful | Fear | 12.50 | 9.38 | 12.50 | 9.38 | 25.00 | 31.25 | 68.75 | 31.25 |
| seq7_fearful | Fear | 0.00 | 12.12 | 15.15 | 6.06 | 27.27 | 39.39 | 60.61 | 39.39 |
| seq8_fearful | Fear | 0.00 | 6.06 | 21.21 | 0.00 | 30.30 | 42.42 | 57.58 | 42.42 |
| seq9_fearful | Fear | 3.23 | 25.81 | 6.45 | 9.68 | 41.94 | 12.90 | 87.10 | 12.90 |

**Table 4** (continued)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| seq1_joy | Joy | 0.00 | 25.81 | 9.68 | 12.90 | 16.13 | 35.48 | 87.10 | 12.90 |
| seq2_joy | Joy | 13.64 | 9.09 | 4.55 | 27.27 | 36.36 | 9.09 | 72.73 | 27.27 |
| seq3_joy | Joy | 13.89 | 16.67 | 8.33 | 5.56 | 27.78 | 27.78 | 94.44 | 5.56 |
| seq4_joy | Joy | 0.00 | 27.78 | 5.56 | 19.44 | 36.11 | 11.11 | 80.56 | 19.44 |
| seq5_joy | Joy | 3.13 | 31.25 | 3.13 | 34.38 | 18.75 | 9.38 | 65.63 | 34.38 |
| seq6_joy | Joy | 9.38 | 15.63 | 0.00 | 6.25 | 53.13 | 15.63 | 93.75 | 6.25 |
| seq7_joy | Joy | 9.09 | 21.21 | 3.03 | 12.12 | 33.33 | 21.21 | 87.88 | 12.12 |
| seq8_joy | Joy | 9.09 | 12.12 | 9.09 | 36.36 | 15.15 | 18.18 | 63.64 | 36.36 |
| seq9_joy | Joy | 12.90 | 12.90 | 6.45 | 58.06 | 9.68 | 0.00 | 41.94 | 58.06 |
| seq1_neutral | Neutral | 3.23 | 29.03 | 9.68 | 3.23 | 38.71 | 16.13 | 61.29 | 38.71 |
| seq2_neutral | Neutral | 6.82 | 20.45 | 4.55 | 18.94 | 43.18 | 6.06 | 56.82 | 43.18 |
| seq3_neutral | Neutral | 0.00 | 25.00 | 2.78 | 0.00 | 52.78 | 19.44 | 47.22 | 52.78 |
| seq4_neutral | Neutral | 0.00 | 13.89 | 0.00 | 0.00 | 69.44 | 16.67 | 30.56 | 69.44 |
| seq5_neutral | Neutral | 0.00 | 31.25 | 6.25 | 0.00 | 56.25 | 6.25 | 43.75 | 56.25 |
| seq6_neutral | Neutral | 0.00 | 15.63 | 9.38 | 6.25 | 50.00 | 18.75 | 50.00 | 50.00 |
| seq7_neutral | Neutral | 3.03 | 21.21 | 0.00 | 3.03 | 51.52 | 21.21 | 48.48 | 51.52 |
| seq8_neutral | Neutral | 0.00 | 18.18 | 12.12 | 3.03 | 42.42 | 24.24 | 57.58 | 42.42 |
| seq9_neutral | Neutral | 6.45 | 16.13 | 3.23 | 12.90 | 58.06 | 3.23 | 41.94 | 58.06 |
| seq1_sad | Sad | 0.00 | 16.13 | 9.68 | 0.00 | 3.23 | 70.97 | 29.03 | 70.97 |
| seq2_sad | Sad | 3.03 | 20.45 | 3.03 | 9.85 | 29.55 | 34.09 | 65.91 | 34.09 |
| seq3_sad | Sad | 2.78 | 16.67 | 5.56 | 5.56 | 22.22 | 47.22 | 52.78 | 47.22 |
| seq4_sad | Sad | 2.78 | 13.89 | 2.78 | 2.78 | 30.56 | 47.22 | 52.78 | 47.22 |
| seq5_sad | Sad | 0.00 | 18.75 | 6.25 | 0.00 | 25.00 | 50.00 | 50.00 | 50.00 |
| seq6_sad | Sad | 0.00 | 6.25 | 9.38 | 3.13 | 9.38 | 71.88 | 28.13 | 71.88 |
| seq7_sad | Sad | 0.00 | 15.15 | 0.00 | 9.09 | 12.12 | 63.64 | 36.36 | 63.64 |
| seq8_sad | Sad | 0.00 | 12.12 | 6.06 | 0.00 | 6.06 | 75.76 | 24.24 | 75.76 |
| seq9_sad | Sad | 0.00 | 6.45 | 3.23 | 12.90 | 35.48 | 41.94 | 58.06 | 41.94 |

Avatar stimuli. Percent of responses as a function of intended emotion of 54 avatar stimuli. *Left side columns* indicate clip name and the intended emotional expression by the dancer when performing the movement ("Intended Emotion by dancer"). The percentages of emotion judgments made by the human observers in the experiment are represented in each of the columns under the label "Decoded emotion". The confusion matrix illustrates the distribution of the given emotion judgment for each of the different emotion options. The *grey shading* indicates when intended and decoded emotion correspond. The *two columns on the right* set out the total incorrect and correct responses for each clip (i.e., when the intended and the decoded emotion where the same -correct-, or not -incorrect-)

**Table 5** Confusion matrix: EMOKINE full-light displays (FLDs) stimuli

| Clip name | emotion by dancer | Angry | Content | Fear | Joy | Neutral | Sad | Incorrect | Correct |
|---|---|---|---|---|---|---|---|---|---|
| seq1_angry | Angry | 80.65 | 6.45 | 0.00 | 3.23 | 6.45 | 3.23 | 19.35 | 80.65 |
| seq2_angry | Angry | 49.24 | 7.58 | 9.09 | 19.70 | 11.36 | 3.03 | 50.76 | 49.24 |
| seq3_angry | Angry | 58.33 | 5.56 | 13.89 | 2.78 | 13.89 | 5.56 | 41.67 | 58.33 |
| seq4_angry | Angry | 83.33 | 0.00 | 5.56 | 2.78 | 5.56 | 2.78 | 16.67 | 83.33 |
| seq5_angry | Angry | 75.00 | 6.25 | 0.00 | 0.00 | 15.63 | 3.13 | 25.00 | 75.00 |
| seq6_angry | Angry | 40.63 | 3.13 | 6.25 | 3.13 | 25.00 | 21.88 | 59.38 | 40.63 |
| seq7_angry | Angry | 48.48 | 6.06 | 9.09 | 3.03 | 24.24 | 9.09 | 51.52 | 48.48 |
| Seq8_angry | Angry | 60.61 | 6.06 | 6.06 | 9.09 | 15.15 | 3.03 | 39.39 | 60.61 |
| seq9_angry | Angry | 48.39 | 3.23 | 6.45 | 12.90 | 29.03 | 0.00 | 51.61 | 48.39 |
| seq1_content | Content | 3.23 | 25.81 | 3.23 | 19.35 | 19.35 | 29.03 | 74.19 | 25.81 |
| seq2_content | Content | 14.39 | 14.39 | 3.03 | 34.85 | 25.00 | 8.33 | 85.61 | 14.39 |
| seq3_content | Content | 0.00 | 11.11 | 19.44 | 11.11 | 13.89 | 44.44 | 88.89 | 11.11 |
| seq4_content | Content | 2.78 | 16.67 | 25.00 | 5.56 | 11.11 | 38.89 | 83.33 | 16.67 |
| seq5_content | Content | 0.00 | 31.25 | 0.00 | 6.25 | 21.88 | 40.63 | 68.75 | 31.25 |
| seq6_content | Content | 0.00 | 6.25 | 9.38 | 15.63 | 25.00 | 43.75 | 93.75 | 6.25 |
| seq7_content | Content | 9.09 | 24.24 | 6.06 | 30.30 | 27.27 | 3.03 | 75.76 | 24.24 |
| seq8_content | Content | 3.03 | 33.33 | 6.06 | 12.12 | 30.30 | 15.15 | 66.67 | 33.33 |
| seq9_content | Content | 3.23 | 41.94 | 3.23 | 19.35 | 19.35 | 12.90 | 58.06 | 41.94 |
| seq1_fearful | Fear | 10.00 | 16.67 | 13.33 | 3.33 | 26.67 | 30.00 | 70.00 | 30.00 |
| seq2_fearful | Fear | 12.12 | 9.85 | 18.18 | 16.67 | 14.39 | 28.79 | 71.21 | 28.79 |
| seq3_fearful | Fear | 0.00 | 8.33 | 41.67 | 0.00 | 5.56 | 44.44 | 55.56 | 44.44 |
| seq4_fearful | Fear | 5.56 | 19.44 | 33.33 | 0.00 | 16.67 | 25.00 | 75.00 | 25.00 |
| seq5_fearful | Fear | 9.38 | 6.25 | 18.75 | 0.00 | 37.50 | 28.13 | 71.88 | 28.13 |
| seq6_fearful | Fear | 15.63 | 3.13 | 15.63 | 0.00 | 21.88 | 43.75 | 56.25 | 43.75 |
| seq7_fearful | Fear | 0.00 | 21.21 | 27.27 | 0.00 | 33.33 | 18.18 | 81.82 | 18.18 |
| seq8_fearful | Fear | 0.00 | 15.15 | 30.30 | 0.00 | 15.15 | 39.39 | 60.61 | 39.39 |
| seq9_fearful | Fear | 6.45 | 25.81 | 6.45 | 12.90 | 19.35 | 29.03 | 70.97 | 29.03 |
| seq1_joy | Joy | 9.68 | 22.58 | 0.00 | 12.90 | 22.58 | 32.26 | 87.10 | 12.90 |

