

Neural and Neuromimetic Perception: A Comparative Study of Gender Classification from Human Gait

Viswadeep Sarangi^{1,▲}, Adar Pelah¹, William Edward Hahn², Elan Barenholtz²

¹Department of Electronic Engineering, University of York, York, UK; ²Center for Complex Systems & Brain Sciences, Florida Atlantic University, Florida, USA
E-mail: adar.pelah@york.ac.uk

Abstract. Humans are adept at perceiving biological motion for purposes such as the discrimination of gender. Observers classify the gender of a walker at significantly above chance levels from a point-light distribution of joint trajectories. However, performance drops to chance level or below for vertically inverted stimuli, a phenomenon known as the inversion effect. This lack of robustness may reflect either a generic learning mechanism that has been exposed to insufficient instances of inverted stimuli or the activation of specialized mechanisms that are pre-tuned to upright stimuli. To address this issue, the authors compare the psychophysical performance of humans with the computational performance of neuromimetic machine-learning models in the classification of gender from gait by using the same biological motion stimulus set. Experimental results demonstrate significant similarities, which include those in the predominance of kinematic motion cues over structural cues in classification accuracy. Second, learning is expressed in the presence of the inversion effect in the models as in humans, suggesting that humans may use generic learning systems in the perception of biological motion in this task. Finally, modifications are applied to the model based on human perception, which mitigates the inversion effect and improves performance accuracy. The study proposes a paradigm for the investigation of human gender perception from gait and makes use of perceptual characteristics to develop a robust artificial gait classifier for potential applications such as clinical movement analysis. © 2020 Society for Imaging Science and Technology.

[DOI: 10.2352/J.Percept.Imaging.2020.3.1.010402]

1. INTRODUCTION

A significant body of research has investigated the information that can be extracted from human gait. Johansson [24–26] first introduced the point-light display (PLD) animation technique, in which points of light attached to the limb joints of walkers were recorded in the dark, a type of stimulus that came to be known as “biological motion.” This basic paradigm led to the extensive study of human visual perception of these dynamic patterns [1, 28, 30, 41, 42]. Human observers are not only able to distinguish biological motion from sparse noise, but they can also recognize the identity of walkers [4, 15] and generic attributes from strategically placed sparse PLDs

such as gender [1, 22, 31, 44, 50], emotion [17, 41, 42], and walking direction [10]. The ability to perceive certain properties from biological motion has been found to be highly dependent on whether the stimulus conforms to a canonical upright viewpoint. When a point-light walker is inverted vertically, gender classification performance drops to below chance levels [1, 35], a phenomenon known as the inversion effect [51]. Neuromimetics is a system in which computational models or methods apply underlying concepts of neural processes [54]. The framework of this approach refers to the study of perception, action, learning, memory and cognition within neuroscience. Examples of this category include all the perceptron-based artificial neural networks (ANNs) detailed in Section 2.2. This study evaluates the extent to which these models mimic certain aspects of human performance. In addition, by conducting computational experiments on a neuromimetic machine-learning model (NM) we assess whether generic learning mechanisms can account for the inversion effect observed in humans.

Experiment 1 is conducted to establish overall similarity between humans (H) and the NM through parallel computation and psychophysical experiments on gender perception from a PLD representation of gait. Results demonstrate a conformity in gender classification performance with increasing duration of stimulus exposure between human observers and the NM. Experiment 2 tests for the presence of human-like susceptibility to inversion in the NM by training and testing ten randomly initialized instances of the model. Results demonstrate vulnerability in all the model versions, indicating that the inversion effect is an emergent property from generic learning of the associations between gender and the PLD-based biological motion. This argues against the need for specialized mechanisms, pre-tuned to upright stimuli, for the explanation of the effect in human perception. This is further strengthened by the emergence of the effect in the generic learning model trained on upright stimuli, which suggests that the effect is a result of insufficient training instances of inverted stimuli.

The second motivation of the study is the creation of a practical, high-performing artificial gait classifier that overcomes the observed limitations in humans. Experiment 3 tests the dependence of the model on structural cues by re-

▲ IS&T Member.

Received July 15, 2019; accepted for publication July 23, 2020; published online Sep. 1, 2020. Associate Editor: Ruth Rosenholtz.

2575-8144/2020/3(1)/010402/11/\$00.00

moving them and forcing the model to learn the association between gender and explicit motion cues. Results indicate an improved robustness to inversion while maintaining classification accuracy. Experiment 4 further improves performance by providing the model with spatiotemporally modified gait, resulting in a high-performing artificial gait classifier without significant loss of accuracy due to the inversion of stimulus.

2. BACKGROUND

2.1 Gait Classification

Automation of gait classification has been studied for more than two decades [8, 16, 21, 23, 32, 48, 53], with gender classification being the most common evaluation objective. There have been numerous applications of machine-learning and deep-learning models in the recognition of gender from gait. Support vector machines, decision trees, and feedforward artificial neural networks are some of the most common classification models [16, 21, 23, 32, 48]. To conform to the input requirements of the models, these studies represent human gait as a static set of numerical values. Multiple studies have explored the static representation in conjunction with different classification models with encouraging results. Spatiotemporal gait metrics (stride length, joint angles and displacements, gait cycle time, etc.), gait energy images from two-dimensional (2D) silhouettes, and gait energy volume representations of gait [21, 48] have shown the most promise in gender recognition. There has also been a shift of motion capture technology from 2D RGB (video) sensors to more sophisticated three-dimensional (3D) motion capture technology, especially markerless technologies in the recent literature [2, 6, 40]. This has led to dramatic improvements in gender classification accuracy. However, the study of the potential for gait analysis using 3D sensors for automated gait analysis is at an early stage compared to the advances in 2D sensors.

Although gender discrimination from gait has been extensively studied, most studies have looked at human- and machine-based classification independently. There is a lack of literature focusing on a parallel comparison of the learning capability of machines with human performance. Moreover, past studies have prioritized improvements in performance accuracy in the classification of gender from gait. As a result, hyper-customization of features and models for the classification objective discourages adaptability of the solution and applicability to non-gender-based objectives. Humans, on the other hand, are highly versatile in their learning capabilities. In this study, we focus on the training and evaluation of a learning model that mimics the versatile learning approach of humans using ANNs.

Artificial neural networks aim to mimic the processing of information in the biological brain using a network of artificial neurons based on the perceptron model [43]. Through iterative adjustment of their network weights, they incrementally reduce errors in their own input/output pairings, which are repeated until an acceptable error rate is reached; this is then followed by testing on an unseen set

of data [3]. This “supervised learning” process may share similarities with learning processes in humans.

2.2 Neuromimetic Models

While numerous models and approaches are possible [3, 20, 29, 36, 43], the following criteria were applied for the review of neuromimetic models: (1) modeling is based on neural principles and an understanding of neuroscience in the connectionist approach; (2) processing requires minimal human-assisted, hand-crafted feature design; (3) models are capable of processing arbitrarily long data sequences; and (4) models are practical enough to train and classify on the available dataset.

