Classification of Salsa Dance Level using Music and Interaction based Motion Features

Simon Senecal, Niels A. Nijdam and Nadia Magnenat Thalmann
University of Geneva, Geneva, Switzerland

Keywords: Modelling of Natural Scenes and Phenomena, Motion Analysis, Couple Dance, Motion Features, Machine Learning.

Abstract: Learning couple dance such as Salsa is a challenge for the modern human as it requires to assimilate and understand correctly all the dance parameters. Traditionally learned with a teacher, some situation and the variability of dance class environment can impact the learning process. Having a better understanding of what is a good salsa dancer from motion analysis perspective would bring interesting knowledge and can complement better learning. In this paper, we propose a set of music and interaction based motion features to classify salsa dancer couple performance in three learning states (beginner, intermediate and expert). These motion features are an interpretation of components given via interviews from teacher and professionals and other dance features found in systematic review of papers. For the presented study, a motion capture database (SALSA) has been recorded of 26 different couples with three skill levels dancing on 10 different tempos (260 clips). Each recorded clips contains a basic steps sequence and an extended improvisation sequence during two minutes in total at 120 frame per second. Each of the 27 motion features have been computed on a sliding window that corresponds to the 8 beats reference for dance. Different multiclass classifier has been tested, mainly k-nearest neighbours, Random forest and Support Vector Machine, with an accuracy result of classification up to 81% for three levels and 92% for two levels. A later feature analysis validates 23 out of 27 proposed features. The work presented here has profound implications for future studies of motion analysis, couple dance learning and human-human interaction.

1 INTRODUCTION

The analysis and investigation of the effects and intricacies of social dances are ample and find their contributions in many of the sociological, cultural and psychological areas. This comes as no surprise, as social dances already exist for centuries and are embedded in many cultures, ethnic groups and are often related to a social and/or religious context (Powers,). In more recent studies, the attention to social couple dances is also found in the fields of bio-mechanics, Human Robotic Interaction (HRI) and Human Computer Interaction (HCI), examining the features and its application in the digital domain. Within the latter context, we focus on the predominantly cognitive connection between the dancers while performing a social couple dance. The human to human interaction with full-body movements are coordinated and finetuned upon each other, and in most cases attuned to the music, which dictates the rhythm and the 'way' a dance is carried out (e.g. slow vs. energetic). Another aspect of the interaction is the 'lead' and 'follow' roles, which refer to the impulse and response pattern during the dance and the connectivity between the couple. The vastly dynamic and interactive situations of social couple dances brings a plethora of parameters, derived from the physical and cognitive interaction, the musical interpretation and listening (e.g. body "drive"), and represents a tremendous challenge to comprehend and analyse this intricate and interdependent set of parameters.

The objective is to extract a set of musical and interaction based features that can classify the performance of a dancer and in this case two people simultaneously as a dancing couple (as Figure 1). To learn a couple dance, such as Salsa, is a challenge as it requires learning an extensive range of mechanico-cognitivo-interactive parameters and is limited by our modern perils:

- Learning in (large) collective classes, which is less effective to spot errors on individual students.
- The need to practice with another partner on location, meaning the risk of deficient facilities and/or

Senecal, S., Nijdam, N. and Thalmann, N.
Classification of Salsa Dance Level using Music and Interaction based Motion Features.
DOI: 10.5220/0007399701000109
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 100-109
ISBN: 978-989-758-354-4

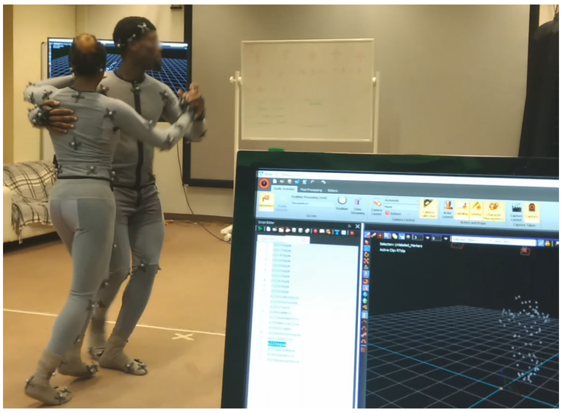


Figure 1: couple performing salsa steps.

not having a partner to practice with (either by lack of dance partners or due to personal time schedules).

- Other parameters can influence the study, such as mood, stress, fatigue and other external social factors.
- Time and location constraints due to other obligations (e.g. studies, work).