**Table 5** (continued)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| seq2_joy | Joy | 18.32 | 8.40 | 5.34 | 37.40 | 22.14 | 8.40 | 62.60 | 37.40 |
| seq3_joy | Joy | 11.11 | 19.44 | 11.11 | 0.00 | 16.67 | 41.67 | 100.00 | 0.00 |
| seq4_joy | Joy | 2.78 | 13.89 | 11.11 | 33.33 | 30.56 | 8.33 | 66.67 | 33.33 |
| seq5_joy | Joy | 0.00 | 6.25 | 12.50 | 53.13 | 21.88 | 6.25 | 46.88 | 53.13 |
| seq6_joy | Joy | 9.38 | 15.63 | 12.50 | 21.88 | 21.88 | 18.75 | 78.13 | 21.88 |
| seq7_joy | Joy | 3.03 | 27.27 | 3.03 | 9.09 | 30.30 | 27.27 | 90.91 | 9.09 |
| seq8_joy | Joy | 6.06 | 18.18 | 3.03 | 36.36 | 15.15 | 21.21 | 63.64 | 36.36 |
| seq9_joy | Joy | 12.90 | 12.90 | 0.00 | 48.39 | 22.58 | 3.23 | 51.61 | 48.39 |
| seq1_neutral | Neutral | 3.23 | 12.90 | 0.00 | 0.00 | 48.39 | 35.48 | 51.61 | 48.39 |
| seq2_neutral | Neutral | 7.58 | 23.48 | 2.27 | 15.15 | 40.91 | 10.61 | 59.09 | 40.91 |
| seq3_neutral | Neutral | 0.00 | 33.33 | 0.00 | 2.78 | 52.78 | 11.11 | 47.22 | 52.78 |
| seq4_neutral | Neutral | 0.00 | 19.44 | 0.00 | 0.00 | 69.44 | 11.11 | 30.56 | 69.44 |
| seq5_neutral | Neutral | 3.13 | 6.25 | 0.00 | 3.13 | 71.88 | 15.63 | 28.13 | 71.88 |
| seq6_neutral | Neutral | 0.00 | 15.63 | 3.13 | 9.38 | 53.13 | 18.75 | 46.88 | 53.13 |
| seq7_neutral | Neutral | 0.00 | 18.18 | 6.06 | 6.06 | 51.52 | 18.18 | 48.48 | 51.52 |
| seq8_neutral | Neutral | 0.00 | 27.27 | 0.00 | 9.09 | 54.55 | 9.09 | 45.45 | 54.55 |
| seq9_neutral | Neutral | 0.00 | 19.35 | 3.23 | 12.90 | 61.29 | 3.23 | 38.71 | 61.29 |
| seq1_sad | Sad | 0.00 | 6.45 | 3.23 | 3.23 | 3.23 | 83.87 | 16.13 | 83.87 |
| seq2_sad | Sad | 4.55 | 11.36 | 5.30 | 13.64 | 22.73 | 42.42 | 57.58 | 42.42 |
| seq3_sad | Sad | 0.00 | 22.22 | 13.89 | 0.00 | 5.56 | 58.33 | 41.67 | 58.33 |
| seq4_sad | Sad | 0.00 | 33.33 | 5.56 | 2.78 | 16.67 | 41.67 | 58.33 | 41.67 |
| seq5_sad | Sad | 0.00 | 21.88 | 6.25 | 3.13 | 18.75 | 50.00 | 50.00 | 50.00 |
| seq6_sad | Sad | 0.00 | 6.25 | 9.38 | 3.13 | 18.75 | 62.50 | 37.50 | 62.50 |
| seq7_sad | Sad | 0.00 | 15.15 | 0.00 | 3.03 | 18.18 | 63.64 | 36.36 | 63.64 |
| seq8_sad | Sad | 0.00 | 15.15 | 0.00 | 0.00 | 6.06 | 78.79 | 21.21 | 78.79 |
| seq9_sad | Sad | 0.00 | 9.68 | 6.45 | 6.45 | 6.45 | 70.97 | 29.03 | 70.97 |

Full-light displays (FLDs) stimuli. Percent of responses as a function of intended emotion of 54 FLD stimuli. *Left side columns* indicate clip name and the intended emotional expression by the dancer when performing the movement ("Intended emotion by dancer"). The percentages of emotion judgments made by the human observers in the experiment are represented in each of the columns under the label "Decoded emotion". The confusion matrix illustrates the distribution of the given emotion judgment for each of the different emotion options. The grey shading indicates when intended and decoded emotion correspond. The *two columns on the right* set out the total incorrect and correct responses for each clip (i.e., when the intended and the decoded emotion where the same -correct-, or not -incorrect-)

**Table 6** Confusion matrix: EMOKINE point-light displays (PLDs) stimuli

| Clip name | Intended emotion by dancer | Decoded emotion by human observers (%) | | | | | | Responses (%) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Angry | Content | Fear | Joy | Neutral | Sad | Incorrect | Correct |
| seq1_angry | Angry | **58.06** | 6.45 | 3.23 | 3.23 | 22.58 | 6.45 | 41.94 | **58.06** |
| seq2_angry | Angry | **43.94** | 4.55 | 6.06 | 22.73 | 15.91 | 6.82 | 56.06 | **43.94** |
| seq3_angry | Angry | **88.89** | 0.00 | 0.00 | 0.00 | 8.33 | 2.78 | 11.11 | **88.89** |
| seq4_angry | Angry | **72.22** | 5.56 | 0.00 | 0.00 | 16.67 | 5.56 | 27.78 | **72.22** |
| seq5_angry | Angry | **40.63** | 9.38 | 9.38 | 9.38 | 21.88 | 9.38 | 59.38 | **40.63** |
| seq6_angry | Angry | **40.63** | 6.25 | 12.50 | 3.13 | 31.25 | 6.25 | 59.38 | **40.63** |
| seq7_angry | Angry | **30.30** | 6.06 | 3.03 | 9.09 | 36.36 | 15.15 | 69.70 | **30.30** |
| Seq8_angry | Angry | **45.45** | 6.06 | 3.03 | 9.09 | 27.27 | 9.09 | 54.55 | **45.45** |
| seq9_angry | Angry | **32.26** | 9.68 | 12.90 | 19.35 | 25.81 | 0.00 | 67.74 | **32.26** |
| seq1_content | Content | 0.00 | **12.90** | 3.23 | 6.45 | 32.26 | 45.16 | 87.10 | **12.90** |
| seq2_content | Content | 8.33 | **12.88** | 6.06 | 25.00 | 40.15 | 7.58 | 87.12 | **12.88** |
| seq3_content | Content | 2.78 | **5.56** | 11.11 | 2.78 | 38.89 | 38.89 | 94.44 | **5.56** |
| seq4_content | Content | 0.00 | **22.22** | 5.56 | 8.33 | 41.67 | 22.22 | 77.78 | **22.22** |
| seq5_content | Content | 0.00 | **15.63** | 6.25 | 9.38 | 43.75 | 25.00 | 84.38 | **15.63** |
| seq6_content | Content | 0.00 | **12.50** | 6.25 | 3.13 | 34.38 | 43.75 | 87.50 | **12.50** |
| seq7_content | Content | 6.06 | **18.18** | 0.00 | 6.06 | 54.55 | 15.15 | 81.82 | **18.18** |
| seq8_content | Content | 3.03 | **15.15** | 9.09 | 3.03 | 45.45 | 24.24 | 84.85 | **15.15** |
| seq9_content | Content | 6.45 | **19.35** | 3.23 | 29.03 | 35.48 | 6.45 | 80.65 | **19.35** |
| seq1_fearful | Fear | 6.45 | 25.81 | **19.35** | 3.23 | 19.35 | 25.81 | 74.19 | **25.81** |
| seq2_fearful | Fear | 6.82 | 12.12 | **9.09** | 19.70 | 34.85 | 17.42 | 82.58 | **17.42** |
| seq3_fearful | Fear | 0.00 | 5.56 | **16.67** | 0.00 | 27.78 | 50.00 | 50.00 | **50.00** |
| seq4_fearful | Fear | 0.00 | 8.33 | **27.78** | 0.00 | 38.89 | 25.00 | 75.00 | **25.00** |
| seq5_fearful | Fear | 0.00 | 21.88 | **3.13** | 6.25 | 34.38 | 34.38 | 65.63 | **34.38** |
| seq6_fearful | Fear | 0.00 | 3.13 | **21.88** | 0.00 | 18.75 | 56.25 | 43.75 | **56.25** |
| seq7_fearful | Fear | 3.03 | 21.21 | **9.09** | 3.03 | 42.42 | 21.21 | 78.79 | **21.21** |
| seq8_fearful | Fear | 6.06 | 0.00 | **24.24** | 3.03 | 33.33 | 33.33 | 66.67 | **33.33** |
| seq9_fearful | Fear | 3.23 | 9.68 | **16.13** | 6.45 | 32.26 | 32.26 | 67.74 | **32.26** |

**Table 6** (continued)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| seq1_joy | Joy | 9.68 | 6.45 | 16.13 | **12.90** | 29.03 | 25.81 | 87.10 | **12.90** |
| seq2_joy | Joy | 14.39 | 7.58 | 6.82 | **25.76** | 35.61 | 9.85 | 74.24 | **25.76** |
| seq3_joy | Joy | 27.78 | 5.56 | 2.78 | **2.78** | 47.22 | 13.89 | 97.22 | **2.78** |
| seq4_joy | Joy | 2.78 | 25.00 | 11.11 | **2.78** | 36.11 | 22.22 | 97.22 | **2.78** |
| seq5_joy | Joy | 9.38 | 15.63 | 6.25 | **18.75** | 28.13 | 21.88 | 81.25 | **18.75** |
| seq6_joy | Joy | 0.00 | 3.13 | 3.13 | **0.00** | 78.13 | 15.63 | 100.00 | **0.00** |
| seq7_joy | Joy | 9.09 | 15.15 | 6.06 | **9.09** | 30.30 | 30.30 | 90.91 | **9.09** |
| seq8_joy | Joy | 18.18 | 30.30 | 3.03 | **12.12** | 33.33 | 3.03 | 87.88 | **12.12** |
| seq9_joy | Joy | 9.68 | 12.90 | 6.45 | **45.16** | 25.81 | 0.00 | 54.84 | **45.16** |
| seq1_neutral | Neutral | 3.23 | 9.68 | 0.00 | 0.00 | **38.71** | 48.39 | 61.29 | **38.71** |
| seq2_neutral | Neutral | 4.55 | 15.91 | 5.30 | 16.67 | **44.70** | 12.88 | 55.30 | **44.70** |
| seq3_neutral | Neutral | 0.00 | 11.11 | 0.00 | 2.78 | **52.78** | 33.33 | 47.22 | **52.78** |
| seq4_neutral | Neutral | 0.00 | 11.11 | 8.33 | 0.00 | **58.33** | 22.22 | 41.67 | **58.33** |
| seq5_neutral | Neutral | 0.00 | 18.75 | 0.00 | 3.13 | **53.13** | 25.00 | 46.88 | **53.13** |
| seq6_neutral | Neutral | 0.00 | 6.25 | 0.00 | 3.13 | **65.63** | 25.00 | 34.38 | **65.63** |
| seq7_neutral | Neutral | 0.00 | 9.09 | 0.00 | 6.06 | **45.45** | 39.39 | 54.55 | **45.45** |
| seq8_neutral | Neutral | 0.00 | 6.06 | 12.12 | 9.09 | **45.45** | 27.27 | 54.55 | **45.45** |
| seq9_neutral | Neutral | 9.68 | 16.13 | 3.23 | 9.68 | **51.61** | 9.68 | 48.39 | **51.61** |
| seq1_sad | Sad | 3.23 | 6.45 | 0.00 | 0.00 | 9.68 | **80.65** | 19.35 | **80.65** |
| seq2_sad | Sad | 8.33 | 15.91 | 6.06 | 16.67 | 41.67 | **11.36** | 88.64 | **11.36** |
| seq3_sad | Sad | 0.00 | 19.44 | 0.00 | 2.78 | 22.22 | **55.56** | 44.44 | **55.56** |
| seq4_sad | Sad | 0.00 | 19.44 | 2.78 | 0.00 | 38.89 | **38.89** | 61.11 | **38.89** |
| seq5_sad | Sad | 0.00 | 21.88 | 12.50 | 6.25 | 25.00 | **34.38** | 65.63 | **34.38** |
| seq6_sad | Sad | 0.00 | 6.25 | 12.50 | 3.13 | 12.50 | **65.63** | 34.38 | **65.63** |
| seq7_sad | Sad | 0.00 | 3.03 | 0.00 | 0.00 | 27.27 | **69.70** | 30.30 | **69.70** |
| seq8_sad | Sad | 0.00 | 9.09 | 9.09 | 3.03 | 24.24 | **54.55** | 45.45 | **54.55** |
| seq9_sad | Sad | 0.00 | 6.45 | 6.45 | 9.68 | 38.71 | **38.71** | 61.29 | **38.71** |