Previous works have proposed a computational neuromimetic model for motion perception through the use of feedforward and recurrent ANNs that aim to emulate the two-fold neural pathway [20, 29]. The model proposed by Giese et al. [20] creates corresponding models for global and local processing modules to provide a one-to-one correspondence to the hypothesized perceptual modules. Specific global and local features are extracted from motion (e.g. optical flow) to parallel the modification of information in the neural pathway. Despite the perceptual correspondence, the practical applications of the model for automated gait classification are subject to the availability of an extensive training data and requirement for high computational capability given the large number of tunable parameters in the models. Moreover, the aforementioned model requires explicit pre-processing of gait information by extracting hand-crafted optical flow features. The limited dataset and computational capability therefore dictate a more practical alternative that still meets the mimetic criterion. Traditional recurrent neural networks (RNNs) used in the models by Giese et al. and Lange et al. [20, 29] suffer from the vanishing and exploding gradient problem, making them ineffective in meeting the criterion of being able to process long sequences [37]. On the other hand, long short-term memory (LSTM) cells, a variety of RNNs, introduce additional gates in the network, which regulate the flow of information into short- and long-term memory, thus enabling them to remember relevant temporal patterns over long periods of time [19]. The LSTM cells also mimic the memory capability in human learning more closely. In particular, their ability to learn multidimensional time series representations captures the dynamic joint trajectories in gait from point-light animations as those used by human perception. Additionally, there is no restriction on the model for the provision of structural information for processing (such as an inverted stimulus). For the purposes of the current article, we consider LSTM models operating on sequences of PLD motion to be neuromimetic machine-learning models under evaluation. In the first half of the article, we focus on the training and evaluation of the model and its comparison with human observers on the same stimulus set under both upright (Experiment 1) and inverted (Experiment 2) conditions. The second half (Experiments 3 and 4) focuses on improving the NM outcome by altering the representation of gait. The

alterations are adapted from the human perception literature to further test the neuromimetic nature of the model.

3. DATA COLLECTION

Forty-one consenting healthy adults (26 male and 15 female) between the ages of 18 and 50 years were recorded walking on a treadmill. Participants volunteered and received credit toward a participation grade for their class. Appropriate consent forms were signed and anonymity was maintained. Gait data were recorded as spatiotemporal three-dimensional joint trajectories for 20 tracked joints of the body. The tracked points on the walker's skeleton included the head, neck, shoulders, elbows, wrists, fingertips, mid-spine, back, hips, knees, ankles, and toes. The collection of the joint positions formed a static frame. Data were captured at 24 frames per second, each frame represented by 60 numbers (3D coordinates of 20 joints) and a corresponding timestamp of capture of the frame. Data were recorded for six sessions per participant. Each session consisted of a minute of walking on the treadmill at a self-selected speed followed by a minute's rest. The joints were extracted utilizing a popular time-of-flight based RGB-D sensor, the Microsoft Kinect v2. The sensor provides an anthropomorphic representation of the human skeleton through 3D joint coordinates. The sensor was placed approximately 1.5 m in front of the treadmill with the front board removed to avoid issues with occlusion. The machine-learning-based skeletal motion capture method mentioned in Ref. [46] is used for capturing the PLD representation of the biological motion of the walkers. When compared with the state-of-the-art optical motion tracking methods (such as Vicon [39]), the anatomical landmarks from the Kinect-generated point clouds can be measured with high test-retest reliability, and the differences in the interclass coefficient correlation between Kinect and Vicon are <0.16 [12–14]. Reference [18] further showed that both systems can effectively capture $>90\%$ variance in full-body segment movements during exergaming. The validity of biological motion captured using the Kinect v2 sensor is established in Ref. [46] with human observers through reflexive attentional orientation and extraction of emotional information from the upright and inverted PLDs.

4. EXPERIMENT 1: VARIATION IN EXPOSURE DURATION IN NEURAL AND NEUROMIMETIC MODELS

Humans have been shown to require no more than two gait cycles to correctly identify gender from gait [1]. This translates to viewer exposure to gait animation lasting less than 2.7 seconds. Although viewers can decipher point-light configurations into a Gestalt of a walking human figure within 200 msec, at least 1.6 sec of gait animation is required for significantly above chance performance. If the neuromimetic models imitate human perception, perhaps a corresponding exposure duration threshold exists for above chance performance.

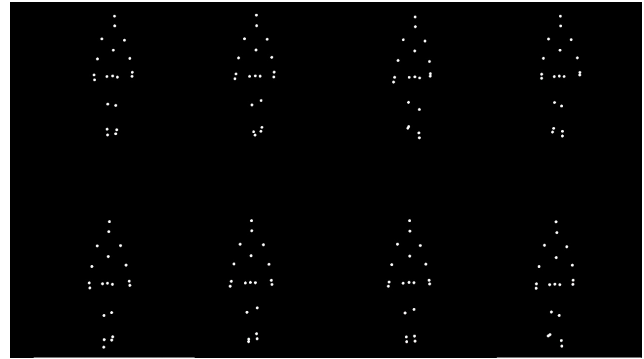


Figure 1. Point-light representation of a walking stimulus at eight different stages of a gait cycle. This acts as the direct visual stimulus to human observers.

4.1 Method

4.1.1 Neural Models

Fifteen female and six male healthy observers with ages ranging from 20 to 43 years participated in the experiment. All had some experience of biological motion displays although none had been required to make judgements about gender.

4.1.2 Stimuli

A PC-compatible computer with a high-performance raster graphics system displayed stimuli on an Iiyama ProLite B2283HS color monitor (1920 × 1080 resolution, 60 Hz refresh rate). Human figures were defined by 20 circular white dots of 5 pixel radius overlaid on a black background, located on the head, neck, shoulders, elbows, wrists, fingertips, back, spine, hips, knees, ankles, and toes. None of the dots were occluded by other subjective parts of the figure. Animated sequences were created by placing the dots at the three-dimensional trajectory of each of the 20 tracked joints and temporally sampling the coordinates to produce 24 static frames per second as shown in Figure 1.

The stimulus size was 6 degrees wide and 8 degrees tall for the whole frame, including zero (black) padding. Here, a degree is defined as the subtended angle at the nodal point of the eye. The actual walking clip was 2.5 degrees wide and 4 degrees tall. When the static frames were played in quick succession, a vivid impression of a walking person emerged. There was no progressive component to the walking animation. Thus the human figure appeared to walk on an unseen treadmill with the walking direction oriented toward the observer. None were notably over- or underweight as shown in Table 1. The x and y components were sampled to display the walker in the coronal plane to emphasize lateral sway and maximize the provision of dynamic cues to the observer [1, 50]. The recorded gait sequences were converted into an animation sequence in the same manner to be presented as visual stimuli. Animation playback was normalized for size [50] and occurred at veridical speed with linear interpolation of joint trajectories between frames. The veridical speed was determined based

on the timestamps attached to each recorded frame. The observers were seated in a well-lit room in front of the monitor and had access to a standard computer mouse for interaction. The randomly chosen walker stimuli were presented for exposure durations of 0.4, 1.5, 2.5, and 3.8 sec followed by an on-screen prompt in the form of two buttons requesting the observer's classification of binary gender through a mouse click on either of the labeled buttons. The order of all the stimuli presented was randomized. Following the response from the observer, the next stimulus was presented. A total of 200 walking clips were shown per observer per exposure duration, and the responses were recorded for each.

