Recognizing Solo Jazz Dance Moves Using a Single Leg-Attached Inertial Wearable Device

Sara Stančin * and Sašo Tomažič

Faculty of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia; saso.tomazic@fe.uni-lj.si
* Correspondence: sara.stancin@fe.uni-lj.si; Tel.: +386-14768-939

Abstract: We present here a method for recognising dance moves in sequences using 3D accelerometer and gyroscope signals, acquired by a single wearable device, attached to the dancer’s leg. The recognition entails dance tempo estimation, temporal scaling, a wearable device orientation-invariant coordinate system transformation, and, finally, sliding correlation-based template matching. The recognition is independent of the orientation of the wearable device and the tempo of dancing, which promotes the usability of the method in a wide range of everyday application scenarios. For experimental validation, we considered the versatile repertoire of solo jazz dance moves. We created a database of 15 authentic solo jazz template moves using the performances of a professional dancer dancing at 120 bpm. We analysed 36 new dance sequences, performed by the professional and five recreational dancers, following six dance tempos, ranging from 120 bpm to 220 bpm with 20 bpm increment steps. The recognition F1 scores, obtained cumulatively for all moves for different tempos, ranged from 0.87 to 0.98. The results indicate that the presented method can be used to recognise repeated dance moves and to assess the dancer’s consistency in performance. In addition, the results confirm the potential of using the presented method to recognise imitated dance moves, supporting the learning process.

Keywords: motion recognition; motion analysis; inertial sensors; 3D accelerometer; 3D gyroscope; solo jazz; dancing

1. Introduction

Dancing is an engaging physical activity that requires high levels of body control, skill, and physical fitness. In addition, dancing is an activity cultivating the capacity for creative expression. Following the rhythmic structure, the dancer, especially the professional, expresses his or her interpretation of music by selecting, performing, and assembling various moves in dance sequences.

The main motivation of the research presented is to provide a mechanism for dance move recognition in performed sequences. In general, it has already been reported that dancers can greatly benefit from various assistive technologies [1–5]. A dance move recognition technology would benefit dancers engaged in the learning process. Namely, as one learns to dance, he or she tends to imitate performances demonstrated by instructors. These imitations are visually assessed, usually using a mirror. In addition, instructors provide their students with feedback on how their performances compare to the ideal template. Providing such feedback is usually more challenging in group sessions, in which more students engage in the learning process at the same time. A technological solution would benefit both the student and the instructor and would support performance assessment and progress monitoring.

More experienced dancers would also benefit. Recognising dance moves would enable further investigation and comparison of the sequences of moves, supporting overall creativity evaluation, individual style classification, and observations of differences when dancing alone, in pairs and/or in the presence of an audience. Additionally combining the
recognised dance moves with information about the accompanying music would illuminate how the dancer interprets and responds to music. Relying on a dance move recognition technology during the creative process of producing a choreography, a professional dancer would also be able to create a transcript of the assembled moves for later reference.

A segment of dance motion analysis approaches made so far relies on signals obtained from optical motion capture systems, incorporating multiple cameras watching the stage and usually requiring reflective markers to be positioned on the subject’s body [6–9]. In [6], the authors present a framework for capturing the posture of the dancer’s body. Using the body posture parameters, signals are temporally segmented, motion patterns are extracted, and motion animation in generated. In [7], the authors present dance pose sequences similarity estimation.

Another segment of approaches relies on signals obtained from a Kinect device [1,2,5,10–12]. The Kinect device incorporates a depth camera together with a standard video camera into a single device, distinguishes human body parts and joints, and estimates their position and orientation. In [1], the authors present a novel framework for the real-time capture, assessment, and visualisation of ballet dance movements. The dance sequence captured is segmented and cross-referenced against a library of gestural components performed by a teacher. Recognition performance was evaluated on a database of six isolated basic ballet positions recordings made by a teacher and a student. Reported average recognition rates are between 90.5% and 99.5%. Bharatanatyam [2] and a salsa dancing coach application [5] have also been developed. In [10], another Kinect-based system is presented enabling Korean pop (K-pop) motion classification by extracting statistical features, reducing dimensionality with Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) and incorporating extreme learning machine. Skeletal joint angles for 200 movement types obtained for four dancers were considered and the reported maximum classification rate is 96.5%. In [11], the authors analysed data on Greek traditional dance postures with the goal of extracting robust summaries and providing end users or dance experts a concise and meaningful abstract of dance movements.

Although valuable results have been reported, using equipment located in the surrounding area, either one or more video cameras or a Kinect device, limits the practical aspect of the methodology, making it difficult or altogether impossible to dance in the ballroom. Moreover, processing video signals is computationally intensive, and although the underlying technology and algorithms are steadily improving, computer vision is still limited by lighting and clutter constraints and subtraction of dynamic background.

The presented method capitalises on the benefits of small and lightweight microelectromechanical (MEMS) inertial sensors. Over the past few years, it has been consistently demonstrated that these sensors are an efficient tool in the broader research area of human motion analysis [13–29]. Their characteristic light weight, small size, low power consumption, portability, ease of use, and low cost pave the way for their omnipresence and indispensability in motion analysis.

Most of the research conducted in this field concerns daily activity recognition [13–19], gait analysis [20–22], gesture recognition [23], and sports activity performance assessment for activities such as golf, tennis, baseball, swimming, and skiing [24–29]. One of the first wearable systems for capturing dance gestures were developed by the MIT Media Lab [30,31], and various assistive solutions have been presented since [3,4,32–39]. In this context, wearable devices are usually used to detect dancer’s steps and provide them with feedback [3,4,32] or enable the dancer to interact or generate sounds [33–35]. In [36], a three-axis inertial measurement unit (IMU) is positioned on the subject’s right forearm and used to analyse one specific hip hop arm move. In particular, the aim is to divide the motion into phases, predict which phase is executed, and give feedback to the subject. In addition, accelerometers are used to estimate torso tilt [37] and assess performance [38] of basic routines in classical ballet.

We focus our analysis on the solo jazz dance style, a rhythmical and playful solo dance in which the dancer depicts jazz music through one’s one movement. Solo jazz evolved
through the first half of the 20th century to include elements of both African and European dance and features the vocabulary and steps of the vernacular jazz tradition. To enable the dancer to follow the jazz song rhythmical structure, as a rule, a single solo jazz move is performed following eight music beats. The music beat is considered as the smallest time interval between two successive notes. Dance moves comprise various motion elements, performed in a specified order. Besides steps, motion elements can be, for example, kicks, taps, and jumps. The variety of motion elements and the order of their execution brings a pleiad of predefined, authentic solo jazz dance moves. Due to this variety, solo jazz fits perfectly into our dance move recognition research polygon.

In general, in dancing, the speed of move execution represents the dance tempo and is directly related to the musical tempo of the song. It is measured as the number of steps, or other motion elements, a dancer performs in a minute. As a rule, solo jazz is usually danced to jazz music with tempo anywhere between 120 and 250 beats per minute (bpm). Tempos between 140 and 160 bpm are considered medium. Tempos above 220 bpm are considered fast and generally prove to be too high for the recreational dancers to execute with accuracy and ease.

Aiming to provide ease of use, with the smallest amount of sensing equipment, our dance move recognition methodology relies on a single wearable unit, comprising a MEMS 3D accelerometer and gyroscope, positioned on the dancer’s leg, and enables dance move recognition, regardless of the device orientation and dance tempo. The methodology presented considers that different moves have different signal shapes in the time domain and that these shapes can be used as dance move fingerprints. Furthermore, since position and attitude estimates obtained using inertial sensors are known to be characterised by different inaccuracies, especially drift, the methodology resides on the shapes of the original acceleration and angular velocity signals instead.

The methodology presented is a continuation of our previous work and features solo jazz dance tempo estimation, as presented in [39]. In addition, the specific technique presented in [20], used for defining a wearable device orientation-invariant coordinate system using the 3D accelerometer signals acquired during walking, is adapted and applied in the research presented in this article.

The remainder of this article is organised as follows. In Section 2, we present the materials and methods used for creating the dance move template database, recognition, and experimental validation. In Section 3, we present and discuss the results. Finally, in Section 4, we summarise our findings and draw conclusions, implying further research directions. In all the subsequent sections, we use the following notation rules: large bold letters denote matrices, small bold letters denote vectors, and large or small italics denote scalars.

