Important feature identification for perceptual sex of point-light walkers using supervised machine learning

Chihiro Asanoi
Graduate School of Human Science, Tokyo Woman’s Christian University, Tokyo, Japan

Koichi Oda
School of Arts and Sciences, Tokyo Woman’s Christian University, Tokyo, Japan

The present study aimed to elucidate the dynamic features that are highly predictive in the biological and perceptual sex classification of point-light walkers (PLWs) and how these features behave in sex classification using supervised machine learning. Fifteen observers judged the sex of 21 PLWs from a side view. A fast Fourier transform was applied to retrieve the spectral components from the multiphasic hip and shoulder movements. An exhaustive search identified the most important features for biological and perceptual sex classifications. An individual conditional expectation (ICE) with a support vector machine (SVM) model was used to interpret the behavior of each important feature. The observers judged the biological sex from side-view PLWs with an accuracy of 62.9% for 10 male PLWs and of 57.0% for 11 female PLWs. The SVM model for biological sex prediction demonstrated that the third harmonic of hip motion played a dominant role in achieving a high predictive accuracy of 90.5% with few feature interactions. In the model of perceptual sex prediction, however, an accurate prediction of 85.7% was achieved using five spectral components of hip and shoulder motions, where the ICE plots of the features followed heterogeneous courses, suggesting feature interactions. The machine learning model suggests that biological sex classification depends mainly on local cues of the PLW. However, the high-performance model of perceptual sex classification involves interactions of various frequency components of hip and shoulder motions, suggesting more complex processes in sex perception.

Introduction

Since Kozlowski and Cutting (1977) described that sex was readily perceived from point-light walkers (PLWs) or biological motion (Johansson, 1973, 1976), many studies have attempted to find cues for the perception of sex in the motion of a small number of dots. Cutting, Proffitt, and Kozlowski (1978) and Barclay, Cutting, & Kozlowski (1978) reported that the center of moment derived from the ratio of shoulder and hip widths served as a structural cue to discriminate between males and females in the frontal view. However, this cue is not readily available for a lateral view because it hides the structural differences in the hip and shoulder widths. Cutting (1978) suggested that the elliptical motion of the hip and shoulder in the lateral view could be a dynamic cue for the biological sex. Mather and Murdoch (1994) also focused on these joints and found that, in the lateral view, the rapid hip movement serves as a dynamic cue of female PLWs. Their findings indicate that local dynamic cues are of particular importance in biological sex classification from lateral PLWs. To examine what PLW information is encoded into patterns that represent sex, Troje (2002) constructed a novel model to classify males and females by applying a linear discriminant analysis to joint movements. Instead of local dynamic cues, he used superimposed dynamic cues using principal component analysis (PCA) and achieved a highly accurate prediction of biological sex with the classification model. However, it remains unclear which dynamic cues, local or holistic, contribute more to the classification of biological sex because no direct comparison of the discriminatory power of these cues has been performed.

Most of these previous studies have assumed that the perception of sex should coincide with the biological sex of PLWs (Kozlowski & Cutting, 1977; Barclay et al., 1978; Cutting et al., 1978; Cutting, 1978; Mather & Murdoch, 1994; Hirashima, 1997; Hirashima, 1999; Sumi, 2000; Troje, 2002; Pollick, Lestou, Ryu, & Cho, 2002; Davis & Gao, 2004; Pollick, Kay, Heim, & Stringer, 2005), and accordingly, the perception of male for a female PLW is considered false. However, previous studies have repeatedly shown that the correct percentage of sex classification of side-view PLWs has not reached the threshold of 75%, although it has exceeded the chance level (Pollick et al., 2005). This low...
accuracy suggests that human observers may obtain some impression related to sex (perceptual sex) rather than the actual biological sex. To solve this problem, we need to compare the important cues and accuracy between two classification models with different targets, perceptual and biological sex. Recently, Roether, Omlor, Christensen, and Giese (2009) examined the perceptual relevance of body movements and found that human observers readily recognize emotions expressed in body movements. Applying the sparse regression method and anechoic demixing, they extracted critical emotion-specific dynamic features and confirmed that the perceptual relevance of these features was supported by artificial walkers having only critical features. These findings suggest the presence of perceptual sex-specific dynamic features, regardless of the biological sex.

The present study aimed to define the important dynamic cues from the lateral view that contribute to the classification of perceptual and biological sexes. To accomplish this study, we extracted local features by applying a fast Fourier transform (FFT) (Unuma et al., 1995) to hip and shoulder joint movements and obtained superimposed features by applying PCA to all joint movements. Using these features and machine learning algorithms, we constructed classification models: first, to identify the important features of biological and perceptual sex classifications; second, to compare the contribution of local and superimposed dynamic cues to these sex classifications; and third, to provide details on how the features impact the prediction of both sexes.

### Methods

#### Experiment

**Stimulus**

Twenty-one undergraduate students (10 male and 11 female) with 13 white markers attached to important positions walked on a treadmill wearing sneakers at a speed of approximately 4 km/h for 1 min. Their movements were digitally recorded using a side-view camera (Sony Digital HandyCam DCR-VX2100; Sony, Tokyo, Japan) from the left side at 30 Hz. The positions of the 13 white markers were as follows: one on the head, two on the shoulders, two on the elbows, two on the wrists, two on the hips, two on the knees, and two on the ankles (Asanoi & Osada, 2005, 2006).