4.1.3 Neuromimetic Model

A standard LSTM cell model consisting of 128 hidden states was designed. The cell state weights were initialized as a random normal distribution. The final cell state was ReLU-activated [33] and connected to an affine output layer, which represented the on-hot encoded gender identity of the walker during training. The "Male" and "Female" labels were encoded as [1,0] and [0,1], respectively. During testing, the output layer represented the classification values. The error of classification was evaluated using a cross-entropy function [7] for updating the weights using an Adam optimizer [27] based on the error differentials and a learning rate of 0.001. The most probable output was taken as the class label during classification. Ten instances of LSTM models were created by randomizing the initial weight matrix before training, thus creating ten different NMs. Each of the NMs could be argued to represent an independent human observer undergoing training. The reason for having the independent NMs was also for obtaining comparable statistics as followed in the standard machine-learning literature.

4.1.4 Data Input

The three-dimensional trajectories of each of the 20 tracked joints were concatenated to form a vector representation of a static frame with a cardinality of 60, representing the location of the head, neck, shoulders, elbows, wrists, fingertips, mid-back, hips, knees, ankles, and toes. Gait input to the model consisted of a sequence of vector representations of subsequent static frames, sampled at 24 frames per second. Joint trajectories were size-normalized [50] and standardized with a zero mean and unit standard deviation. Model training sessions included initialization of the model weights, classification of the output probabilities based on the gait input, propagation of the classification error, and updating the network weights. Model training was executed in batches of 50 and repeated for 100 epochs. Input sequence durations mirrored the exposure durations in the corresponding human perception experiment and varied incrementally for ten durations from 0.4 sec to 3.8 sec in steps of 0.4 sec. Ten-fold cross validation was carried out to ensure model generalizability, and a total of 250 gender classifications were obtained per input sequence duration. The models

Table I. Description of walking subjects taking part in the stimulus set. Both the humans and NM models are evaluated using this dataset.

	Height (cm)	Weight (kg)	Age (years)
Male	176.23 + / - 32.43	80.49 + / - 2.86	26.06 + / - 6.42
Female	128.56 + / - 23.51	73.3 + / - 4.59	21.29 + / - 1.23

trained per session per duration were stored locally for future analyses.

4.2 Results

4.2.1 Neural Perception

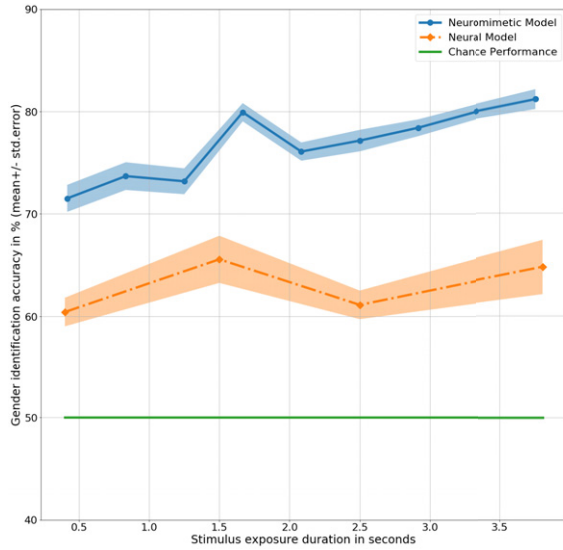
Human observers correctly discriminated 63% of all the trials across all exposure durations, which was significantly greater than the chance performance of 50% ($t_{20} = 7.8$, $p < 0.001$) and was two-tailed (note that all t -tests reported in this article are two-tailed). Correct classification at 0.4 sec, which consisted of approximately a quarter of a step cycle, was above chance at 60% ($t_{20} = 3.7$, $p < 0.01$), which was in disagreement with that in Barclay et al. [1]. This could be attributed to the presentation of the stimulus in the coronal plane as opposed to the sagittal plane [31], leading to higher emphasis on dynamic cues. Performance accuracy at 1.5 sec was 66% ($t_{20} = 3.8$, $p < 0.005$), which was higher than the performance at 2.5 sec of 61% ($t_{20} = 4.8$, $p < 0.001$). Troje et al. explains this anomalous phenomenon due to an additional partial step at 2.5 sec by highlighting the preferred perception of velocity over positional cues, where sensitivity to gender classification decreases mid-swing in the gait cycle [50]. Humans were able to discriminate gender with the highest accuracy at 3.8 sec with 69% ($t_{20} = 3.4$, $p < 0.01$). Details of the results are shown in Table II. Overall, human performance is consistent with that in other perception studies [1, 24, 50], thus providing a reliable baseline for comparison with the neuromimetic accuracy on the same stimulus set.

4.2.2 Neuromimetic Performance

The NMs correctly classified 76% of all the gait inputs presented across all the input durations ($t_9 = 9.2$, $p < 0.001$). Chance performance remains the same at 50%. Correct classification at about a quarter of a step cycle at 0.4 sec was 71% ($t_9 = 5$, $p < 0.001$), higher than the same with human observers ($F_{1,29} = 3.6$, $p < 0.1$). All F -tests in this article are assumed to be one-way analysis of variance (ANOVA) hypothesis tests between two groups. The difference in performance indicates a higher inference capacity from a limited amount of available data. The inference performance increases slightly with increase in the amount of information available from 0.4 to 3.8 sec ($t_{9,9} = 2$, $p < 0.1$). At 3.8 sec, the model correctly classified gender with 81% accuracy ($t_9 = 9.6$, $p < 0.001$), considerably higher than human observers ($F_{1,29} = 9$, $p < 0.01$). Generalizing across all the input (or exposure) durations, the NM classified gender with a significantly higher accuracy than the human observers

Table II. Gender classification accuracy as a function of exposure duration of the stimulus.

Stimulus Duration/Model	0.4 sec	1.5 sec	2.5 sec	3.8 sec
Neural (Human)	60% ($p < 0.01$)	66% ($p < 0.005$)	61% ($p < 0.001$)	69% ($p < 0.05$)
Neuromimetic (LSTM)	71% ($p < 0.001$)	73% ($p < 0.001$)	77% ($p < 0.001$)	81% ($p < 0.001$)

**Figure 2.** Gender classification performance in mean +/- standard error % by the models as a function of exposure duration in seconds.

($F_{1,29} = 39.9$, $p < 0.001$). Details of results obtained for the LSTM model have been presented in Table II with the corresponding trend plotted in Figure 2. As shown in the figure, mean performance peaks temporarily at 1.6 sec (about halfway completion of one gait step) with 79% accuracy ($t_9 = 10.1$, $p < 0.001$), suggesting a dependence on dynamic and velocity cues similar to humans at 1.5 sec. Notably, performance at all durations was above chance.

In summary, human observers were able to discriminate gender from gait with significantly above chance performance from moving dot presentations of joints, while conforming to the existing human perception literature. The NM, when presented with moving point representation, performed significantly higher than the human observers, indicating a higher learning and inference capability. Like humans, NM showed a general trend toward better performance with greater stimulus duration as well as a dip in performance in the middle range.