In addition, when the student is reaching a similar skill level as its teacher, the student may oppose the advises given by the teacher as to what is ‘correct’. The status of an *expert* in social dance can be a source of confusion as there is no universal recognised diplomas but rather a public recognition of skills by pairs. In both cases the learning process can be less effective, halted or reconsidered depending on the relationship between student and teacher.

These challenges are not easily overcome, if at all. However, a solution that can provide some relieve would be to set up a virtual coach that utilises the set of features within an Artificial Intelligence (AI) framework, such that it can analyse the movements of the dancer(s) and provide a positive feedback to improve their skills. In a previous work (Senecal et al., 2018), a first analysis was made and is build upon and further improved within this paper. A specific set of features for dance are proposed and investigated using a database of 3D movements of dancers, synchronised with music.

2 STATE OF THE ART

Motion in dance has been investigated through multiple scientific studies. Health studies show the benefit of social dances for balance and cognition for elderly (Merom et al., 2013; Merom et al., 2016a; Merom et al., 2016b). Moreover, the interactive aspect has been touched upon by the HRI domain, where through Inertial Measurement Unit (IMU) detection the user’s movements were transcribed into an intermediary data set to generate poetry (Cuykendall et al., 2016a; Cuykendall et al., 2016b). Human to human interaction has been explored via a setup of patches (Shum et al., 2008) and scene ranking (Won et al., 2014) in the context of animated character. Another example is the use of robots acquiring the knowledge and skills to perform a dance (Paez Granados et al., 2016). However, the research is limited to single instances of a dancer, thus not taking into account the simultaneous act of dancing. The interaction between performers themselves has been studied in the psychological domain (Ozcimder et al., 2016; Whyatt and Torres, 2017), even with the audience, that take part in the performing process (Theodorou et al., 2016).

To take into account the uncertainty of observations, the judgement process by a human coach is based on experience, historical knowledge and making assumptions about the state, intentions and methods of the students. It is at this point that bias can appear in decision-making: ”*fatigue, stress, stakes, prejudices, errors, beliefs, intuitions, the tendency to partiality through ignorance, similarity decision, random correlation belief, great influence of the first time, finding before the evidence, contradictions with unfulfilled beliefs, unjustified emphasis of information interpreted as more egregious.*” (Hicks et al., 2004). These human deficiencies, mostly due to infobesity (i.e. information overload), can be corrected by a virtual coach. This idea is developing recently in parallel of increasing of virtual reality application such as (DanceVirtual, 2018).

Extracting the motion features from continuous movement is a key element for describing, evaluating and understanding dance and movement in general. The use of Laban Motion Analysis (LMA)-based motion retrieval and indexing for motion features is a solution that has proved to work well in different situations (Aristidou et al., 2014), and is therefore ideal to be used as a base to build a machine learning classifier, as demonstrated for theatre emotional expression (Senecal et al., 2016) or evaluating the performer’s emotion using LMA features (Aristidou et al., 2015). Some studies focused on a specific motion feature, for example the fluidity of the movement is an important dance parameter investigated in (Piana, 2016). In this particular study, it is proposed to see how fluidity can help describing and classifying dance performance. Through interdisciplinary research including bio-mechanic, psychology and experiments with choreographers and dancers, they propose a definition that takes specifically the minimum energy dissipation when looking at the human body as a kinematic chain. Another work (Alborno et al., 2016), elaborated upon the expressive qualities, such as rigidity, fluidity and impulsiveness, to investigate intra-personal synchronisation for full body movement classification. More recently, several motion features for social dance (Forro) have been proposed, taking the music component into account (dos Santos et al., 2017). These proposed features are computed with the user's motion data on one hand and the music data, e.g. Beats Per Minute (BPM), on the other hand. First the "*Rhythm BPM: We calculate the average beats per minute.*", then the "*Rhythm consistency: we calculate the coefficient of variation of the student's BPM across the full dancing exercise*". This study brings interesting insights on characterising social dancing but the weakness is the accuracy due to the sensor (a simplified IMU for the full body, representing a single point in space). An equal high percentage classification of 96% accuracy of motion have been recently achieved by using a long short term memory based neural network from full body motion data to only sparse data captured by two separate inherent wearable, showing the importance of proper algorithm and possible reduction of motion measurement (Drumond et al., 2018).

In comparison to the previous mentioned approaches, our work takes two persons dancing together and defines this as the input entity for analysis, indexing and classification. The work is further set in the context of Salsa social dance. Prior to establishing the input entity, we first reflected upon the most relevant motion features extraction method from literature (BPM rhythm and consistency from (dos Santos et al., 2017)) and reviewed and discussed these through interviews and focus groups of experts in dance (teachers and choreographers). As for the acquisition of the data, a motion capture high precision system was utilised to ensure a maximum accuracy on the movements. Finally we propose a music-related motion feature from the processing of motion and audio file to classify salsa dance.