Point-light display (PLD) stimuli. Percent of responses as a function of intended emotion of 54 PLD stimuli. *Left side columns* indicate clip name and the intended emotional expression by the dancer when performing the movement ("Intended emotion by dancer"). The percentages of emotion judgments made by the human observers in the experiment are represented in each of the columns under the label "Decoded emotion". The confusion matrix illustrates the distribution of the given emotion judgment for each of the different emotion options. The *grey shading* indicates when intended and decoded emotion correspond. The *two columns on the right* set out the total incorrect and correct responses for each clip (i.e., when the intended and the decoded emotion where the same -correct-, or not -incorrect-)

**Table 7** Confusion matrix: EMOKINE silhouette stimuli

| Clip name | Intended emotion by dancer | Decoded emotion by human observers (%) | | | | | | Responses (%) | |
|---|---|---|---|---|---|---|---|---|---|
| | | Angry | Content | Fear | Joy | Neutral | Sad | Incorrect | Correct |
| seq1_angry | Angry | **77.42** | 6.45 | 0.00 | 0.00 | 6.45 | 9.68 | 22.58 | **77.42** |
| seq2_angry | Angry | **40.15** | 4.55 | 11.36 | 28.03 | 12.12 | 3.79 | 59.85 | **40.15** |
| seq3_angry | Angry | **44.44** | 5.56 | 11.11 | 13.89 | 13.89 | 11.11 | 55.56 | **44.44** |
| seq4_angry | Angry | **75.00** | 2.78 | 2.78 | 11.11 | 2.78 | 5.56 | 25.00 | **75.00** |
| seq5_angry | Angry | **75.00** | 0.00 | 3.13 | 9.38 | 12.50 | 0.00 | 25.00 | **75.00** |
| seq6_angry | Angry | **34.38** | 6.25 | 3.13 | 12.50 | 28.13 | 15.63 | 65.63 | **34.38** |
| seq7_angry | Angry | **42.42** | 24.24 | 0.00 | 9.09 | 24.24 | 0.00 | 57.58 | **42.42** |
| Seq8_angry | Angry | **51.52** | 6.06 | 12.12 | 18.18 | 12.12 | 0.00 | 48.48 | **51.52** |
| seq9_angry | Angry | **32.26** | 9.68 | 12.90 | 29.03 | 12.90 | 3.23 | 67.74 | **32.26** |
| seq1_content | Content | 3.23 | **35.48** | 0.00 | 9.68 | 6.45 | 45.16 | 64.52 | **35.48** |
| seq2_content | Content | 8.33 | **12.12** | 9.09 | 39.39 | 23.48 | 7.58 | 87.88 | **12.12** |
| seq3_content | Content | 0.00 | **22.22** | 16.67 | 5.56 | 22.22 | 33.33 | 77.78 | **22.22** |
| seq4_content | Content | 0.00 | **22.22** | 0.00 | 19.44 | 55.56 | 2.78 | 77.78 | **22.22** |
| seq5_content | Content | 0.00 | **34.38** | 3.13 | 15.63 | 18.75 | 28.13 | 65.63 | **34.38** |
| seq6_content | Content | 0.00 | **28.13** | 0.00 | 15.63 | 18.75 | 37.50 | 71.88 | **28.13** |
| seq7_content | Content | 12.12 | **39.39** | 0.00 | 21.21 | 27.27 | 0.00 | 60.61 | **39.39** |
| seq8_content | Content | 0.00 | **27.27** | 0.00 | 12.12 | 36.36 | 24.24 | 72.73 | **27.27** |
| seq9_content | Content | 0.00 | **25.81** | 3.23 | 22.58 | 25.81 | 22.58 | 74.19 | **25.81** |
| seq1_fearful | Fear | 3.23 | 9.68 | **22.58** | 0.00 | 29.03 | 35.48 | 64.52 | **35.48** |
| seq2_fearful | Fear | 6.82 | 16.67 | **16.67** | 22.73 | 20.45 | 16.67 | 83.33 | **16.67** |
| seq3_fearful | Fear | 8.33 | 27.78 | **25.00** | 2.78 | 8.33 | 27.78 | 72.22 | **27.78** |
| seq4_fearful | Fear | 0.00 | 11.11 | **27.78** | 5.56 | 19.44 | 36.11 | 63.89 | **36.11** |
| seq5_fearful | Fear | 0.00 | 28.13 | **6.25** | 9.38 | 37.50 | 18.75 | 81.25 | **18.75** |
| seq6_fearful | Fear | 18.75 | 6.25 | **15.63** | 9.38 | 9.38 | 40.63 | 59.38 | **40.63** |
| seq7_fearful | Fear | 0.00 | 24.24 | **21.21** | 15.15 | 21.21 | 18.18 | 81.82 | **18.18** |
| seq8_fearful | Fear | 0.00 | 15.15 | **18.18** | 3.03 | 24.24 | 39.39 | 60.61 | **39.39** |
| seq9_fearful | Fear | 0.00 | 29.03 | **6.45** | 25.81 | 22.58 | 16.13 | 83.87 | **16.13** |

**Table 7** (continued)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| seq1_joy | Joy | 9.68 | 16.13 | 3.23 | **25.81** | 25.81 | 19.35 | 74.19 | **25.81** |
| seq2_joy | Joy | 18.94 | 12.88 | 3.79 | **43.94** | 15.91 | 4.55 | 56.06 | **43.94** |
| seq3_joy | Joy | 8.33 | 19.44 | 8.33 | **8.33** | 25.00 | 30.56 | 91.67 | **8.33** |
| seq4_joy | Joy | 2.78 | 25.00 | 8.33 | **11.11** | 33.33 | 19.44 | 88.89 | **11.11** |
| seq5_joy | Joy | 6.25 | 3.13 | 3.13 | **65.63** | 18.75 | 3.13 | 34.38 | **65.63** |
| seq6_joy | Joy | 15.63 | 15.63 | 3.13 | **15.63** | 40.63 | 9.38 | 84.38 | **15.63** |
| seq7_joy | Joy | 12.12 | 33.33 | 0.00 | **15.15** | 27.27 | 12.12 | 84.85 | **15.15** |
| seq8_joy | Joy | 6.06 | 36.36 | 0.00 | **45.45** | 6.06 | 6.06 | 54.55 | **45.45** |
| seq9_joy | Joy | 3.23 | 19.35 | 0.00 | **58.06** | 16.13 | 3.23 | 41.94 | **58.06** |
| seq1_neutral | Neutral | 3.23 | 29.03 | 3.23 | 3.23 | **38.71** | 22.58 | 61.29 | **38.71** |
| seq2_neutral | Neutral | 5.30 | 15.91 | 4.55 | 25.00 | **42.42** | 6.82 | 57.58 | **42.42** |
| seq3_neutral | Neutral | 0.00 | 33.33 | 0.00 | 2.78 | **47.22** | 16.67 | 52.78 | **47.22** |
| seq4_neutral | Neutral | 0.00 | 25.00 | 5.56 | 2.78 | **44.44** | 22.22 | 55.56 | **44.44** |
| seq5_neutral | Neutral | 3.13 | 28.13 | 0.00 | 3.13 | **40.63** | 25.00 | 59.38 | **40.63** |
| seq6_neutral | Neutral | 0.00 | 31.25 | 9.38 | 6.25 | **43.75** | 9.38 | 56.25 | **43.75** |
| seq7_neutral | Neutral | 0.00 | 12.12 | 3.03 | 6.06 | **60.61** | 18.18 | 39.39 | **60.61** |
| seq8_neutral | Neutral | 0.00 | 33.33 | 12.12 | 3.03 | **42.42** | 9.09 | 57.58 | **42.42** |
| seq9_neutral | Neutral | 0.00 | 19.35 | 3.23 | 6.45 | **70.97** | 0.00 | 29.03 | **70.97** |
| seq1_sad | Sad | 3.23 | 12.90 | 3.23 | 3.23 | 12.90 | **64.52** | 35.48 | **64.52** |
| seq2_sad | Sad | 6.82 | 13.64 | 1.52 | 21.21 | 25.00 | **31.82** | 68.18 | **31.82** |
| seq3_sad | Sad | 2.78 | 22.22 | 13.89 | 2.78 | 19.44 | **38.89** | 61.11 | **38.89** |
| seq4_sad | Sad | 0.00 | 27.78 | 0.00 | 5.56 | 13.89 | **52.78** | 47.22 | **52.78** |
| seq5_sad | Sad | 0.00 | 34.38 | 0.00 | 12.50 | 15.63 | **37.50** | 62.50 | **37.50** |
| seq6_sad | Sad | 0.00 | 15.63 | 0.00 | 9.38 | 15.63 | **59.38** | 40.63 | **59.38** |
| seq7_sad | Sad | 0.00 | 24.24 | 3.03 | 3.03 | 12.12 | **57.58** | 42.42 | **57.58** |
| seq8_sad | Sad | 0.00 | 18.18 | 9.09 | 3.03 | 0.00 | **69.70** | 30.30 | **69.70** |
| seq9_sad | Sad | 0.00 | 25.81 | 3.23 | 12.90 | 19.35 | **38.71** | 61.29 | **38.71** |

Silhouette stimuli. Percent of responses as a function of intended emotion of 54 silhouette stimuli. *Left side columns* indicate clip name and the intended emotional expression by the dancer when performing the movement ("Intended emotion by dancer"). The percentages of emotion judgments made by the human observers in the experiment are represented in each of the columns under the label "Decoded emotion". The confusion matrix illustrates the distribution of the given emotion judgment for each of the different emotion options. The *grey shading* indicates when intended and decoded emotion correspond. The *two columns on the right set* out the total incorrect and correct responses for each clip (i.e., when the intended and the decoded emotion where the same -correct-, or not -incorrect-)

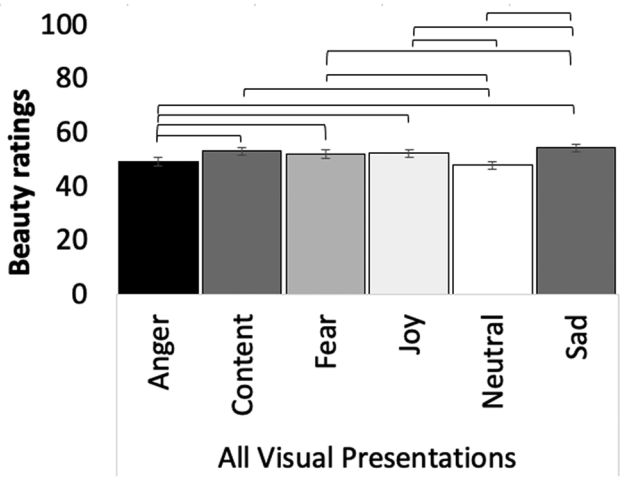



**Fig. 14** Beauty ratings: Main effect of emotion. *Note*: The *bars* show observer beauty rating averages across *visual presentations* (0; 100) were highest for sad stimuli (EMM = 54.46; SE = 1.33) > content stimuli (EMM = 53.23; SE = 1.31) > joyful stimuli (EMM = 52.46; SE = 1.30) > fearful stimuli (EMM = 52.10; SE = 1.33) > anger stimuli (EMM = 49.32; SE = 1.31) > neutral stimuli (EMM = 48.01; SE = 1.31). *Brackets* show significant differences in beauty ratings between emotion categories, all *p*s > .001, except for the comparison joy-sad, which was p = .017 (Bonferroni corrected). Error bars represent SE

that some of these differences were significant. FLDs were recognized above all others (all *p*s > .018). Avatars and silhouettes were recognized equally well ($p = 1.00$), and emotions expressed in the PLDs were recognized below all other visual presentations (all *p*s > .001). These results are illustrated in Fig. 10.

