Available online at www.sciencedirect.com

ScienceDirect

International Journal of Human-Computer Studies 00 (2022) 1–22

Enriching Teachers' Assessments of Rhythmic Forró Dance Skills by Modelling Motion Sensor Data

Augusto Dias Pereira dos Santos[a,b,*], Lian Loke[c], Kalina Yacef[c], Roberto Martinez-Maldonado[d]

[a] *University of New South Wales, Sydney, Australia*
[b] *Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil*
[c] *The University of Sydney, Sydney, Australia*
[d] *Monash University, Melbourne, Australia*

Abstract

Rhythmic skills are fundamental in social partner dance. Yet, large class sizes and time constraints often make it challenging for teachers to assess learners' rhythmic skills to provide constructive feedback. As motion sensors are becoming widely available and embedded in most smartphones, it is also becoming feasible to create models of anatomical movements of the human body that could be used to support such teachers' assessments. Current solutions aimed at modelling motor skills using motion sensors and artificial intelligence (AI) have enabled the detection of some dimensions of rhythm, but they are either technically too complex to be used 'in-the-wild' or have not been designed based on dance teachers' assessment needs. This paper presents an approach for modelling rhythmic movements, via a single smartphone, while learners learn how to dance Forró. The purpose is to enrich teachers' rhythmic assessments with data. Following a user-centred approach, we first elicited from teachers the dance elements that they commonly focus their rhythm assessments on (i.e., tempo, pause, step size and weight transfer). Then, selected features were extracted from raw motion sensor data related to the rhythmic patterns of learners dancing, their synchronisation with the beat of the music, and particular characteristics of the song being played. Finally, machine learning (ML) algorithms were used to create predictive computational models using these features. The modelling approach was validated through two studies: 1) a quantitative comparison between the ML outputs and dance teachers assessments of learners' dance performance; and 2) a qualitative analysis of the potential pedagogical uses of the outputs of the ML models envisioned by dance teachers.

Keywords: dance rhythm detection, accelerometer, dance education, motor learning, activity recognition, user study

1. Introduction

The Brazilian partner dance style called Forró is performed to the songs of Forró music styles. In Forró music, the zabumba (drum) is the instrument that, most of the time, keeps the strongest beat of the song and its tempo (Fernandes, 2005). Forró songs have a quaternary tempo (1-4) to which the dancers synchronise their steps; a common pattern present in the music of other popular partner dance styles such as salsa, bachata, merenge and kizomba. Yet, learning to dance to the rhythm of the music (in Forró and the other dance styles mentioned above) is a complex and often challenging activity that involves the development of cognitive and motor skills (Schmidt and Wrisberg, 2008) aligned with constant auditory stimuli (Phillips-Silver and Trainor, 2007). Whilst dancing requires the development of several

*Corresponding author
Email address: augusto.dias@sydney.edu.au (Augusto Dias Pereira dos Santos)

skills, *rhythmic* skills are fundamental because they relate to the synchronisation of body movements and the music (Erkert, 2003). Dance teachers have developed different pedagogical strategies to help people develop rhythmic skills according to the particular difficulties each may face (Erkert, 2003; McCutchen, 2006). Yet, in practice, dance teachers commonly have to manage large classes and cannot easily assess each learner's progress (e.g. in the dance studio or at home) in order to provide constructive and personalised feedback (Hsia et al., 2016; Goldstein, 2007).

Recent improvements to wearable technologies (e.g. smartphones, activity trackers and motion sensors) and artificial intelligence (AI) algorithms offer new opportunities to model movement, which can enable the creation of support tools for dance teaching and learning. For example, some researchers have attempted to model rhythm by assessing *tempo* (i.e. the synchrony between the body movement and the music beat) using accelerometer sensors (e.g. dos Santos et al., 2017; Faridee et al., 2018; Lee et al., 2007) or optical motion trackers (e.g. Camurri et al., 2016b; Chan et al., 2011; Trajkova and Cafaro, 2018; Romano et al., 2019; Senecal et al., 2018). Other researchers have focused on automatically warning learners if they get out of rhythm (e.g. Drobny et al., 2009) or modelling other related dance skills such as dynamic symmetry (Camurri et al., 2016a), dance movements from ballet (Hinton-Lewis et al., 2016) and classical Indian steps (Faridee et al., 2018). In short, there have been some attempts to model rhythm, and some attempts to support learning to dance using resulting models. However, previous approaches are too complex to be used in authentic dance settings (e.g. requiring multiple wearable devices to be attached to the various body parts or large room setups) and have not been designed based on dance teachers' assessment needs (i.e. modelling other aspects related to rhythm learning besides tempo).

This paper addresses this gap in the context of the Forró dance style. We present an approach for modelling learners' rhythmic movements for the purpose of enriching teachers' rhythmic assessments with data. We followed a user-centred approach, to investigate: 1) how to model rhythm in the context of social dance, using a single smartphone; and 2) how teachers of social dance can use the AI outputs from the models to enrich their assessments of rhythmic skills. From previous literature (dos Santos et al., 2018b; McCutchen, 2006; dos Santos et al., 2018a), four elements of rhythmic skills were identified: tempo, weight transfer, pause and step size. Based on these four elements we developed STEP, an algorithm that extracts music-related motion features (MMFs) from motion sensors. These features were used to model those four rhythmic skills using predictive machine learning (ML) algorithms. The modelling approach was validated through two studies. In Study 1, the resulting classification models were validated using a ground-truth dataset collected from four dance teachers assessing video-recorded data from 16 learners enrolled in a 3-day dance course. The outputs from the resulting predictive models and the dance teachers' assessments of learners' dance performance were quantitatively compared to identify the accuracy of the models. In Study 2, eight dance teachers were interviewed while assessing learners' video-recorded performance, with and without the support of the automated assessment metrics generated using the ML models from Study 1.

The rest of the paper is structured as follows. The next section presents foundations of rhythm in dance, common assessment strategies used in dance education, the particularities of the Forró dance style and the related work in the area of rhythm modelling using sensors and AI for supporting learning. Section 3 presents the proposed modelling approach and describes the studies. Section 4 presents the systems used in the studies. Section 5 presents the quantitative results of our algorithm evaluation (Study 1) and Section 6 presents the qualitative evaluation with teachers (Study 2). Section 7 discusses the implications of both studies and concludes with recommendations for future work.

2. Background and Related Work

2.1. Foundations of Rhythm in Dance Learning

Although there is not a single accepted definition of rhythm, it can be defined as a regular flow of expected events, which in dance correspond to body movements (Fraisse, 1982). Rhythm is usually taught in dance classes as part of a broader topic called *musicality* (Côté-Laurence, 2000). *Musicality* refers to the quality of the dancer's movement and its connection with the various elements of the music. For example, rhythmic expression includes the flow of arms and legs following the music; the bodily interaction with time, space and gravity; the extent to which weight and energy are used to express movements; and the use of the ground to land, push or absorb the impact of the body (Erkert, 2003). It is therefore important for learners of all ages to develop rhythmic skills, that is, develop the psychomotor and perceptual abilities to perform and synchronise to music (Chatzopoulos et al., 2019).

In the context of social partner dance, teachers use a broad array of concepts such as synchronicity, weight transfer, limbs-joints, quality of movement, posture and gaze to assess musicality, and more specifically, learners' rhythm (dos

Santos et al., 2018b). Teachers commonly scaffold the development of rhythmic skills by creating standard exercises (Côté-Laurence, 2000). These exercises not only focus on the temporal aspects of rhythm, but also on the accuracy of the motor response, *feeling the music*, building a phrase (a sequence of movements), breathing, or even *playing* with the song (i.e. improvising) (Côté-Laurence, 2000).

2.2. Assessment Strategies in Dance Education

Strategies used by teachers and instructors to assess learners' performance are somewhat similar across dance genres. Teachers commonly use visual observation to assess learners in the classroom (Jarmolow and Selck, 2011; Vecchi, 2012; DeMers, 2013; Erkert, 2003). Another common strategy is to perform kinaesthetic assessments by physically sensing and feeling the learners' movement (Vecchi, 2012). This method is particularly used in partner dance classes as the nature of the dance (in pairs) enables teachers to take one of the roles of the dancers (leader or follower) and assess the learner. These two types of assessments are also in line with the broader literature on motor learning (Schmidt and Wrisberg, 2008).

Self-assessment and peer-assessment are strategies less commonly used to assess learners' progress in dance contexts (McCutchen, 2006; Hsia et al., 2016). In the context of formal dance classes, it is common for the teacher to objectively assess learners using guidelines and rubrics (Erkert, 2003; McCutchen, 2006), which in some countries are established by national bodies (Ross, 1994). However, dance teachers often face a large number of learners and it is not feasible for them to assess the progress of each of them (Goldstein, 2007). These large classes makes it challenging for teachers to be aware of the individual needs to provide personalised attention (Hsia et al., 2016).

2.3. The learning context: Forró Social Partner Dance Style

The Forró social partner dance style frames a unique set of requirements compared to other dance practices and relevant works. The main and first movement that learners need to learn when starting Forró classes is the Básico 1 exercise. We use this movement pattern in this paper to illustrate our approach. Learners require several skills to perform this movement pattern such as rhythm, balance and coordination. When using this movement pattern as an exercise, learners have to repeat it continuously while listening to a song. Básico 1 requires the learner to perform six movements during eight beats of the song. During two beats of the song (4th and 8th) the learner has to pause their movement. The full description of the movement pattern is illustrated in Figure 1.

We chose this exercise because it is the first step taught to learners: it involves the foundational dance skills that learners need to master for evolving in their dance learning pathway. Two properties of this exercise are critical for creating computational models of rhythm: i) the **8 beat** cycle in which the exercise must be performed, and ii) the **6 movements** that the learners need to perform. When using this exercise to evaluate learners' performance, the teacher can observe a number of elements and distinguish between correct and incorrect execution. This evaluation process occurs mostly inside the head of the teacher; it is rarely described in written protocols and seldom involves the use of any tool to assist the assessment. In this paper, we propose that computational approaches to modelling rhythm can be used to assist and augment this assessment process, instead of replacing teachers' judgement. In the next subsection we discuss state-of-the-art developments in rhythm modelling and motivate the need for pedagogically-informed, lightweight rhythm modelling approaches.

2.4. Related Work: Rhythm Modelling in Dance

Emerging sensing technologies are allowing researchers to investigate learning in a wide range of motor learning scenarios (Santos, 2016; Santos and Eddy, 2017). Researchers have been using motion sensors to model and support rhythm learning in scenarios such as percussion (Matsumura et al., 2011; Kawakami and Fujinami, 2008), piano (Hadjakos et al., 2008) and drumming (Ochi, 2018) playing. This growing interest in using accelerometer data for modelling rhythm in the context of musical instrument-playing has also been reflected in dance education.