3 METHODOLOGY

3.1 Field Study on Criteria Improvement

A field study has been conducted in order to improve the motion features from the literature. This study was conducted in Geneva, a dynamic city for social dancing with an official number of 15 active Latin dance schools and hosting international dance congresses; making it a very important central dance area in Switzerland and also in Europe. Experts in social dances are persons with a high-level of expertise and skills, with a subsequent level of reputation, and/or recognised by pairs to be expert as there is no official diploma or formation for social dances (albeit some private schools, and in some countries they do provide a diploma). We therefore define a person as 'an expert' in social dance if it belongs to one of the following definitions:
Jury of international competition, Champion of International championships, Invited dancer in international congress, Director of major dance schools or *Professor of dance*.

Several professors and directors of the Latin dance schools in Geneva have been contacted and invited for an interview. They have been asked about what would be the criteria to teach or evaluate a dance student and indicate per criteria its importance. In addition, a questionnaire was filled out, with additional annotations on which questions were not clear and/or required further explanations. Initially, the questionnaire contained only the motion feature extracted from the literature and was updated with the feedback of the first expert (extending and improving upon further features), then suggested to the next expert and so on. This led to a final list of six features, listed in table 1, ranked by overall importance. Strong importance means that the criteria is essential for dancing, whereas little importance means to be less important (especially at beginner level).

3.2 Motion Features Algorithm

Three of the six criteria (Rhythm, Driving and Style) from table 1 are analysed along three axis: *(1)* the motion data itself (3D points), *(2)* the relation between motion data and music data, *(3)* and finally the relation between the two dancers. For each component, different parameters have been extracted and tested.

Salsa dance steps and figures are always performed within a reference of eight musical beats (binary tempo). This dance structure is taken into account for the computation that is proposed on a sliding window of corresponding width. Besides to be useful as a motion segmentation, this dance structure is making our computation normalised and independent of the BPM. The proposed motion features are based on the velocity profile analysis of the basic step (Mambo), represented in figure 2.

This particular velocity profile (extracted from a

Table 1: Retained criteria definitions and relative importance on a scale from 1 to 10.

Proposed criteria	Definition	Importance
Dancing on the rhythm	Being synchronised with the music's tempo.	10
Lead and Follow (Guidance)	Being able to guide / follow his / her partner.	7
Fluidity	Being able to move smoothly on the music.	6
Style and Variation	Adding your own variation to the basic movement.	5
Intention and Sharing	Being able to share the moment and enjoy the dance.	7
Musicality	Using your own dance movement with the music's variation.	3

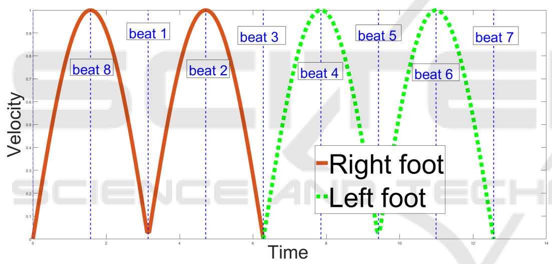


Figure 2: Model of velocity norm over time of both feet when dancing the basic step "mambo". The left foot makes two velocity peaks during the beat 8 and 2 whereas the right foot have velocity peaks for the beat 4 and 6.

motion capture of the basic steps of salsa experts) can be slightly different for the steps variation performed by the dancer during the song, but is a good base to compute temporally normalised features. In the couple during the dance, the follower will have a similar profile but with the right foot first.

3.2.1 Rhythm

The definition of the rhythm component can be interpreted as the regularity of each dancer to move his feet in synchronisation with the tempo. To detect such feature, the velocity peak of each foot is considered for both follower and leader. The absolute difference between the true musical beat location *Tdb* and the dancer beat *Trb* is proposed as the rhythmic error of each dancer. Indeed as the foot stop moving on the beat (for the beat 1, 3, 5 and 7), its velocity will decrease and reach a minimum when the dancer is marking the beat. For easier calculation, the previous beat (respectively 8, 2, 4 and 6) is considered as the reference points for the rhythmic as it corresponds to a velocity peak. Therefore a peak detection algorithm is used on two sliding windows for each foot, between the beat 7 to 5 for the left foot (right feet for the follower) and between the beat 3 to 1 for the right foot (left feet for the follower). This finally give us 8 features defined in equation 1.