The X–Y coordinates of 11 visible white markers (2 markers were not visible because they were on the other side of the body) were digitized frame-by-frame for each walker, and a PLW stimulus animation was generated according to these coordinates as a sequence of white dots in the dark background. The sequence presented a complete stride cycle in 60 frames and lasted for 2 s. Figure 1A shows the trajectories of all the 11 markers in the stimulus animation. At an observation distance of 60 cm from the display, we scaled each display and set each PLW to have the same height and width using this scaling factor. This size corresponded to 11° of visual angle for the screen of a laptop computer (Windows 8.1 Dell Inspiron 11 3000 series; Dell, Round Rock, TX, USA).

**Procedure**

Fifteen Japanese observers (8 females and 7 males, aged 52.3 ± 19.2 years) participated in the experiment. All the participants had normal or corrected-to-normal visual acuity. After providing informed consent, the observers sat 60 cm away from the touchscreen in a light room. To familiarize themselves with the trial, all the observers practiced answering the sex of 21 PLWs twice while watching them on the screen. In the main session, each PLW was displayed twice for 2 s in random order at the center of the monitor, and the observers answered whether the walker looked like a male or female by touching the screen. In the present study, we defined perceptual sex as the sex perceived by the observer, regardless of the actual biological sex. After the observers finished responding to each stimulus, the trial moved on to the next stage. Each observer performed one session responding to each of the 21 different PLWs and repeated the session 10 times in random order of PLW presentation. In total, 150 judgments for each PLW were made by the 15 observers. The study was conducted in accordance with the Declaration of Helsinki (revised in October 2013) and approved by the College Ethics Committee of Tokyo Women’s Christian University (Ethics ID: A2018-06).

**Study 1: Sex perception rate and local features**

**Sex perception rate**

Each observer was asked to classify whether the PLW looked male or female every time the PLW was randomly presented on the screen. We averaged the classification results of the 150 judgments on each PLW, which were obtained from the 10 trials for each PLW carried out by the 15 observers, and examined the percentage of biological sex classification.

**Time- and frequency-domain analyses**

We defined the trajectories of all joint markers of the lateral PLWs used in the sex classification task. Because the time series of the joint motions had a periodic nature around the head position, as shown
Figure 1. Extraction of joint movements from a lateral point-light walker. (A) Eleven point-lights observable from the left-side view were detected frame-by-frame to trace the trajectories of lateral point-light movements. (B) The distance of each joint from the head along the horizontal axis of the walking direction was measured and displayed within a two-dimensional framework of time and distance.

In Figure 1A, we measured the distance from the head to each joint along the horizontal axis of the walking direction. The cyclic changes in the distance from the head position demonstrated that limb movements, including elbows, wrists, knees, and ankles, represented a typical pendulum motion, where the upper and lower extremities of the same side were moving in an opposite direction, nearly as a mirror image, in both males and females (Figure 1B). In contrast to the simple monophasic pendulum motion of the limbs, the hip and shoulder movements demonstrated multiphasic movements with smaller amplitudes than those of the limbs.

Almost all the spectral components of the walking joint movements were found within 6 Hz, including the third harmonic, as shown in Figures 1B and 2. To delete the high-frequency noise above 6 Hz, we applied a zero-phase digital filter that performed moving average filtering by processing the input data in both the forward and reverse directions, resulting in no phase shift. As the ankles had the most typical pendulum motion, we extracted two cycles of the left ankle pendulum motion from peak to peak and reconstructed the time series by connecting two subsets (four cycles) of the same data with the same phase. Within the same time range for ankle data extraction, the same procedure was applied.
to the time series of all joint movements. We used the zero-phase digital filter again for these connected cycles, so that the end of one cycle was seamlessly looped to the start of the next cycle. To retrieve the spectral components from joint movements, we applied FFT to the time series (30 Hz) of their movements. Because the amplitudes of higher-order spectral components were negligible (Troje, 2002; Westhoff & Troje, 2007), we adopted the fundamental frequency (the first harmonic) corresponding to one gait cycle and the next second and third harmonics (two and three times the frequencies of the fundamental one) within 6 Hz and measured their amplitudes as features of sex classification (Figure 2, lower panel).

Statistics

Before calculating the average perceptual sex rate for each PLW, we observed the distribution of individual sex perception rates for each PLW and performed a Shapiro–Wilk test to examine whether the observers’ responses to each PLW were distributed normally. The test revealed a normal distribution in 16 of the 21 PLWs. Therefore, we expressed the correct recognition rate for sex as means ± standard deviations. The difference from the chance level in the observers’ ability to determine sex was examined using the binominal test. Unpaired t-tests were used to assess the differences in the spectral components of hip and shoulder motions between male and female. The statistical significance was set at \( p < 0.05 \). We calculated the response bias (C) to determine whether there was an overall bias toward males or females based on the signal detection theory.