### Chance level analysis: Emotion category

Chi-square *t* tests were used to determine whether observer recognition rates were above chance for the six emotion categories (chance level: 100% / 6 emotion categories = 16.67%). The stimuli of the emotion categories anger, content, joy, neutral, sad (regardless of visual presentation) had been recognized above chance level (all *p*s > .014), while fear had not ($p = .645$).

To explore whether there were differences between recognition rates as a function of the different emotion categories intended by the dancer, we performed a 1x6 RM ANOVA with the factor *emotion category* (six levels; anger, content, fear, joy, neutral, sad), and the dependent variable *percent of correct responses* ('correct responses' = when observers guessed the intended emotion). There was a main effect of *emotion category* ($F(5,655) = 99.457$, $p < .001$, partial $\eta^2 = .432$). The observer recognition rates were highest for angry videos (EMM = 53.91%; SE = 1.98) > neutral videos (EMM = 49.432%; SE = 2.17) > sad videos (EMM = 48.04%; SE = 1.86) > joyful video (EMM = 26.39%; SE = 1.32) > content videos (EMM = 19.76; SE = 1.24) > fearful videos (EMM = 17.30%; SE = 1.36). Bonferroni corrected pairwise comparisons showed that recognition rates for angry, neutral and sad stimuli were highest and did not differ between each other (all *p*s < .227) but differed significantly from all other emotions (all *p*s > .001). Further, emotion recognition rates for joyful videos was higher than for fearful videos ($p > .001$). All comparisons are illustrated in Fig. 11.

### Chance level analysis: Visual presentation x emotion category

For each visual presentation, Chi-square *t* tests were used to compare emotion recognition rates for each emotion against the chance level of 16.67%. For the avatars, all emotions were recognized above chance (all *p*s > .050), except for fear ($p = .363$). For the FLDs, all emotions were recognized above chance (all *p*s > .016), except for content ($p = .071$). For the PLDs, emotions were recognized above chance (all *p*s > .001), except for content ($p = .226$), fear ($p = .315$), and joy ($p = .898$). For the silhouettes, emotions were recognized above chance (all *p*s > .050), except for fear ($p = .643$). For an overview of the results see Fig. 12.

### Confusion matrices emotion ratings

Four confusion matrices were computed, one for each visual presentation. They represent the observers' emotion judgments as a function of intended and decoded emotion. The advantage of confusion matrices is that the 'confused' responses (i.e., the wrong emotion judgments) for a stimulus can be compared across all emotion categories at a glance (see Banse & Scherer, 1996; Scherer & Scherer, (2011), for a detailed explanation). These matrices are set in Tables 3 and 4 for avatars, in Table 5 for FLDs, in Table 6 for PLDs and in Table 7 for silhouettes.

### Observer experiment beauty ratings

A 1 × 4 RM ANOVA was conducted with the factor *visual presentation* (four levels; avatars, full-light displays (FLDs), point-light displays (PLDs), silhouettes). The dependent variable was '*Beauty rating*' on a scale from 0 (not beautiful) to 100 (very beautiful). There was a main effect of *visual presentation* ($F(3,393) = 35.336$, $p < .001$, partial $\eta^2 = .212$). Estimated marginal means (EMM) showed that silhouette stimuli were rated as most beautiful (EMM = 54.35; SE = 1.33), followed by FLD stimuli (EMM = 53.91; SE = 1.33), then avatar stimuli

(EMM = 50.17; SE = 1.34), and, finally, PLDs (EMM = 47.93; SE = .93). Bonferroni corrected pair-wise comparisons showed that FLDs and silhouette beauty ratings did not differ significantly between each other ($p = .280$), but all other comparisons were significant (all $p$s > .001). FLDs and silhouettes being rated as more beautiful than avatars and PLDs, PLDs being rated the least beautiful (see Fig. 13).

A 1 × 4 RM ANOVA was conducted with the factor *emotion* (six levels; anger, content, fear, joy, neutral, sad). The dependent variable was '*beauty rating*' on a scale from 0 (not beautiful) to 100 (very beautiful). There was a main effect of *emotion* ($F(5,655) = 32.562$, $p < .001$, partial $\eta^2 = .199$). Descriptively, the observers' beauty ratings were highest for sad stimuli (EMM = 54.46; SE = 1.33) > content stimuli (EMM = 53.23; SE = 1.31) > joyful stimuli (EMM = 52.46; SE = 1.30) > fearful stimuli (EMM = 52.10; SE = 1.33) > anger stimuli (EMM = 49.32; SE = 1.31) > neutral stimuli (EMM = 48.01; SE = 1.31). Bonferroni corrected pair-wise comparisons showed that there were significant differences in beauty ratings between most categories (all $p$s > .001, except for the comparison joy-sad, which was $p = .017$; Bonferroni corrected). There was *no* significant difference between anger and neutral, nor between fear and joy. Also, the beauty ratings to the emotion contentment did not differ from those for fear, joy, and sadness (see Fig. 14 for an illustration of these results).

## Discussion and conclusion

We provide the EMOKINE software, computational framework, and pilot dataset of emotional movement for research in experimental psychology, affective neuroscience and computer vision. The key contribution of the project to the wider community is a computational framework comprising a detailed plan, software, and code for creation of highly controlled emotional body movement datasets at scale in the future. Comprehensive procedure instructions and kinematic feature extraction code are provided via releases on GitHub. The pilot dataset and its renderings into four different visual presentations (avatars, full-light-displays, point-light-displays, and silhouettes), along with observer ratings and the kinematic data are available on Zenodo under a Creative Commons Attribution 4.0 International License.

A series of computational validations and an observer experiment confirmed the validity of the EMOKINE pilot dataset and the creation procedure. Besides these validations provided here, the dataset has been shown to be useful to assess research questions in health psychology, for example, about how dance breaks during work hours may improve mood and motivation (Schmidt et al., 2023). Datasets created following the EMOKINE suite may be particularly useful for addressing questions about which kinematic features drive high emotion recognition and/or misclassifications. Yet for future large scale experiments, we remind researchers that the EMOKINE pilot dataset was only created with a single dancer and only contains nine movement sequences. For generalizability and scaling up, creating datasets with several dancers as models and more sequences would be advisable.

Foreground statistics showed that stimuli were balanced and within frames, qualitative histograms (silhouettes) confirmed that stimuli were aligned with each other, kinematic histograms indicated that the stimuli and features yielded meaningful signals, not random noise.

The observer experiment confirmed that emotional expression was recognized above chance level in the pilot dataset. Emotion recognition was highest for full-light displays (FLDs), than for all other visual presentations, while point-light displays (PLDs) had the lowest emotion recognition rates. Observer recognition rates for avatars and silhouettes were higher than recognition rates of PLDs. Observer recognition rates for avatars and silhouettes did not differ between each other. With regards to the emotion categories, observer emotion recognition rates were highest for angry videos (> neutral videos > sad videos > joyful video > content videos > fearful videos). We present confusion matrices, one for each visual presentation, which represent the observers' emotion judgments as a function of intended and decoded emotions. Confusion matrices allows to compare 'confused' responses (i.e., the wrong emotion judgments) for a stimulus across all emotion categories at a glance, following previous work in the field (Scherer & Scherer, 2011). With regards to aesthetic judgment, observer beauty ratings were highest for silhouette stimuli followed by FLDs, followed by avatar stimuli, and, finally, PLDs. Furthermore, aesthetic judgment was highest for sad stimuli (> content stimuli > joyful stimuli > fearful stimuli > anger stimuli > neutral stimuli).

Pilot stimuli intended to express fear were hardest to recognize for observers, and average recognition across fearful stimuli was not above chance for any of the visual presentations except in the FLDs; an effect that has previously been reported with other stimuli sets (Atkinson et al., 2007; Camurri et al., 2003; Christensen et al., 2023; Christensen et al., major revisions; Dahl & Friberg, 2007; Pasch & Poppe, 2007; Smith & Cross, 2022). Contentment, the emotion added to this dataset was recognized above chance in the visual presentations avatar and silhouettes, but surprisingly not in FLDs and PLDs.

Importantly, as described in previous work, also in the EMOKINE pilot dataset, aesthetic judgment (i.e., beauty) differed significantly between all emotions. This adds another datapoint to previous findings that suggest that aesthetic judgment can be an implicit emotion recognition task (Christensen et al., 2019; Christensen et al., 2023; Christensen et al., major revisions).

Creating future datasets based on the procedure set out above has three main advantages. First, as shown with the EMOKINE pilot dataset creation procedure, we propose to use complex movements. A dancer repeated several choreographies six times each, maintaining the same movements, but expressing different emotional intentions at each repetition. Traditionally, emotional 'actions' are often used in emotion datasets (e.g., jumping of joy or recoiling in fear), which makes the emotion rather obvious. For EMOKINE, the dancer used exactly the same dance choreography to express six different emotional intentions, thus, increasing the usefulness of the dataset to assess subtle kinematic features in emotional movement that is not emotional actions. Second, the EMOKINE dataset creation procedure proposes to include more emotional intentions. Here, we included six emotional intentions, namely anger, contentment, fear, joy, neutrality, and sadness. Classically, datasets rarely contain the emotion contentment (Ekman, 1973/2015; Ekman & Friesen, 1971), which increases the usefulness of EMOKINE. 'Contentment' is another positively valenced emotion like joy, yet of low arousal; symmetrical to what anger (negative valence, high arousal) is to sad (negative valence, low arousal). Third, the EMOKINE software provides, for the first time, thirty-two statistics from twelve kinematic features that can be obtained from one same dataset, namely, speed, acceleration, angular speed, angular acceleration, limb contraction, distance to center of mass, quantity of motion, dimensionless jerk (integral), head angle (with regards to vertical axis and to back), and space (convex hull 2D and 3D). Average, median absolute deviation (MAD) and maximum value were computed for each.

Future iterations of the dataset creation plan may take into account that the four visual presentations were not parametrically varied, but could be, using the kinematic data to vary the visual presentation of the stimuli, and also to control the exact length of the stimuli. Further, the XSENS avatar rendering was not 100% overlapping with the other visual presentations because the positioning of the legs was not fixed by the software, causing the avatar to move slightly unnaturally with less common arm and leg movement combinations. Finally, we acknowledge the WEIRD focus of this dataset creation, and suggest exploring non-Western dance with the same procedure as e.g., Christensen et al. (major revisions).

The EMOKINE software and pilot dataset is the outcome of a proof-of-principle dataset creation procedure for highly controlled kinematic video datasets of emotionally expressive full-body movement sequences. The pilot data for EMOKINE was recorded via the XSENS® system, however, with small alterations that we have outlined above and on the GitHub repository, the software can be used with data obtained from other motion capture systems too. The novelty of EMOKINE lies in the successful integration of the experimental control requirements for psychology and affective neuroscience research involving human participants and, simultaneously, ensuring the technical intricacies required for datasets in computer vision and related fields.

## Appendix

This section contains several additional histograms from our Computational Tests Section "Computational tests".

See Figs. 15, 16, 17, 18, 19, 20, 21, 22.