Note that in the context of couple dance, the beat number *one* is marked by the *lead* dancer with the left leg and by the *follow* dancer with the right leg. It is then possible to apply our proposed algorithm to the *follow* dancer as well, taking the right leg for the Ec computation on the beat *one* and *three*.

3.2.2 Drive - Lead/Follow Interaction

This section focuses more on the relationship between the two partners of the dancing couple, the leader and follower. Indeed the leader have to guide the follower mechanically through the dance via anticipation of movement whereas the follower must respond to the guidance. Two sets of parameters are proposed: *the linear correlation of legs motion and the temporal difference for both dancers when marking the beat.*

Linear Correlation of Legs Motion. In order to investigate the relationship between the movement of the foot from the leader and follower, a motion feature is proposed as the linear correlation between the velocity profile of the foot of both dancer during an 8 beat time frame window. Similarly to the previous computations, the calculation is made between the left foot of the leader and the right foot of the follower for the first sliding window (beat 7 to 5) and then the other foot for the second window (beat 3 to 1). The features are defined in the equation 2.

Temporal Difference Man and Woman. The rhythm marked by the follower and leader can have some minor difference due to the different role of anticipating and responding to the music and can be proposed as relevant feature for the connection between partners (also due to natural imprecision). As previously, the peak location of each beat is computed within the two sliding windows. If Tl is the temporal location of the leader's beat and Tf the temporal location of the woman's beat, then the temporal difference Td for one beat is proposed within the equations 3.

Table 2: Summary of the proposed features. Fs is the feature number, SW the sliding window used, criteria the category and a brief explanation.

Fs		SW	Criteria	Detail
$f_1 = abs(Tdb - Trb)_{leader/beat1}$	(1)	2	Rhythm	Rhythm error b1 - Leader
$f_2 = abs(Tdb - Trb)_{follower/beat1}$		2	Rhythm	Rhythm error b1 - Follower
$f_3 = abs(Tdb - Trb)_{leader/beat3}$		2	Rhythm	Rhythm error b3 - Leader
$f_4 = abs(Tdb - Trb)_{follower/beat3}$		2	Rhythm	Rhythm error b3 - Follower
$f_5 = abs(Tdb - Trb)_{leader/beat5}$		1	Rhythm	Rhythm error b5 - Leader
$f_6 = abs(Tdb - Trb)_{follower/beat5}$		1	Rhythm	Rhythm error b5 - Follower
$f_7 = abs(Tdb - Trb)_{leader/beat7}$		1	Rhythm	Rhythm error b7 - Leader
$f_8 = abs(Tdb - Trb)_{follower/beat7}$		1	Rhythm	Rhythm error b7 - Follower
$f_9 = corr2D(VfootRM, VfootLW)_{beat1,3}$	(2)	2	Guidance	2D correlation velocity first half
$f_{10} = corr2D(VfootLM, VfootRW)_{beat5,7}$		1	Guidance	2D corr. velocity second half
$f_{11} = abs(Tl - Tf)_{beat1}$	(3)	2	Guidance	Temporal difference beat 1
$f_{12} = abs(Tl - Tf)_{beat3}$		2	Guidance	Temporal difference beat 3
$f_{13} = abs(Tl - Tf)_{beat5}$		1	Guidance	Temporal difference beat 5
$f_{14} = abs(Tl - Tf)_{beat7}$		1	Guidance	Temporal difference beat 7
$f_{15} = \int_{b3}^{b1} abs(V(t))dt_{leader/leftfoot}$	(4)	1	Style	Area under acc. curve - leader left foot
$f_{16} = \int_{b7}^{b5} abs(V(t))dt_{leader/rightfoot}$		1	Style	Area under acc. curve - leader right foot
$f_{17} = \iint_{b3}^{b1} abs(dV(t))dt^2_{leader/leftfoot}$		1	Style	Area under jerk curve - leader left foot
$f_{18} = \iint_{b7}^{b5} abs(dV(t))dt^2_{leader/rightfoot}$		1	Style	Area under jerk curve - leader right foot
$f_{19} = \int_{b3}^{b1} abs(V(t))dt_{follower/rightfoot}$		1	Style	Area under acc. curve - follower left foot
$f_{20} = \int_{b7}^{b5} abs(V(t))dt_{follower/leftfoot}$		1	Style	Area under acc. curve - follower left foot
$f_{21} = \iint_{b3}^{b1} abs(dV(t))dt^2_{follower/rightfoot}$		1	Style	Area under jerk curve - follower left foot
$f_{22} = \iint_{b7}^{b5} abs(dV(t))dt^2_{follower/leftfoot}$		1	Style	Area under jerk curve - follower left foot
$f_{23} = avg[dist(P_{righthand}, P_{hips})]_{leader}$	(5)	1	Style	Mean dist. hips-right hand - Leader
$f_{24} = avg[dist(P_{lefthand}, P_{hips})]_{leader}$		1	Style	Mean dist. hips-left hand - Leader
$f_{25} = avg[dist(P_{righthand}, P_{hips})]_{follower}$		1	Style	Mean dist. hips-right hand - Follower
$f_{26} = avg[dist(P_{lefthand}, P_{hips})]_{follower}$		1	Style	Mean dist. hips-left hand - Follower
$f_{27} = BPM$		1	Rhythm	BPM