Study 2: Important feature contribution to sex classification with machine learning

To evaluate the importance and contribution of PLW features, we constructed two classification models with different targets: biological sex (male = 1, female = 0) and perceptual sex (male = 1, female = 0). Reducing the number of features is important in machine learning because many features may not produce a desired outcome by overfitting the learning algorithm to noise (Brink, Richards, & Fetherolf, 2016). In the present study, we performed feature selection using forward and backward sequential methods and exhaustive search and feature transformation using PCA. The optimal machine learning algorithm was determined from the performance of six algorithms: discriminant analysis, \( k \)-nearest neighbors, naive Bayes, random forest, Gaussian kernel support vector machine (kernel SVM), and ensemble algorithms. In \( k \)-nearest neighbors, there is always a trade-off in setting the value of \( k \). When \( k \) is low (\( k = 1 \)), the algorithm becomes sensitive to noise in
In fact, as Troje (2002) summarized, Cutting (1978) iteratively finds the irrelevant features and iteratively finds the predictive features from all features. To ensure that all features were considered equally, we normalized the data to be in the range of 0 to 1 (Theodoridis, Pikrakis, Koutroumbas, & Cavouras, 2010). We performed machine learning modeling using a MATLAB-based library (Statistics and Machine Learning Toolbox 2021a; The MathWorks, Inc., Natick, MA, USA).

**Feature selection and feature transformation**

Feature selection can serve multiple purposes, including increasing the interpretability of our classifications, improving the computational performance, and strengthening the quality of machine learning classification. One of our goals was to identify a subset of features with different effects on prediction. Although it may be ideal to consider all joint spectral components for sex classification, it is certainly impractical to include a large number of input variables. In the present study, all side-view PLWs displayed on the screen were set to have the same height and width, proportional to the correcting height. Consequently, there was little difference in the displayed size of each walker. Furthermore, the ankle movement consisted of a simple sine curve (Figure 1B) of the first harmonic corresponding to the gait cycle. Similarly, other joint movements of the extremities (knees, wrists, and elbows) were highly correlated with ankle movements. On the contrary, hips and shoulders have multiphasic movements with relatively low correlation (correlation coefficient: $-0.51 \pm 0.06$ in males, $-0.51 \pm 0.11$ in females) with each other, suggesting that more information is included in these movements (Figure 2). In fact, as Troje (2002) summarized, Cutting (1978) identified the elliptical motion of the shoulder and hip in the sagittal plane as an important cue to sex, and Mather and Murdoch (1994) reported a rapid lateral sway of hip movement as an important cue to sex, height, and the query (Kelleher, Mac Namee, & D’arcy, 2020). To determine the optimal machine learning algorithm, we used an exhaustive search with each algorithm to examine all the combinations of features selected by forward and backward sequential approaches. In an exhaustive feature search, there are $nCr$ unique combinations of $r$ features: $nCr = n! / r!(n – r)!$. We applied each algorithm to features selected by sequential approaches and assessed all feature combinations, expecting that the performance would initially increase with the inclusion of more informative variables and eventually decrease as a high number of less informative variables were added. All the combinations were ranked according to their cross-validated predictive accuracy ($1 –$ misclassification error), the area under the receiver operating characteristic (ROC) curve (AUC), precision, recall, and F-measure (Kelleher et al., 2020). These feature engineering methods allow the determination of the best feature combination for each machine learning algorithm.

Another type of dimensionality reduction is the feature transformation approach, which approximates point-light trajectories with a linear superimposition of a small number of basis components. Several previous studies have applied dimensional reduction techniques for modeling, including PCA (Troje, 2002) and anechoic demixing (Roether et al., 2009). Both the PCA and independent component analysis (ICA) approximate sets of time signals by weighted linear superimposition of source signals that are statistically independent in ICA and uncorrelated in PCA. In the present study, we used PCA for dimensionality reduction in the same way applied by Troje (2002) to biological sex classification. Uncorrelated components were extracted from all the detrended 11 joint movements along the horizontal axis of the walking direction. In this study, we used PCA to reduce dimensionality and generated a new data set of linearly superimposed features for sex classification. In the first PCA, we adopted the first to sixth principal components (Figure 3) because they explained more than 98% of the total variance and included frequency components of the first to third harmonics obtained by FFT. The first to sixth component coefficients (loadings) of 11 joint motions were calculated in each PLW and served as the 21-by-66 matrix for the second PCA, which retrieved the first to 10th principal component scores as transformed features for machine learning (Troje, 2002). In the present study, we constructed sex classification models using the first six principal components.

**Machine learning modeling**

To determine the optimal machine learning algorithm for sex classification, we compared the performance of six previously mentioned machine learning algorithms. To determine the best performance of each algorithm
for biological and perceptual sex classification, we constructed machine learning models by applying each algorithm to its best feature combination with hyperparameter tuning and 21 randomly determined initializing numbers, which had been fixed in the exhaustive search. To obtain a good estimate of our error rate for new data, we used leave-one-out cross-validation because of the small number of data sets. The performance of the machine learning models constructed using different algorithms was compared in terms of the highest accuracy, AUC, precision, recall, and F-measure. Once the optimal machine learning algorithm was identified through these procedures, we compared the best performance of the models between biological and perceptual sex classification, as well as between frequency components and principal components.

Interpretation of feature contribution

Using the biological and perceptual sex classification models with the best performance, we attempted to interpret the quantitative relationship between the important features and posterior probability of the prediction. First, to explain how each feature behaves in sex classification, we examined the individual conditional expectation (ICE), which allows us to define the relationship between each feature and the posterior probability in an individual PLW. The values for an ICE line can be computed by keeping all other features the same, creating a variant of this PLW by replacing the feature’s value with values from a grid and making predictions with the machine learning model (Molnar, 2020). This study employed centered ICE plots that showed the influence of each feature on the posterior probability from the same starting point at 0 because these plots make it easy to compare the ICE curves of an individual PLW. An increase in the ICE curve with an increase in the feature value indicates an increased male probability, whereas a decrease in the ICE curve indicates an increased female probability. When all ICE curves seem to follow the same course, there are few obvious interactions. However, in the case of mutual interactions, ICE plots with different courses and wide dispersions uncover heterogeneous relationships. We also averaged the ICE curves of all PLWs as a partial dependence curve, which is a summary of the relationship between the displayed feature and posterior probability of the prediction.