In one of the pioneering studies in rhythm modelling, Lee et al. (2007) used several accelerometers attached to participants to model rhythm in general human motion. The approach was useful for identifying some features of the movements, such as the time length of the movement pattern, but ineffective for assessing rhythm in dance movement. Importantly, the relationship between the system's output and the experts' mental models was suggested as future work. In another study, Drobny et al. (2009) used a transducer attached to learners' shoes to help them in learning rhythm while dancing. The study reported that some learners liked the system as it helped them to stay on the correct

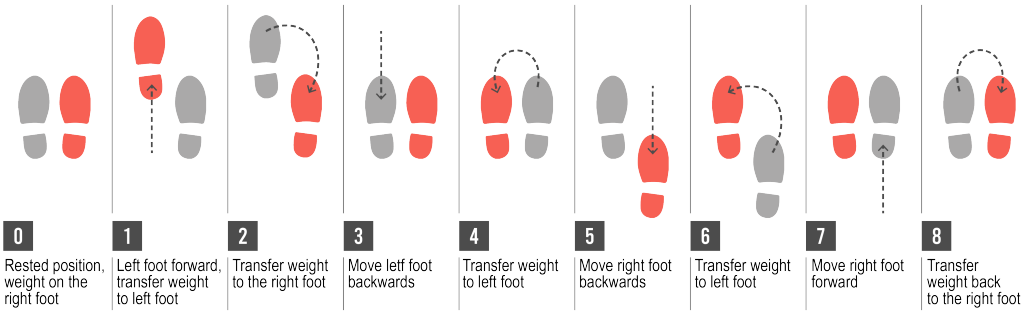

Figure 1. Dance notation for the Forró Básico 1 exercise. The red (darker colour) footprint represents the foot where the weight should be on the corresponding beat of the music.

beat. However, this approach was limited to identifying mistakes that happened on the strong beats of the song. Other aspects related to the development of rhythmic skills were not evaluated. In a more recent study Senecal et al. (2018) used a marker-based motion capture system to model rhythm in terms of style and drive. Rhythm was measured as following the tempo of the song (i.e., the speed of the underlying beat at which a piece of music is played). A question remains in regards to which other rhythmic skills beyond tempo can be monitored by using sensors.

Some authors have explored the provision of direct feedback to learners. While some have used sound to help learners to keep on the beat (Drobny et al., 2009; Dong et al., 2017), most researchers have used real-time visual cues for students to learn how to keep the rhythm (Romano et al., 2019; Trajkova and Cafaro, 2018), and *post-hoc* feedback in the form of charts or textual messages (dos Santos et al., 2017; Yang et al., 2013). Although these studies have provided important advances to the field of dance technologies, the actual benefits for learners in receiving the automated feedback were not evaluated. Moreover, none of these works have proposed to support dance teachers, who could use processed forms of rhythm and movement data to inform their pedagogical strategies.

Most of the work in dance/rhythm modelling has involved the use of multiple Inertial Measurement Units (IMUs) or accelerometer devices (Lee et al., 2007; Kawakami and Fujinami, 2008; Kikhia et al., 2014; Camurri et al., 2016a; Hinton-Lewis et al., 2016; Faridee et al., 2018). A few authors explored the use of other types of sensors. For example, Drobny et al. (2009) used force sensors in the sole of learners' shoes, while other authors have used whole-body computer-vision tracking (Kitsikidis et al., 2015; Kim et al., 2017; Romano et al., 2019) or infrared Motion Capture systems (Piana et al., 2016; Senecal et al., 2018). However, these systems are either expensive or hard to use in authentic dance settings.

In our previous work, we presented a proof-of-concept of an approach for extracting metrics that can identify whether the learner is in the correct tempo of the song (dos Santos et al., 2018b), similar to what other authors had attempted before (e.g. Drobny et al., 2009; Matsumura et al., 2011; Senecal et al., 2018). However, we identified that an algorithmic assessment of tempo alone could be misleading, as teachers commonly assess multiple motor skills when evaluating the learner's rhythm such as tempo, pause, step size and weight transfer. We also explored how that system could be used to provide automated feedback to dance learners using a visualisation dashboard (dos Santos et al., 2017). Results showed that learners were able to track their performance and gain some insights, but they did not know how to use the information to improve their performance.

In short, the work presented in this paper goes beyond previous studies by proposing an approach for building more comprehensive models of rhythm based on the ways dance teachers currently assess it, yet using only the sensors embedded in a single smartphone. The rationale for this is to facilitate the future adoption of the technology in dance classrooms, studios or at home. We also investigate how teachers would adopt and use the outputs from the rhythm models to enrich their assessments of learners.

3. Studies

Against the research gap identified in the previous section, we defined two main aims to guide the design and validation of an approach for modelling learners' rhythmic dance movements, using a single smartphone, with the purpose of enriching teachers' rhythmic assessments with data.

- Aim 1: Design and evaluate an approach to model from low-level motion/song data of a single smartphone to rhythm-related constructs that dance teachers understand.
- Aim 2: Investigate how teachers can use the outputs from the resulting models to enrich their assessments on learners' rhythm.

The next subsections present methodological details of studies 1 and 2 that address Aims 1 and 2, respectively. Study 1 is focused on evaluating the performance of classification models to be presented in Section 5. These models can then be used to process new motion data and generate automatic assessments that can be used to enrich teachers' assessments. Study 2 is focused on investigating how teachers react to outputs from these models while learners dance, with results presented in Section 6.

3.1. Study 1: Generating and Evaluating Music-related Motion Features

3.1.1. Participants

We recruited 16 participants (ten males and six females) to join as learners in a free individual private dance course consisting of three classes. Their ages ranged from 18 to 54 years. Two participants had no dancing experience, seven participants identified themselves as beginners, two were confident dancing simple steps, and five reported feeling confident dancing a variety of dance steps/styles. Nine participants did not practise dance regularly and eight of the participants had experience with the Forró style. Four dance teachers were invited to assess video-recordings of the learners dancing (details in next subsection). All teachers were teaching in Brazil (two males and two females) with a teaching experience ranging from 7 to 12 years (average of 10 years ($\pm$2.4)). Their age ranged from 26 to 28 years. Three teachers were full-time teachers, and one was part-time. The teachers did not know the learners in the videos.

3.1.2. Procedure

In each class, the learners were video-recorded individually performing the Básico 1 exercise (Fig.1) while wearing a smartphone on their left hip, placed in the same orientation (y-axis upright and screen facing outside the body), using a mobile phone band strap. In our study, we used the same model of mobile phone to enable comparison across models. It is also feasible to use different models/brands of smartphones provided that differences across devices regarding accelerometer readings are addressed using automated or user-driven calibration (Siirtola and Röning, 2018). Learners underwent 10 recordings per class, split into two blocks: five before and five after the class. Each recording was 1 minute long. In each block of recordings, the participant was asked to perform the same dance exercise using five different songs that had different paces. The pace of each song gradually increased from the first to the last song. This design ensured that learners were exposed to various levels of difficulty. All learners were recorded under the exact same conditions and all performed the five exercises with the songs played in the same order. After the end of the course, the learners' videos were sent to the dance teachers who assessed the performances.

A total of 480 videos (16 learners $\times$ 3 classes $\times$ 10 exercises) were recorded and tracked with the Forró Trainer mobile app (details to be provided in the next section). All sessions were pre-annotated by a researcher during the classes using one of the following labels: correct, tempo problem, pause problem, weight transfer problem and other problem. The list of rhythmic skills used in this paper is on aspects previously identified in the literature (dos Santos et al., 2018b; McCutchen, 2006). Based on this pre-annotation, a total of 70 videos were selected to be assessed by the teachers; 14 for each label to create a *balanced dataset*. Teachers annotated the videos remotely using an online video annotation tool, purposely built for this study, without any intervention or observations from the researchers (see annotation details below).

Assessment of learners' dance through video annotation. To relieve the assessment process load for the teacher, a batch of 30 videos was allocated to the four teachers to assess so that inter-rater agreement could be generated, and an additional batch of 10 unique videos was allocated to each teacher. This means that a total of 70 different videos were assessed by the dance teachers. Based on social dance literature (Schmidt and Wrisberg, 2008; McCutchen,

Table 1. List of parameters tuned for each classification method

Classification Method	Parameters Tuned
Neural Network	activation, solver, regularisation alpha
SVM	kernel, gamma constant, constant c0, degree
Random Forest	Number of trees, tree depth
Decision Tree	Number of instances in leaves, max depth
Logistic Regression	Regularization type, strength
Naive Bayes	none
kNN	k-Neighbours, metric, weight

2006) and our previous pilot design study in which Forró teachers were asked about the critical aspects they consider when assessing their students' dance progress (dos Santos et al., 2018a), we identified the following aspects of each learner's dance to be annotated by each teacher:

- Tempo: Slower, Correct, Faster
- Pause: Wrong Beat, Correct, No Pause
- Weight transfer: Too Few, Correct, Too Much
- Step size: Too Small, Normal, Too Large
- Other mistakes: Jumping, Stepping Strong, Hip Twist
- Comments section. 'Write here additional comments about this video if necessary'.

In such a study (dos Santos et al., 2018a), we found that, in some cases, teachers considered that the quality of a learner's performance can vary even for a 1 minute performance. As a result, in the present study teachers could label the performance of students several times while playing each 1 minute video. Yet, teachers were instructed for their annotations to refer to the whole 1-minute recording for each learner session. Thus, we considered the annotations that were repeated the most as the teacher's assessment of each dance aspect for the whole 1 minute video. Each of the 70 videos were associated to the corresponding motion and song's data using a unique ID.

3.1.3. Analysis

Using ML classification algorithms (Ravi et al., 2005; Kwapisz et al., 2011), teachers' annotations were used as labels and processed using STEP-generated features (details in Section 4). Classification algorithms were used with a combination of different parameters (inputs) to evaluate which features were more relevant during the classification. Seven types of classification methods were used to evaluate the features: Neural Network, kNN, Decision Tree, Support Vector Machine (SVM) Leaner, Random Forest, Naive Bayer and Logistic Regressions. Table 1 presents the parameters used when tuning each classification method. The Neural Network used is a Multi-layer Perceptron (MLP) with a single hidden layer and one added bias unit, no feature normalisation for the input and the weights are randomly initialised. The *Baseline* model assigns the most frequent class of a problem to all prediction cases. We used a stratified 5-fold cross-validation. This means that each fold contains roughly the same proportions of each class (Demšar et al., 2013).The reported accuracy of each model is calculated using the average accuracy across all validation folds. Due to the size of the dataset no test set was used to calculate the final accuracy.