3.2.3 Style - Variation

Beyond the pure rhythmic features, an important point is the style and variation expressed by the dancers within their dance. Multiple criteria can be taken into account, as shown in the several successful dance style studies using LMA. Among them, two parameters have been proposed for investigation; The area covered and the quantity of hand movement using the hand to hips distance. These parameters are directly inspired by LMA model.

Area Covered. The area covered by the dancer within a time range can help differentiate the level of expertise, as part of the style component. To compute it, the integration of both legs velocity profiles is proposed. These features are calculated within the sliding window range as in equations 4. The integration of the derivative of the velocity is also taken into account.

Mean Movement Quantity Hand to Hips. During salsa dance, the movement of the upper limbs are also important for styling, additionally to the guidance action. This is taken into account through the proposed feature of quantity of hand movement. For this feature the average distance between hand and hips during the 8 beat window is considered. The 3D location of the hips and both hands are computed and then the distance between hips and each hand is averaged, as shown in equations 5.

Summary of Features. In total, 26 features are proposed: 8 feature related to rhythm, 6 related to guidance and 12 related to style elements. The song's BPM is also added as a 27^{th} feature. A summary of the features can be found in the table 2. Please note that the data has been processed twice, on two different sliding window corresponding to a beat-to-beat time frame: From beat 7 to beat 5 for the 'Sliding Window 1' and from beat 3 to beat 1 for 'Sliding Window 2'.

4 EXPERIMENT

In order to validate the proposed music-related motion features as relevant parameter for the classification of dance learning level, a motion capture database of salsa dance is constructed.

4.1 Motion Capture Data

To train the supervised classifiers, we have established a database of motion captures (position and rotation of the body's joints in 3D) of couples dancing the basic salsa moves (SALSA database). A total of twenty six different dancer couples were recorded, of different skill levels (beginner, intermediate and expert), using a set of computer generated music with different beats per minutes (from 100 BPM to 280 BPM, with increments of 20 BPM). The level of dancers have been determined according to their experience: the experts are dance school directors and teachers (As shown in the figure 1), the beginners started to dance less than six month ago and the intermediate have more than one year and a half of dancing experience (Figure 3 illustrates the recording session with different couples).

The *variable tempo* was determined by performing a BPM study based on commercial salsa songs and music commonly used for teaching in dance schools as well as playlist of notorious salsa deejay. The used tempos cover different BPMs from the selection of music which are perceived as the most comfortable to dance to (refined by expert feedback). It has been asked to each couple to perform three basic steps, the *Mambo step*, the *Rumba step* and the *Guapea step*, prior to an improvisation part. A Vicon motion capture system with eight cameras was used for recording (at 120fps). The standard template from Vicon for the placement of the markers was used in each articulation of the body for a total of 52 markers per person. For each couple, we asked them to perform a first sequence of basic steps, followed by a sequence of improvisation. Is it important to mention that the numerous occlusion occurring during certain dance moves (mainly the closed position) made the capture very difficult (and especially the labelling process). In order to ensure the exact and systematic synchronisation of the music and the captured performance, the music was started simultaneously with the capture through the Vicon software interface.

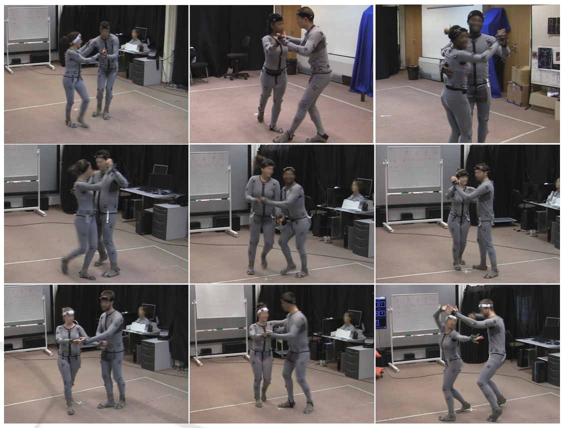


Figure 3: Different couples dancing salsa basic steps.