One could argue that the higher performance of NM was a result of a provision of explicit depth information of the joints, which the humans had to infer from the two-dimensional planar display of stimuli on a PC monitor, thus leading to unfair comparison between the models. An additional experiment was conducted to train the NM with 2D data only (by removing the depth, z , component) and compare the results with the NM trained with 3D data. No significant difference in performance behavior was

found. Results on 2D data exhibited similar neuromimetic properties as the model trained on 3D data, including classification accuracy values and the gender sensitivity profile with increase in exposure duration.

The stimuli presented to the models in Experiment 1 conform to how walkers would normally be seen in everyday life, that is, with an upright skeleton. When presented with a vertically inverted (upside down) representation of the walker, humans typically misjudge the gender and invert their decision of gender of the same walker in an upright orientation [1]. To understand the overlap between neural and neuromimetic perception, it is useful to consider the known human inversion effect for evaluating the NM. The next experiment explores the contribution of the vertical inversion of stimuli to compare robustness in gender classification.

5. EXPERIMENT 2: INVERSION EFFECT

The inversion effect is an extensively studied phenomenon in human perception. The effect has been studied through multiple input methods including face inversion [11, 34, 47, 52] and biological motion inversion. When biological motion is presented upside down, perception is strongly impaired [38, 49, 51]. The effect seemed to occur irrespective of the experimental task and affected the detection of a point-light walker [5, 38]. In the case of gender classification from gait, when presented with the vertically inverted stimuli, humans performed significantly below chance with performance varying from 37% to 41%, with significantly higher classification confidence when responding incorrectly. In most cases, humans changed their classification of gender for the same walker when presented with the inverted stimuli [1]. Although Ref. [1] maintains the coherent shape of the walker, it has invited criticism from subsequent works because of the synthetic nature of the stimulus [45], which seems to omit local motion information. Given the importance of local motion, Ref. [35] utilized motion-captured data on human walkers to test for the inversion effect on gender perception, resulting in a chance or near chance performance on the inverted stimuli. Retaining the same theme, this experiment evaluates the neuromimetic model on metrics of accuracy and classification confidence when predicting inverted gait inputs along with introducing the metric *classification inversion probability*.

5.1 Method

5.1.1 Neuromimetic Model

The trained models stored locally in Experiment 1 are evaluated for the inversion effect established in human perception in this experiment. The models are evaluated on the same walkers as in Experiment 1 but through gender classification of vertically inverted (upside down) three-dimensional trajectories of the joints.

5.1.2 Data Input

The test dataset was generated by vertically mirroring the three-dimensional joint trajectories of the walkers on a horizontal plane. Essentially, the y component of the trajectory was mirrored while maintaining the values of the x and z coordinates in the generated dataset. This resulted in a mirrored center-of-mass motion as well. The joint trajectories were further processed through size normalization followed by standardization with a zero mean and unit standard deviation. The most probable output was taken as the class label during classification, and the absolute difference in the classification values between the output nodes was regarded as classification confidence. Here, *classification inversion probability* is defined as the ratio between the number of walkers with opposing gender classifications between upright and inverted orientations to the total number of walkers.

5.2 Results

Across all input durations, the classification performance for the neuromimetic models was below chance at 37% ($t_9 = -3.7$, $p < 0.005$). Performance across different durations remained stable without any significant difference. Overall, classification confidence levels for correctly and incorrectly identified genders showed no significant difference ($F_{1,19} = 0.03$, $0.75 < p < 1.0$). However, at durations above 3.6 sec, classification confidence for incorrectly identified genders was higher than correct responses, with the difference in confidence levels being considerably below zero ($t_9 = -2.7$, $p < 0.05$). The probability for inversion of gender classification remained close to a chance performance of 50% overall. A higher probability for inversion was observed at 3.6 and 3.7 sec duration at 60%, with scope for further investigation for statistical significance.

Both the human and NM demonstrated similar inversion effects, resulting in very similar gender classification outcomes. For NM during shorter durations, classification confidence levels overlap for incorrect and correct classifications. In longer durations, incorrect classifications are made with higher confidence by both the models. The concept of classification confidence is reserved to standard practice in evaluating ANN models without necessarily drawing an exact parallel with human confidence. However, the similar pattern of confidence values between human and machine models might suggest a conceptual similarity.

The overlapping tendency of bias for both the models could be attributed to an over-reliance on the hip and shoulder motions as a result of the skeletal structural

differences between men and women [1, 24, 50]. The objective of the next experiment is to condition the NM models on dynamic elements of human gait using a gender-neutral structure of the walker to observe the change in bias and gender classification performance.

However, evaluation of human perception trained purely on dynamic motion is unattainable as humans need to see the anthropomorphic structure to derive the motion. However, the neuromimetic model can be created and trained on a synthetically generated gender-neutral structure of the walker for learning discriminating features solely from the dynamic cues.

One could argue that humans might have an inversion bias due to lack of access to depth information, which was available to the computational model. Humans had to infer depth from the two-dimensional planar display of stimuli on a PC monitor, potentially leading to unfair comparison between the models. An additional experiment was conducted to train the NM with 2D data (by removing the depth, z , component) of the upright (right side up) stimuli and test on vertically inverted stimuli. Following the protocols of the experiment, the inverted stimuli were created by mirroring the y component of the joint trajectories. A significant reduction in classification accuracy was observed between the NMs tested on the upright stimuli and the inverted stimuli ($p < 0.05$). When tested on the upright 2D stimuli, NM classified gender with a mean accuracy of 77.6% ($t_9 = 8.0$, $p < 0.05$), with a chance accuracy of 50%. When tested on the inverted 2D stimuli, the classification accuracy was decreased to a mean accuracy of 41.1% ($t_9 = -1.8$, $0.05 < p < 0.1$). This demonstrates the emergence of an effect similar to the inversion effect in the absence of depth information as well.

6. EXPERIMENT 3: CONTRIBUTION OF STRUCTURAL CUES FOR NEUROMIMETIC

Troje et al. generated a gender-neutral posture by averaging postures across all the participants of the study [50]. The resulting walker possesses a generic anthropomorphic posture within the variance of the participating subjects, which is then used for a human observation study. Although this approach is perfectly suited for human observers, given their *a priori* assumption of an anthropomorphic model, the neuromimetic models provide us with greater flexibility of experimentation because of their ability to learn a generic spatiotemporal stimulus. This allows for a higher range of postural modifications, which can generalize beyond the variance of postures available in the dataset at hand. In this experiment, a new neuromimetic model is created for comparison with the model used in previous experiments and evaluated for gender classification accuracy and robustness to vertical inversion of the walker. For the purpose of this experiment, the model trained with veridical walkers shall be referred to as NM₁ and the new model as NM₂. NM₂ is trained on a gait input sequence synthesized by modifying the structure of the walkers in the existing data to reflect a gender-neutral body structure.