2. Materials and Methods

2.1. Data Acquisition

2.1.1. Materials

We captured dance motion using a single mbientlab MetaMotionR (MMR) wearable device [40], including a MEMS 3D accelerometer and gyroscope with respective measurement ranges ±16 g and ±2000°/s, placed directly above the dancer’s right leg ankle. We hypothesised that analysing the motion of a single leg is sufficient to distinguish between various dance moves. An alternative solution would be to place the device on the dancer’s lower back, since leg motion of high-quality dance performances is usually initiated from the pelvic region, torso, or even the shoulders. However, preliminary investigations have shown that this significantly reduces the recognition performance of the designed solution. The micro-position and orientation of the device are arbitrary.

We set the sampling frequency to 200 Hz, which proved to be sufficient for the problem at hand using empirical evidence. The 3D accelerometer and gyroscope together provided 6 inertial signals, each representing one variable in a 6D space.
In addition, a software metronome with an established beat frequency error of 1 bpm was used to simulate a steady quarter note music tempo and dictate the tempo of dancing.

2.1.2. Measurements

Six female dancers participated in the study—five recreational dancers (age: 33 ± 5 years, height: 163 ± 5 cm) with over three years of experience in solo jazz dancing and one professional (age: 35 years, height: 164 cm).

Altogether, \( I = 15 \) authentic solo jazz moves were considered: (1) Tackie Annie, (2) Fall of the log, (3) Kicks, (4) Half break, (5) Struttin’, (6) Savoy kick, (7) 20s Charleston, (8) Knee slaps, (9) Fishtails, (10) Apple Jacks, (11) Boogie back, (12) Boogie forward, (13) Crazy leg, (14) Cross step, and (15) Shorty George. For an informative overview of how these moves are performed, the reader is referred to various sources available online, e.g., [41]. Each move was performed on an 8-beat basis.

Measurements were performed in two sessions. Following the obtained instructions, participants attached the wearable device above their right leg ankle, voluntarily setting its micro-position and orientation.

The first measurement session was conducted only for the professional dancer. The dancer performed several consecutive repetitions for each of the 15 considered moves, as consistently as possible, following a fixed reference tempo \( \nu_{ref} = 120 \text{ bpm} \), dictated by the metronome. For each move, the dancer started dancing from a still position. We refer to the obtained set of six inertial signals as the learning sequence.

The second measurement session was conducted for all six participants. Each was given the task of performing 5 repetitions of each of the 15 considered moves in a prescribed order. The recreational dancers were instructed to mimic the professional’s execution of moves as closely as possible. This task was repeated for 6 different dance tempos, ranging from 120 bpm to 220 bpm with 20 bpm increments. For each dancer, we obtained 6 sets of inertial signals. We refer to these 36 sets as test sequences.

All measurements were supplemented with video recordings. The study was conducted in Ljubljana, Slovenia. It followed the Code of Ethics of the University of Ljubljana, which provides guidelines for studies involving human beings and is in accordance with the Declaration of Helsinki. All participants gave written informed consent.

2.2. Signal Processing Overview

The implemented signal processing workflow is presented in Figure 1. For each acquired sequence (1), signal pre-processing (2) is applied. The pre-processed learning sequence is used to extract template moves and create the template database (3), following a semi-automatic correlation-based segmentation procedure. The pre-processed test sequences are analysed and searched for template moves (4), according to the following steps:

4.1. Dance tempo estimation and signal temporal scaling, achieved using a bank of enhanced comb filters as presented in [39];
4.2. Initial template matching, performed on a sliding correlation basis, using the magnitudes of the temporally scaled acceleration and angular velocity;
4.3. Signal transformation to the templates’ coordinate system; and
4.4. Final template matching, performed again on a sliding correlation basis, but by using the acceleration and angular velocity 3D projections of on the template coordinate system axes instead of their magnitudes.

For both the initial and final template matching, correlation is used as the only feature for recognition. In the following, all the components of the proposed method are presented in detail. All signal processing was executed offline in the MATLAB 2021 environment [42].
2.3. Signal Pre-Processing

Following the calibration procedure presented in [43], we compensated all signals of the learning and test sequences for sensor inaccuracies. Since the device itself does not provide outputs at exactly equidistant time samples, we interpolated and decimated the acquired signals as necessary, considering the associated measurement timestamp values, to provide for uniform sampling at exactly 200 Hz. To remove motion artefacts and noise, we applied a low-pass filter with a cut-off frequency $f_{co} = 50$ Hz and finally performed downsampling to $f_s = 100$ Hz, obtaining 3D acceleration and angular velocities at equidistant time samples $T = 1/f_s = 0.01$ s.

2.4. Templates’ Database

2.4.1. Template Extraction

Using all $N_s$ acquired samples of the pre-processed learning sequence, we formed two $N_s \times 3$ signal matrices, $A_s$ and $\Omega_s$, representing 3D acceleration and angular velocity, respectively. Columns of these two matrices are equal to the respective signal projections on the device-intrinsic coordinate system axes $x$, $y$, and $z$, while rows represent the time samples. The specific orientation of axes $x$, $y$, and $z$ in a reference coordinate system, defined by the set position of the wearable device, is arbitrary. We further combined $A_s$ and $\Omega_s$ in a common $N_s \times 6$ signal matrix $S$:

$$S = \begin{bmatrix} A_s & \Omega_s \end{bmatrix}. \tag{1}$$

Columns of $S$ are equal to the 3D acceleration and angular velocity $x$, $y$, and $z$ projections, while rows represent the time dimension. Comprising the inertial signals of the entire learning sequence, matrix $S$ comprises all performed repetitions of all considered dance moves, used for creating the templates’ database.

We extracted the comprised dance move repetitions by partitioning $S$ into submatrices in the vertical dimension. We achieved this by estimating the repetitions’ onsets and offsets, for each $1 \leq i \leq I = 15$ dance move separately, following a semi-automatic, correlation-based segmentation procedure as follows. First, since for each dance move, the consecutive repetitions are performed from a still position, we defined the onset of the first repetition by visual inspection of the six inertial signals and the accompanying video. We denote the matrix $S$ row index corresponding to this onset as $n_0$. 

![Figure 1. Dance move recognition processing workflow.](image-url)
For defining the offset of the first repetition, we consider, given the known reference dance tempo \( \nu_{\text{ref}} \), the expected length, expressed as the number of samples \( N_T \), of a single dance move. Since each considered solo jazz move is performed on an 8-beat basis, at \( f_s = 100 \) Hz sampling and precise 120 bpm dancing \( (T_{\text{beat}} = 0.5 \) s), the expected length is \( N_T = 8 \times 0.5 \times 100 = 400 \) samples. Since dancing is rarely this precise, we allow for a 2\% length deviation and set the expected length to be in between 0.98 \times N_T = 392 and 1.02 \times N_T = 408 samples. Each length from this interval, i.e., \( N_n = N_0 + n \), where \( N_0 = 392 \) and \( 0 \leq n \leq 0.04 \times N_T = 16 \), gives a candidate for the first repetition offset, \( n_0 + N_n - 1 \).

Since the repetitions are consecutive, for each offset candidate for the first repetition, \( n_0 + N_n - 1 \), there is a single onset candidate for the second repetition, \( n_0 + 2N_n - 1 \). Finally considering both repetitions to be equal in length gives a single offset candidate for the second repetition, \( n_0 + 2N_n - 1 \). For each \( n \), we obtained one candidate pair of the first two repetitions of a move, represented with adjacent, same-size \( N_n \times 6 \) submatrices of \( S \).

We extracted and standardised to zero mean and unit standard deviation these two submatrices and performed column-wise vectorisation, obtaining, for each \( n \), two 6\( (N_0 + n) \) long column vectors. Pursuing the highest similarity, for each candidate pair \( n \), we calculated the correlation coefficient between these two vectors. This calculation, supported with implementation equations, is described in detail in Appendix A.2.

We set the first and the second repetition of the considered move to be equal to the repetitions of the candidate pair with the highest correlation coefficient. We finally extracted all possible additional repetitions of the particular move on an equal correlation-based search basis using the onset of the last determined repetition as the new \( n_0 \).