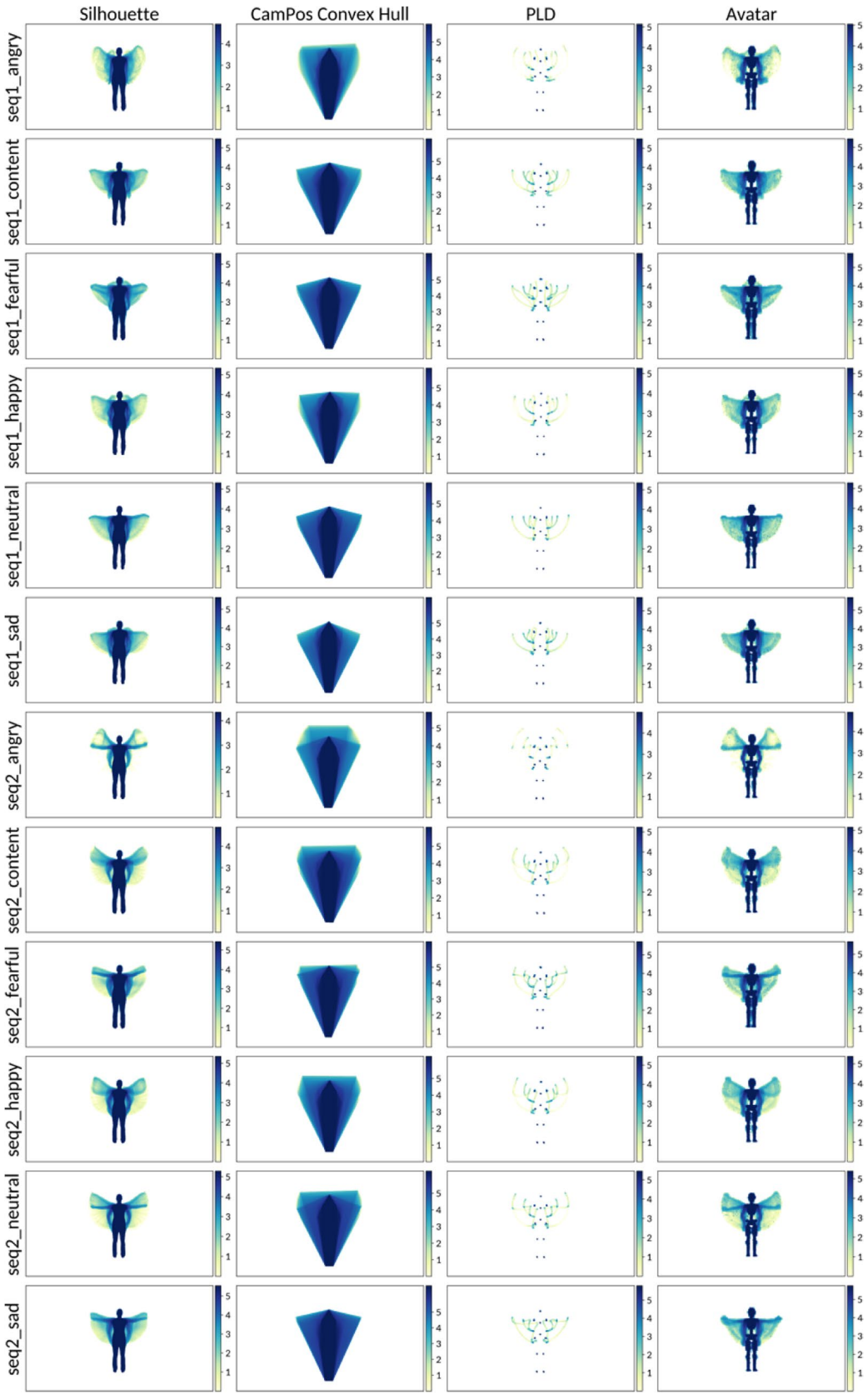



**Fig. 15** Histograms of kinematics that depend on silhouette images (Sequence 1–2)

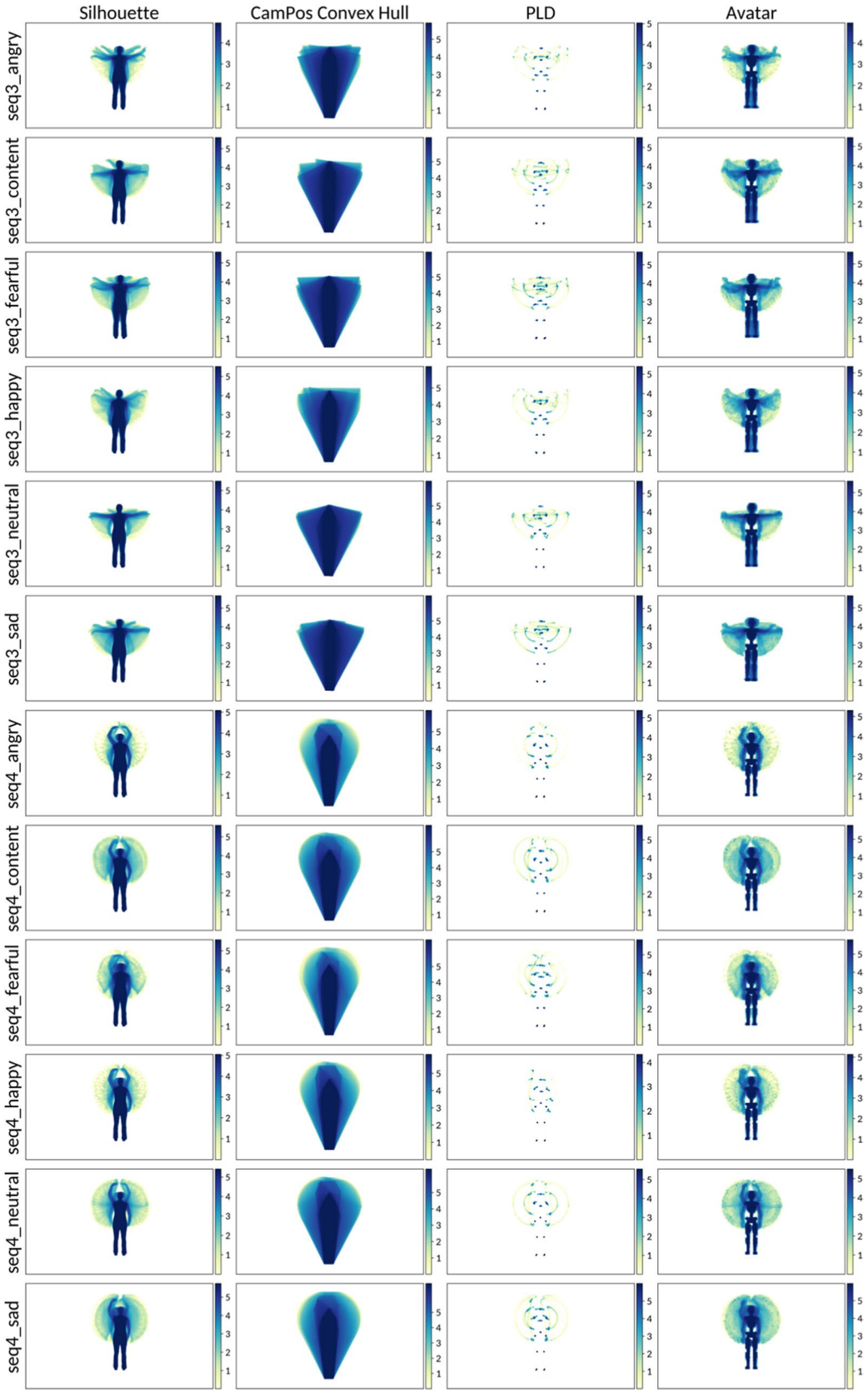



**Fig. 16** Histograms of kinematics that depend on silhouette images (Sequence 3–4)




**Fig. 17** Histograms of kinematics that depend on silhouette images (Sequence 5–6)

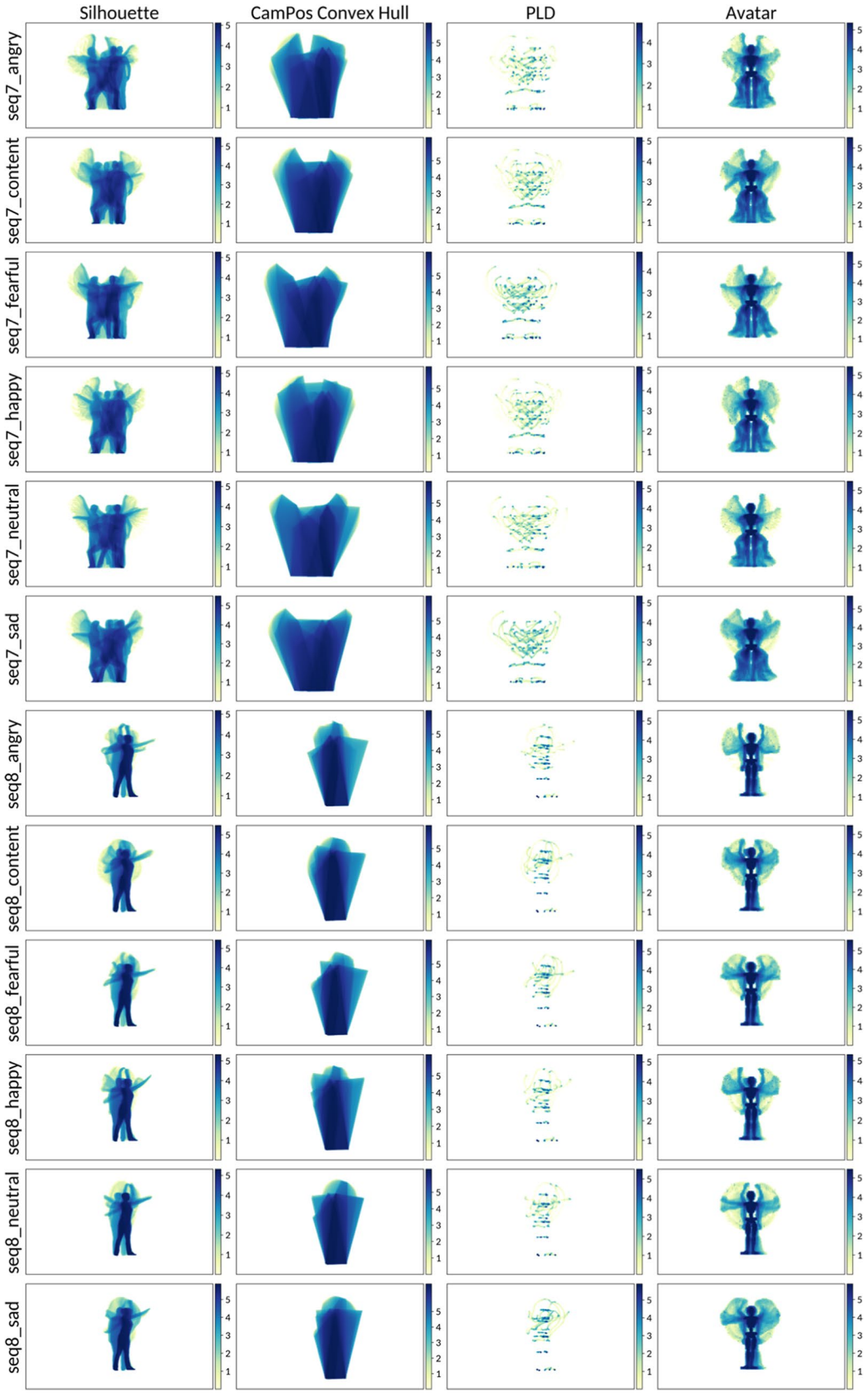



**Fig. 18** Histograms of kinematics that depend on silhouette images (Sequence 7–8)

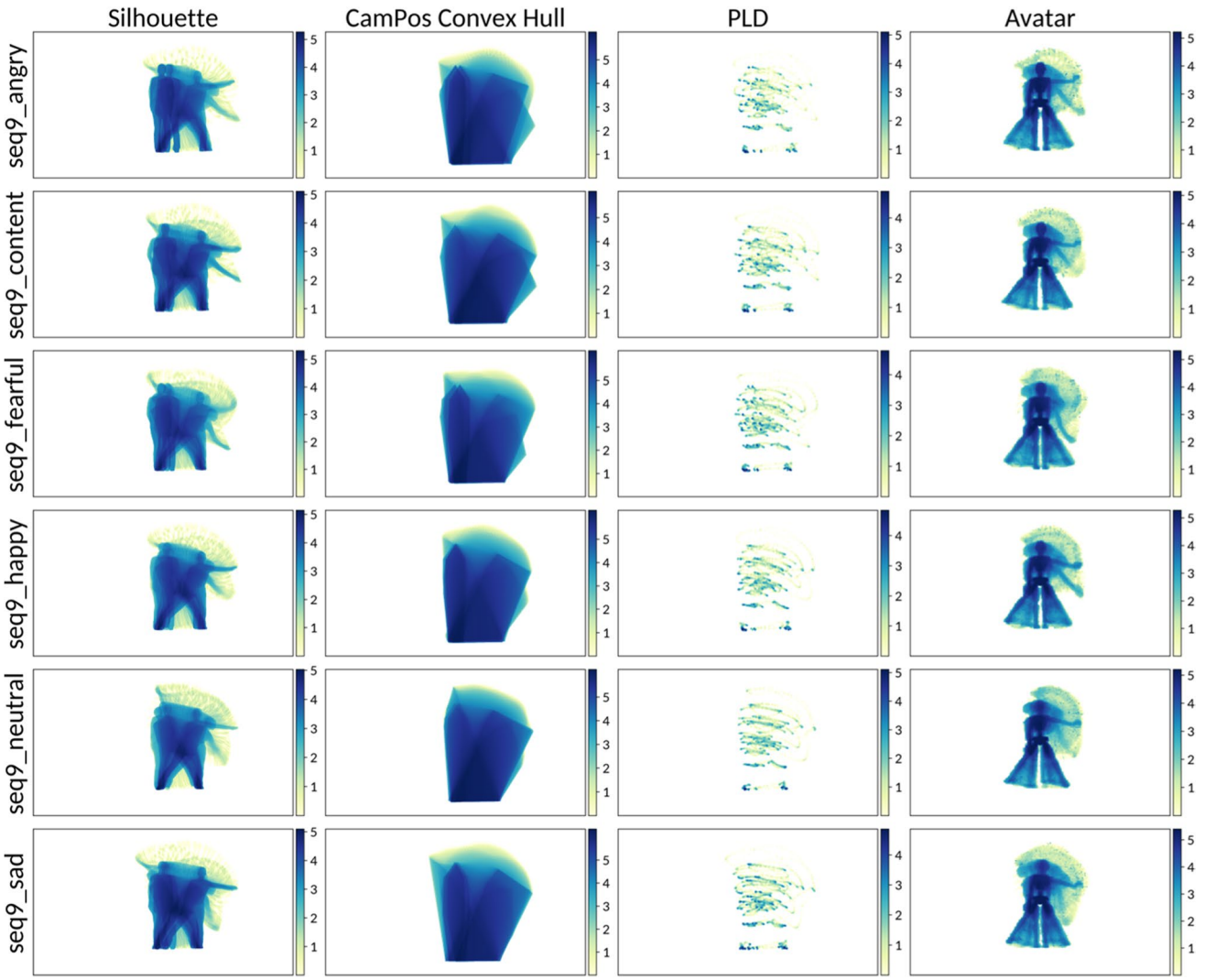



**Fig. 19** Histograms of kinematics that depend on silhouette images (Sequence 9)

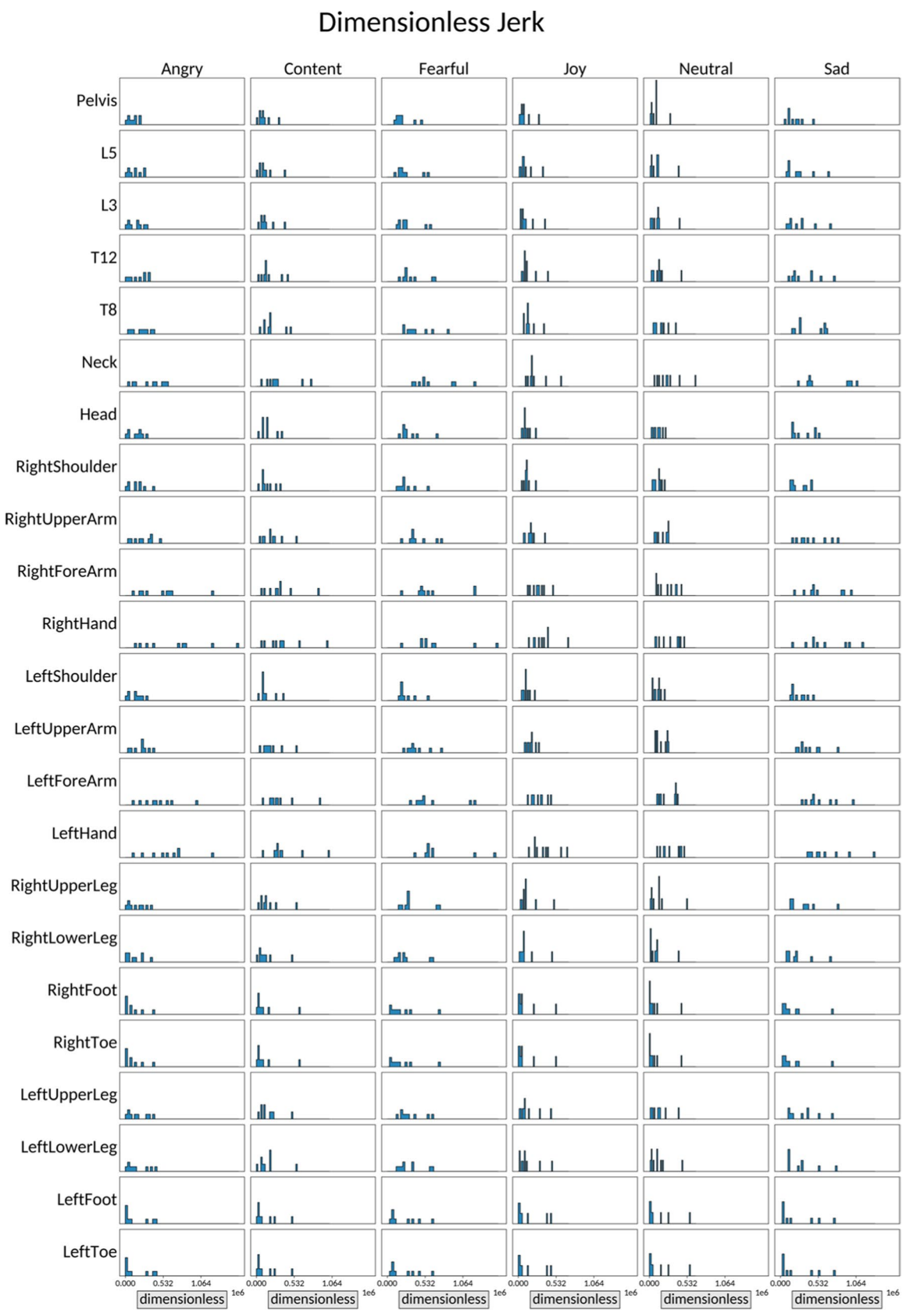



**Fig. 20** Dimensionless jerk

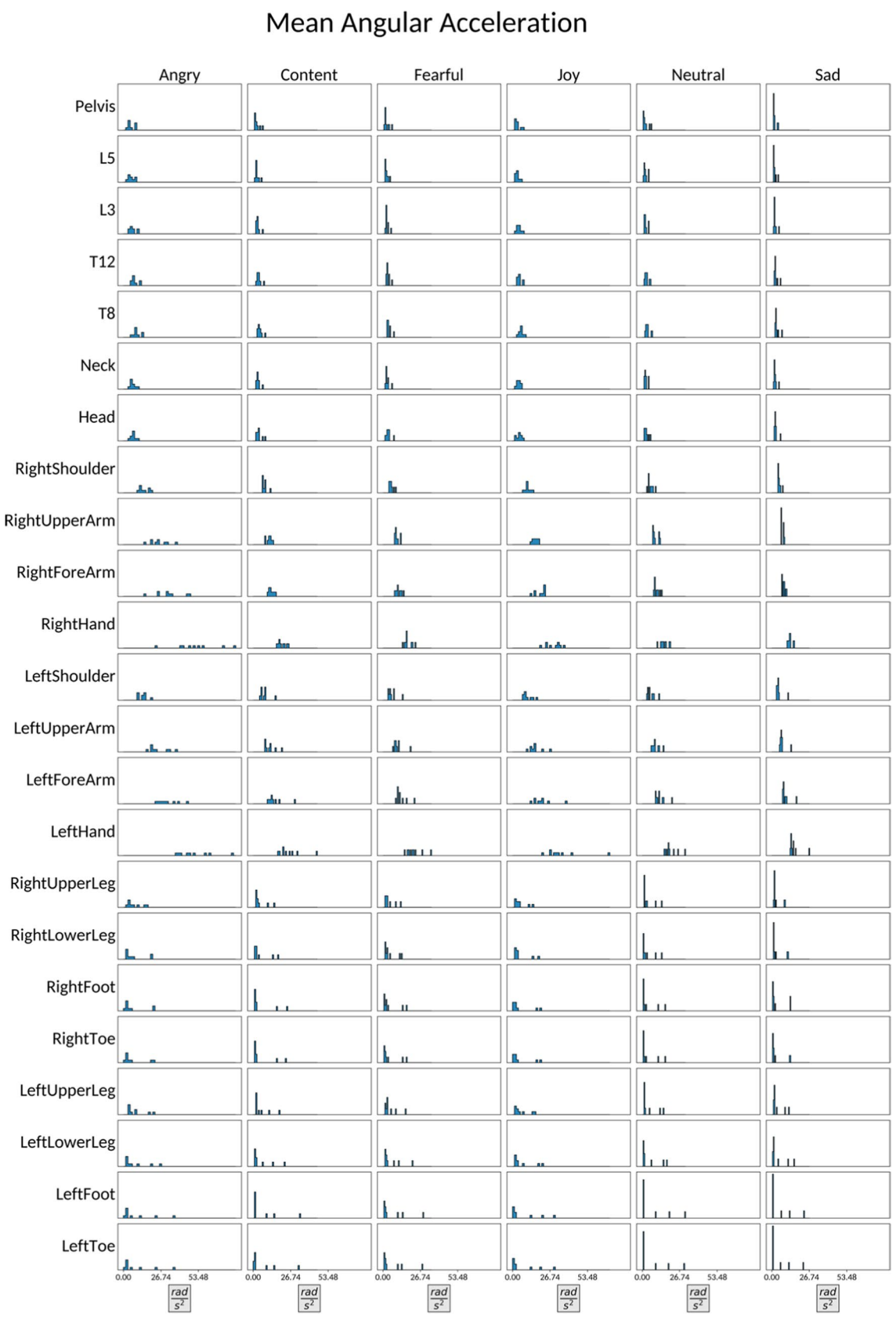



**Fig. 21** Angular acceleration

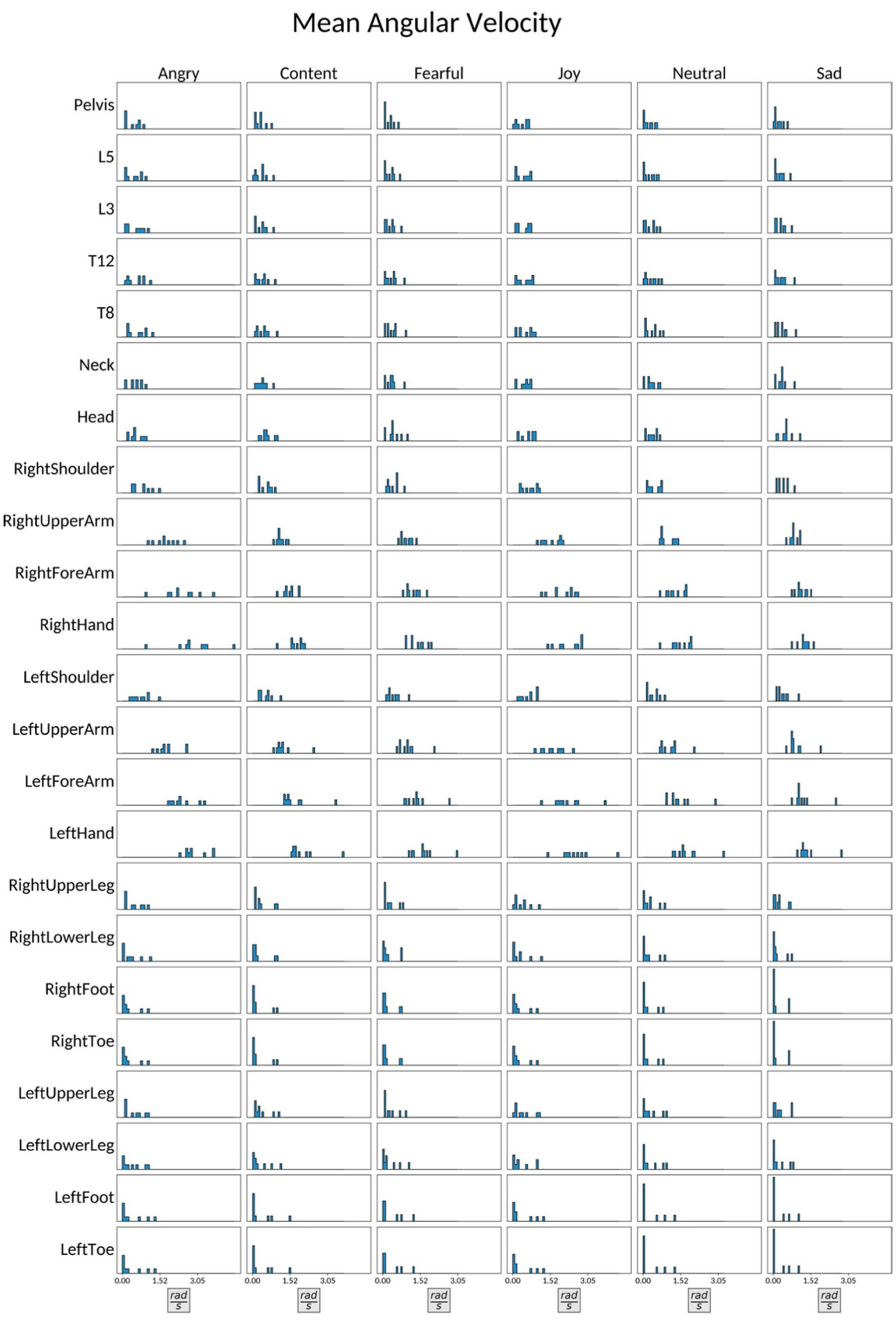



**Fig. 22** Angular velocity

**Acknowledgements** All authors of this project were funded by the Max Planck Society, Germany. In addition, AF was supported by the International Max Planck Research School for Intelligent Systems (IMPRS-IS). RAS was funded by The Economic and Social Research Council (ESRC), UK. We would like to express our gratitude to Professor Melanie Wald-Fuhrmann, director of the Music Department at the Max Planck Institute for Empirical Aesthetics (MPIEA), for her strong support and for lending us the XSENS® Motion Capture suit. We would also like to thank Rainer Pollack, Stefan Redeker, Holger Stenschke, Nikita Kudakov, Lena Blis, Stefan Strien, Nancy Schön, Norbert Bergermann, Patrick Ulrich, and the team from the MPIEA graphics department, all for invaluable support at various stages of this project. Finally, and importantly, we are indebted to the directors at the MPIEA, Professor Fredrik Ullén, and emeritus Professor Winfried Menninghaus for their kind support at different stages of this project.

**Author contributions** Julia F. Christensen: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. Andrés Fernández: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – review & editing, Visualization. Rebecca A. Smith: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization. Georgios Michalareas: Conceptualization, Methodology, Software, Writing – review & editing. Sina H.N. Yazdi: Conceptualization, Methodology, Software, Validation, Resources. Fahima Farahi: Conceptualization, Methodology, Software, Validation, Resources. Eva-Madeleine Schmidt: Methodology, Formal analysis, Data Curation. Nasimeh Bahmanian: Methodology, Formal analysis, Data Curation. Gemma Roig: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization, Supervision.

**Funding** Open Access funding enabled and organized by Projekt DEAL. International Max Planck Research School for Intelligent Systems (IMPRS-IS), Max-Planck-Gesellschaft, Economic and Social Research Council.