The importance of the features in modelling each skill was evaluated using the feature selection metrics: gain ratio, Gini, X^2 and reliefF. To evaluate how powerful the features were for modelling each skill we used classification accuracy (CA) and F1 score as performance metrics (since F1 is more suitable for datasets with imbalanced classes, such as the one used in the present study). The McNemar's test (Dietterich, 1998; Keller et al., 2006) was used to identify the *best model*. The value of McNemar's test is zero if the models have the same errors in classifying the instances and returns higher values if the models have different error rates.

To better evaluate the power of classification of the STEP-generated features, other features used in previous research (Ahmadi et al., 2014; Gupta and Dallas, 2014; dos Santos et al., 2018b) were also extracted from the sensor data. The set of additional features were:

- {x,y,z}Min / {x,y,z}Max: the minimum and maximum raw acceleration values from each wave.
- {x,y,z}RawStd: standard deviation of the raw values of the waves {x,y,x} (Gupta and Dallas, 2014).
- {x,y,z}FilteredStd: standard deviation of the filtered values of the waves {x,y,x} (Gupta and Dallas, 2014).
- {x,y,z}MeanDiffAccWindowed: average value of the differences of min and max of the raw acceleration values in a time window of a 8-beat size (Gupta and Dallas, 2014).
- {x,y,z}SdDiffAccWindowed: averaged standard deviation values of the differences of min and max of the raw acceleration values in a time window of a 8-beat size.
- rmseRemainder / meanTrend: root mean square error of the remainder data and mean value of the trend data of the seasonal decomposition of time series by Loess (Cleveland et al., 1990).
- userBPM: User BPM based on (dos Santos et al., 2018b)
- 1-Move Score: User consistency based on (dos Santos et al., 2018b)
- bpmRatio: Ratio between userBPM and the song BPM (dos Santos et al., 2018b)

Several of the features were calculated over each of the three axes of the accelerometer data, and thus the axis name was included as a prefix of the features. For example, xK-Move score, yK-Move score, zK-Move score, xMin, yMin and zMin.

3.2. Study 2: Validating the Enhancement of Rhythm Assessment by Dance Teachers

3.2.1. Participants

Eight Forró teachers took part in this qualitative study. Four of these had participated in Study 1 (T1, T2, T3 and T4) and had been exposed to the four metrics used to enhance their assessment task. The other four had not seen the metrics before (T5, T6, T7 and T8). This mix of participants enabled us to both give teachers a voice in the design and evaluation process and also consider perspectives from new teachers. They were four males and four females, with teaching experience varying from 7 to 24 years and an average of 11.4 years (±5.9). Five of them were teaching in Brazil and three were teaching in other countries. Their age range was from 25 to 47 years, with an average of 30.2 years (±7.6). Seven teachers reported to be working exclusively as dance teachers.

3.2.2. Procedure

A total of 10 videos were taken from the dataset recorded in Study 1, each containing a learner performing the Básico 1 exercise for 1 minute. These 10 videos were different from the 70 videos used in Study 1 to avoid bias from participants who already participated in Study 1. The teachers were asked to watch and assess the first five videos (1 to 5) without any assistance and the last five videos (6 to 10) with the automated assessment information generated using STEP. The videos were selected based on the automatic assessment results. Videos 1 and 9 were instances of good performance; videos 2 and 8 presented *step size* problems; videos 3 and 7 presented *weight transfer* problems; videos 4 and 6 presented *pause* problems; and videos 5 and 10 presented *tempo* problems.

Semi-structured interviews were conducted with teachers using the following protocol:

1. The teacher assessed five learners using just the videos and answered the following questions for each video:
    - What is the evaluation for this learner?
    - What feedback would you give to this learner?
    - How easy/hard was it to assess this learner? Comment on this please.

2. The teacher assessed five learners using videos that contained information generated using the automatic assessment of each learner's performance and for each video they answered the same three questions as above.

3. The teacher was asked to share their experience using the rhythm metric-enhanced assessment information during the video assessment.

4. The teacher was asked to formulate hypothetical scenarios where they would see benefit from using the automatic assessment of rhythm.

3.2.3. Analysis

The data analysis followed an inductive approach to prioritise the practices of the participants in the generation of analytical themes, compared to imposing an existing coding system (Lazar et al., 2017a). Interviews were transcribed and processed with open coding of the teachers' answers using the questions to guide the creation of emerging categories. During axial coding, categories were connected to establish relationships between them. The resulting categories and their connections formed a conceptual map (Elo and Kyngäs, 2008; Lazar et al., 2017b). Resulting coded statements were examined by the authors who had several discussions to select instances that effectively illustrate how teachers of social dance can use the outputs from this modelling to enrich their assessments of rhythmic skills.

4. Approach

In any learning context where a teacher is assessing their learners, they basically require two pieces of information: a reference (guide or rubric) and the learner's performance (Hsia et al., 2016). The same applies to rhythm learning. The teacher looks for how well the learner is synchronised with the beat of the song. Then, the teacher uses the song as a reference and compares it to the learner's body movements (Erkert, 2003). The rationale behind the algorithm proposed in this paper also relies on this strategy. Figure 2 presents an overview of our proposed modelling approach, which directly addresses Aim 1. The input data for the modelling consists of motion data captured by a smartphone (i.e. x, y, z accelerometer waves) and features from the song playing while a learner is dancing (i.e. its beat rate and the positions of stronger beats). The STEP algorithm extracts music-related motion features (MMFs) from these data. These features are then used to model four rhythmic skills using predictive machine learning (ML) algorithms, based on human assessments. The following subsections describe the music-related motion features (MMFs), extracted from both the song and dance learners' movement data, and the STEP algorithm created for modelling dance rhythm.

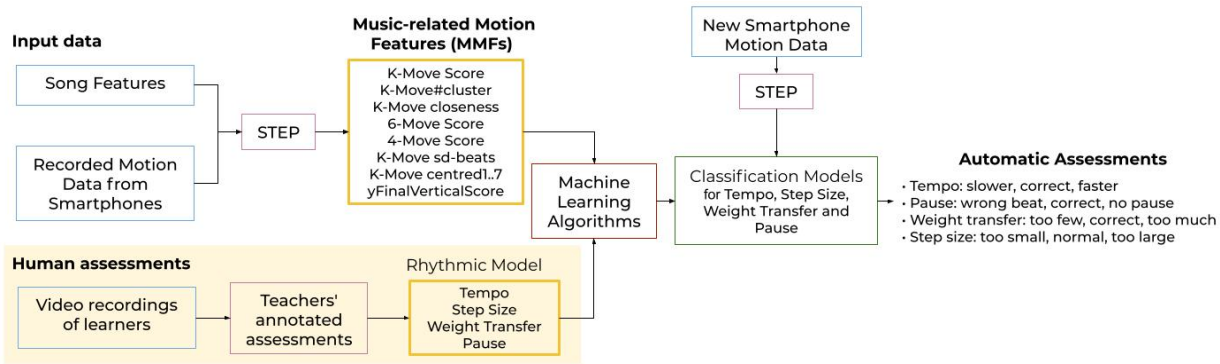

Figure 2. A Computational Model of Rhythm

4.1. Extracting Features from the Songs

The song is a valuable source of information to be used as a reference when people are dancing. Several elements of the song can be used by the dancer to guide their dance movements (e.g. the strong beats and beat rate of the song).These elements can be extracted from the song using music information retrieval algorithms (Böck et al., 2016; Lartillot and Toiviainen, 2007). Below, each song's feature will be described.

4.1.1. Beats per Minute (BPM)

An important element of dance is the music. The BPM feature extracts from the song its beat rate. This feature can be extracted by automated tools or by manual annotation. It is possible to compare the beat rate of the song to the beat rate of the learner's movements, and use this comparison to inform learners about whether their movements are in time, too slow or too fast. In this paper, the BPM was automatically extracted from the song using the Madmom algorithm (Böck et al., 2016).

4.1.2. Strong beat

In Forró, the strong beat of the song is the moment that the learner must start their movement. Most of the time, the strong beat matches the beginning of each bar in the song's score. This feature identifies from the song when each strong beat occurs. We need only two pieces of information to compute all the strong beats of the song; namely, the occurrence of the first strong beat and the period of time (in seconds) that the strong beat is repeated. This information can be used to assess whether learners are synchronised with the song's beat and derive other metrics as presented in the next subsection.

4.2. The STEP algorithm: Extracting Music-related Motion Features (MMF)

This subsection describes the algorithm and the features extracted, using the Básico 1 exercise for illustrative purposes. This exercise requires the learner to perform **six movements** during each 8-beat cycle of the song. This can be formulated as a pattern recognition problem, where the algorithm analyses the data obtained from the sensors in search of a pattern that contains six consecutive observations that equally repeat across the data stream. The algorithm uses the song's features (i.e., strong beats) as an additional source of data to compare the learner's movement with the song's pattern. The synchronisation of motion and song data is achieved using the internal clock of the smartphone which is playing the song. The algorithm first segments the learner's movement, using the strong beat cycle length as the window size. Then, it compares if the segments are similar or if the segment data vary. The algorithm presented does not combine the 3 axis into one signal. Instead, it uses features extracted from the 3 axis separately. The steps to extract the features are illustrated in Figure 3 and are described below:

1. The first step performed by STEP is to **identify the peaks and valleys** in the accelerometer data. The accelerometer data is processed with a low-pass filter (see Fig. 3a) and the local maxima formula was used to find the peaks/valleys having the window sizes defined by the song's beat interval length (Section 4.1.1). The window size is eight times shorter than the window size used in previous research (dos Santos et al., 2018b; Romano et al., 2019). The peaks/valleys represent moments of the movement where the learner was changing the acceleration polarity (Fig. 3b).

2. Next, **the motion wave is divided into segments** with a window size of 8 beats (Fig. 3c). Depending on the song, the window size will be different. If the dance was performed correctly, these intervals should contain six points (three peaks and three valleys), which represent six movements. These points are two-dimensional (time and acceleration).

3. Then, the time value of each point (peaks and valleys) in the segment **is converted** from being relative to the beginning of the song to a time value that is relative to the beginning of the segment it belongs to (Fig. 3d). The time dimension is then converted to beats so that the resulting information is the same disregarding the song, and could be compared with sessions from different songs. For example, the song *Beijo Bom - Trio Dona Zefa* has the first strong beat in the 1.3 seconds position and has a strong beat cycle length of 1.6 seconds where each beat is separated by 0.4 seconds. A peak that is detected at 3.2 seconds from the start of the song will be converted to 4.8 beats, to indicate that it occurred 4.8 beats after the start of the 1st segment ($(3.23 - 1.3)/0.4 = 4.82$). Figure 3d to 3i show the time dimensions (x axis) varying from 1 to 8 beats. By converting the values to a relative position, it is possible to compare the points across different segments.