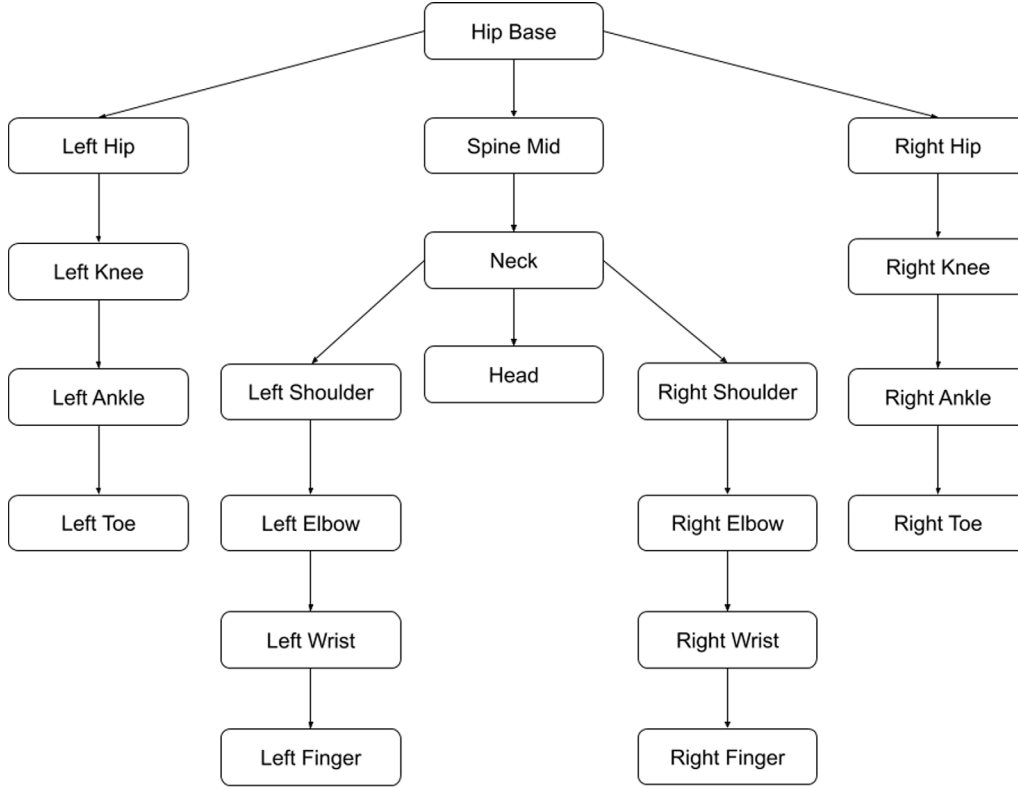


Figure 3. Joint dependency tree of the human body representing the parent and child joints originating from the hip base.

6.1 Method

6.1.1 Neuromimetic Model

The model architecture was the same as that of the previous experiments. The training and testing of NM₂ adhered to the same protocols as followed by the previous experiments, including creation of ten LSTMs with randomly generated initial weights, providing nine degrees of freedom during the evaluation of results.

6.1.2 Data Input

The input dataset was generated by changing the limb lengths of each of the 19 limbs connecting the 20 joints to have unit length. Figure 3 describes the joint dependency tree of the human body as a hierarchy of attached joints, where each node in a subsequent layer of the tree is dependent on its parent node. The new three-dimensional joint trajectory of the gender-neutral structured walker was determined by adjusting the limb length to one and calculating the new trajectory in the direction of the limb. The hip base is taken as the reference joint for calculating the new joint trajectories. The new joint trajectories were determined using the following steps:

$$L = |i_{\text{pos}} - x_{\text{pos}}| \quad (1)$$

$$\hat{D} = (i_{\text{pos}} - x_{\text{pos}})/L \quad (2)$$

$$i'_{\text{pos}} = x_{\text{pos}} + \hat{D}, \quad (3)$$

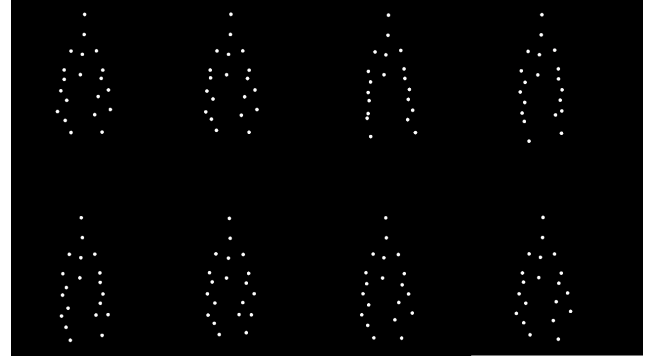


Figure 4. PointLight representation of a walker with unit limb lengths at different stages of the gait cycle. This stimulus dataset is used for training and testing NM₂ to evaluate for its dependence on structural cues of the walker.

where L is the limb length between the parent joint x and the dependent joint i , with their trajectories being represented by x_{pos} and i_{pos} , respectively. $\hat{D}$ is the unit vector in the direction of the limb vector and i'_{pos} is the new three-dimensional trajectory of the dependent joint i after the structural correction. The process is repeated for each of the 19 limbs of the body. The result of structural corrections in the static frames is demonstrated in Figure 4. The model is trained and tested with the new dataset similarly to Experiments 1 and 2.

6.2 Results

The NM₂ performed at 75% mean accuracy in gender classification across all the input durations ($t_9 = 7$, $p < 0.001$), with the highest mean accuracy of 77.3% at 3.8 sec exposure duration ($t_9 = 6$, $p < 0.05$). There was no significant difference in gender classification between the models NM₁ and NM₂ ($F_{1,19} = 0.004$, $0.75 < p < 1.0$), signifying alternative dynamic gender discriminatory cues available for the model to learn to classify gender with similar accuracy in the case of vertically inverted walkers. NM₂ performed slightly above chance performance at 53% ($t_9 = 2$, $p < 0.1$).

However, there was a significant improvement in gender classification of vertically inverted walkers between the two neuromimetic models ($F_{1,19} = 138$, $p < 0.001$), suggesting a high contribution of structural cues toward the bias, leading to poor performance in NM₁ and increase in robustness with dynamic cues in NM₂. The difference in classification confidence values between correct and incorrect responses was however not significant.

A corresponding 2D version of the experiment was performed by removing depth information from the joint trajectories (as described in Experiment 2) to evaluate the validity of the proposed strategy. In the case of upright stimuli, the 2D version of NM₂ classified gender with the highest mean accuracy of 81.2% at 3.8 sec of exposure duration ($t_9 = 9.9$, $p < 0.05$). This was significantly higher than the chance performance of 50%. However, results from NM₂ and NM₁ ($F_{1,19} = 0.6$, $p > 0.4$) were not significantly different.

In the case of inverted 2D stimuli, NM₂ classified gender with the highest mean accuracy of 49.6% ($t_9 = -0.1$, $p > 0.9$) at 3.8 sec exposure duration. However, there was no significant difference between NM₂ and the chance performance of 50%. Additionally, there was no significant difference between NM₂ and NM₁ in terms of gender classification from inverted 2D stimuli.