**Data availability** The dataset and all materials can be downloaded from Zenodo:https://zenodo.org/record/7821844 (Zenodo DOI:https://doi.org/10.5281/zenodo.7821844). The software is available on GitHub: https://github.com/andres-fr/emokine Comprehensive Readme files accompany the data and software. None of the experiments was preregistered.

## Declarations

**Open practices statement** The dataset and all materials can be downloaded from Zenodo: https://zenodo.org/record/7821844 (Zenodo https://doi.org/10.5281/zenodo.7821844). The software is available on GitHub: https://github.com/andres-fr/emokine Comprehensive Readme files accompany the data and software. None of the experiments was preregistered.

**Competing interest** The authors have no conflict of interest to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*(6), 717–746. https://doi.org/10.1068/p5096

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition, 104*(1), 59–72. https://doi.org/10.1016/j.cognition.2006.05.005

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225–1229. https://doi.org/10.1126/science.1224313

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion, 9*(5), 691–704. https://doi.org/10.1037/a0017088

Bellot, E., Garnier-Crussard, A., Pongan, E., Delphin-Combe, F., Coste, M.-H., Gentil, C., Rouch, I., Hénaff, M.-A., Schmitz, C., Tillmann, B., & Krolak-Salmon, P. (2021). Blunted emotion judgments of body movements in Parkinson's disease. *Scientific Reports, 11*(1), 18575. https://doi.org/10.1038/s41598-021-97788-1

Bernhardt, D., & Robinson, P., September). (pp. ). Berlin, Heidelberg. (2007). *Detecting affect from non-stylised body motions. In International conference on affective computing and intelligent interaction* In International conference on affective computing and intelligent interaction Berlin.

Blender Community, B. O. (2018). Blender - a 3D modelling and rendering package. Stichting Blender Foundation. Amsterdam. Retrieved from. http://www.blender.org. Accessed 20 May 2022

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology, 34*(5), 1007–1016.

Boone, R. T., & Cunningham, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of Nonverbal Behavior, 25*(1), 21–41. https://doi.org/10.1023/a:1006733123708

Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Byron, K., Terranova, S., & Nowicki, S., Jr. (2007). Nonverbal emotion recognition and salespersons: Linking ability to perceived and actual success. *Journal of Applied Social Psychology, 37*(11), 2600–2619. https://doi.org/10.1111/j.1559-1816.2007.00272.x

Camurri, A., Lagerlof, I., & Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies, 59*(1–2), 213–225.

Castellano, G., Villalba, S., & Camurri, A. (2007). Recognizing human emotions from body movement and gesture dynamics. *Lecture Notes in Computer Science, 4738*, 71.

Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T. S., & Zhang, L. (2019). HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation. arXiv:1908.10357 [cs.CV].

Christensen, J. F., Bruhn, L., Schmidt, E. M., Bahmanian, N., Yazdi, S. H. N., Farahi, F., Sancho-Escanero, L., & Menninghaus, W. (2023). A 5-emotions stimuli set for emotion perception research