4. Once the previous step is completed, every point from every segment has a time value between 0 and 8. STEP then **plots all the points together**. If the learner's movements were consistent, then the points corresponding to the same movement in the step would be close to each other in this 1-8 beats space (Fig. 3d).

5. At this stage, even though similar points are generally close to each other, the two-dimensional space cannot capture the continuous nature of the data. For example, a point on beat 8 should be close to a point on beat 1, but in this space they will be far apart. To avoid discontinuity in the data, the points are mapped in a three-dimensional space. Specifically, the points are mapped onto a three-dimensional cylinder, where the original acceleration value corresponds to the point's height, and the time value corresponds to the radial angle on the cylinder (Fig. 3e). Thus, points with the same acceleration that were closer to beat 0 or 8 occupy the same region in the new three-dimensional space.

6. In this cylindrical space, a **K-means clustering** algorithm was applied to split the points into groups (Fig. 3f and 3g).

7. The Coefficient of Variation (CoV) is calculated for each group (excluding outliers) and then computed the average CoV of all groups (Fig. 3h).

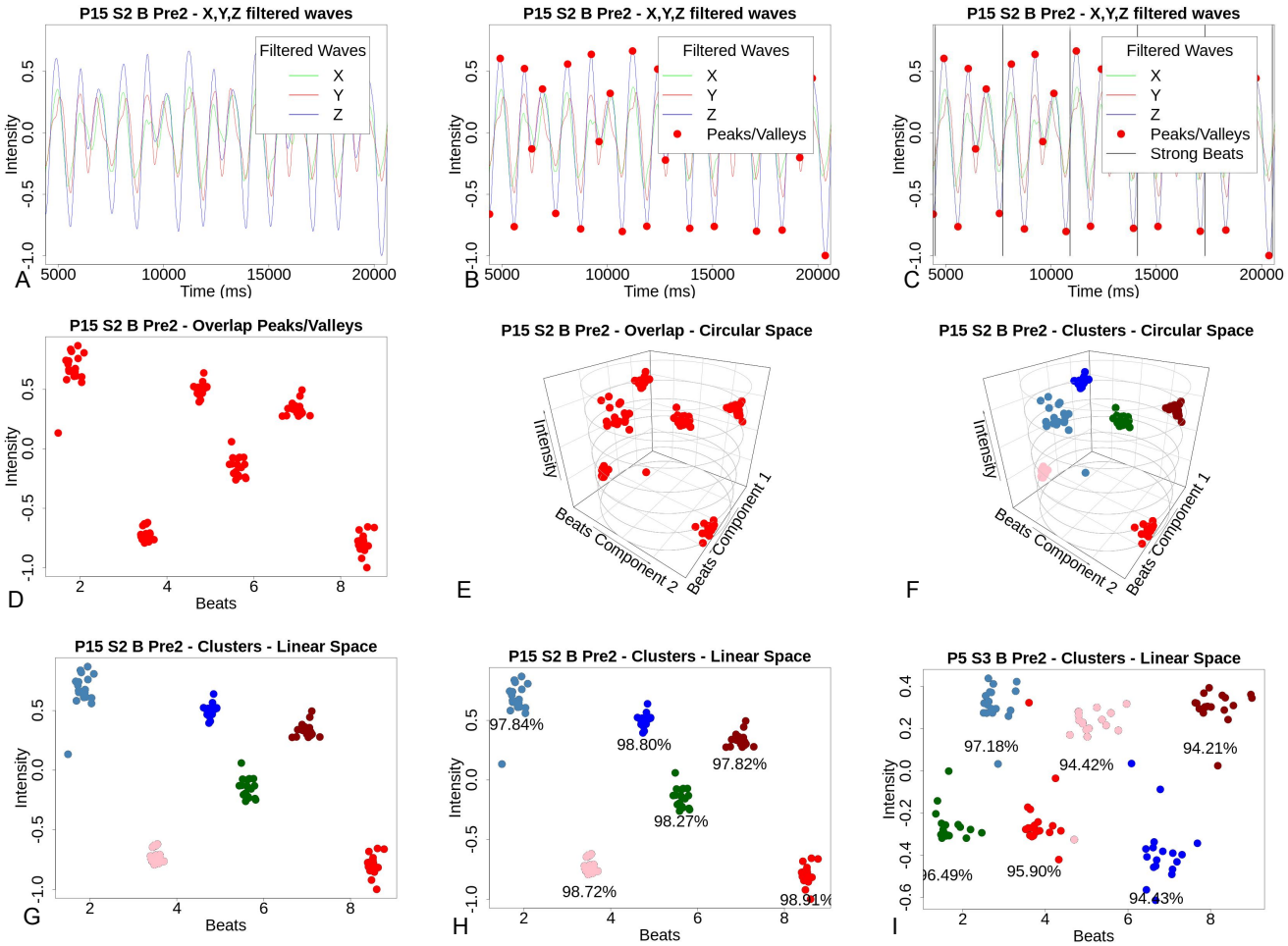

Figure 3. Workflow of the algorithm. a) Filtered accelerometer waves (X, Y, Z). b) Peaks/valleys identified in the motion data. c) Using the strong beat of the song to slice the acceletometer waves. d) Overlapping the wave slices in one 1–8-beat space. e) Creating a circular space to keep the time relation between beats 8 and 1. f) Using K-means to find the clusters. g) Clusters in the linear space. h) Retrieving the score of each cluster. i) An example of a low score attempt.

STEP Outputs. From this process, we extracted several features, all of which are averaged across the clusters:

- Cluster score: CoV of the beat's dimension (removing outliers)
- Cluster closeness (between-cluster sum of squares ÷ total sum of squares.)
- Standard deviation of the differences between the beat's dimensions and the closest beat
- Number of clusters (K of K-means)

4.2.1. K-Move Score

This feature models the number of steps the learner is doing. Students are meant to perform a 6-step movements but in some cases, they incorrectly perform 5-step or 4-step movements. The STEP algorithm uses K-means to identify the number of steps actually performed that appear in the accelerometer data as clusters. To select which n-step better describes the learners' movements, steps 6 and 7 (above) are preformed with different K numbers. **K is used as an additional feature** to classify different performance patterns the learners may have.

4.2.2. 6-move Score

This feature follows a similar strategy as that of the K-score, without the creation of a cyclical space to find the best cluster. Instead, the algorithm looks for six peaks/valleys in each window (of 8-beat size). The peaks/valleys are selected based on how close/apart they are from each other. They must be at least 1/6 of the window size apart from each other. The other steps to calculate the final value of the 6-move score are the same as that for the K-score, where the average CoV of the six groups of peaks/valleys was calculated. In the example of Básico 1, the rationale is that the 6-move score would be associated with good learners, who performed the Básico 1 using the six movements, which is the correct form for performing the exercise.

4.2.3. 4-move Score

Obtaining more information can help us derive features that are related to mistake patterns. If the learner does not properly transfer their weight at movements 1, 3, 5 and 7 of the Básico 1 exercise, the movement will appear to have a 4-step movement. By fixing the K value of K-means to 4, at the 6th step of K-score, we can assess whether the learner was performing the Básico 1 without doing the pause.

4.2.4. Move x beat Interval

For this feature, the movements of the learner, extracted from the accelerometer data are compared with the beats of the song. This is calculated comparing the timestamp from the peak/valley with the song's beat timestamp. Most body moves in dance movements should match the beats of the song. Thus, this feature provides an estimate as to whether the learner was synchronised with the song and, if not, where in the song the learner was out-of-sync.

4.2.5. Vertical Fluctuation

The y-axis of the accelerometer provided us with information on how much the learner fluctuated vertically when performing the movement. If the learner hit the ground too hard with their foot the y-axis would also be affected; therefore, we could identify how much weight the learner transferred into the ground between each step. Ideally, Forró dance learners should have a stable height (i.e., avoid jumping) when dancing and land smoothly on the floor with their feet.

The dataset and the algorithm developed are available in GitHub for replication purposes [1].

[1] The final link will be added in the final version of the article

5. Quantitative Results: Evaluation of STEP

Table 2 presents the accuracy of the ML algorithms used to model the four rhythm skills. The performance metrics used are Classification Accuracy (CA) and F1-measure. The results present the averaged performance of each ML algorithm across the multiple classes of each rhythm skill. The features were ordered four times based on the feature selection metrics: gain ratio, Gini, X^2 and reliefF. For each iteration, one metric was used to order the features, from best to worst, and the ML algorithm were tested from having 1 to 20 features in the model. Table 3 shows the rank of the features used in the best ML algorithm model for each rhythm skill.

Table 2. The average Classification Accuracy (CA) and F1-measure metrics, across the possible classes, for each ML algorithm used to model each rhythm skill. **Best models in bold**.

	Tempo		Pause		Step Size		Weight Transfer	
Method	CA	F1	CA	F1	CA	F1	CA	F1
Neural Network	**0.743**	**0.741**	**0.771**	**0.744**	0.686	0.682	0.614	0.577
SVM	0.729	0.714	0.7	0.673	**0.771**	**0.756**	0.6	0.583
Random Forest	0.629	0.632	0.643	0.632	0.714	0.680	0.614	0.597
Decision Tree	0.629	0.602	0.586	0.578	0.586	0.577	0.6	0.552
Logistic Regression	0.629	0.583	0.571	0.466	0.543	0.386	**0.671**	**0.629**
Naive Bayes	0.6	0.622	0.529	0.555	0.5	0.49	0.514	0.558
kNN	0.414	0.405	0.6	0.564	0.7	0.658	0.586	0.504
Baseline	*0.371*	*0.201*	*0.457*	*0.287*	*0.543*	*0.382*	*0.443*	*0.272*

Average of the **best models**: CA = 0.739 F1 = 0.719

5.1. Tempo

The possible mistakes the learner can make regarding tempo are to be *Slower* or *Faster* than the song. The best ML model for modelling tempo using four classes (*Slower*, *Correct*, *Faster* and *Mixed*) was the one described in the confusion matrix in Table 4a. The best model was trained using the Neural Network method (100 neurons, ReLu activation, Adam solver, Regularisation a=0.0008, 200 iterations) using the 12 first features as ordered by X^2. Table 3 presents the 12 features for modelling the tempo skill. The results from McNemar's test between the Neural Network model and the Baseline model was 17.45 ($p < 0.01$), which represents a great improvement when compared to the Baseline model. The average F1 score of the best ML model was 0.741, having an even balance between the average precision (0.741) and the average recall (0.743).

The features extracted by STEP appeared as the top eight features, demonstrating that the extraction algorithm is more robust than statistical features and features proposed in previous work. Table 2 presents the performance of the different ML models tested when using 12 features. The performance of a few models were between 0.6 and 0.7 when using from two to four features. This performance decreases when more features are added but increases again when the number of features reaches 10 or more. The top three models using the F1 score as a reference were the Neural Network with 12 features (0.741), kNN with two features (0.73) and SVM Learner with 12 features (0.714).