Due to slight variations in execution duration, we unified in length all extracted repetitions by temporal scaling to the exact expected length \( N_T = 400 \) samples. For each move \( i \), we composed a set of consistent repetitions, eliminating repetition outliers, determined by velocity, respectively. Each \( N_n \) column-wise vectorisation of the resulting matrix. For each move, we obtained a 2D \( N_n \times 6 \) matrix \( T_i \). Denoting the first three columns, representing the 3D acceleration, with \( A_i \) and the second three, representing 3D angular velocity, with \( \Omega_i \), for each \( 1 \leq i \leq I = 15 \), we can write:

\[
T_i = \begin{bmatrix}
A_i \\
\Omega_i
\end{bmatrix}
\]  

The 15 matrices \( T_i \) of the acceleration and angular velocity projections on the device intrinsic coordinate system axes \( x, y, \) and \( z \) represent our database of template moves. The specific orientation of the axes \( x, y, \) and \( z \) in the reference coordinate system, defined by the arbitrarily set position of the wearable device during the learning sequence acquisition, defines the coordinate system of the templates.

In addition to \( T_i \), for each move, \( i \) we also calculated an \( N_T \times 2 \) matrix \( \overline{T}_i \) having the first and second column equal to the magnitudes of the 3D acceleration and angular velocity, respectively. Each \( k \)-th row of \( \overline{T}_i \) is obtained according to:

\[
\overline{T}_{ik, *} = \left[ \sqrt{A_{ik,1}^2 + A_{ik,2}^2 + A_{ik,3}^2}, \sqrt{\Omega_{ik,1}^2 + \Omega_{ik,2}^2 + \Omega_{ik,3}^2} \right].
\]  

In (3), \( * \) denotes all columns.

2.4.2. Templates Similarity Measures

To estimate the similarity between template moves, we used two correlation-based measures as follows. Firstly, we compared the template moves along the acceleration and angular velocity magnitudes. To achieve this, we standardised the columns of \( \overline{T}_i \) to zero mean and unit standard deviation for each \( 1 \leq i \leq I = 15 \) move and performed column-wise vectorisation of the resulting matrix. For each move, we obtained a \( 2N_T \)-long vector of magnitudes. For each pair of template moves \( i \) and \( j \), \( 1 \leq i, j \leq I = 15 \),
we calculated the correlation between the associated vectors as a function of cyclical shift $0 \leq \tau < N_T$ of the acceleration and angular velocity vector parts. We define the maximum value of this correlation for each $i,j$ pair of template moves, denoted with $r_{T_iT_j}$, as the magnitudes’ similarity. This calculation is supported in detail with implementation equations in Appendix A.3.

We also compared the template moves along the acceleration and angular velocity 3D projections. We standardised the columns of $T_i$ to zero mean and unit standard deviation for each $1 \leq i \leq I = 15$ move and performed column-wise vectorisation of the resulting matrix. For each move, we obtained a $6N_T$-long vector. For each pair of template moves $i$ and $j$, $1 \leq i,j \leq I = 15$, we again calculated the correlation between the associated vectors as a function of cyclical shift $0 \leq \tau < N_T$ of the projections vector parts. We define the maximum value of this correlation for each $i,j$ pair of template moves, denoted with $r_{T_iT_j}$, as the projections’ similarity. This calculation is supported in detail with implementation equations in Appendix A.4.

In the same way as the correlation coefficient calculated along one dimension, $r_{T_iT_j}$ and $r_{T_iT_j}$ can take any value from the range $[-1,1]$, where 1 indicates identical moves, 0 orthogonal moves, and –1 opposite moves. High $r_{T_iT_j}$ values suggest common leg activation patterns, regardless of the actual direction of motion in 3D. On the other hand, high $r_{T_iT_j}$ suggests that two moves also match with respect to the direction of the executing motion. Considering this, the similarity measure $r_{T_iT_j}$ is expected to have a higher discriminative potential. Note that by performing standardisation column-wise, equal weight is given to each dimension of comparison, $x$, $y$, and $z$, regardless of the intensity of motion along that dimension.

2.5. Dance Move Recognition

2.5.1. Dance Tempo Estimation and Temporal Scaling

For each pre-processed test sequence, we first estimate the dance tempo $v_{est}$, using a method based on multiple resonators, implemented with enhanced comb feedback filters, as presented in [6].

We temporally scale all acquired inertial signals of the sequence by a factor of $v_{ref}/v_{est}$ to obtain a test sequence with dance moves that match the template moves in tempo. We denote with $N_X$ the final length of the test sequence.

Using all $N_X$ samples, we form two $N_X \times 3$ signal matrices of 3D acceleration and angular velocity, denoted with $A_X$ and $\Omega_X$, respectively. The columns of these two matrices correspond to the respective inertial signal projections on the intrinsic coordinate system axes of the wearable device. Since the orientation of the wearable device is different for each measurement session, these axes and the axes of the coordinate system of the template moves are generally not aligned.

We further form a common $N_X \times 6$ test sequence signal matrix $X$:

$$X = \begin{bmatrix} A_X & \Omega_X \end{bmatrix}. \quad (4)$$

Columns of $X$ correspond to the acceleration and angular velocity 3D projections while rows represent the time dimension.

We also form an $N_X \times 2$ matrix of acceleration and angular velocity magnitudes $\bar{X}$. Each $k$-th row of $\bar{X}$ is calculated according to:

$$\bar{X}_{k,*} = \left[ \begin{array}{l} \|A_{X,k,*}\| \|\Omega_{X,k,*}\| \\ \sqrt{A_{X,k,1}^2 + A_{X,k,2}^2 + A_{X,k,3}^2} \sqrt{\Omega_{X,k,1}^2 + \Omega_{X,k,2}^2 + \Omega_{X,k,3}^2} \end{array} \right]. \quad (5)$$

In (5), $*$ denotes all columns.
2.5.2. Initial Template Matching

We search the time-scaled test sequence for template moves by applying sliding correlation-based template matching, initially considering the acceleration and angular velocity magnitudes. For the template \( i \) and the test sequence, these are comprised in matrices \( \overline{r}_i(3) \) and \( \overline{x}(5) \), respectively.

For each template move \( 1 \leq i \leq I = 15 \), we first standardise the columns of \( \overline{r}_i \) to zero mean and unit standard deviation. By vectorising the standardised matrices column-wise, for each template move \( i \), we obtain a \( 2N_T \)-long vector \( \overline{r}_i \) of the acceleration and angular velocity magnitudes.

We further apply a 2D sliding window of size \( N_T \times 2 \) to the test sequence signal matrix \( \overline{x} \). For each \( 1 \leq n \leq N_X - N_T + 1 \), we extract an \( N_T \times 2 \) submatrix of \( \overline{x} \), i.e., \( \overline{x}_{n:n+N_T-1,\ast} \). After standardising and vectorising \( \overline{x}_{n:n+N_T-1,\ast} \) we obtain a \( 2N_T \)-long vector of the associated acceleration and angular velocity magnitudes. We then calculate the correlation coefficient between this vector and \( \overline{r}_i \). Sliding the 2D window vertically, we repeat this process for all \( n \) and obtain \( N_X - N_T + 1 \) correlation coefficients. We store the results for each template move \( i \) in a vector denoted with \( r_{\overline{r}x} \). This calculation is supported in detail with implementation equations in Appendix A.5.

Indicating high similarity between \( \overline{r}_i \) and \( \overline{x} \), the local maximums of \( r_{\overline{r}x} \) reveal the possible onsets of template move \( i \) executions in the test sequence. For each dance move \( i \), we identify the local maximums of \( r_{\overline{r}x} \) that exceed a certain threshold value, denoted with \( r_{\text{min}} \). We define the indices of these local maximums as move \( i \) onsets in the sequence. We further define the associated values of \( r_{\overline{r}x} \) local maximums as the recognition confidence.

After searching through the entire sequence for all template moves, we performed a final correction in the sense of false positives detection and elimination: if any two onsets of moves were closer than \( 0.7N_T \), we discarded the one with the lower recognition confidence.