The result from the capture sessions is a database of 52 people as 26 couples (figure 3), dancing averagely two minutes, representing nearly $26\ couples\ x\ 10\ songs\ x\ 120\ sec\ x\ 120Hz = 3,700,000$ time frames of 104 points. The results have been exported as two fully labelled skeletal entities in *C3D* formatted files.

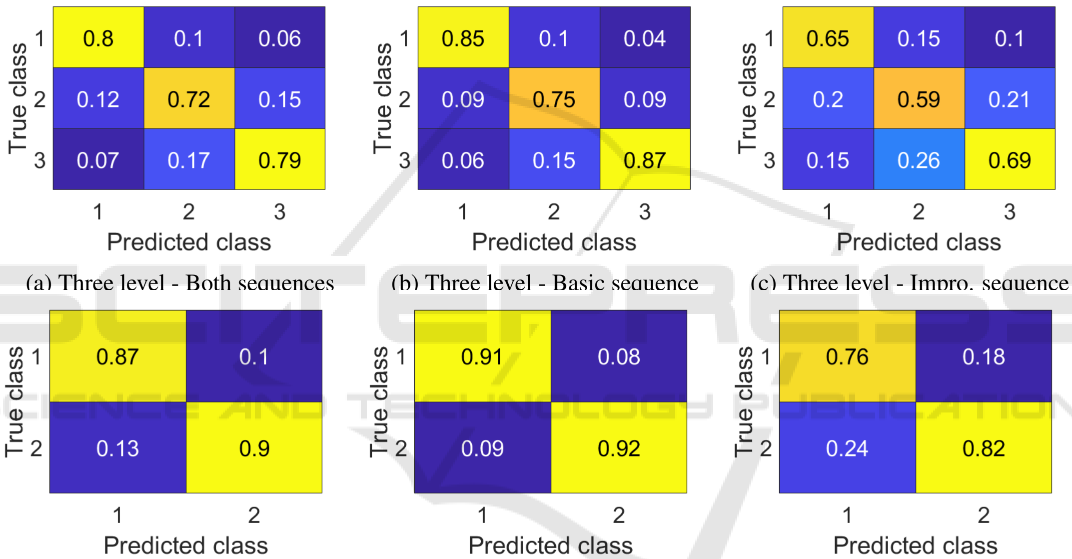
4.2 Audio Processing & Segmentation

In parallel to the motion capture, it is needed to identify clearly the beat temporal location on the music. Audacity was used to extract the beat temporal location from the audio files under the form of a dual column array containing textual annotations with timestamps, describing the number of each beat (one to eight repetitively). Thereafter, all the regular beats are marked from the music with the related labels along the duration of the song. Due to the synchronisation, we can directly compare point to point any temporal location of a musical event given by the music analysis with the motion data.

The feature extraction have been done for each song and each dancer, trough the basic sequence segment, the improvisation segment and the combination of both. Each 27 features has been computed within a sliding window of various length. This length directly depends on the BPM of the music and so the time length between musical beat.

Table 3: Summary of the resulting statistics from all classifiers. R is recall, P precision and A accuracy.

	Three level classification (beginner - intermediate - expert)								
	KNN			SVM			RF		
Seq.	*R*	*P*	*A*	*R*	*P*	*A*	*R*	*P*	*A*
Basics	81.10	82.07	**81.33**	75.12	73.91	74.00	76.26	77.48	78.87
Impro.	59.13	64.09	63.94	56.75	56.29	59.44	58.75	63.63	**64.14**
Both	76.32	77.33	**76.65**	74.25	73.50	73.87	78.84	79.75	76.54
	Two level classification (beginner - expert)								
Basics	86.89	90.78	**92.23**	89.63	68.24	79.17	82.16	89.51	90.32
Impro.	49.30	75.97	**84.62**	66.85	54.79	74.19	54.04	79.51	82.49
Both	83.63	86.89	**90.03**	86.68	74.98	83.58	86.17	89.26	89.33



(a) Three level - Both sequences (b) Three level - Basic sequence (c) Three level - Impro. sequence

(d) Two level - Both sequences (e) Two level - Basic sequence (f) Two level - Impro. sequence

Figure 4: Confusion matrices.

4.3 Results

In this section we present the classification method and results from the extraction of three sequences: *(1)* the basic steps, *(2)* the improvisation sequence and *(3)* the combination of both. For each sequence, the data processing and segmentation produce a vector of 27 features (numbers) with a variable sample size between one thousand to eight thousand. This data is introduced into different multiclass classifiers to try distinguishing between firstly the three learning levels (beginner, intermediate and expert) and in a second time between beginner and expert only.