The removal of structural cues led to an improvement in the mean gender classification accuracy. However, given the number of models for comparison, the improvement cannot be deemed statistically significant, warranting the application of additional feature extraction steps for enhancing the outcome. Providing dynamic and velocity cues to human observers has led to an improvement in gender classification [50]. The next experiment leverages this result to perform an extensive evaluation of dynamic cues of joint velocities and acceleration derived from veridical and structurally corrected walkers with the aim of establishing the best model training and testing strategy for accuracy, robustness, and generalizability to non-gender-related gait classification tasks.

7. EXPERIMENT 4: SPATIOTEMPORAL FEATURE EXTRACTION STRATEGIES FOR NEUROMIMETIC MODELS

This experiment aims to evaluate NM performance on datasets that have been synthetically generated through the

application of various spatiotemporal feature engineering pre-processing steps on the veridical walkers' dataset. The spatial pre-processing includes, (1) veridical walkers' structure, and (2) structurally corrected walkers to have unit limb lengths, as mentioned in Experiment 3. Temporal pre-processing includes, (1) Position, (2) Velocity, and (3) Acceleration of the three-dimensional joint trajectories. The objective of the analysis is to establish a strategy of choosing the appropriate model and pre-processing steps with a given threshold and priority of performance measures.

7.1 Method

7.1.1 Neuromimetic Model

The neuromimetic architecture for building the framework of the model, training and testing protocol is the same as that in the previous experiments. However, differences in the spatiotemporal feature engineering pre-processing steps would result in six different models. The nomenclature of the models is established according to the following pre-processing steps:

Spatial\Temporal	Position	Velocity	Acceleration
Veridical structure	NM _{pos}	NM _{vel}	NM _{acc}
Unit limb length structure	NM _{pos,ull}	NM _{vel,ull}	NM _{acc,ull}

NM_{pos} and NM_{pos,ull} refer to the models NM₁ and NM₂ in Experiment 3, respectively.

7.1.2 Data Input

Temporal derivatives of the gait of the veridical walkers and unit limb length walkers were used for generating the corresponding velocity and acceleration values from the position joint trajectories. The positional data was smoothed using a five-frame moving average filter before calculating the derivatives for the subsequent frames. The data underwent size normalization and standardization following the guidelines from the previous experiments before training and testing of the models. The procedure was replicated to generate the data for the vertically inverted walkers, by mirroring the joint trajectories on a horizontal plane passing through the center of mass of the body, as described in Experiment 2. Subsequently, temporal derivatives of the trajectories provide the corresponding velocity and acceleration of the joints. The trained models are tested on the upright walkers' data for accuracy and on inverted walkers for robustness. The corresponding classification confidence values are stored for further analysis.

7.2 Results

The NM_{acc,ull} model trained on joint accelerations of walkers with unit limb lengths possessed the highest overall gender classification accuracy of 84% ($t_9 = 9.4$, $p < 0.001$), with the performance reaching 87% at 3.8 sec ($t_9 = 12$, $p < 0.001$). Coincidentally, NM_{acc,ull} also had the highest performance in classification in vertically inverted walkers, with an accuracy of 77.4% ($t_9 = 12$, $p < 0.001$), with a

Table III. Gender classification accuracy (%) of the neuromimetic model across all the stimulus exposure durations (seconds).

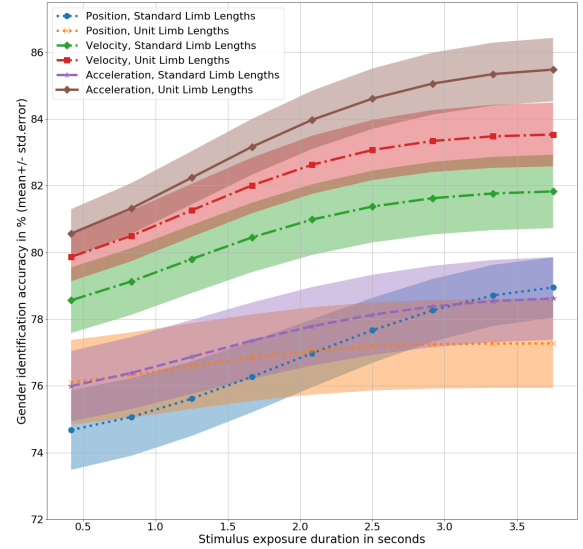
Exposure Duration/Neuromimetic Model	0.4	0.8	1.2	1.5	2.0	2.5	2.9	3.3	3.8
NM _{pos}	71	74	73	80	76	77	78	80	81
NM _{vel}	79	80	81	82	80	82	83	81	82
NM _{acc}	76	77	77	77	78	79	77	79	79
NM _{pos, ull}	73	76	79	75	79	76	78	76	77
NM _{vel, ull}	78	80	82	84	84	83	85	83	83
NM _{acc, ull}	78	83	83	85	84	86	85	85	87

Table IV. Average Classification inversion probability across all the durations of the neuromimetic models.

Neuromimetic Model	Classification Inversion Probability
NM _{pos}	0.52
NM _{vel}	0.38
NM _{acc}	0.46
NM _{pos, ull}	0.36
NM _{vel, ull}	0.12
NM _{acc, ull}	0.12

significant difference in classification confidence levels in favor of correct gender classification ($F_{1,19} = 8$, $p < 0.05$), supporting the argument of robustness of dynamic cues as opposed to structural cues. The performance of the model improved with increasing input duration available from 0.4 to 3.8 sec for the upright walker orientation ($F_{1,19} = 5.1$, $p < 0.05$) as well as the inverted orientation ($F_{1,19} = 20$, $p < 0.001$). Results from the NM_{acc, ull} model further demonstrate the presence of gender-specific, distinct and robust features in the dynamics, specifically acceleration, of the joint motion trajectories of the walker with a unit limb length structure. Detailed results of the models have been presented in Table III and Figure 5, with NM_{acc, ull} closely followed by NM_{vel, ull} with an accuracy of 82% ($t_9 = 12$, $p < 0.001$) on the upright orientation and 78% ($t_9 = 14$, $p < 0.001$) on the inverted orientation. The results presented in Fig. 5 have been enhanced using a Gaussian filter (with sigma of 3) to emphasize the trends in a visual manner. Additionally, NM_{acc, ull} has the lowest overall classification inversion probability of 0.12 across all the durations as shown in Table IV. The models trained on veridical walker body structures, NM_{acc} and NM_{vel}, performed similarly on the inverted walkers with an overall accuracy of 61%, $p < 0.001$. However, a marked difference was observed in their performance on the upright orientation ($F_{1,19} = 3.5$, $p < 0.1$). Structural and temporal processing achieved higher accuracy and robustness.

A corresponding 2D version of the experiment was performed by removing depth information from the joint trajectories (as described in Experiment 2) to evaluate the validity of the proposed strategy. In the case of upright

**Figure 5.** Gender classification performance in mean $\pm$ standard error in % of all the models as a function of exposure duration in seconds. The performance values are filtered through a one-dimensional Gaussian filter with a standard deviation of 3 for the Gaussian kernel.

stimuli, the 2D version of NM_{acc, ull} classified gender with a mean accuracy of 86.8% at 3.8 sec of exposure duration ($t_9 = 8$, $p < 0.05$). This was a significantly higher accuracy than that of NM_{pos} ($F_{1,19} = 4.4$, $p < 0.05$) with a mean accuracy of 77.6% ($t_9 = 8$, $p < 0.05$) at 3.8 sec exposure duration. In the case of inverted stimuli, NM_{acc, ull} was able to classify gender with a mean accuracy of 66.3% ($t_9 = 4.9$, $p < 0.05$) at 3.8 sec exposure duration, which is significantly higher than that of NM_{pos} ($F_{1,19} = 20$, $p < 0.05$). This validates the proposed approach both in terms of improvement in classification accuracy and robustness to inversion in the absence of depth information as well.