2.5.3. Signal Transformation

To transform the signals of the test sequence from the device intrinsic to the coordinate system of the templates, we first identify from all onsets found in the previous step, for all \( i \) and for all \( n \), the one with the highest recognition confidence. This onset defines the strongest matching pair: for \( i \) and \( n \), for which the recognition confidence is the highest, the magnitude matrices of the template and subsequence, i.e., \( \overline{r}_i \) and \( \overline{x}_{n:n+N_T-1,\ast} \), respectively, match the most. We denote the corresponding matrices of the acceleration 3D projections with \( A_{\overline{TM}} \) and \( A_{\overline{XM}} \) and the combined matrices of the accelerations and angular velocity 3D projections with \( \overline{TM} \) and \( \overline{XM} \).

Since this template–subsequence pair represents the same move, but in different coordinate systems, it is reasonable to assume that a rotation that aligns them can be defined. Moreover, if we determine such a rotation, we can apply it to the entire test sequence, obtaining the acceleration and angular velocity 3D projections in the coordinate system of the templates.

To define this rotation, we adapted and applied an orientation-independent transformation, as presented in [20]. The original procedure estimates a wearable device orientation-independent coordinate system by calculating three orthogonal unit vectors from the 3D accelerometer signals, acquired using a smartphone in the front pocket of the user’s trousers during walking. The first unit vector, denoted with \( \zeta \), is defined by the gravity direction, and is calculated as the mean acceleration within a motion (walking) cycle. The second unit vector, denoted with \( \xi \), is considered to be aligned with the direction of motion and is determined using PCA as the direction with the highest variance once the projection on \( \zeta \) has been subtracted. The third unit vector, denoted with \( \psi \), is defined as orthogonal to the first two and as such covers lateral motion during the motion cycle.

In our dance analysis context, both the template and subsequence of the best-matching pair represent one full dance move, i.e., motion cycle. Considering this, following the
abovementioned, from $A_{TM}$, we obtain three orthogonal unit vectors $\zeta_T$, $\xi_T$, and $\psi_T$ while from $A_{XM}$, we obtain three orthogonal unit vectors $\zeta_X$, $\xi_X$, and $\psi_X$. Extracting unit vectors $\zeta$, $\xi$, and $\psi$ from the acceleration 3D projections has already been presented in detail in [20]; we provide details specific to this study in Appendix B.

Combining both triplets of unit vectors, $\zeta_T$, $\xi_T$, $\psi_T$ and $\zeta_X$, $\xi_X$, $\psi_X$, we can now define a rotation matrix $R$ that aligns the coordinate system of the test sequence with that of the templates as:

$$R = \begin{bmatrix} \zeta_X & \xi_X & \psi_X \\ \zeta_T & \xi_T & \psi_T \end{bmatrix}^{-1}. \quad (6)$$

Using $R$ (6), we can transform the original acceleration and angular velocity 3D projections, comprised in matrices $A_X$ and $\Omega_X$, respectively, from the test sequence’s device intrinsic to the coordinate system of the template moves. Denoting the resulting matrices with $A_X^{(R)}$ and $\Omega_X^{(R)}$, we write:

$$A_X^{(R)} = A_X R$$
$$\Omega_X^{(R)} = \Omega_X R. \quad (7)$$

The common $N_X \times 6$ matrix of the rotated acceleration and angular velocity 3D projections of the test sequence, denoted with $X^{(R)}$, given in the coordinate system of the templates, is then:

$$X^{(R)} = \begin{bmatrix} A_X^{(R)} & \Omega_X^{(R)} \end{bmatrix}. \quad (8)$$

2.5.4. Final Template Matching

We can now search the test sequence for template moves, following a similar procedure as presented in Section 2.5.2., only this time considering the acceleration and angular velocity 3D projections instead of their magnitudes. For the template moves and the test sequence, these are comprised in matrices $T_i(2)$ and $X^{(R)}$ (6), respectively.

For each template move $1 \leq i \leq I = 15$, we first standardise to zero mean and unit standard deviation the columns of $T_i$. Vectorising the standardised matrices column-wise, for each template move $i$, we obtain a $6N_T$-long vector of acceleration and angular velocity projections $t_i$.

We further use a 2D sliding window of size $N_T \times 6$, and for each $1 \leq n \leq N_X - N_T + 1$, we extract an $N_T \times 6$ submatrix of $X^{(R)}$, i.e., $X^{(R)}_{n:n+N_T-1,n}$. After standardisation and vectorisation, we obtain a $6N_T$-long vector of the associated acceleration and angular velocity projections. We then calculate the correlation coefficient between this vector and $t_i$.

Sliding the 2D window vertically, we repeated this process for all $n$ and obtain $N_X - N_T + 1$ correlation coefficients. We store the results for each template move $i$ in a vector denoted with $r_{T_i X^{(R)}}$. This calculation is supported in detail with implementation equations in Appendix A.6.

Indicating high similarity between $T_i$ and $X^{(R)}$, local maximums of $r_{T_i X^{(R)}}$ reveal the possible presence of move $i$ in the test sequence. For each dance move $i$, we identify the local maximums of $r_{T_i X^{(R)}}$ that exceed a certain threshold value, $r_{min}$. We define the indices of these local maximums as move $i$ onsets in the sequence and the associated $r_{T_i X^{(R)}}$ values as the recognition confidence.

After the entire sequence has been searched through, we performed a final correction for false positives detection and elimination: if any two onsets were closer than $0.7$ $N_T$, we discarded the one with the lower recognition confidence.

2.6. Recognition Performance Assessment

All performances of dancers for all test sequences were visually inspected by the professional using the accompanying videos. All moves determined to be improperly performed were discarded from analysis. All proper performances were appropriately
labelled as one of the 15 moves and considered for assessing the recognition ability of the presented method.

The presented recognition method is non-binary and we assess its recognition ability by considering each move that is correctly recognised as a true positive (TP). Each move that is incorrectly recognised we consider as a false positive (FP). Using cumulative TP and FP rates for all \( I = 15 \) moves, we calculate the sensitivity and precision scores for each dance tempo considered. We calculate these scores for threshold values \( r_{\text{min}} \) ranging from 0.25 to 0.80 with 0.05 increment steps. We finally use the F1 score, calculated from sensitivity and precision, as the unified measure of recognition ability.

In the specific dance motion recognition context, the sensitivity scores are mainly influenced by the similarity between the analysed moves and the corresponding templates in the database. As such, the sensitivity scores obtained for the professional dancer indicate how consistent her performance is, while those obtained for the recreational dancers show how well they imitate the professional.

Precision scores are additionally influenced by the similarities between different moves, indicated by nondiagonal elements of \( r_{\text{TT}} \) for the initial template matching and of \( r_{TT} \) for the final template matching. Performed on acceleration and angular velocity 3D projections, final template matching is expected to have a higher discriminative potential than the initial template matching, performed using the acceleration and angular velocity magnitudes.

3. Results and Discussion

3.1. Database of Template Moves

Figure 2 shows the acquired acceleration and angular velocity signals for two illustrative template moves: (3) Kicks and (7) 20s Charleston. The first row depicts the acceleration signals, while the second row shows the angular velocities. Even from pure visual inspection, we can see that both the 3D projections and magnitudes have distinctive signal shapes. It is therefore reasonable to assume that these signal shapes can be used as fingerprints enabling dance move recognition.

![Figure 2](image-url)

Figure 2. Examples of acceleration (first row) and angular velocity (second row) of right leg motion during two solo jazz moves. The first three columns show the 3D projections and the last column illustrates the magnitudes.
The similarities between different template moves $r_{T_i T_j}$, calculated using the magnitudes of acceleration and angular velocity, are presented in Table 1. As expected, some dance moves show greater pairwise similarity while others are more idiosyncratic and show greater distinction from the rest. The highest similarity is $r_{T_7 T_8} = 0.61$, obtained for the pair of moves (7) 20s Charleston and (8) Knee slaps, reflecting a leg activation pattern these two moves have in common. Both of these moves are executed mainly with the right leg during the first four music beats and with the left leg during the second four music beats.