4.3.1 Machine Learning Classifier

All the 27 features have been extracted for each sliding window along all songs. The data have been then segmented into three sequences, via manual video analysis and motion capture visualisation:

- Basic steps sequence - 3000 samples.
- Improvisation sequence ('free style') - 4000 samples.
- Both sequences, 8000 samples.

The three sequences have been inserted successively into machine learning classifiers with the 27 features as inputs and the dancer's level as target (a 1-dim vector with the number 1,2,3 that corresponds to the three levels). Three of the most popular of the families of multiclass classifier have been tested, namely k-nearest neighbours (KNN) algorithm, weighted Support Vector Machine (SVM) with city block metric and Random Forest (RF) algorithm. The classification has been tested for three levels and then two levels of dance, giving a total of 18 results. The figure 4 shows the confusion matrices for the classifier showing the best results.

The table 3 provides a summary of the collected statistical results among all classifier. The evaluation of classifiers is made through the following statistic: Recall (*R*), the proportion of motion parts of specific level which have been identified to be the correct level, Precision (*P*) the proportion of the motion classified as a specific level, whose true class label was that level and Accuracy (*A*) the global proportion of data classified correctly. The definition is:

$$R = TP/(TP + FN)$$
$$P = TP/(TP + FP)$$
$$A = (TP + TN)/(TP + TN + FP + FN)$$

Where TP is *true positive*, TN *true negative*, FP *false positive* and FN *false negative*. The KNN algorithm seems to be performing better with the highest accuracy.

The classifier that shows the best result is a k-nearest neighbours algorithm with a distance metric as 'cityblock', 10 nearest neighbours and an inverse squared weight distance function, with a maximum of 90% accuracy for basic sequence between the two extreme levels. In all cases the improvisation sequence presents a lower accuracy, indicating that the diversity of produced movement increase the difficulty of the classification task. Indeed a combination of basic and specific movements such as spin and other rotation as well as subtle rhythm variation would bring more noise to the data and so is constituent with lower accuracy.

Studying these results, it appears that Two level classification have a much higher accuracy than three level classification. This can be explained by the fact that the intermediate dancer can be of very different skills, making the category harder to define than expert or beginner. Also the Improvisation category have smaller score as well. Since it's a more free dance sequence, we can expect to have more difficulties to distinguish levels as it would requires more analysis components.

4.3.2 Features Importance

The importance of each feature have been investigated through the computation of their relative weight for the classification, that is shown from the KNN classifier on the figure 5 and figure 6 for the three levels and then two levels respectively.

Clearly the feature 11 to 14 are not contributing to classification and can be removed for future optimisation. It is important to note that the lesser data for dual classification may impact the feature importance as well. The importance varies very much for the improvisation part, which is understandable and suggest that additional features are needed for this sequence.

5 CONCLUSIONS AND FUTURE WORK

From literature and interviews with professionals, a set of 6 main components were identified. Among them, 3 were interpreted and expressed as 27 interactions based musical-related motion parameters. A database of salsa dance in synchronisation with music was realised. The proposed parameters were computed for each couple on each song within a sliding window, and inserted into classifiers. The results show an accuracy of up to 90% and mostly above 75%, validating most of the parameters and our sliding window method. A latter analyse on the feature importance shows that 23 features out of 27 are relevant for learning level classification, allowing to have a complementary evaluation of salsa dancer during couple performance.

This study is a first step toward an artificial intelligence based virtual coach that use automatic analysis of learning states for Salsa to improve the dancer's skills. Our proposed music and interaction based motion features shows some success to classify social couple dance performance. This first approach defines a building block for a framework that could be utilised within the other couple dances and the different domains, such as in robot interaction, emotional recognition or bio-mechanic studies, but also in the domain of virtual reality, avatar systems and generally serious game topics can be of interest and lastly to further our general understanding of motion analysis. The SALSA database and the extracted motion feature can also serve as base for other studies and cultural heritage conservation examples.