Second, to determine why the model produced such a prediction with the feature of interest, we employed the Shapley value based on the cooperative game theory. The Shapley value of each feature for the PLW of interest expresses the contribution of each feature (marginal contribution) to the difference between the prediction in the PLW and the average prediction for all the PLWs (Molnar, 2020). The Shapley values of each feature are obtained by averaging all the marginal contributions of the feature to the prediction across all possible coalitions of features, considering that changing the sequence in which the features join the coalition may change the respective prediction (Gianfagna & Cecco, 2021). In the present study, we ranked the contribution of features to the prediction by averaging all the absolute Shapley values of all the PLWs in each feature.

Results

Study 1: Sex perception rate and local features

Sex perception rate

Figure 4 shows the overall performance of the biological sex classification answered by the 15 observers for each PLW. Interindvidual differences are indicated by the standard deviation of the average sex perception rate of each PLW. The percentage of correct biological sex classification for 10 male PLWs was $62.9 \pm 16.6\%$ (45.7–91.4%) and $57.0 \pm 18.4\%$ (25.7–78.6%) for 11 female PLWs. For 12 (5 male and 7 female) out of 21 PLWs, the biological sex of PLWs was classified with significantly higher accuracy above the chance level ($zs > 2.23$). In contrast, two female PLWs (No. 11 and No. 16) were classified as male above the chance level (No. 11: $z = -4.30$ and No. 16: $z = -4.46$). We compared the influence of observer’s sex on sex classification between eight females and seven males. The correct perception rates were confirmed normally distributed in both male and female observers by the Shapiro–Wilk test (male observer: $w = 0.963$, $p = 0.715$; female observer: $w = 0.932$, $p = 0.155$). Thus, we used an unpaired $t$-test
and found that the observers’ sex did not affect the sex classification of either male or female PLWs \((t(40) = 0.05, p = 0.96, d = 0.02)\). The response bias examined by the C-value was 0.12, suggesting the presence of a slight male bias.

**Sex-specific spectral components**

Gait cycles and the related amplitudes of the first, second, and third harmonics of the joint movements were detected by FFT. The first spectral component corresponds to a one-step cycle (first harmonics). Table 1 compares the amplitudes of the first, second, and third harmonics of the shoulder and hip motions between males and females. When compared with the biological sex, the amplitudes of the hip-first, hip-second, hip-third, and shoulder-first harmonics were significantly greater in females than in males (hip-first: \(t(19) = 3.32, p = 3.57E-03, d = 1.47\); hip-second: \(t(19) = 3.13, p = 5.42E-03, d = 1.38\); hip-third: \(t(19) = 6.09, p = 7.37E-06, d = 2.69\); shoulder-first: \(t(19) = 3.60, p = 1.91E-03, d = 1.60\)). However, there were no sex differences in the shoulder-second (hip-first: \(t(19) = 1.43, p = 0.17, d = 0.63\) and shoulder-third (hip-second: \(t(19) = 1.84, p = 0.08, d = 0.82\)) harmonics. In perceptual sex, the amplitude differences in hip and shoulder harmonics between males and females were not as large as in biological sex (Table 1). Significant differences were found only in the hip-second and hip-third harmonics (hip-second: \(t(19) = 2.19, p = 0.04, d = 0.96\); hip-third: \(t(19) = 2.86, p = 0.01, d = 1.26\)). Figure 5 shows the box plots of the spectral amplitude ranges (median, first, and third quantiles) of male and female hip motions. All amplitudes of the male hip-first, hip-second, and
Study 2: Important feature contribution to sex classification with machine learning

Best feature combination and feature importance

Forward and backward sequential feature selection identified four features (hip-first, hip-second, hip-third, and shoulder-first harmonics) as an important subset of biological sex and all six features (hip-first, hip-second, and hip-third harmonics and shoulder-first, shoulder-second, and shoulder-third harmonics) for perceptual sex. After the forward and backward sequential approaches for important feature selection, we applied an exhaustive search to determine the best feature combination for each of the six algorithms with hyperparameter tuning, while the initializing number was fixed to facilitate the time-consuming process of the exhaustive search. Tables 2 and 3 indicate how to determine the best feature combination for each algorithm through this process, taking a concrete example when the SVM algorithm was applied to all 15 different combinations of four features in the biological sex classification and to all 63 different combinations of six features in the perceptual sex classification. In the biological sex classification, the performance of four different feature combinations demonstrated that the hip-third harmonic was the best in all the features; the hip-third and shoulder-first harmonics were the best in all the two-feature combinations; the hip-first, hip-second, and hip-third harmonics were the best in all the three-feature combinations; and the hip-first, hip-second, hip-third, and shoulder-first harmonics were the best in all the four-feature combinations. Among these, the least feature combination that achieved the best performance was the two-feature subset of the hip-third and shoulder-first harmonics (predictive accuracy: 90.5%, AUC: 0.93). Similarly, in perceptual sex classification, a five-feature subset of hip-first, hip-second, and hip-third harmonics.
Biologicalsex

| Combination | Accuracy | AUC | Sensitivity | Specificity | Precision | Recall | F-means |
|-------------|----------|-----|-------------|-------------|-----------|--------|---------|
| 1           | 90.5     | 0.91| 90          | 91          | 0.90      | 0.90   | 0.90    |
| 2           | 90.5     | 0.93| 90          | 91          | 0.90      | 0.90   | 0.90    |
| 3           | 90.5     | 0.93| 90          | 91          | 0.90      | 0.90   | 0.90    |
| 4           | 85.7     | 0.89| 90          | 91          | 0.82      | 0.90   | 0.86    |