Table 1. Solo jazz template moves’ pairwise similarities calculated using the magnitudes of acceleration and angular velocity. Depicted in red are the highest similarities obtained for each move, excluding self-similarity.

| Move | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|
| (1)  | 1.00| 0.22| 0.45| 0.29| 0.39| 0.29| 0.37| 0.27| 0.35| 0.35  | 0.32  | 0.46  | 0.27  | 0.31  | 0.31 |
| (2)  | 1.00| 0.19| 0.21| 0.35| 0.47| 0.29| 0.44| 0.34| 0.28| 0.33  | 0.33  | 0.27  | 0.30  | 0.37 |
| (3)  | 1.00| 0.42| 0.39| 0.28| 0.30| 0.33| 0.35| 0.26| 0.36| 0.55  | 0.33  | 0.43  | 0.30  |
| (4)  | 1.00| 0.37| 0.29| 0.31| 0.36| 0.39| 0.24| 0.43| 0.55| 0.32  | 0.27  | 0.43  |
| (5)  | 1.00| 0.33| 0.37| 0.37| 0.32| 0.29| 0.26| 0.29| 0.48| 0.23  | 0.48  | 0.38  |
| (6)  | 1.00| 0.38| 0.42| 0.33| 0.32| 0.44| 0.38| 0.24| 0.36| 0.44  |
| (7)  | 1.00| 0.61| 0.20| 0.37| 0.31| 0.32| 0.25| 0.33| 0.28 |
| (8)  | 1.00| 0.27| 0.34| 0.47| 0.39| 0.34| 0.30| 0.42 |
| (9)  | 1.00| 0.22| 0.29| 0.24| 0.30| 0.43| 0.31 |
| (10) | 1.00| 0.29| 0.35| 0.31| 0.27| 0.30 |
| (11) | 1.00| 0.47| 0.54| 0.26| 0.35 |
| (12) | 1.00| 0.39| 0.31| 0.43 |
| (13) | 1.00| 0.19| 0.32 |
| (14) | 1.00| 0.43 |
| (15) | 1.00 |

The move with the lowest similarity with respect to all other moves, i.e., with the lowest maximum $r_{T_i T_j}$ value, is Apple Jacks (10): $r_{T_{10} T_j} \leq 0.37$ for $1 \leq j \leq 15, j \neq 10$.

The presented values indicate the discriminative potential of using the acceleration and angular velocity magnitudes for recognising different solo jazz dance moves. In particular, for threshold values $r_{\text{min}} > 0.61$, high precision scores are expected; however, the higher $r_{\text{min}}$ is, the lower is the expected sensitivity, diminishing the model’s overall recognition ability.

The similarities between different template moves $r_{T_i T_j}$, calculated using the 3D projections, are presented in Table 2. We can see that these values are generally lower than the corresponding values of $r_{T_i T_j}$, and for all pairs of moves, $r_{T_i T_j}$ is below 0.5.

The highest $r_{T_i T_j}$ value is obtained for the pair of moves (4) Half break and (11) Boogie back: $r_{T_4 T_{11}} = 0.43$. The lowest maximum $r_{T_i T_j}$ value is 0.24 and is obtained for moves (8) Knee slaps and (14) Cross step. The presented values indicate the expected performance of the final template matching: for threshold values $r_{\text{min}} > 0.43$, high precision scores are expected.

Lower values of $r_{T_i T_j}$ obtained with respect to $r_{T_i T_j}$ confirm the natural assumption that comparing dance moves using the 3D projections of signals instead of their magnitudes, considering the direction of motion instead of only the general leg activation pattern, has a higher discriminative potential. Therefore, template matching achieved using the 3D projections instead of magnitudes is expected to provide for higher precision scores, better overall recognition ability, and lower values of $r_{\text{min}}$. 
### Table 2. Solo jazz template moves’ pairwise similarities calculated using the 3D projections of the acceleration and angular velocity. Depicted in red are the highest similarities obtained for each move, excluding self-similarity.

| Move | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|-----|
| (1)  | 1.00| 0.18| 0.28| 0.21| 0.23| 0.15| 0.23| 0.15| 0.26| 0.32| 0.25| 0.19| 0.10| 0.23| 0.25 |
| (2)  | 1.00| 0.22| 0.10| 0.22| 0.23| 0.20| 0.23| 0.12| 0.21| 0.15| 0.22| 0.14| 0.29| 0.19| 0.21 |
| (3)  | 1.00| 0.21| 0.40| 0.18| 0.25| 0.10| 0.19| 0.186| 0.18| 0.32| 0.13| 0.38| 0.18| 0.38| 0.18 |
| (4)  | 1.00| 0.26| 0.19| 0.13| 0.12| 0.26| 0.17| 0.43| 0.34| 0.11| 0.16| 0.40| 0.24| 0.18| 0.21 |
| (5)  | 1.00| 0.23| 0.35| 0.14| 0.17| 0.20| 0.26| 0.33| 0.10| 0.34| 0.25| 0.26| 0.26| 0.26| 0.26 |
| (6)  | 1.00| 0.33| 0.22| 0.17| 0.18| 0.18| 0.16| 0.13| 0.23| 0.20| 0.26| 0.26| 0.26| 0.26| 0.26 |
| (7)  | 1.00| 0.23| 0.15| 0.11| 0.17| 0.17| 0.17| 0.17| 0.17| 0.17| 0.17| 0.17| 0.17| 0.17| 0.17 |
| (8)  | 1.00| 0.18| 0.21| 0.29| 0.17| 0.29| 0.17| 0.29| 0.17| 0.29| 0.17| 0.29| 0.17| 0.29| 0.17 |
| (9)  | 1.00| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23 |
| (10) | 1.00| 0.18| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29 |
| (11) | 1.00| 0.18| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29 |
| (12) | 1.00| 0.18| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29 |
| (13) | 1.00| 0.18| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29 |
| (14) | 1.00| 0.18| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29 |
| (15) | 1.00| 0.18| 0.28| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29| 0.23| 0.14| 0.29 |

#### 3.2. Dance Move Recognition

For all 36 test sequences, the estimated dance tempo is accurate up to 1 bpm, allowing us to temporally scale all sequences and perform correlation-based template matching, as presented in the previous section.

**3.2.1. Validation Using the Professional Dancer’s Test Sequences**

For the professional dancer, for each dance tempo 120–200 bpm, 75 moves were acquired. For tempo 220 bpm, 70 moves were acquired. The sensitivity and precision scores are presented in Figures 3 and 4.

*Figure 3. Sensitivity scores obtained for the professional dancer’s test sequences.*

**initial recognition**

**final recognition**
For the initial template matching, the highest sensitivity scores range from 0.80 to 0.91 and are obtained for \( r_{\text{min}} \leq 0.4 \). For the final template matching, they range from 0.84 to 0.96 and are also obtained for \( r_{\text{min}} \leq 0.4 \).

For all dance tempos considered, the precision scores are equal to 1 for \( r_{\text{min}} \geq 0.65 \) for the initial template matching and \( r_{\text{min}} \geq 0.50 \) for the final template matching, indicating that for \( r_{\text{min}} \) this high, there are no FPs. These results are consistent with the template moves’ pair-wise similarities presented in Tables 1 and 2. The improvement in precision for the final template matching is a direct consequence of using the 3D projections of the inertial signals instead of their magnitudes for template matching. As expected, considering the direction of motion provides for a better discrimination of moves.

Figure 5 shows the \( F1 \) scores. For each dance tempo considered, the maximum \( F1 \) score indicates the best recognition ability. Lower \( F1 \) scores to the left of the maximum indicate higher FP rates and consequently lower precision. Lower \( F1 \) scores to the right of the maximum indicate lower TP rates and consequently, lower sensitivity.

For the initial template matching, \( F1 \) maximums are obtained for \( r_{\text{min}} \leq 0.50 \). The highest maximum is 0.91, obtained for tempos 160 and 180 bpm for \( r_{\text{min}} = 0.30 \) and 0.35. The lowest maximum is 0.82, obtained for the slowest tempo considered, i.e., 120 bpm, for \( 0.30 \leq r_{\text{min}} \leq 0.40 \). We can also observe that for \( r_{\text{min}} > 0.60 \), \( F1 \) scores drop sharply.

Using the criteria of the highest minimum \( F1 \) score at a specific \( r_{\text{min}} \), we can conclude that the overall best recognition performance is achieved for \( r_{\text{min}} = 0.30 \) and 0.35, for which the \( F1 \) scores range from 0.82 for 120 bpm to 0.91 for 180 bpm dance tempos.