In summary, gender classification accuracy and robustness increased with every additional spatiotemporal feature processing step, revealing more readily available gender discriminating cues with additional pre-processing. The aforementioned processing steps include converting from veridical limb length to unit limb lengths and going from the computing position to velocity to acceleration of the joint trajectories. The lack of structural influence on the joint trajectories could exaggerate the behavioral differences expressed through motion between male and female walkers, leading to robustness in the inversion of the walker. The NM also demonstrated capacity for self-learning the relevant cues without the need for hand-crafted features. The difference in performance accuracy between NM_{pos} and NM_{acc, ull} is significant ($F_{1,19} = 13$, $p < 0.001$) despite the basic nature of the spatiotemporal feature extraction. Thus as a corollary, the difference in performance accuracy between NM_{pos} and human perception is significant as well ($F_{1,29} = 113$, $p < 0.001$).

Considering the similarity in inversion effect between NM_{pos} and humans, the difference in robustness could be extended toward the perceptual bias of humans as well.

8. DISCUSSION

The psychophysical results are consistent with previous studies in visual perception [1, 24–26, 50], suggesting a plausible biological baseline for comparing human performance with that of neuromimetic models (NM). Experiment 1 shows a consistently significant quantitative difference in performance in gender discrimination between humans and NM yet a qualitative similarity in changes in performance with stimulus duration. Experiment 2 reveals that a common inversion effect is shared between humans and NM in the misclassification of inverted stimuli. The emergence of the inversion effect in the neuromimetic models as a result of learning the association between gender and biological motion supports the hypothesis that one may not need to resort to specialized mechanisms to explain the inversion effect in humans. The commonality also suggests that humans may operate similar generic learning mechanisms as NM in processing biological motion. That the neuromimetic models tested exhibit the same characteristic after training (even when initialized with random weights) suggests that the behavior emerges due to convergence toward a set of weights that are optimized for gender classification, and this provides additional support for the behavior being a result of the training itself.

Experiments 3 and 4 were directed at understanding and improving the NM to develop a high-performing gait classifier that is more robust than human perception. The absence of preconceived anthropomorphics in the model is utilized in Experiment 3 through a novel gender-neutral representation of the body structure to mitigate the inversion effect and improve classification performance to progressively higher levels. Experiment 4 leveraged the observation that humans use dynamics in preference to positional cues for identifying gender from gait [1] to train the neuromimetic models on either velocity or acceleration of the skeletal joints. Results show that higher temporal derivatives improve the accuracy, robustness, and efficiency of the models. The improvement in performance, despite the potential loss of relevant information through the removal of structural cues, highlights the redundant nature of the information extracted by both humans and machine-learning models. However, the advantage of machine models lies in the ability to restrict the type of information used, which is not necessarily the case in humans. For example, although acceleration of foot trajectories is perceptually important [9], humans may find it comparatively difficult to isolate such a feature from the general context to improve performance. Machine models, on the other hand, allow for testing on arbitrary modifications to biological motion vectors, whether it is to improve model performance or to extend understanding of human perception.

9. CONCLUSION

Gender classification from human gait was used to evaluate differences and perceptual commonalities between human and artificial learners. A neuromimetic machine-learning model that does not require hand-crafted features shares aspects of biological motion perception similar to human perception. Furthermore, additional modifications guided by human perception are shown to exceed the ability of humans in classifying gender from gait. The results provide support for a generic, rather than a pre-tuned, learning system in human visual perception, potentially precluding one from requiring special mechanisms to explain the inversion effect in humans. The effect can be argued to have originated given insufficient training instances of inverted stimuli. This approach may allow for robust gait classification in other applications. Other attributes of the walker such as age, weight, emotional state, and personality traits could be treated in a similar way. Given an extended dataset, it is straightforward and analogous to the gender classification problem to train and test the model for other attributes that may be represented in walking patterns.

Aspects of human perception that are shared by the machine models include dominance of dynamic information, presence of an inversion effect, and improved performance with availability of more data. As with any predictive model, neuromimetic models can possess biases of their own. However, the significantly higher performance and difference in known biases allow the models to be utilized either in isolation or in combination to compensate for each other's biases.

The study demonstrates the use of neuromimetic models as a paradigm for studying human perception and for developing automatic gait classifiers that mitigate perceptual characteristics selectively to significantly exceed human performance. Because mobility through gait is a cross-cutting manifestation of disease across many healthcare conditions, future work could include adapting the proposed neuromimetic model for the diagnosis and assessment of interventions in gait-impairing conditions such as stroke, Parkinson's disease, and osteoarthritis.

ACKNOWLEDGMENT

This research was partially supported by a grant [G0041501] from the Medical Research Council (UK) titled "Clinical movement markers of osteoarthritis for treatment with regenerative medicine".

REFERENCES

- 1 C. D. Barclay, J. E. Cutting, and L. T. Kozlowski, "Temporal and spatial factors in gait perception that influence gender recognition," *Perception Psychophysics* **23**, 145–152 (1978).
- 2 R. Barrett, M. VonkNoordegraaf, and S. Morrison, "Gender differences in the variability of lower extremity kinematics during treadmill locomotion," *J. Motor Behavior* **40**, 62–70 (2008).
- 3 I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *J. Microbiological Methods* **43**, 3–31 (2000).