The tempo skill was the one that the teachers annotated the most mixed cases (12 in total). Two of those cases were flagged as mixed because there was no agreement in the teachers' annotations, and the other 10 cases were mixed because teachers annotated multiple times the same videos with different annotations. An example is presented in Figure 4, where T7, T6 and T15 annotated the video multiple times with distinct annotations. Teacher T12 only annotated once. The number of annotations for each tempo skill option were *Correct* (6), *Faster* (3) and *Slower* (3). Because of the complexity of such cases, the ML models could not accurately model mixed cases.

The features extracted by STEP added value to the model of the tempo skill as the best ML model had good score when modelling the four classes (F1 score = 0.741). Using more details of the accelerometer wave and the song's beat position as a reference to extract features proved to be very relevant when modelling the tempo skill of Forró dance learners.

Table 3. Top features used in the best models for each of the skills. *Best model used features ordered using X^2 metric. †Best model used features ordered using Gini metric

Skill	Tempo	Pause	Step Size	Weight Transfer
Best Model	Neural Network*	Neural Network*	SVM*	Logistic Regression†
Best Features	xK-Move closeness zK-Move centered1 zK-Move centered7 zK-Move closeness zK-Move score xK-Move centered7 xK-Move centered1 6-Move score 1-Move score bpmRatio yBPM zBPM	zK-Move centered1 zK-Move score zK-Move closeness xK-Move closeness zK-Move centered7 6-Move score	zMin zFilteredStd zRawStd zMax yFinalVerticalScore yFilteredStd zMeanSize ymeanDiffAccWindowed	yRawStd zK-Move centered1 zK-Move centered7 xK-Move closeness yMeanDiffAccWindowed zMax yFinalVerticalScore zK-Move closeness zK-Move score2 -zBPM

Table 4. Confusion matrices of the best ML models when using four classes. Comparing the teachers (Actual) assessment and the ML using STEP-generated features (Predicted). Colours indicate correctly classified instances (blue) and incorrect instances (red).

(a) Tempo skill

Actual \ Predicted	Slower	Correct	Faster	Mix.	Total
Slower	**9**	2	0	3	14
Correct	2	**22**	0	3	26
Faster	2	0	**16**	0	18
Mixed	3	2	2	**5**	12
Total	16	26	18	10	**70**

(b) Pause Skill

Actual \ Predicted	Wrong Beat	Correct	No Pause	Mix.	Total
Wrong B.	**15**	3	1	0	19
Correct	1	**29**	2	0	32
No Pause	3	1	**10**	0	14
Mixed	4	1	0	**0**	5
Total	23	34	13	0	**70**

(c) Step Size Skill

Actual \ Predicted	Too Small	Normal	Too Large	Mix.	Total
Too Small	**7**	2	0	0	9
Normal	1	**36**	1	0	38
Too Large	0	9	**8**	0	17
Mixed	1	1	1	**3**	6
Total	9	48	10	3	**70**

(d) Weight Transfer Skill

Actual \ Predicted	Too Few	Correct	Too Much	Mix.	Total
Too Few	**16**	4	1	0	21
Correct	2	**27**	2	0	31
Too Much	0	8	**4**	0	12
Mixed	2	4	0	0	6
Total	20	43	7	0	**70**

5.2. *Pause*

The pause skill refers to the ability of the student to pause or slow down their movement during movements 4 and 8 of the Básico. There are two types of common mistakes: 1) the learner does not pause at all during the Básico 1, or 2) the pause is at the wrong time, for instance, at movements 1 or 5. The pause skill was expected to have good results when modelled using the STEP algorithm, because the algorithm uses more information from the song's profile to match the accelerometer data. Table 3 contains the performance of the top six features for modelling pause, ordered by X^2. Selecting the top six features by their X^2 importance produced the best results for most classification algorithms. Table 2 presents the performance of the models. The Neural Network had the best performance and its resulting confusion matrix is presented in Table 4b. McNemar's test result between the Neural Network and the Baseline was 15.75 ($p < 0.01$), which represents a good improvement when compared to the Baseline model. The model presented

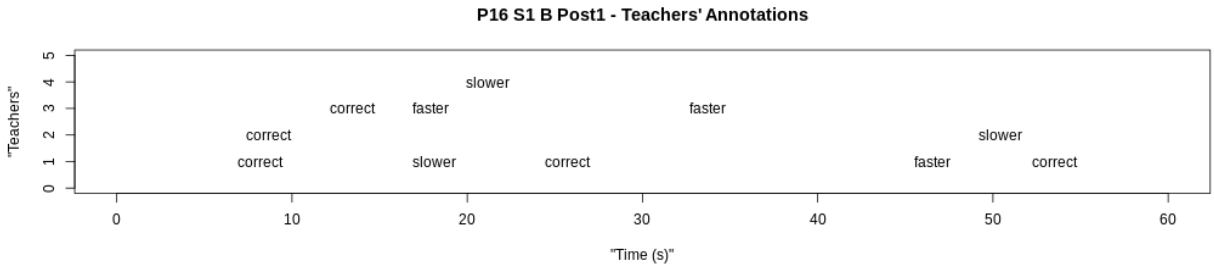

Figure 4. Tempo Mixed Case – Example of a learner case where teachers did multiple and different annotations in the same video. The Y-axis represents the teachers' annotators 1(T7), 2(T6), 3(T15 and 4(T12). The X-axis represents the time position in which the annotation was recorded.

an average F1 score of 0.744, which was very similar to the best ML model for the tempo skill (0.741), although the model had a higher average recall (0.771) compared to the average precision (0.721). The improvement occurred because the ML model classified correctly more instances in the pause skill than the previous ML model for the tempo skill. A possible explanation for this is that the most representative class, *Correct – Pause*, was annotated 32 times by teachers compared to 22 instances for *Correct – Tempo*. This imbalance of class tends to favour the ML model for pause to classify correctly more instances. As the teachers observed fewer pause skill mistakes than that of the tempo skill, it would be beneficial to validate this ML model in additional instances where learners had pause mistakes.

STEP-generated features were again essential for modelling the pause skill. The best ML model had a great result, even when modelling four classes. The best ML model required fewer features (six features) to achieve a good F1 score compared to the best ML model for tempo (12 features). This result is positive because the resultant model is less complex. Combining the detection of the pause skill with the tempo skill could provide a good advantage for learners and teachers to assess tempo problems.

5.3. Step Size

A common mistake for beginner students is stepping as far as when they are walking. Potentially, the accelerometer would record higher values when the learner has a long step size and lower values if the learner is doing too small a step. The step size skill had the best modelling performance among the four skills. The confusion matrix presented in Table 4c shows that the number of *Correct* cases was much higher than that of the other classes, which explained the model's higher performance. The best model was generated using SVM Learner (polynomial kernel, g=0.15, c=0.07, d=3.0) with the nine top features ranked by X^2. McNemar's test result between the SVM Learner model and the Baseline model was 11.25 ($p < 0.01$). The model had a F1 score of 0.756, having a higher average precision (0.787) than that of the average recall (0.771). Similar to the best ML for the pause skill, it is important to validate the best ML model for step size skill with other instances that contain step mistakes.

The top nine features are presented in Table 3. The top features demonstrate that the *z-axis* and the *y-axis* are the ones that most represent the step size skill. The *z-axis* is the one that records the back and forward movement of the Básico exercise, and the *y-axis* records the vertical fluctuations. The step size skill described how much variations existed in the z- and y-axes. Table 2 presents the rank of the different ML methods ordered by the F1 score.

The step size skill is a very important skill in Forró, especially because beginner learners commonly have long step sizes. Being able to correctly assess the step size skill is an important addition to the pool of skills assessed. Even though STEP-generated features were not relevant in modelling the step size skill, the results showed a positive indication towards creating systems that can support social partner dances.

5.4. Weight Transfer

Another skill that is hard for beginner students to learn is weight transfer. Related mistakes include 1) stepping too hard on the floor, transferring too much weight; and 2) transferring too little weight or not at all. Depending on how soft or hard the learner lands their feet on the ground, this is going to change the amplitude of the y-axis. Both y-axis features and STEP-generated features were ranked as the relevant features to model the weight transfer skill (Table 3). If the learner does not transfer their weight properly, then they will not generate a wave pattern similar to

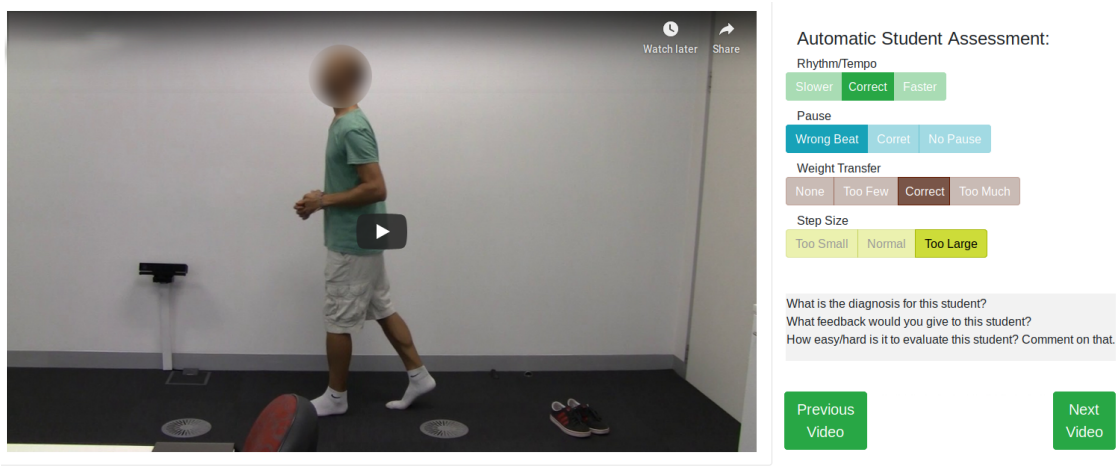

Figure 5. Tool presenting the learner's video performance together with the enhanced assessment. Labels in bold colour represent the automatic assessment.

the one presented in Figure 3a. This skill was the one with the least cases correctly classified (47 out of 70 cases). Even though the models for weight transfer had lower performance compared to the models of the other skills, the best ML model had better performance than that of the Baseline model. The performance of the models is presented in Table 2.