Table 2. An example of the best feature combination by an exhaustive search with SVM. Notes: Best feature combination: Combination1 feature = Hip3; Combination2 features = Hip3, Shoulder1; Combination3 features = Hip1, Hip2, Hip3; Combination4 features = Hip1, Hip2, Hip3, Shoulder1. The bold indicates the most important and fewer feature combination selected among four.

Perceptual sex

| Combination | Accuracy | AUC | Sensitivity | Specificity | Precision | Recall | F-means |
|-------------|----------|-----|-------------|-------------|-----------|--------|---------|
| 1           | 71.4     | 0.81| 92          | 67          | 0.75      | 0.75   | 0.75    |
| 2           | 76.2     | 0.83| 75          | 89          | 0.82      | 0.75   | 0.78    |
| 3           | 80.9     | 0.79| 83          | 78          | 0.83      | 0.83   | 0.83    |
| 4           | 80.9     | 0.76| 83          | 78          | 0.79      | 0.92   | 0.85    |
| 5           | 85.7     | 0.78| 100         | 78          | 0.80      | 1      | 0.89    |
| 6           | 71.4     | 0.69| 92          | 56          | 0.68      | 0.92   | 0.79    |

Table 3. An example of the best feature combination by an exhaustive search with SVM. Notes: Best feature combination: Combination1 feature = Hip3; Combination2 features = Shoulder1, Shoulder3; Combination3 features = Hip1, Shoulder1, Shoulder3; Combination4 features = Hip1, Hip3, Shoulder1, Shoulder3; Combination5 features = Hip1, Hip2, Hip3, Shoulder1, Shoulder3; Combination6 features = Hip1, Hip2, Hip3, Shoulder1, Shoulder2, Shoulder3. The bold indicates the most important and fewer feature combination selected among four.

and shoulder-first and shoulder-third harmonics achieved the best performance (predictive accuracy: 85.7%, AUC: 0.81). Using the same method as in the example with the SVM, we applied an exhaustive search with the other five algorithms and found the best feature combination for each algorithm.

**Best machine learning algorithm for sex classification**

Using the best feature subset for each of the six algorithms, we finally constructed six machine learning models with hyperparameter tuning and 21 randomly selected variable initializing numbers. Table 4 lists the performance of the final model constructed using each algorithm and its best feature subsets. In biological sex classification, three algorithms, kernel SVM, k-nearest neighbors, and naive Bayes, achieved 90.5% classification accuracy. In perceptual sex classification, however, only kernel SVM attained the highest accuracy of 85.7%. Through these procedures, we finally identified the kernel SVM as the optimal algorithm for our data to compare the model performance between biological and perceptual sex classification and between the features of frequency components and those of principal components.

| Algorithm | Accuracy | AUC | Precision | Recall | F-means |
|-----------|----------|-----|-----------|--------|---------|
| Biological sex | | | | | |
| SVM       | 90.5     | 0.92| 0.90      | 0.90   | 0.90    |
| Discriminant | 85.7 | 0.90| 0.82      | 0.90   | 0.85    |
| KNN       | 90.5     | 0.91| 0.90      | 0.90   | 0.90    |
| Naive Bayes | 90.5   | 0.95| 0.90      | 0.90   | 0.90    |
| Random forest | 85.7 | 0.87| 0.82      | 0.90   | 0.86    |
| Ensemble  | 85.7     | 0.88| 0.82      | 0.90   | 0.86    |
| Perceptual sex | | | | | |
| SVM       | 85.7     | 0.81| 0.80      | 1      | 0.89    |
| Discriminant | 76.2   | 0.69| 0.82      | 0.75   | 0.78    |
| KNN       | 66.7     | 0.68| 0.78      | 0.58   | 0.67    |
| Naive Bayes | 80.9   | 0.70| 0.83      | 0.83   | 0.83    |
| Random forest | 61.9  | 0.63| 0.67      | 0.67   | 0.67    |
| Ensemble  | 71.4     | 0.75| 0.75      | 0.50   | 0.60    |

Table 4. Performance of machine learning algorithms. Note: KNN = k-nearest neighborhood. The bold indicates the machine learning algorithm with best performance for both biological and perceptual sex classification.

**Comparison between feature selection and feature transformation**

Using classification accuracy, ROC curve, AUC, precision, recall, and F-measure, the model performance
constructed by the kernel SVM algorithm was compared between the features of spectral components by FFT and those of principal components by PCA (Table 5). The classification accuracy with principal components was the lowest (57.1%) in the model of biological sex, while it reached 71.4% in the model of perceptual sex with relatively higher AUC, precision, recall, and F-measure. In contrast, as previously mentioned, the classification accuracy with spectral components reached 90.5% in biological sex and 85.7% in perceptual sex, with further increases in AUC, precision, recall, and F-measure as compared to those in principal components.