with full-body dance movements. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-33656-4

Christensen, J. F., & Calvo-Merino, B. (2013). Dance as a subject for empirical aesthetics. *Psychology of Aesthetics, Creativity, and the Arts, 7*(1), 76–88. https://doi.org/10.1037/a0031827

Christensen, J. F., Gaigg, S. B., & Calvo-Merino, B. (2017). I can feel my heartbeat: Dancers have increased interoceptive accuracy. *Psychophysiology, 55*(4), 1–14. https://doi.org/10.1111/psyp.13008

Christensen, J. F., & Jola, C. (2015). Towards ecological validity in empirical aesthetics of dance. In M. Nadal, J. P. Huston, L. Agnati, F. Mora, & C. J. Cela-Conde (Eds.), *Art, Aesthetics, and the Brain*. Oxford University Press.

Christensen, J. F., Frieler, K., Vartanian, M., Khorsandi, S., Yazdi, S. H. N., Farahi, F., Smith, R.A., Walsh, W. (major revisions). A joy bias: Perception of expressive body language is modulated by enculturation.

Christensen, J. F., Lambrechts, A., & Tsakiris, M. (2019). The Warburg Dance Movement Library—The WADAMO Library: A validation study. *Perception, 48*(1), 26–57. https://doi.org/10.1177/0301006618816631

Christensen, J. F., Nadal, M., Cela-Conde, C. J., & Gomila, A. (2014). A norming study and library of 203 dance movements. *Perception, 43*(2/3), 178–206. https://doi.org/10.1068/p7581

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates Inc.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and emotions. In M. L. J. M. Haviland-Jones (Ed.), *Handbook of emotions* (pp. 91–115). Guilford.

Crane, E., & Gross, M. (2007). Motion Capture and Emotion: Affect Detection in Whole Body Movement. Affective Computing and Intelligent Interaction, Berlin, Heidelberg.

Crane, E. A., & Gross, M. M. (2013). Effort-shape characteristics of emotion-related body movement. *Journal of Nonverbal Behavior, 37*, 91–105. https://doi.org/10.1007/s10919-013-0144-2

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion, 12*(5), 1085–1101. https://doi.org/10.1037/a0025737

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*(5), 433–454. https://doi.org/10.1525/mp.2007.24.5.433

Darwin, C. (1872/2009). *The Expression of the Emotions in Man and Animals*. Oxford University Press, Anniversary edition.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*(85).

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7*(3), 242–249. https://doi.org/10.1038/nrn1872

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society London Series B: Biological Sciences, 364*(1535), 3475–3484. https://doi.org/10.1098/rstb.2009.0190

Dekeyser, M., Verfaillie, K., & Vanrie, J. (2002). Creating stimuli for the study of biological-motion perception. *Behavior Research Methods, Instruments, & Computers, 34*(3), 375–382. https://doi.org/10.3758/BF03195465

Dittrich, W. H., Troscianko, T., Lea, S., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25*(6), 727–738.