As expected, the \( F1 \) scores for the final template matching exceed those obtained for the initial template matching. The \( F1 \) maximums range from 0.87 to 0.98. The overall best recognition performance is obtained for \( r_{\text{min}} = 0.35 \) and 0.40.

These results confirm the potential of using the presented method to recognise repeated dance moves and to assess dancers’ consistency in performance.
3.2.2. Validation Using the Recreational Dancer’s Test Sequences

For the recreational dancers, 293–311 dance moves were collected for each dance tempo considered. The sensitivity, precision, and \( F1 \) scores are presented in Figures 6–8, respectively.

![Graph showing initial and final recognition for professional dancer's test sequences](image1)

**Figure 5.** \( F1 \) scores obtained for the professional dancer’s test sequences.

![Graph showing sensitivity scores for recreational dancer’s test sequences](image2)

**Figure 6.** Sensitivity scores obtained for the recreational dancers’ test sequences.
For the initial template matching, the sensitivity maximums range from 0.74 to 0.82 and are obtained for $r_{\text{min}} \leq 0.40$. For the final template matching, they range from 0.89 to 0.98 and are obtained for $r_{\text{min}} \leq 0.30$. For all dance tempos considered, the precision scores are equal to 1 for $r_{\text{min}} \geq 0.70$ for the initial template matching and $r_{\text{min}} \geq 0.60$ for the final template matching.
For the initial template matching, the $F_1$ maximums range from 0.76 to 0.84 and are obtained for $r_{\text{min}} \leq 0.45$. For the final template matching, the $F_1$ maximums range from 0.89 to 0.98 and are obtained $r_{\text{min}} \leq 0.30$.

The main difference when compared to the professional dancer’s results is that the maximum $F_1$ scores are obtained for lower values of $r_{\text{min}}$. In addition, after rising to their maximum value, $F_1$ scores fall more sharply. These observations are an expected consequence of individual style and subtle differences in move performances of different dancers.

The results confirm the potential of using the presented method to recognise imitated dance moves. Similar recognition ability can be achieved for the template professional dancer, but at lower values of $r_{\text{min}}$.

Finally, the recognition ability of the final template matching proves that extracting three device-independent coordinate system axes from the acquired dance motion acceleration signals as presented in the previous section is valid and enables reliable transformation of the inertial signals from the device-intrinsic to the coordinate system of the templates.

4. Conclusions

We have shown that a single wearable device, capturing 3D acceleration and angular velocity of a dancer’s right leg motion, provides for recognising 15 solo jazz dance moves. We have demonstrated that the analysed sequences can be reliably temporally scaled and that dance recognition can be achieved using the same templates, independent of the tempo of dancing. Considering this makes the creation of the template database easier.

Residing on a single device makes the solution highly practical. The practical aspect is further enhanced with the recognition’s independency of the device orientation. The results confirm that even for such a dynamic motion as solo jazz dancing, the acquired acceleration and angular velocity 3D projections can be reliably transformed to a device orientation-invariant system. Exploiting the 3D projections of the signals has a higher discriminative potential and provides for better recognition ability than does considering only the signals’ magnitudes.

The presented solution fills the gap of the existing state of the art: since the presented solution avoids using video cameras or IR imaging sensors, it is not computationally expensive or limited to specific locations; residing on a single wearable device and not being dependent on the orientation of the wearer makes the methodology suitable for unsupervised everyday use and applicable for a variety of dancing situations—whether the dancer is dancing alone, in the crowd, or in front of an audience.

Activity recognition methods relying on wearable devices, developed for ubiquitous measurement scenarios, usually try to balance the opposing requirements of device life autonomy and recognition performance. By relying on a limited set of features extracted from the inertial signals, these methods reduce the computational complexity and prolong the device life autonomy. Since the methodology presented in this article is primarily intended to be used for relatively short dancing sequences, e.g., per song basis, as opposed to throughout-the-day use, more computational power needed for performing the correlation-based template matching, can be invested in the analysis.

The advantages for the dancer and/or dancing instructor are straightforward: for each performed dance sequence, dance moves can be extracted together with recognition confidence levels, indicating how well the execution matches the reference template. This way, the methodology supports progress monitoring. In addition, the recognised moves can be further analysed and compared to the reference moves in the database to evaluate the overall performance. Finally, while tested on solo jazz dance moves, the methodology presented in this article can be extended to other dance styles, as well.

The presented methodology demonstrates a high recognition ability for the preliminary set of measurements obtained for six dancers. However, a standalone study is needed to further explore in-depth the variability between dancers. Further analysis is also necessary to assess the methodology recognition performance in real dancing scenarios, when the dancer is dancing freely, combining various predefined and improvised moves.
in sequences. Further studies can also capitalise on this study and assess the dancer’s creativity and response to music, investigating the crucial connection between dancing and music. The benefits of using additional devices, positioned on other body segments, can be investigated. Finally, the presented implementation can be explored further with the aim of optimisation for real-time execution.

**Author Contributions:** Conceptualisation, S.S.; methodology, S.S.; software, S.S.; validation, S.S.; formal analysis, S.S. and S.T.; investigation, S.S. and S.T.; resources, S.S.; writing, S.S.; visualisation, S.S.; project administration, S.S.; funding acquisition, S.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Slovenian Research Agency, grant number P2-0246 ICT4QoL—Information and communication technologies for quality of life.

**Institutional Review Board Statement:** This study was conducted during regular dance training hours with the participants performing exactly their regular training activities. More than minimal risks were not identified, and ethical review and approval were waived. The study followed the Code of Ethics of the University of Ljubljana, which provides guidelines for studies involving human beings and is in accordance with the Declaration of Helsinki.

**Informed Consent Statement:** Informed written consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors give special thanks to Špela Premelč for contributing her professional dancing experience.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**Appendix A**

**Appendix A.1 Signal Matrices Similarity**

Let \( \mathbf{M} \) be an arbitrary \( N \times K \) matrix. Let \( \mathbf{\hat{M}} \) further be a matrix with columns equal to the columns of matrix \( \mathbf{M} \), standardised to zero mean and unit standard deviation:

\[
\mathbf{\hat{M}}_{s,k} = \frac{\mathbf{M}_{s,k} - \mu_{\mathbf{M}_{s,k}}}{\sigma_{\mathbf{M}_{s,k}}}, \quad 1 \leq k \leq K,
\]  

(A1)

where \( \mu \) and \( \sigma \) denote the mean and standard deviation, respectively, while * denotes all rows.

Let now \( \mathbf{m} \) be an \( LK \)-long vector obtained by vectorising matrix \( \mathbf{\hat{M}} \) column-wise:

\[
\mathbf{m} = \left[ \mathbf{\hat{M}}_{s,1}^T \mathbf{\hat{M}}_{s,2}^T \cdots \mathbf{\hat{M}}_{s,K}^T \right]^T.
\]  

(A2)

In (A2), \( ^T \) denotes the transpose operator.

Following (A1) and (A2), for any two matrices \( \mathbf{M} \) and \( \mathbf{N} \) of equal size \( N \times K \), we can now define a similarity measure \( r_{MN} \), calculated as the correlation coefficient across the matrices’ \( K \) dimension as:

\[
r_{MN} = \frac{\mathbf{m}^T \mathbf{n}}{(NK)}.
\]  

(A3)

In particular, instead of \( \mathbf{M} \) and \( \mathbf{N} \), we use signal matrices with columns representing \( 1 \leq k \leq K \) signals and rows representing \( 1 \leq n \leq N \) time samples. The calculated similarity measure (A3) represents the correlation coefficient between two sets of \( K \) signals.
Appendix A.2 Dance Move Repetition Extraction

Let $S$ be an $Ns \times 6$ signal matrix, with columns representing the 6 inertial signals and rows representing the signals’ $Ns$ time samples. For a defined starting row $n_0$ and an initial length $Nn_0$, let us extract two adjacent $Nn_0 \times 6$ submatrices of $S$. By standardising these two submatrices according to (A1) and applying (A2), we obtain their respective vector forms. Further applying (A3), we obtain the correlation coefficient, indicating the submatrices’ similarity.