The high performance of the model with spectral components (Figure 6, right panel) was also reflected by shifts in the ROC curves near the top left corner and high AUCs in both biological and perceptual sex classification. However, in the model with principal components (Figure 6, left panel), the ROC curve of biological sex deviated left upward slightly away from the random line with a low AUC, whereas the ROC curve of perceptual sex showed an upward shift with a higher AUC than that of biological sex.

**Interpretation of feature contribution**

In the biological sex model (Figure 7A), the centered ICE curves demonstrated that the male posterior probability declined with increasing amplitude of the hip-third harmonic. All of these responses followed a similar homogeneous course, suggesting that the other features had little influence. Further, the ICE curves of shoulder-first harmonics were more widely distributed; some declined and others remained unchanged, suggesting interactions with the other feature, hip-third harmonic. The partial dependence curve obtained by averaging the ICE curves of all PLWs (red lines in Figure 7B) demonstrated that the posterior probability of biological sex prediction depended exclusively on hip-third harmonic, as reflected by the steeper decline with an increase in this harmonic than with an increase in the shoulder-first harmonic.

In perceptual sex (Figure 7B), the partial dependence plots (averaged ICE curves) indicated that the male posterior probability declined with increases in the hip-second, hip-third, and shoulder-third harmonics. Their ICE curves displayed gradual dispersion as the features increased, whereas the ICE curves of the shoulder-third harmonics followed different and widely distributed courses. In contrast to the biological sex, the ICE curves displayed gradual increases with dispersion as the shoulder-first harmonics increased. Although it seemed that hip-first harmonics had little influence on the posterior probability, as shown by the behavior of the partial dependence plot, the ICE curves clearly

| Variable | FFT 1-3 harmonics | PCA 1-6 scores |
|----------|------------------|----------------|
|          | Biological | Perceptual | Biological | Perceptual |
| Accuracy (%) | 90.5 | 85.7 | 57.1 | 71.4 |
| AUC | 0.92 | 0.81 | 0.58 | 0.74 |
| Precision | 0.90 | 0.80 | 0.56 | 0.75 |
| Recall | 0.90 | 1.00 | 0.50 | 0.75 |
| F-means | 0.90 | 0.89 | 0.53 | 0.75 |

Table 5. The best model performance for biological and perceptual sex classification.
Figure 7. (A) Centered ICE curves of posterior probability in biological sex classification. The lines show the changes in posterior probability of biological sex classification for each PLW, when normalized amplitudes of the hip-third and shoulder-first harmonics varied from minimum to maximum values. All centered ICE plots of the hip-third harmonic followed a similar course, suggesting little obvious interaction. The ICE curves of the shoulder-first harmonic were more widely distributed; some declined and others remained unchanged, suggesting some interaction with the other feature. The partial dependence curves (red lines) obtained by averaging the ICE curves of all PLWs indicate that the posterior probability depends exclusively on the hip-third harmonic, as reflected by a steep decline along with an increase in this harmonic. The dot on each line represents the actual value of the feature of each PLW. (B) Centered ICE curves of posterior probability in perceptual sex prediction. The posterior probability predicted by the machine learning model with each normalized amplitude of the hip-first, hip-third, shoulder-first, and shoulder-third harmonics displayed gradual dispersion as the features increased, whereas the ICE curves of the shoulder-third harmonic followed different courses and distributed widely. The heterogeneity or dispersion of ICE curves suggests the presence of interactions with other features. The dot on each line represents the actual value of the feature of each PLW. The red lines indicate the partial dependence curves obtained by averaging the ICE curves of all the PLWs.
Figure 8. Mean Shapley values of each feature. All absolute Shapley values of individual PLWs were averaged for each feature to obtain the mean contribution of each feature to the prediction. In the biological sex classification (left panel), the mean Shapley value of the hip-third harmonic was greater than that of the shoulder-first harmonic. Although the mean Shapley values of the hip-third harmonic were dominant in perceptual sex (right panel), the shoulder-first, shoulder-third, and hip-first harmonics also contributed significantly to perceptual sex prediction with relatively high Shapley values.

demonstrated that the unchanged partial dependence plots resulted from the offset by the heterogeneity of the ICE curves; some ICE curves increased, but others remained unchanged or declined. The heterogeneity or dispersion of ICE curves suggests the presence of interactions with other features.

All the absolute Shapley values calculated for individual PLWs were averaged for each feature to obtain the mean contribution of each feature to the prediction (Figure 8). In the biological sex classification, the mean Shapley value of hip-third harmonics was greater than that of shoulder-first harmonics (Figure 8, left panel). This coincided with the results of the ICE plots observed for biological sex. In the perceptual sex, however, other features, as well as hip-third harmonics, also contributed significantly to the prediction; the contribution expressed as a percentage of that of hip-third harmonics was over 42% in shoulder-first harmonics, 40% in shoulder-third harmonics, and 21% in hip-first harmonics (Figure 8, right panel).