The best ML model for the weight transfer skill was the one using Logistic Regression with the top 10 features as ranked by Gini. The resulting confusing matrix is presented in Table 4d. McNemar's test result between the Logistic Regression model and the Baseline model was 9.375 ($p < 0.01$), the lowest among the four skills. The model had a F1 score of 0.629, having a higher recall (0.671) than precision (0.616). This indicates that further analysis of the motion data is required to better model the weight transfer skill. According to the ranking of the features (Table 3), there is an indication that the y-axis can be a good starting point for extracting new features. STEP focused on extracting features from the Básico movement characteristics that are more related to the z- and x-axes.

The relevance of STEP-generated features was tested using ML models over the four skills related to rhythm learning. Out of the 70 cases annotated by teachers, 27 were classified correctly by the ML models in all four skills and 43 were classified incorrectly in at least one skill. A total of 10 of the 70 cases were misclassified in two skills at the same time, and nine cases were incorrectly classified in three skills at the same time.

McNemar's test results of the best models compared with the Baseline models were the following: rhythm (17.45), pause (15.75), step size (11.25) and weight transfer (9.375), all with statistical significance ($p < 0.01$). Rhythm had the best performance compared to the Baseline model and weight transfer had the worst. All the ML models, the best model for each skill, were better than the Baseline model. However, the weight transfer model had an F1 score that was less than 0.7. Therefore, the model should be used with caution to predict the weight transfer skills of the learners. The models of the other skills had F1 scores greater than 0.74; therefore, these models were more reliable and could be used by teachers to support the assessment of their learners.

6. Qualitative Results: Enriching Dance Teachers' Assessments

This section presents results from Study 2 which aimed at understanding how teachers can benefit from the automatic assessment of rhythm.

The interviews included four parts: *i*) the assessment of five videos without any support, *ii*) the assessment of a different set of five videos with the automated assessment , *iii* a discussion about the automatic assessment , and iv) a discussion about how to embed this technology into teachers' practice.

We provided the teachers with a web-based annotation tool to complete parts i and ii (see Figure 5).A video of the student is presented at the left of the interface. A list of buttons at the right allows the teacher to tag the video with preset labels. Each skill is represented through a group of buttons with a colour set. The options for each skill are the

following: Rhythm-Tempo; Pause; Weight Transfer; and Step Size. For videos that have been automatically assessed, the respective labels are represented already by the buttons. At the bottom-right, a large text box is available for the teacher if they want to give an overall feedback for the student's performance. After assessing each video, teachers used the green button, located at the bottom of the interface, to go to the next video. Each interview took an average of 1 hour and 53 minutes (±27 minutes).

6.1. How do Teachers Benefit from Automatic Assessment when Assessing Learners' Video Performance?

Regarding the automatic assessment results, teachers mentioned that these helped to structure their assessment (T3, T6 and T8), as a confirmation/comparison of what they had already perceived (T1, T2 and T4) and as a pre-assessment that helped them imagine what the learner's performance would be like (T4 and T5). T4 mentioned,*"The videos I saw the automatic evaluation before, I already started the video wondering how the learner would dance. The ones I saw the automatic evaluation after, it was more of a confirmation."* Teachers found the automatic assessment very accurate, reporting that the automatic assessment had only 1–3 errors out of 20 items assessed (T2, T4, T5, T7 and T8). Teachers reported that they would increase their trust in the automatic assessment if they were more experienced and had more exposure to it (T1, T5 and T8) or if the system was trained solely with their own data (T2). T5 said, *"There would have to be more videos with more results for me to fully rely on the automatic evaluation"*. T2 demonstrated a deep understanding of how the technology worked - models are trained with annotated data. T2 commented, *"Each teacher will have a different way of evaluating, so if only I did the evaluations and after that it was automatic in my data, then it would be based on my reality"*.

In summary, the first benefit that the teachers perceived was that the automated assessment created a structured way to assess their learners. Most teachers did not use any tools to support their teaching practice. Also, it was shown that teachers required more time to understand the automatic assessment and how to obtain benefits from it. Even though the teachers had limited experience with the use of technology, they did have some ideas on how automatic assessment could be used in their context as dance teacher, which is explained in the following section.

6.2. Hypothetical Scenarios for Using Automatic Assessment

Teachers were asked to imagine scenarios in which they could see the benefit of using the automatic assessment, in their context, to assess their own learners. The scenarios that teachers imagined can be divided into the following groups: **time**: use it before, after or during classes; **scale**: for private, small or large classes; **proximity**: regular, workshop or distance learners; and **access**: to teachers, learners or the school. Most teachers referred to more than one of these four aspects when imagining the scenarios. All teachers reacted positively to thinking about how an automated assessment could improve their teaching practices.

Timing – when to use the assessment information. Before the class, learners could practice with the automatic assessment to improve their skills (T1) and, based on the assessment results, be more aware on which skills they needed to focus when taking the class (T4). T4 said, *"Using the automatic assessment before class is interesting because the learner would already have a sense of their learning stage, be aware of their difficulties and of interesting aspects to practice in that class. He will be more attentive, seeking that knowledge because he knows where his difficulty is"*. Some teachers said they would use an assessment tool during the classes to let learners practise by themselves while they gave individual feedback to other learners (T4 and T8). The main benefit that teachers imagined was to allow learners to practise as homework and receive automated results (T1, T2, T4, T5 and T7). T5 commented: *"As a form of homework it could be good. For the learners to get feedback a little faster, without having the help of the teacher"*. T4 and T5 mentioned the need to allow learners to view the teacher's feedback when practising by themselves so they could recall the teacher's recommendation as the automated assessment would not suffice to provide a clear message. For the teachers, the assessment information could also be used to assess the effectiveness of the class plan (T4 and T8) and the learners' overall performance and difficulties encountered (T4). T4 commented: *"This [automated assessment] turns out to be a feedback to the teacher on what he is teaching effectively or not. It is a guide to what learners are absorbing"*.

Addressing the scale of the class. Teachers were unanimous in saying that large cohorts were the scenario in which the automatic assessment would be the most useful. T1 said: *"Some schools have so many learners that it is difficult, even with a large team, to give feedback to each learner. But if each learner has the app he can do his own self-assessment"*. The automatic assessment could guide and speed up the learners' assessment by the teachers, and learners would have access to a pre-assessment before talking with the teacher (T1, T4, T5 and T8). T4 mentioned: *"With this type of software the learner can already have some idea of his performance until the teacher arrives to give feedback"*. Accessing aggregated information of learners was mentioned by teachers (T8), which could be used also to assess learners from private classes (T4).

Knowing more about learners current state and progress. Teachers reported that they were more interested in using the automatic assessment for regular learners as they would be continuously working with them and would be able to work on the feedback for these learners (T1, T4 and T8). T8 said: *"I would use it more with regular learners because I will have time to work on that information with the learner"*. Some teachers mentioned that, for workshop learners who they usually do not know, the automatic assessment could help allocate the learners to the most appropriate workshop level, matching the learners' assessment and the workshop's level of difficulty (T4 and T5). Additionally, T6 mentioned that this information would be very useful for distance learners. T6 said: *"For those who are far away I think it is one of the most interesting scenarios. Because you have at least a technical assessment of the learner's dance in these items. This already facilitates strategies for how to improve"*.

Access to the assessment information. Learners, teachers and schools could benefit from accessing the automatic assessment information. For the learners, teachers reported that this information would allow learners to recall their progress, recognise this evolution and be more aware of their dance learning (T1, T2, T4, T5 and T7). T1 mentioned: *"The learner starts today and in three months he has progressed but he no longer remembers what he was like, the difficulties he had. With this information he will become aware of his own dance skills"*. Additionally, learners would be more independent and more motivated to undertake their learning as they would spend more time engaged with the learning content (T1 and T4). In the same way that the teachers filtered their feedback to learners, giving limited and precise information when compared to their complete and holistic evaluation, teachers mentioned that the automatic assessment could also be filtered based on the needs of the learners or their stage of development (T4). T1 wondered how this new tool would change the relationship between teachers and learners. T1 commented: *"This can lead to a closer relation between teachers and learners. It shows a greater interest from teachers wanting to take better care of their learners"*. Additional studies are required to assess how the automatic assessment would change the relationship between teachers and learners and other aspects of the dance learning. T2 also mentioned the automatic assessment information could be used for learner competition or comparison with other learners from the same class or other places, increase the interaction among learners, and thus creating a social network where the learners would not feel alone when practising.

For teachers, the automatic assessment information could help them to improve the assessment of the learners' evolution (T1 and T4), be more efficient (T1 and T5) and plan and evaluate the teachers' work (T2, T4 and T5). T1 said, *"It would make us teachers have a more effective job in monitoring the learner's evolution"*. Teachers also mentioned that this tool could allow teachers pre-assess beginner learners and help decide on their class levels (T2 and T8). T8 mentioned: *"For example, when I'm going to receive a new learner who says he doesn't know how to dance, this can be a way to pre-evaluate my learner. If he has a serious rhythm problem he can only go up to level 2"*.

For the school, the automatic assessment could be a resource to assess the learners' progress, the overall development of the class, the teachers' performances and the efficiency of teaching methodologies (T4 and T7). Some teachers mentioned that it could help to train and support teachers at the beginning of their teaching career or help teachers that have difficulty assessing learners' development (T1 and T4). It would also demonstrate that the school is progressive compared to other school (T3 and T6). T6 commented: *"This is a way of showing that your school is modern, that it uses technology"*.

7. Discussion and Conclusion

The ultimate purpose of this work is to augment and enrich dance teachers' assessment of dance skills related to rhythm, using motion sensors from smartphones. Although the results from Study 1 cannot directly support teaching practices, the findings demonstrated that it is indeed possible to extract rhythmic information from a smartphone, with expected caveats. In Study 2 we rendered visible the outputs from the models through a teacher-facing interface.

From this study, we learnt that such information collected can spark teachers' ideas about potential pedagogical uses in their actual educational contexts. However, much care must be taken when using technology to model dance, as an algorithmic approach to assessing dance skills and qualities is not as complete as a human expert evaluating a dancer's performance. This section further discusses the findings of the studies and the main lessons learnt.

7.1. Accuracy of Models and Challenges in Modelling Rhythm

The results from Study 1 showed that STEP-generated features were more relevant when modelling Tempo, Pause and Weight Transfer skills, but not for Step Size skill. The best machine learning models for each skill had the following F1 score: Tempo (0.741), Pause (0.744), Step Size (0.756) and Weight Transfer (0.629). However, McNemar's test shows that the Tempo model was most effective in reducing the model's error when compared with the Baseline model. Compared to previous studies in which multiple motion sensors were used in alternative dance contexts (e.g., Kuhn et al., 2011; Faridee et al., 2018, reporting accuracies of 0.89 and 0.94, respectively), and as a new field of study with not many classifiers of this type, an accuracy close to 0.80 using a single motion sensor can be considered acceptable for the ultimate purpose of supporting teachers' assessments instead of replacing them.