Repeating the aforementioned process for different lengths of the submatrices, $Nn_0 = N_0 + n$, where, in particular for the problem at hand, $N_0 = 392$ and $0 \leq n \leq 16$, we obtain a vector of correlation coefficients $r_s$. Finally, by identifying the index of the $r_s$ maximum, denoted with $n_{max}$, we obtain the number of rows $N_{n_{max}}$:

$$N_{n_{max}} = N_0 + n_{max}$$

for which the adjacent submatrices match the most. These two submatrices, i.e., $S_{n_0;N_0+N_{n_{max}}-1,*}$ and $S_{n_0+N_{n_{max}}+2N_{n_{max}}-1,*}$, are used as signal matrices of two consecutive repetitions of a dance move.

Appendix A.3 Magnitudes Similarity

Let $T_i$ and $T_j$ denote $N_T \times 2$ matrices of acceleration and angular velocity magnitudes for template moves $i$ and $j$. By inserting $T_i$ instead of $M$ in (A1)–(A2), we perform matrix normalisation and vectorisation. We denote the resulting $2N_T$-long vector with $t_i$.

For each $0 \leq \tau < N_T$, we cyclically shift matrix $T_j$ rows:

$$T_{j,\tau} = \begin{bmatrix} T_{j,\tau+1:N_T,*} \\ T_{j,1:\tau,*} \end{bmatrix}.$$  (A5)

By inserting $T_{j,\tau}$ instead of $M$ in (A1)–(A2), we perform normalisation and vectorisation. For each $\tau$, we denote the resulting $2N_T$-long vector with $t_{i,\tau}$. By further inserting $t_i$ and $t_{j,\tau}$ instead of $m$ and $n$ in (A3) for each $\tau$, we calculate a component of the similarity vector, denoted with $r_{T_iT_j,\tau}$. We use the maximum value of this vector, denoted with $r_{T_iT_j}$, as the two moves’ magnitude similarity measure.

Appendix A.4 Projections Similarity

Let $T_i$ and $T_j$ denote two $N_T \times 6$ matrices of 3D acceleration and angular velocity projections for template moves $i$ and $j$. By inserting $T_i$ instead of $M$ in (A1)–(A2), we perform normalisation and vectorisation. We denote the resulting $6N_T$-long vector with $t_i$.

For each $0 \leq \tau < N_T$, we cyclically shift matrix $T_j$ rows:

$$T_{j,\tau} = \begin{bmatrix} T_{j,\tau+1:N_T,*} \\ T_{j,1:\tau,*} \end{bmatrix}.$$  (A6)

By inserting $T_{j,\tau}$ instead of $M$ in (A1)–(A2), we perform normalisation and vectorisation. For each $\tau$, we denote the resulting $6N_T$-long vector with $t_{j,\tau}$. By further inserting $t_i$ and $t_{j,\tau}$ instead of $m$ and $N$ in (A3) for each $\tau$, we calculate a component of the similarity vector, denoted with $r_{T_iT_j,\tau}$. We use the maximum value of this vector, denoted with $r_{T_iT_j}$, as the two moves’ projection similarity measure.

Appendix A.5 Initial Template Matching

Let $X$ denote an $N_T \times 2$ matrix of the test sequence 3D acceleration and angular velocity magnitudes. For each $1 \leq n \leq N_X - N_T + 1$, we extract a submatrix of $\overline{X}$, i.e.,
By inserting this submatrix instead of $M$ in (A1)–(A2), we perform column-wise standardisation and matrix vectorisation, obtaining a $2N_T$-long vector denoted with $\overline{x}_t$ for each $n$.

With the signals prepared, we can now calculate the vector of correlation coefficients between the magnitude signals of the sequence and the templates. For each template move $i$, we denote this vector with $\mathbf{r}_{t_X}^{T}$. For each time lag $n$, we obtain an element of vector $\mathbf{r}_{t_X}^{T}$ by inserting $t_i$ and $\overline{x}_n$ instead of $M$ and $N$ in (A3):

$$r_{t_X}^{T}(n) = t_i^T x_n / (2N_T).$$  (A7)

Indicating high similarity between $T_i$ and $X$, local maximums of $\mathbf{r}_{t_X}^{T}$ reveal the possible onsets of template move $i$ execution in the test sequence.

**Appendix A.6 Final Template Matching**

Let $X^{(R)}$ denote an $N_T \times 6$ matrix of the test sequence 3D accelerations and angular velocities projections on the templates’ coordinate system axes. For each time lag $1 \leq n \leq N_X - N_T + 1$, we extract a submatrix of $X^{(R)}$, i.e., $X^{(R)}_{n:n+N_T-1,s}$. By inserting this submatrix instead of $M$ in (A1)–(A2), we perform column-wise standardisation and matrix vectorisation, obtaining a 6$N_T$-long vector, denoted with $x^{(R)}_n$ for each $n$.

We can now calculate the vector of the correlation coefficients between the projection signals of the sequence and the templates. For each template move $i$, we denote this vector with $\mathbf{r}_{t_X(R)}^{T}$. For each time lag $n$, we obtain an element of $\mathbf{r}_{t_X(R)}^{T}$ by inserting $t_i$ and $x^{(R)}_n$ instead of $M$ and $N$ in (A3):

$$r_{t_X(R)}^{T}(n) = t_i^T x_n^{(R)} / (6N_T).$$  (A8)

Indicating high similarity between $T_i$ and $X^{(R)}$, local maximums of $\mathbf{r}_{t_X(R)}^{T}$ reveal the possible presence of move $i$ execution in the test sequence.

**Appendix B**

**Orientation Independent Transformation**

Let $A$ be an $N_T \times 3$ matrix of 3D acceleration signals of a single dance move. The columns of $A$ represent the acceleration projections on the device-intrinsic coordinate system axes and rows represent the time dimension. Following the procedure presented in [20], we extract three orthogonal $3 \times 1$ unit vectors from $A$ as follows. The orientation of the first unit vector, denoted with $\zeta$, is defined as the gravity direction, being parallel to the user’s torso, and is calculated as the mean acceleration direction within the motion cycle, i.e., dance move. Denoting $a$ as the mean direction of acceleration during one dance move, we can write:

$$a = \left[ \mu_{A_{z,1}} \mu_{A_{z,2}} \mu_{A_{z,3}} \right]^T$$

$$\zeta = a / \|a\|. \quad (A9)$$

In (A9), $\mu$ and $\|\|$ denote the vector mean and magnitude calculation, respectively, while $^T$ denotes the transpose operator. Acceleration projection on axis $\zeta$ is then calculated:

$$A_{\zeta} = A \zeta \quad (A10)$$

and subtracted from the original 3D acceleration signals, giving acceleration in the plane orthogonal to the gravity direction:

$$A_{\zeta}^f = A - \zeta A_{\zeta}^T. \quad (A11)$$
The orientation of the second unit vector $\xi$ is defined as the direction covering most of $A^T \xi$ variance. Normalised eigenvectors of $A^T \xi$ are determined using Principal Component Analysis (PCA). The one associated with the maximum eigenvalue is set as $\xi$. Finally, the remaining unit vector is defined as being orthogonal to the first two:

$$\psi = \zeta \times \xi.$$  \hspace{1cm} (A12)

In (A12), $x$ denotes the cross product.

We apply this method to our dance signals. By inserting the 3D acceleration signal matrices of the template and subsequence of the best matching pair, $A_{TM}$ and $A_{XM}$, respectively, instead of $A$ in (A9)–(A12), we obtain two triplets of orthogonal unit vectors. From $A_{TM}$ we obtain unit vectors $\zeta_T$, $\xi_T$, and $\psi_T$, while from $A_{XM}$ we obtain three orthogonal unit vectors $\zeta_X$, $\xi_X$, and $\psi_X$.