Discussion

In our study, the observers’ correct perception rate of biological sex averaged 57.0% for male PLWs and 62.9% for female PLWs. This accuracy is consistent with that previously reported by many investigators (Kozlowski & Cutting, 1977; Barclay et al., 1978; Cutting et al., 1978; Mather & Murdoch, 1994; Hirashima, 1997, 1999; Sumi, 2000; Troje, 2002; Pollick et al., 2002; Davis & Gao, 2004; Pollick et al., 2005). The FFT analysis revealed that the amplitudes of the hip-first, hip-second, hip-third, and shoulder-first harmonics were significantly larger in female PLWs than in male PLWs. These prominent hip-second and hip-third harmonics may correspond to the rapid hip movement previously defined by Mather and Murdoch (1994) as a dynamically important cue of female PLWs. However, in the present study, two female PLWs were judged to be male. Although this seems to be misclassified, the spectral density of their hip movements was similar to that of males rather than that of females. These findings suggest that human observers may obtain some impression related to sex (perceptual sex) rather than the biological sex. To solve this problem, we constructed a supervised machine learning model to predict two different prediction targets: biological and perceptual sexes. In reference to predictive accuracy, we tried to find the important cues (features) to predict biological and perceptual sexes and how these important features influence these targets.

As compared with the low perception rate of biological sex by observers, the kernel SVM model with two important spectral components (hip-third and shoulder-first harmonics) classified biological sex with a predictive accuracy over 90% in the present study. The k-nearest neighbors and naive Bayes algorithms achieved the same high accuracy. These findings indicate that the lateral view intrinsically has sufficient dynamic information to classify the biological sex. In human observers, however, their attention may not be paid only to such a specific motion as the hip-third harmonic but also to other joint movements or may not be able to notice a so rapid and small hip motion, resulting in the low classification accuracy of biological sex. The observer’s ability to classify biological sex was similar to
the predictive accuracy of the machine learning model using the features of principal components. This can be explained by the fact that principal components are not specific to the localized joint but are composed of all the joint movements that human observers may pay attention to.

When the features of spectral components were used in perceptual sex classification, the kernel SVM model was superior to other machine learning algorithms, achieving the highest classification accuracy. Despite the high classification accuracy of the kernel SVM in both biological and perceptual sex, there were significant differences between them in the important features selected for modeling. In the prediction model of perceptual sex, the sequential selection and exhaustive search adopted five features (hip-first, hip-second, hip-third harmonics, and shoulder-first and shoulder-third harmonics) as the best feature combination. The Shapley values of the features revealed their relative contributions to the prediction, and the ICE curves indicated that these features interacted with each other to achieve high performance. In the prediction model of biological sex, however, hip-third harmonic contributed significantly to the accurate prediction, as confirmed by the ICE curves and the Shapley value of hip-third harmonics. The present findings suggest that biological sex classification in side-view PLWs depends on specific localized motions, whereas perceptual sex classification is rather complex and is composed of various feature interactions. Additionally, this nature of perceptual sex classification is also suggested by the model performance using the principal components of all joint motions; the model constructed from the features superimposed by PCA achieved a higher performance in perceptual sex classification than in biological sex classification.

Many studies have examined the perceptual sex of PLWs from local or holistic viewpoints (Cutting, 1978; Mather, 1994; Sumi, 2000; Troje, 2002; Pollick, 2005; Jordan et al., 2006; Troje & Westhoff, 2006). They used inverted PLWs (Sumi, 2000), dephased PLWs (Jordan et al., 2006), or PCA (Troje, 2002) and found that humans are apt to perceive sex by holistic rather than local information from PLWs. Jordan et al. (2006) showed that gender-specific adaptation of viewing a female PLW increased the probability of judging a subsequent ambiguous PLW as male and vice versa. This adaptation effect was reduced significantly when they used dephased stimuli that disrupted the global coherence but left the local motion of each point-light unchanged. These findings are consistent with the current finding that human perception does not utilize specific localized information such as the third harmonic of hip motion, which proved very effective in the classification of biological sex as machine learning has shown but relies on more global or combined information at the cost of accuracy. In the present study, we applied model-agnostic interpretation tools to supervised machine learning and obtained new insights into how the global information of human sex perception consists in terms of significant localized components and their relative contributions.

Under the conditions for multidimensional and interactive processes, such as sex perception, machine learning modeling is a suitable approach because it can adopt many features and the complexity of the data in real-world situations (Brink et al., 2016). Generally, machine learning has two goals: highly accurate prediction and interpretation. However, there is a trade-off between performance and explainability: Increasing performance with complex models may reduce explainability. In the present study, we first aimed to predict perceptual sex as accurately as possible by using complex machine learning models. Consequently, we made post hoc explanations, that is, the explainability was achieved later, after model creation. Feature engineering, such as forward and backward sequential methods and an exhaustive search, enabled us to identify the important features of sex perception with which the model performed well. ICE, partial dependence plots, and the Shapley value explained how and why our model achieved high performance.

Limitations

First, the current study was based on a limited number of walking individuals, and hence it cannot readily be generalized. In this study, we focused not only on accurate prediction but also on inference using classification algorithms of supervised machine learning. Therefore, we focused on feature engineering and model-agnostic interpretation to examine the differences in feature importance between biological and perceptual sex classification. Second, we focused on hip and shoulder movements within the spectral range of 6 Hz because these joints had multiple informative harmonics (first, second, and third harmonics) within 6 Hz. Although excluded in the present study, other joints, such as the foot with a typical pendular motion or higher than 6 Hz but much smaller spectral components, might contribute to sex perception. To clarify the generalized processes of sex perception, further investigation is required, including a larger number of PLWs, more joint movements, and much higher spectral and principal components. Third, there was a slight difference between the SVM model performance presented in Tables 2, 3, and 4. This could be attributed to the initializing number, which was fixed in the exhaustive search and varied randomly in the final model. Finally, the present classification of biological and perceptual sex is not based on a real human model but on a machine
learning model. If the ability of humans to perceive sex could be similar to that of machine learning models, machine learning would provide a new tool to gain insight into understanding complex “behind-the-scenes” mechanisms for human sex perception from PLWs.