Nonetheless, teachers had different opinions when assessing the learners. We identified at least the following two main challenges in attempting to model rhythm in dance.

Disagreement among teachers. Forró is a social practice and art thus, the interpretations of musicality and rhythm are closely linked to the cultural values of the practitioners (Trehub et al., 2015). In our study, this is reflected in the occasionally divergent opinions of teachers on learners' performance as each has exposure to different styles of the Forró dance. It is therefore expected that teachers' interpretations will not be all aligned.

In Study 1, the evidence shows that dance teachers had different interpretations about the correct tempo, weight transfer, pause and step size. For instance, teachers' inter-rater agreement on learners' performance varied from 0.167 (Kappa Fleiss) to 0.75. These different interpretations may be related to the personal teachers' expectations, culture and experiences. Teachers' personal beliefs may influence systems that use the teacher's opinions to train models of dance movements, skills and qualities. There is a growing body of research seeking to understand and prevent biases in machine learning (ML) and artificial intelligence (AI) (Zafar et al., 2017; Angwin et al., 2016; Saxena et al., 2019). Systems can potentially discriminate people based on age, gender, race and socioeconomic status. The studies reported in this paper sought to minimise such problems by evenly recruiting teacher annotators across gender and geography. Also, learners had equal opportunities to participate in the data collection, disregarding age, gender and race. However, this intrinsic issue of disagreement among teachers may also suggest the need for alternative modelling approaches that could create personalised models for each teacher to match their individual ways to assess learners. Although an analysis of the reasons behind the differences between teachers interpretations goes beyond the scope of this paper, we recognise this is worth investigating in a further qualitative study.

Interpretability. In this paper, the decision of selecting certain features from the motion sensor data and from the songs was informed by the characteristics of the context (the song – eight-beat cycle and the Básico movement – six movements) instead of using statistical features as in other research that uses motion sensor data. This strategy follows recent research that promotes the creation of ML and AI algorithms that can be explained and interpreted (Wang et al., 2019; Guidotti et al., 2018; Miller, 2018). For instance, deep learning algorithms are becoming increasingly popular for their superior performance when compared to traditional algorithms, but they can create black box models that prevent people from understanding how the system made decisions (Došilović et al., 2018). In our approach, features were selected based on foundations of dance: this enable the translation of the outputs into user interfaces with a vocabulary that teachers can understand since the motion data are connected to the learners' rhythmic skills.

7.2. Implications for Dance Education and Other Fields

In Study 2, the automatic assessment helped teachers to guide their assessment, confirm their assumptions and anticipate learners' performances. The study also reported that teachers could imagine several situations within their current teaching strategies that would benefit from the use of automatic rhythm assessment. Study 2 confirmed that the educational context of teaching to large cohorts is where teachers see more benefits of using the automatic assessment. The results are novel in the field of ubiquitous computing and dance technology, as this is the first research study that we are aware of that has attempted to use motion sensing technology to improve dance teachers' activities. However, technology has been successfully used to improve teaching in other educational scenarios such as tertiary education.

For instance, technology is already used for learners to practise and receive feedback outside the classroom (Ihantola et al., 2010; McBroom et al., 2016), help teachers to improve their teaching practices (Sergis and Sampson, 2017), and help teachers to cope with large cohorts (Saunders and Hutt, 2015). It is expected that dance teachers could receive similar benefits if strategies used to support teachers from tertiary education were combined with motion data from dance learners. Study 2 highlighted the fact that teachers would like to use the tools they interacted with to train and support beginner teachers, as well as to compare and evaluate their teaching methodologies. This points at the potential of developing professional development strategies that can support communities of dance practice to train teachers based on data.

As discussed in Section 2, other fields of research also use sensors to model human movement and use this information to generate information not available before. Fields such as martial arts (Corbí and Santos, 2018), musical instruments (Hadjakos et al., 2008), sports (Hassan et al., 2017), handwriting (Amma et al., 2014), cooking (Stein and McKenna, 2013) and virtual reality (Hu et al., 2016). The strategies used in this paper, to derive motion features using the context information, could also be useful in those fields to improve the modelling of human movement. More precisely, the rhythmic features developed in this paper can be directly used in fields such as sports and musical instruments learning. Rhythm is fundamental for sports motor skills performance (MacPherson et al., 2009).

Additionally, many other partner dance styles share the same basic components of rhythm captured in the conceptual model, and can be considered permutations of the Básico 1 dance exercise that we studied. The corresponding computational model can thus be used for each style's dance step, provided the style has a 6-step movement in 8 beats of the song; otherwise the model parameters can be adapted to suit the movement pattern of the dance style.

7.3. Limitations

It is important to state, and it was observed by teachers during the studies, that learners were being assessed while dancing alone (T2, T7) and they were performing only the Básico step (T1). T1 said, *"It is difficult to give an evaluation only within the base"* and T2 added, *"I could not assess the correct posture as he is practising the step individually"*. It is a common strategy found in the social partner dance literature (Jarmolow and Selck, 2011; Flippin, 2013) and for Forró teachers to ask learners to practise individually before partnering up. Although they were performing an exercise used in dance classes, Básico 1, the results given by the technology and the assessment of the teachers could differ if participants were dancing in pairs. Also, the studies were carried out in a university room, which might also give different results than in a dance classroom environment. Another teacher mentioned that it was their first time assessing the learners from videos, in comparison to assessing their learners in the classroom (T7). T7 mentioned, *"It is difficult to assess other learners that are not from my classroom"*.

The BPM of the songs used in this study (all of them above 140 BPM) enables for the y-axis to be used to identify impact force for the weight transfer skill. However, in extremely slow songs (i.e, lower than 140 BPM), the impact force would not create an abrupt change in the signal enough to be detected. This is a limitation of this approach that could be addressed by applying additional filters to the raw signal to identify subtle changes in the body weight transfer.

7.4. Final Remarks and Future Work

The contribution of this paper is a modelling approach comprising a computational model of dance rhythm that includes four components: tempo, pause, step size and weight transfer. The model is based on information extracted from accelerometer motion data of a single smartphone. Our model of rhythm goes beyond existing computational models of dance rhythm, as it involves not only the *tempo* aspect of rhythm, but also pause, step size and weight transfer; all important elements of partner dancing. Additionally, we presented a thorough exploration of the different scenarios in which STEP can enhance dance teachers' rhythmic assessment. The novelty of this research and the rapid development of technology enables a number of possible directions for future work. These include 1) revising teachers' discrepancies in their judgements, which could contribute to improve the accuracy of our current models; 2) validating the current technology in a dance classroom and extending the model for two people dancing together to include other features that could model particular aspects of partner dance, such as synchrony; and 3) using unsupervised machine learning techniques (e.g., hidden-markov models) on the raw signals could be further explored to identify other features from the accelerometer signals. Finally, investigating ways in which feedback can be provided directly to learners is a recommended next step in this line of work.

Declarations of interest

None.

Acknowledgements

This work was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil) under Grant No.: 207539/2014-6, and UFRGS (Universidade Federal do Rio Grande do Sul). Roberto Martinez-Maldonado's research is partly funded by Jacobs Foundation.

References

Ahmadi, A., Mitchell, E., Destelle, F., Gowing, M., OConnor, N.E., Richter, C., Moran, K., 2014. Automatic activity classification and movement assessment during a sports training session using wearable inertial sensors, in: 2014 11th International Conference on Wearable and Implantable Body Sensor Networks, IEEE. pp. 98–103.

Amma, C., Georgi, M., Schultz, T., 2014. Airwriting: a wearable handwriting recognition system. Personal and ubiquitous computing 18, 191–203.

Angwin, J., Larson, J., Mattu, S., Kirchner, L., 2016. Machine bias. ProPublica, May 23.

Böck, S., Korzeniowski, F., Schlüter, J., Krebs, F., Widmer, G., 2016. Madmom: A new python audio and music signal processing library, in: Proceedings of the 24th ACM international conference on Multimedia, ACM. pp. 1174–1178.

Camurri, A., Canepa, C., Ferrari, N., Mancini, M., Niewiadomski, R., Piana, S., Volpe, G., Matos, J.M., Palacio, P., Romero, M., 2016a. A system to support the learning of movement qualities in dance: A case study on dynamic symmetry, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, ACM, New York, NY, USA. pp. 973–976. doi:10.1145/2968219.2968261.

Camurri, A., El Raheb, K., Even-Zohar, O., Ioannidis, Y., Markatzi, A., Matos, J.M., Morley-Fletcher, E., Palacio, P., Romero, M., Sarti, A., Di Pietro, S., Viro, V., Whatley, S., 2016b. Wholodance: Towards a methodology for selecting motion capture data across different dance learning practice, in: Proceedings of the 3rd International Symposium on Movement and Computing, ACM, New York, NY, USA. pp. 43:1–43:2. doi:10.1145/2948910.2948912.

Chan, J.C., Leung, H., Tang, J.K., Komura, T., 2011. A virtual reality dance training system using motion capture technology. IEEE Transactions on Learning Technologies 4, 187–195.

Chatzopoulos, D., Doganis, G., Kollias, I., 2019. Effects of creative dance on proprioception, rhythm and balance of preschool children. Early Child Development and Care 189, 1943–1953. doi:10.1080/03004430.2017.1423484.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. Stl: A seasonal-trend decomposition. Journal of official statistics 6, 3–73.

Corbí, A., Santos, O.C., 2018. Myshikko: Modelling knee walking in aikido practice, in: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization, ACM. pp. 217–218.

Côté-Laurence, P., 2000. The role of rhythm in ballet training. Research in Dance Education 1, 173–191.

DeMers, J.D., 2013. Frame matching and δp t ed: a framework for teaching swing and blues dance partner connection. Research in Dance Education 14, 71–80.

Demšar, J., Curk, T., Erjavec, A., Črt Gorup, Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., Zupan, B., 2013. Orange: Data mining toolbox in python. Journal of Machine Learning Research 14, 2349–2353.

Dietterich, T.G., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural computation 10, 1895–1923.

Dong, Y., Liu, J., Chen, Y., Lee, W.Y., 2017. Salsaasst: Beat counting system empowered by mobile devices to assist salsa dancers, in: 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), IEEE. pp. 81–89.

Došilović, F.K., Brčić, M., Hlupić, N., 2018. Explainable artificial intelligence: A survey, in: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 0210–0215. doi:10.23919/MIPRO.2018.8400040.

Drobny, D., Weiss, M., Borchers, J., 2009. Saltate!: a sensor-based system to support dance beginners, in: CHI'09 Extended Abstracts on Human Factors in Computing Systems, ACM. pp. 3943–3948.

Elo, S., Kyngäs, H., 2008. The qualitative content analysis process. Journal of advanced nursing 62, 107–115.