Ekman, P. (1973/2015). *Darwin and Facial Expression: A Century of Research in Review*. Malor Books.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124.

Elfenbein, H. A., & Ambady, N. (2002). Predicting workplace outcomes from the ability to eavesdrop on feelings. *Journal of Applied Psychology, 87*(5), 963–971.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/bf03193146

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior, 34*(4), 223–248. https://doi.org/10.1007/s10919-010-0094-x

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2012). Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement Science, 31*(1), 202–221. https://doi.org/10.1016/j.humov.2011.05.001

Halovic, S., & Kroos, C. (2018). Not all is noticed: Kinematic cues of emotion-specific gait. *Human Movement Science, 57*, 478–488. https://doi.org/10.1016/j.humov.2017.11.008

Heberlein, A. S., Adolphs, R., Tranel, D., & Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience, 16*(7), 1143–1158. https://doi.org/10.1162/0898929041920423

Hogan, N., & Sternad, D. (2009). Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of Motor Behavior, 41*(6), 529–534. https://doi.org/10.3200/35-09-004-rc

Kainz, H., Graham, D., Edwards, J., Walsh, H. P. J., Maine, S., Boyd, R. N., Lloyd, D. G., Modenese, L., & Carty, C. P. (2017). Reliability of four models for clinical gait analysis. *Gait Posture, 54*, 325–331. https://doi.org/10.1016/j.gaitpost.2017.04.001

Karin, J. (2016). Recontextualizing dance skills: Overcoming impediments to motor learning and expressivity in ballet dancers. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00431

Karin, J., Haggard, P., & Christensen, J. F. (2016). Mental Training. In V. Wilmerding & D. Krasnow (Eds.), *Dancer Wellness*. Human Kinetics.

Keck, J., Zabicki, A., Bachmann, J., Munzert, J., & Krüger, B. (2022). Decoding spatiotemporal features of emotional body language in social interactions. *Scientific Reports, 12*(1), 15088. https://doi.org/10.1038/s41598-022-19267-5

Kirsch, L. P., Urgesi, C., & Cross, E. S. (2016). Shaping and reshaping the aesthetic brain: Emerging perspectives on the neurobiology of embodied aesthetics. *Neuroscience & Biobehavioral Reviews, 62*, 56–68. https://doi.org/10.1016/j.neubiorev.2015.12.005

Krüger, B., Kaletsch, M., Pilgramm, S., Schwippert, S. S., Hennig, J., Stark, R., Lis, S., Gallhofer, B., Sammer, G., Zentgraf, K., & Munzert, J. (2018). Perceived intensity of emotional point-light displays is reduced in subjects with ASD. *Journal of Autism and Developmental Disorders, 48*(1), 1–11. https://doi.org/10.1007/s10803-017-3286-y

Lim, S., Case, A., & D'Souza, C. (2016). Comparative Analysis of Inertial Sensor to Optical Motion Capture System Performance in Push-Pull Exertion Postures. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting. Human Factors and Ergonomics Society. Annual Meeting*, *60*(1), 970–974. https://doi.org/10.1177/1541931213601224

Ma, Y., Paterson, H. M., & Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods, 38*(1), 134–141. https://doi.org/10.3758/bf03192758

Masuda, M., Kato, S., & Itoh, H. (2010). *A laban-based approach to emotional motion rendering for human-robot interaction*. Paper presented at the Entertainment Computing - ICEC 2010, Berlin, Heidelberg.

McCarty, K., Darwin, H., Cornelissen, P. L., Saxton, T. K., Tovée, M. J., Caplan, N., & Neave, N. (2017). Optimal asymmetry and other motion parameters that characterise high-quality female dance. *Scientific Reports, 7*, 42435. https://doi.org/10.1038/srep42435

Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior, 23*(2), 133–152. https://doi.org/10.1023/A:1021435526134

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers,

E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Nirkin, Y., Masi, I., Tran, A. T., Hassner, T., & Medioni, G. (2017). On Face Segmentation, Face Swapping, and Face Perception. arXiv: 1704.06729 [cs.CV].

O'Boyle, E. H., Jr., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior, 32*(5), 788–818. https://doi.org/10.1002/job.714

Orgs, G., Caspersen, D., & Haggard, P. (2016). You move, I watch, it matters: Aesthetic Communication in Dance. In Sukhvinder S. Obhi & E. S. Cross (Eds.), Shared representations: Sensorimotor foundations of social life. Cambridge University Press

Pasch, M., & Poppe, R. (2007). Person or Puppet? The Role of Stimulus Realism in Attributing Emotion to Static Body Postures. Affective Computing and Intelligent Interaction, Berlin, Heidelberg.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1–20. https://doi.org/10.3758/s13428-021-01694-3

Piwek, L., Petrini, K., & Pollick, F. (2016). A dyadic stimulus set of audiovisual affective displays for the study of multisensory, emotional, social interactions. *Behavior Research Methods, 48*(4), 1285–1295. https://doi.org/10.3758/s13428-015-0654-4

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51-61. https://doi.org/10.1016/s0010-0277(01)00147-0

Poyo Solanas, M., Vaessen, M. J., & de Gelder, B. (2020). The role of computational and subjective features in emotional body expressions. *Scientific Reports, 10*(1), 6202–6202. https://doi.org/10.1038/s41598-020-63125-1

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, *9*(6), 15.11–32. https://doi.org/10.1167/9.6.15

Rosete, D., & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadership & Organization Development Journal, 26*(5), 388–399. https://doi.org/10.1108/01437730510607871

Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal, 48*(5), 845–858. https://doi.org/10.5465/AMJ.2005.18803926

Sawada, M., Suda, K., & Ishii, M. (2003). Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and Motor Skills, 97*(3), 697–708.

Schepers, M., & Giuberti, M. (2018). Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing. *Technical Report*.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index. *Journal of Nonverbal Behavior, 35*(4), 305. https://doi.org/10.1007/s10919-011-0115-4

Scherer, K. R., Sundberg, J., Fantini, B., Trznadel, S., & Eyben, F. (2017). The expression of emotion in the singing voice: Acoustic patterns in vocal performance. *J Acoust Soc Am, 142*(4), 1805. https://doi.org/10.1121/1.5002886

Schlotz, W., Wallot, S., Omigie, D., Masucci, M. D., Hoelzmann, S. C., & Vessel, E. A. (2020). The Aesthetic Responsiveness Assessment (AReA): A screening tool to assess individual differences in responsiveness to art in English and German. *Psychology of Aesthetics, Creativity, and the Arts, 15*(4), 682–696. https://doi.org/10.1037/aca0000348

Schmidt, E.-M., Smith, R. A., Fernández, A., Emmermann, B., & Christensen, J. F. (2023). Mood induction through imitation of full-body movements with different affective intentions. *British Journal of Psychology, 115*(1), 148–180. https://doi.org/10.1111/bjop.12681

Shafir, T. (2016). Using movement to regulate emotion: Neurophysiological findings and their application in psychotherapy. *Frontiers in Psychology, 7*, 1451–1451. https://doi.org/10.3389/fpsyg.2016.01451

Shafir, T., Taylor, S. F., Atkinson, A. P., Langenecker, S. A., & Zubieta, J. K. (2013). Emotion regulation through execution, observation, and imagery of emotional movements. *Brain and Cognition, 82*(2), 219–227. https://doi.org/10.1016/j.bandc.2013.03.001

Shafir, T., Tsachor, R. P., & Welch, K. B. (2016). Emotion regulation through movement: Unique sets of movement characteristics are associated with and enhance basic emotions. *Frontiers in Psychology, 6*, 2030–2030. https://doi.org/10.3389/fpsyg.2015.02030

Shikanai, N., Sawada, M., & Ishii, M. (2013). Development of the Movements Impressions Emotions Model: Evaluation of movements and impressions related to the perception of emotions in dance [journal article]. *Journal of Nonverbal Behavior, 37*(2), 107–121. https://doi.org/10.1007/s10919-013-0148-y

Skogstad, S. A., Nymoen, K., & Høvin, M. (2011). Comparing Inertial and Optical MoCap Technologies for Synthesis Control. *Proceedings of SMC 2011 8th Sound and Music Computing Conference "Creativity rethinks science*, 421–426.

Smith, R. A., & Cross, E. S. (2022). The McNorm library: creating and validating a new library of emotionally expressive whole body dance movements. *Psychological Research*. https://doi.org/10.1007/s00426-022-01669-9

Smith, R. A., & Pollick, F. E. (2022). The role of dance experience, visual processing strategies, and quantitative movement features in recognition of emotion from whole-body movements. In C. Fernandes, V. Evola, & C. Ribeiro (Eds.) *Dance Data, Cognition, and Multimodal Communication*. London: Routledge.

Stanton, K., Carpenter, R. W., Nance, M., Sturgeon, T., & Villalongo Andino, M. (2022). A multisample demonstration of using the Prolific platform for repeated assessment and psychometric substance use research. *Experimental and Clinical Psychopharmacology*. https://doi.org/10.1037/pha0000545

Vaessen, M. J., Abassi, E., Mancini, M., Camurri, A., & de Gelder, B. (2018). Computational feature analysis of body movements reveals hierarchical brain organization. *Cerebral Cortex, 29*(8), 3551–3560. https://doi.org/10.1093/cercor/bhy228

Van Dyck, E., Burger, B., & Orlandatou, K. (2017). The communication of emotions in dance. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 122–130). Routledge. https://doi.org/10.4324/9781315621364-14

Van Dyck, E., Maes, P.-J., Hargreaves, J., Lesaffre, M., & Leman, M. (2013). Expressing induced emotions through free dance movement. *Journal of Nonverbal Behavior, 37*(3), 175–190. https://doi.org/10.1007/s10919-013-0153-1

Van Meel, J., Verburgh, H., & De Meijer, M. (1993). Children's interpretations of dance expressions. *Empirical Studies of the Arts, 11*(2), 117–133.

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers, 36*(4), 625–629. https://doi.org/10.3758/BF03206542

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*(6), 879–896. https://doi.org/10.1002/(SICI)1099-0992(1998110)28:6%3c879::AID-EJSP901%3e3.0.CO;2-W

Walter, F., Cole, M. S., van der Vegt, G. S., Rubin, R. S., & Bommer, W. H. (2012). Emotion recognition and emergent leadership: Unraveling mediating mechanisms and boundary conditions. *The Leadership Quarterly, 23*(5), 977–991. https://doi.org/10.1016/j.leaqua.2012.06.007

Wu, Y., Kirillov, A., Massa, F., Lo, W., & Girshick, R. (2019). Detectron2. Retrievend 16-May 2022 https://github.com/facebookresearch/detectron2.

XSENS. (2023). *MVN Biomechanical Model*. XSENS. Retrieved 16-May-2023 from https://www.movella.com/applications/entertainment

XSENS Manual, 2020). MVN User Manual. In Xsens (Ed.). https://www.xsens.com/hubfs/Downloads/usermanual/MVN_User_Manual.pdf
Zuskin, E., Schachter, E. N., Mustajbegovic, J., Pucarin-Cvetkovic, J., & Lipozencic, J. (2007). Occupational health hazards of artists. *Acta Dermatovenerologica Croatica, 15*(3), 167–177.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.