References

1. Kyan, M.; Sun, G.; Li, H.; Zhong, L.; Muneesawang, P.; Dong, N.; Elder, B.; Guan, L. An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment. ACM Trans. Intell. Syst. Technol. 2015, 6, 1–37. [CrossRef]

2. Aich, A.; Mallick, T.; Bhuyan, H.B.G.S.; Das, P.; Majumdar, A.K. NrityaGuru: A Dance Tutoring System for Bharatanatyam using Kinect. In Computer Vision, Pattern Recognition, Image Processing, and Graphics; Rameshan, R., Arora, C., Dutta Roy, S., Eds.; Springer: Singapore, 2018; pp. 481–493. [CrossRef]

3. Dos Santos, A.D.P.; Yacef, K.; Martinez-Maldonado, R. Let’s dance: How to build a user model for dance students using wearable technology. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 9–12 July 2017; pp. 183–191. [CrossRef]

4. Drobny, D.; Weiss, M.; Borchers, J. Saltate!: A sensor-based system to support dance beginners. In Proceedings of the 27th Annual CHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 3943–3948. [CrossRef]

5. Romano, G.; Schneider, J.; Drachsler, H. Dancing Salsa with Machines—Filling the Gap of Dancing Learning Solutions. Sensors 2019, 19, 3661. [CrossRef] [PubMed]

6. Ofli, F.; Erzin, E.; Yemez, Y.; Tekalp, A.M.; Erdem, C.E.; Erdem, A.T.; Abaci, T.; Ozkan, M.K. Unsupervised dance figure analysis from video for dancing Avatar animation. In Proceedings of the 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1484–1487. [CrossRef]

7. Yamane, R.; Shakunaga, T. Dance motion analysis by correlation matrix between pose sequences. In Proceedings of the 25th International Conference of Image and Vision Computing, Queenstown, New Zealand, 8–9 November 2010; pp. 1–6. [CrossRef]

8. Shikanai, N.; Hachimura, K. The Effects of the Presence of an Audience on the Emotions and Movements of Dancers. Procedia Technol. 2014, 18, 32–36. [CrossRef]

9. Shikanai, N. Relations between Femininity and the Movements in Japanese Traditional Dance. In Proceedings of the IEEE International Conference on Consumer Electronics–Asia (ICCE–Asia), Bangkok, Thailand, 12–14 June 2019; pp. 146–148. [CrossRef]

10. Kim, D.; Kim, D.H.; Kwak, K.C. Classification of K-Pop Dance Movements Based on Skeleton Information Obtained by a Kinect Sensor. Sensors 2017, 17, 1261. [CrossRef] [PubMed]

11. Bakalos, N.; Protopapadakis, E.; Doulamis, A.; Doulamis, N. Dance Posture/Steps Classification Using 3D Joints from the Kinect Sensor. In Proceedings of the IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, 12–15 August 2018; pp. 868–873. [CrossRef]

12. Ho, C.; Tsai, W.; Lin, K.; Chen, H.H. Extraction and alignment evaluation of motion beats for street dance. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2429–2433. [CrossRef]

13. Cornacchia, M.; Ozcan, K.; Zheng, Y.; Velipasalar, S. A Survey on Activity Detection and Classification Using Wearable Sensors. IEEE Sens. J. 2017, 17, 386–403. [CrossRef]

14. Siirtola, P.; Rönng, J. Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data. Int. J. Artif. Intell. Interact. Multimed. 2012, 1, 38–45. [CrossRef]

15. Long, X.; Yin, B.; Aarts, R.M. Single-accelerometer-based daily physical activity classification. In Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009), Minneapolis, MN, USA, 3–6 September 2009; pp. 6107–6110. [CrossRef]

16. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. IEEE Commun. Surv. Tutor. 2013, 15, 1192–1209. [CrossRef]
17. Niemann, F.; Reining, C.; Moya Rueda, F.; Nair, N.R.; Steffens, J.A.; Fink, G.A.; ten Hompel, M. LARa: Creating a Dataset for Human Activity Recognition in Logistics Using Semantic Attributes. *Sensors* **2020**, *20*, 4083. [CrossRef]
18. Sousa Lima, W.; Souto, E.; El-Khatib, K.; Jalali, R.; Gama, J. Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview. *Sensors* **2019**, *19*, 3213. [CrossRef]
19. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors. *Sensors* **2016**, *16*, 426. [CrossRef] [PubMed]
20. Gadaleta, M.; Rossi, M. IDNet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognit.*** **2018**, *74*, 25–37. [CrossRef]
21. Sprager, S.; Juric, M.B. Inertial Sensor-Based Gait Recognition: A Review. *Sensors* **2015**, *15*, 22089–22127. [CrossRef]
22. Cola, G.; Avvenuti, M.; Vecchio, A. Real-time identification using gait pattern analysis on a standalone wearable accelerometer. *Comput. J.* **2017**, *60*, 1173–1186. [CrossRef]
23. Junker, H.; Amft, O.; Lukowicz, P.; Tröster, G. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognit.* **2008**, *41*, 2010–2024. [CrossRef]
24. Martínez, A.; Jahnel, R.; Bucheger, M.; Snyder, C.; Brunauer, R.; Stöggl, T. Development of an Automatic Alpine Skiing Turn Detection Algorithm Based on a Simple Sensor Setup. *Sensors* **2019**, *19*, 902. [CrossRef] [PubMed]
25. Kos, A.; Umek, A. Smart sport equipment: SmartSki prototype for biofeedback applications in skiing. *Pers. Ubiquitous Comput.* **2018**, *22*, 535–544. [CrossRef]
26. Benages Pardo, L.; Buldain Perez, D.; Orrite Uruñuela, C. Detection of Tennis Activities with Wearable Sensors. *Sensors* **2019**, *19*, 5004. [CrossRef]
27. Ghasemzadeh, H.; Jafari, R. Coordination Analysis of Human Movements With Body Sensor Networks: A Signal Processing Model to Evaluate Baseball Swings. *IEEE Sens. J.* **2011**, *3*, 603–610. [CrossRef]
28. Paradiso, J.A.; Hsiao, K.; Benbasat, A.Y.; Teegarden, Z. Design and implementation of expressive footwear. *IBM Syst. J.* **2000**, *39*, 511–529. [CrossRef]
29. Aylward, R.; Lovell, S.D.; Paradiso, J.A. A Compact, Wireless, Wearable Sensor Network for Interactive Dance Ensembles. In Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, Cambridge, MA, USA, 3–5 April 2006. [CrossRef]
30. Dadashi, F.; Millet, G.P.; Aminian, K. A Bayesian approach for pervasive estimation of breaststroke velocity using a wearable IMU. *Pervasive Mob. Comput.* **2015**, *19*, 37–46. [CrossRef]
31. Yamaguchi, T.; Ariga, A.; Kobayashi, T.; Hashimoto, S. TwinkleBall: A Wireless Musical Interface for Embodied Sound Media. In Proceedings of the 17th Brazilian Symposium on Computer Music (SBCM 2019), São João del-Rei, Brazil, 25–27 September 2019. [CrossRef]
32. Tragtenberg, J.; Calegario, F.; Cabral, G.; Ramalho, G. TumTá and Pisada: Two Foot-controlled Digital Dance and Music Instruments Inspired by Popular Brazilian Traditions. In Proceedings of the 17th Brazilian Symposium on Computer Music (SBCM 2019), São João del-Rei, Brazil, 25–27 September 2019. [CrossRef]
33. Kim, Y.; Jung, D.; Park, S.; Chi, J.; Kim, T.; Lee, S. The Shadow Dancer: A New Dance Interface with Interactive Shoes. In Proceedings of the 2008 International Conference on Cyberworlds, Hangzhou, China, 22–24 September 2008; pp. 745–748. [CrossRef]
34. Hinten-Lewis, C.W.; McDonough, E.; Moyle, G.M.; Thiel, D.V. An Assessment of Postural Sway in Ballet Dancers During First Position, Relevé and Sauté with Accelometers. *Procedia Eng.* **2016**, *147*, 127–132. [CrossRef]
35. Thiel, D.V.; Quandt, J.; Carter, S.J.L.; Moyle, G.M. Accelerometer based performance assessment of basic routines in classical ballet. *Procedia Eng.* **2014**, *72*, 14–19. [CrossRef]
36. Stančin, S.; Tomažić, S. Dance Tempo Estimation Using a Single Leg-Attached 3D Accelerometer. *Sensors* **2021**, *21*, 8066. [CrossRef] [PubMed]
37. Mbitentlab MMR. Available online: https://mbientlab.com/metamotionr/ (accessed on 21 December 2021).
38. Alphabetical Jazz Steps 3. Available online: https://www.youtube.com/watch?v=jAIwJd2tQo0&list=PLpLDojUPSMvcYMA7 (accessed on 21 December 2021).