Erkert, J., 2003. Musicality and rhythm, in: Erkert, J. (Ed.), Harnessing the wind: The art of teaching modern dance. Human Kinetics, Champaign, IL. chapter 6, pp. 266–290.

Faridee, A.Z.M., Ramamurthy, S.R., Hossain, H., Roy, N., 2018. Happyfeet: Recognizing and assessing dance on the floor, in: Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications, ACM. pp. 49–54.

Fernandes, A., 2005. Music, Migrancy, and Modernity: A Study of Brazilian Forró. Ph.D. thesis. University of Illinois at Urbana-Champaign.

Flippin, L.L., 2013. Salsa remixed: Learning language, culture, and identity in the classroom. Working Papers in Educational Linguistics (WPEL) 28, 77–91.

Fraisse, P., 1982. Rhythm and tempo. The psychology of music 1, 149–180.

Goldstein, J., 2007. Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review. American Journal of Education 113, 479–508.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., 2018. A survey of methods for explaining black box models. ACM Comput. Surv. 51. doi:10.1145/3236009.

Gupta, P., Dallas, T., 2014. Feature selection and activity recognition system using a single triaxial accelerometer. IEEE Transactions on Biomedical Engineering 61, 1780–1786.

Hadjakos, A., Aitenbichler, E., Mühlhäuser, M., 2008. Syssomo: A pedagogical tool for analyzing movement variants between different pianists, in: Proceedings of the 5th International Conference on Enactive Interfaces, Citeseer. pp. 75–80.

Hassan, M., Daiber, F., Wiehr, F., Kosmalla, F., Krüger, A., 2017. Footstriker: An ems-based foot strike assistant for running. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1.

Hinton-Lewis, C.W., McDonough, E., Moyle, G.M., Thiel, D.V., 2016. An assessment of postural sway in ballet dancers during first position, relevé and sauté with accelerometers. Procedia Engineering 147, 127–132.

Hsia, L.H., Huang, I., Hwang, G.J., 2016. Effects of different online peer-feedback approaches on students' performance skills, motivation and self-efficacy in a dance course. Computers & Education 96, 55–71. doi:10.1016/j.compedu.2016.02.004.

Hu, F., Hao, Q., Sun, Q., Cao, X., Ma, R., Zhang, T., Patil, Y., Lu, J., 2016. Cyberphysical system with virtual reality for intelligent motion recognition and training. IEEE Transactions on Systems, Man, and Cybernetics: Systems 47, 347–363.

Ihantola, P., Ahoniemi, T., Karavirta, V., Seppälä, O., 2010. Review of recent systems for automatic assessment of programming assignments, in: Proceedings of the 10th Koli calling international conference on computing education research, ACM. pp. 86–93.

Jarmolow, D., Selck, B., 2011. Teach Like a Pro: The Ultimate Guide for Ballroom Dance Instructors. Ballroom Dance Teachers College.

Kawakami, M., Fujinami, T., 2008. On the effect of performance evaluation in acquiring samba rhythm, in: 10th Internatinal Conference on Music Perception and Cognition, pp. 132–139.

Keller, M., Bengio, S., Wong, S.Y., 2006. Benchmarking non-parametric statistical tests, in: Advances in neural information processing systems, pp. 651–658.

Kikhia, B., Gomez, M., Jiménez, L.L., Hallberg, J., Karvonen, N., Synnes, K., 2014. Analyzing body movements within the laban effort framework using a single accelerometer. Sensors 14, 5725–5741.

Kim, D., Kim, D.H., Kwak, K.C., 2017. Classification of k-pop dance movements based on skeleton information obtained by a kinect sensor. Sensors 17. doi:10.3390/s17061261.

Kitsikidis, A., Dimitropoulos, K., Uğurca, D., Bayçay, C., Yilmaz, E., Tsalakanidou, F., Douka, S., Grammalidis, N., 2015. A game-like application for dance learning using a natural human computer interface, in: International Conference on Universal Access in Human-Computer Interaction, Springer. pp. 472–482.

Kuhn, M., Wattenhofer, R., Wirz, M., Fluckiger, M., Troster, G., 2011. Sensing dance engagement for collaborative music control, in: 2011 15th Annual International Symposium on Wearable Computers, IEEE. pp. 51–54.

Kwapisz, J.R., Weiss, G.M., Moore, S.A., 2011. Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter 12, 74–82.

Lartillot, O., Toiviainen, P., 2007. A matlab toolbox for musical feature extraction from audio, in: International conference on digital audio effects, Bordeaux, FR. pp. 237–244.

Lazar, J., Feng, J.H., Hochheiser, H., 2017a. Chapter 11 - analyzing qualitative data, in: Lazar, J., Feng, J.H., Hochheiser, H. (Eds.), Research Methods in Human Computer Interaction (Second Edition). second edition ed.. Morgan Kaufmann, Boston, pp. 299 – 327. doi:10.1016/B978-0-12-805390-4.00011-X.

Lazar, J., Feng, J.H., Hochheiser, H., 2017b. Research methods in human-computer interaction. Morgan Kaufmann.

Lee, E., Enke, U., Borchers, J., de Jong, L., 2007. Towards rhythmic analysis of human motion using acceleration-onset times, in: Proceedings of the 7th International Conference on New Interfaces for Musical Expression, ACM, New York, NY, USA. pp. 136–141. doi:10.1145/1279740.1279765.

MacPherson, A.C., Collins, D., Obhi, S.S., 2009. The importance of temporal structure and rhythm for the optimum performance of motor skills: A new focus for practitioners of sport psychology. Journal of Applied Sport Psychology 21, S48–S61.

Matsumura, K., Yamamoto, T., Fujinami, T., 2011. The role of body movement in learning to play the shaker to a samba rhythm: An exploratory study. Research Studies in Music Education 33, 31–45. doi:10.1177/1321103x11400513.

McBroom, J., Jeffries, B., Koprinska, I., Yacef, K., 2016. Mining behaviours of students in autograding submission system logs. International Educational Data Mining Society .

McCutchen, B.P., 2006. Dancing and performing, in: McCutchen, B.P. (Ed.), Teaching dance as art in education. Human Kinetics, Champaign, IL. chapter 6, pp. 125–169.

Miller, T., 2018. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence .

Ochi, Y.I.Y., 2018. Hitting arm detection for drum performance learning system using kinect. Proceedings of The Seventh International Conference on Informatics and Applications 1, 8–11.

Phillips-Silver, J., Trainor, L.J., 2007. Hearing what the body feels: Auditory encoding of rhythmic movement. Cognition 105, 533–546.

Piana, S., Coletta, P., Ghisio, S., Niewiadomski, R., Mancini, M., Sagoleo, R., Volpe, G., Camurri, A., 2016. Towards a multimodal repository of expressive movement qualities in dance, in: Proceedings of the 3rd International Symposium on Movement and Computing, ACM, New York, NY, USA. pp. 10:1–10:8. doi:10.1145/2948910.2948931.

Ravi, N., Dandekar, N., Mysore, P., Littman, M.L., 2005. Activity recognition from accelerometer data, in: Aaai, pp. 1541–1546.

Romano, G., Schneider, J., Drachsler, H., 2019. Dancing salsa with machines—filling the gap of dancing learning solutions. Sensors 19. doi:10.3390/s19173661.

Ross, J., 1994. The right moves: Challenges of dance assessment. Arts Education Policy Review 96, 11–17.

dos Santos, A.D.P., Loke, L., Martinez-Maldonado, R., 2018a. Exploring video annotation as a tool to support dance teaching, in: Proceedings of the 30th Australian Conference on Computer-Human Interaction, ACM. pp. 448–452.

dos Santos, A.D.P., Tang, L.M., Loke, L., Martinez-Maldonado, R., 2018b. You are off the beat! is accelerometer data enough for measuring dance rhythm?, in: Proceedings of the 5rd International Symposium on Movement and Computing, pp. 1–8.

dos Santos, A.D.P., Yacef, K., Martinez-Maldonado, R., 2017. 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, ACM. pp. 183–191.

Santos, O.C., 2016. Training the body: The potential of aied to support personalized motor skills learning. International Journal of Artificial Intelligence in Education 26, 730–755.

Santos, O.C., Eddy, M.H., 2017. Modeling psychomotor activity: Current approaches and open issues, in: Adjunct Publication of the 25th

Conference on User Modeling, Adaptation and Personalization, ACM. pp. 305–310.

Saunders, F., Hutt, I., 2015. Enhancing large-class teaching: a systematic comparison of rich-media materials. Higher Education Research & Development 34, 1233–1250.

Saxena, N.A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D.C., Liu, Y., 2019. How do fairness definitions fare? examining public attitudes towards algorithmic definitions of fairness, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery, New York, NY, USA. p. 99–106. doi:10.1145/3306618.3314248.

Schmidt, R.A., Wrisberg, C.A., 2008. Motor learning and performance: A situation-based learning approach. Human kinetics.

Senecal, S., Nijdam, N.A., Thalmann, N.M., 2018. Motion analysis and classification of salsa dance using music-related motion features, in: Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games, ACM. pp. 1–11.

Sergis, S., Sampson, D.G., 2017. Teaching and learning analytics to support teacher inquiry: A systematic literature review, in: Learning analytics: Fundaments, applications, and trends. Springer, pp. 25–63.

Siirtola, P., Röning, J., 2018. Revisiting" recognizing human activities user-independently on smartphones based on accelerometer data"-what has happened since 2012? International Journal of Interactive Multimedia & Artificial Intelligence 5.

Stein, S., McKenna, S.J., 2013. User-adaptive models for recognizing food preparation activities, in: Proceedings of the 5th international workshop on Multimedia for cooking & eating activities, ACM. pp. 39–44.

Trajkova, M., Cafaro, F., 2018. Takes tutu to ballet: Designing visual and verbal feedback for augmented mirrors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2. doi:10.1145/3191770.

Trehub, S.E., Becker, J., Morley, I., 2015. Cross-cultural perspectives on music and musicality. Philosophical Transactions of the Royal Society B: Biological Sciences 370, 1–9.

Vecchi, R., 2012. O ensino da Dança de Salão pautado na teoria do "Teaching for Understading". Ph.D. thesis. Tese de Doutorado. Universidade São Judas Tadeu, São Paulo, 2012.

Wang, D., Yang, Q., Abdul, A., Lim, B.Y., 2019. Designing theory-driven user-centric explainable ai, in: Proceedings of the 2019 CHI conference on human factors in computing systems, pp. 1–15.

Yang, Y., Leung, H., Yue, L., Deng, L., 2013. Generating a two-phase lesson for guiding beginners to learn basic dance movements. Computers & Education 61, 1–20.

Zafar, M.B., Valera, I., Gomez Rodriguez, M., Gummadi, K.P., 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE. p. 1171–1180. doi:10.1145/3038912.3052660.