- ⁴ T. Beardsworth and T. Buckner, "The ability to recognize oneself from a video recording of one's movements without seeing one's body," *Bull. Psychonomic Society* **18**, 19–22 (1981).
- ⁵ B. I. Bertenthal and J. Pinto, "Global processing of biological motions," *Psychological Science* **5**, 221–225 (1994).
- ⁶ R. Borràs, À. Lapedriza, and L. Igual, "Depth information in human gait analysis: An experimental study on gender recognition," *Int'l. Conf. Image Analysis and Recognition* (Springer, Berlin, Heidelberg, 2012) pp. 98–105.
- ⁷ A. de Brébisson and P. Vincent, "An exploration of softmax alternatives belonging to the spherical loss family," Preprint arXiv:1511.05042 (2015).
- ⁸ L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," *Proc. 16th ACM Int'l. Conf. on Multimedia* (ACM, New York, NY, 2008) pp. 725–728.
- ⁹ D. H. F. Chang and N. F. Troje, "Acceleration carries the local inversion effect in biological motion perception," *J. Vision* **9**, 19–19 (2009).
- ¹⁰ D. H. F. Chang and N. F. Troje, "Characterizing global and local mechanisms in biological motion perception," *J. Vision* **9**, 8–8 (2009).
- ¹¹ C. Civile, R. P. McLaren, and I. P. L. McLaren, "The face inversion effect—Parts and wholes: Individual features and their configuration," *Quarterly J. Experimental Psychology* **67**, 728–746 (2014).
- ¹² R. A. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. E. Webster, L. Denehy, and A. L. Bryant, "Validity of the Microsoft Kinect for assessment of postural control," *Gait Posture* **36**, 372–377 (2012).
- ¹³ R. A. Clark, Y.-H. Pua, A. L. Bryant, and M. A. Hunt, "Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining," *Gait Posture* **38**, 1064–1066 (2013).
- ¹⁴ R. A. Clark, K. J. Bower, B. F. Mentiplay, K. Paterson, and Y.-H. Pua, "Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables," *J. Biomech.* **46**, 2722–2725 (2013).
- ¹⁵ J. E. Cutting and L. T. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society* **9**, 353–356 (1977).
- ¹⁶ J. W. Davis and H. Gao, "An expressive three-mode principal components model for gender recognition," *J. Vision* **4**, 2–2 (2004).
- ¹⁷ W. H. Dittrich, T. Troscianko, S. E. G. Lea, and D. Morgan, "Perception of emotion from dynamic point-light displays represented in dance," *Perception* **25**, 727–738 (1996).
- ¹⁸ M. van Diest, J. Stegenga, H. J. Wörtche, K. Postema, G. J. Verkerke, and C. J. C. Lamoth, "Suitability of Kinect for measuring whole body movement patterns during exergaming," *J. Biomech.* **47**, 2925–2932 (2014).
- ¹⁹ F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," (1999): 850–855.
- ²⁰ M. A. Giese and T. Poggio, "Cognitive neuroscience: neural mechanisms for the recognition of biological movements," *Nature Reviews Neuroscience* **4**, 179 (2003).
- ²¹ J. Han and B. Bir, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 316–322 (2006).
- ²² H. Hill and A. Johnston, "Categorizing sex and identity from the biological motion of faces," *Current Biology* **11**, 880–885 (2001).
- ²³ G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," *Asian Conf. Computer Vision*. (Springer, Berlin, Heidelberg, 2007).
- ²⁴ G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception Psychophysics* **14**, 201–211 (1973).
- ²⁵ G. Johansson, "Visual motion perception," *Sci. Am.* **232**, 76–89 (1975).
- ²⁶ G. Johansson, "Spatio-temporal differentiation and integration in visual motion perception," *Psychological Res.* **38**, 379–393 (1976).
- ²⁷ D. Kingma and J. Ba, Adam: A method for stochastic optimization. Preprint arXiv:1412.6980 (2014).
- ²⁸ L. T. Kozlowski and J. E. Cutting, "Recognizing the sex of a walker from a dynamic point-light display," *Perception Psychophysics* **21**, 575–580 (1977).
- ²⁹ J. Lange and M. Lappe, "A model of biological motion perception from configural form cues," *J. Neurosci.* **26**, 2894–2906 (2006).
- ³⁰ L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," *Proc. Fifth IEEE Int'l. Conf. on Automatic Face Gesture Recognition* (IEEE, Piscataway, NJ, 2002).
- ³¹ G. Mather and L. Murdoch, "Gender discrimination in biological motion displays based on dynamic cues," *Proc. R. Soc. B* **258**, 273–279 (1994).
- ³² Y. Ma, H. M. Paterson, and F. E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior Res. Methods* **38**, 134–141 (2006).
- ³³ A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *Proc. Icml*. **30** (2013).
- ³⁴ E. McKone and G. Yovel, "Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing," *Psychonomic Bulletin Review* **16**, 778–797 (2009).
- ³⁵ B. McGlothlin, D. Jaconetti, and L. Yandell, "The inversion effect: biological motion and gender recognition," *Psi Chi J. Psychological Res.* **17** (2012).
- ³⁶ K. Stefan, T. Mikolov, M. Karafiát, and L. Burget, "Recurrent neural network based language modeling in meeting recognition," *Twelfth Annual Conf. of the Int'l. Speech Communication Association* (2011).
- ³⁷ R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," *Int'l. Conf. on Machine Learning* (2013).
- ³⁸ M. Pavlova and A. Sokolov, "Orientation specificity in biological motion perception," *Perception Psychophysics* **62**, 889–899 (2000).
- ³⁹ A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis," *J. Medical Engineering Technol.* **38**, 274–280 (2014).
- ⁴⁰ A. Phinyomark, S. T. Osis, B. A. Hettinga, D. Kobsar, and R. Ferber, "Gender differences in gait kinematics for patients with knee osteoarthritis," *BMC Musculoskeletal Disorders* **17**, 12 (2016).
- ⁴¹ F. E. Pollick, L. Vaia, R. Jungwon, and S.-B. Cho, "Estimating the efficiency of recognizing gender and affect from biological motion," *Vision Res.* **42**, 2345–2355 (2002).
- ⁴² F. E. Pollick, J. W. Kay, K. Heim, and R. Stringer, "Gender recognition from point-light walkers," *J. Experimental Psychology: Human Perception and Performance* **31**, 1247 (2005).
- ⁴³ F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review* **65**, 386 (1958).
- ⁴⁴ D. R. Saunders, D. K. Williamson, and N. F. Troje, "Gaze patterns during perception of direction and gender from biological motion," *J. Vision* **10**, 9–9 (2010).
- ⁴⁵ D. R. Saunders, J. Suchan, and N. F. Troje, "Off on the wrong foot: Local features in biological motion," *Perception* **38**, 522–532 (2009).
- ⁴⁶ Y. Shi, M. Xiaochi, M. Zheng, W. Jiahuan, Y. Nailang, G. Quan, C. Wang, and Z. Gao, "Using a Kinect sensor to acquire biological motion: Toolbox and evaluation," *Behavior Research Methods* **50**, 518–529 (2018).
- ⁴⁷ F. Simion, L. Regolin, and H. Bulf, "A predisposition for biological motion in the newborn baby," *Proc. National Academy of Sciences* **105**, 809–813 (2008).
- ⁴⁸ S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "Gait energy volumes and frontal gait recognition using depth images," *2011 Int'l. Joint Conf. on Biometrics (IJCB)* (IEEE, Piscataway, NJ, 2011), pp. 1–6.
- ⁴⁹ S. Sumi, "Upside-down presentation of the Johansson moving light-spot pattern," *Perception* **13**, 283–286 (1984).
- ⁵⁰ N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *J. Vision* **2**, 2–2 (2002).
- ⁵¹ N. F. Troje and C. Westhoff, "The inversion effect in biological motion perception: Evidence for a 'life detector'?", *Current Biol.* **16**, 821–824 (2006).
- ⁵² R. K. Yin, "Looking at upside-down faces," *J. Experimental Psychology* **81**, 141 (1969).
- ⁵³ S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, "A study on gait-based gender classification," *IEEE Trans. Image Process.* **18**, 1905–1910 (2009).
- ⁵⁴ A. A. Palacios, "Neuromimetic intelligence," *Lorraine Laboratory of Research in Computer Science and its Applications* (French Institute for Research in Computer Science and Automation, 2012).