Conclusions

We developed a machine learning model to classify biological and perceptual sex from the spectral components of hip and shoulder motion in lateral PLWs. There were distinct differences in the crucial features for modeling between biological and perceptual sexes. Biological sex classification depends mainly on local cues of the PLW, while the perceptual sex classification rate is achieved by a combination and interaction of various frequency components of hip and shoulder motions. This suggests more complex and organized processes in perceptual sex classification than in biological sex classification.

Keywords: point-light walker, sex perception, fast Fourier transform, machine learning, individual conditional expectation

Acknowledgments

Supported by a research grant from the Japan Society for the Promotion of Science (17J06220).

Commercial relationships: none.
Corresponding author: Chihiro Asanoi.
Email: chihiroa@odalab.org.
Address: Graduate School of Human Science, Tokyo Woman’s Christian University, Tokyo, Japan.

References

Asanoi, C., & Osada, Y. (2005). The perception of point-light walker across a slit: The perception of “human likeness.” Journal of Psychological Science, 24(1), 133–134, https://doi.org/10.14947/psychono.KJ00004348801.

Asanoi, C., & Osada, Y. (2006). The effect of the amplitude of movement of a point-lights walker on perception of gender recognition and perception of “human-likeness.” Rikkyo Psychological Research, 48, 7–14 (in Japanese), https://doi.org/10.14992/00000323.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. Perception & Psychophysics, 23, 145–152, https://doi.org/10.3758/BF03208295.

Brink, H., Richards, J., & Fetherolf, M. (2016). Real-world machine learning. Shelter Island, NY: Manning.

Brink, H., Richards, J. W., Poznanski, D., Bloom, J. S., Rice, J., Negahban, S., ... Wainwright, M. (2013). Using machine learning for discovery in synoptic survey imaging data. Monthly Notices of the Royal Astronomical Society, 435(2), 1047–1060, https://doi.org/10.1093/mnras/stt1306.

Cutting, J. E. (1978). Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. Perception, 7(4), 393–405, https://doi.org/10.1068/p070393.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. Perception & Psychophysics, 21(6), 575–580, https://doi.org/10.3758/BF03198740.

Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. Journal of Experimental Psychology: Human Perception and Performance, 4(3), 357, https://doi.org/10.1037/0096-1523.4.3.357.

Davis, J. W., & Gao, H. (2004). An expressive three-mode principal components model for gender recognition. Journal of Vision, 4(5), 2, https://doi.org/10.1167/4.5.2.

Gianfagna, L., & Di Cecco, A. (2021). Explainable AI with Python. Cham, Switzerland: Springer.

Hirashima, S. (1997). Recognizing the gender of point-light runners. Japanese Journal of Psychonomic Science, 16(2), 76–81 (in Japanese), https://doi.org/10.14947/psychono.KJ00004413473.

Hirashima, S. (1999). Recognition of the gender of point-light walkers moving in different directions. Japanese Journal of Psychology, 70(2), 149–153 (in Japanese), https://doi.org/10.4992/jjpsy.70.149.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 14(2), 201–211, https://doi.org/10.3758/BF03212378.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. Psychological Research, 38, 379–393, https://doi.org/10.1007/BF03090403.

Jordan, H., Fallah, M., & Stoner, G. R. (2006). Adaptation of gender derived from biological motion. Nature Neuroscience, 9(6), 738–739.

Kelleher, J. D., Mac Namee, B., & D’arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: Algorithms, worked
examples, and case studies. Cambridge, MA: MIT Press.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics, 21*, 575–580, https://doi.org/10.3758/BF03198740.

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 258*, 273–279, https://doi.org/10.1098/rspb.1994.0173.

Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models interpretable*. Milton Keynes, UK: Leanpub.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1247, https://doi.org/10.1037/0096-1523.31.6.1247.

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research, 42*(20), 2345–2355, https://doi.org/10.1016/S0042-6989(02)00196-7.

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for Perception of emotion from gait. *Journal of Vision, 9*(6), 15, https://doi.org/10.1167/9.6.15.

Sumi, S. (2000). Perception of point-light walker produced by eight lights attached to the back of the walker. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie, 59*(2), 126, https://doi.org/10.1068/v970256.

Theodoridis, S., Pikrakis, A., Koutroumbas, K., & Cavouras, D. (2010). *Introduction to pattern recognition: A Matlab approach*. New York, NY: Academic Press.

Troje, N. F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait pattern. *Journal of Vision, 2*(5), 371–387, https://doi.org/10.1167/2.5.2.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector”? *Current Biology, 16*(8), 821–824, https://doi.org/10.1016/j.cub.2006.03.022.

Unuma, M., Anjyo, K., & Takeuchi, R. (1995). Fourier principles for emotion-based human figure animation. *Proceedings of the 1995 Conference on Computer Graphics and Interactive Techniques* (pp. 91–96). Los Angeles, CA.

Westhoff, C., & Troje, N. F. (2007). Kinematic cues for person identification from biological motion. *Perception & Psychophysics, 69*(2), 241–253, https://doi.org/10.3758/BF03193746.