Future work includes to improve the classification via the optimisation and fine tuning of the classifier as well as studying feature importance and weight. An implementation to a neural network can be also interesting for classification and real time application as well as reducing the number of features able to discriminate the different levels. The inclusion of the three remaining music related motion features and tested over the previously mentioned larger accumu-

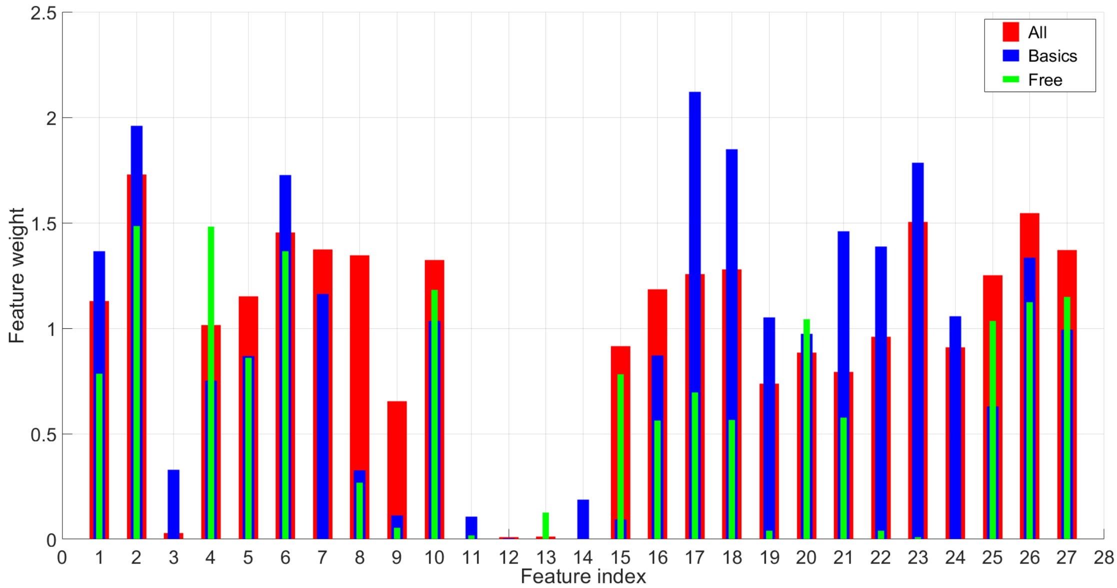


Figure 5: Importance of the different features for three level classification upon the different input sequences.

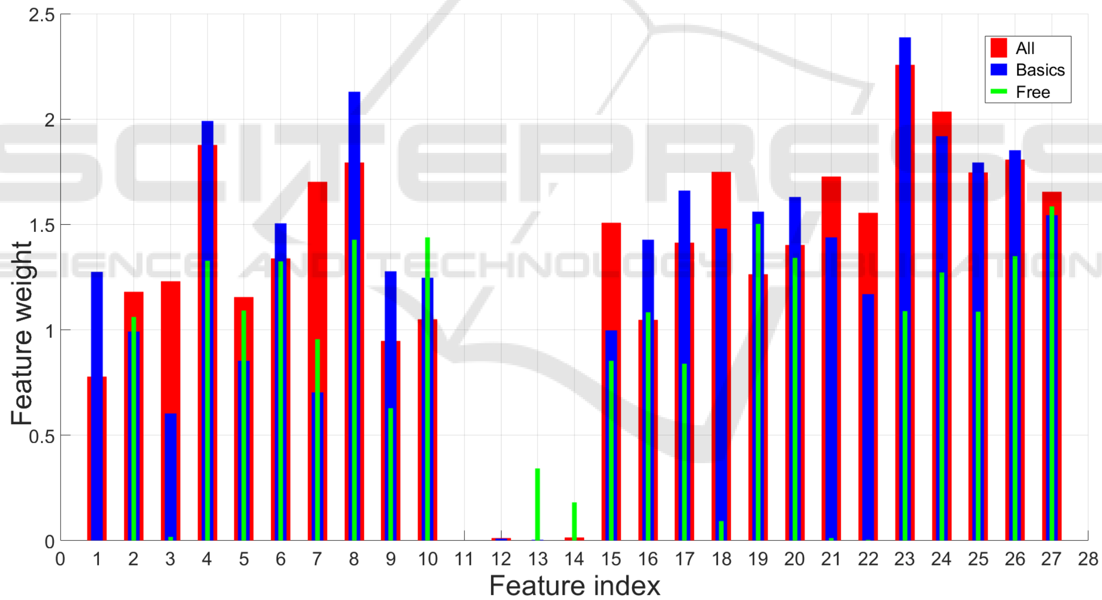


Figure 6: Importance of the different features for two level classification upon the different input sequences.

lated data set. The possibility to include some form of electroencephalogram or facial detection study while dancing as to try detect the emotional states of both participant. A learning study can be developed using the proposed features to investigate how they can have a real impact on the improvement of social dance learning.

ACKNOWLEDGEMENTS

The authors would like to thank Jeremy Patrix for his valuable help in machine learning, the salsa experts for the interesting discussions and all the dancers who performed the salsa dances at our department.

REFERENCES

Alborno, P., Piana, S., and Camurri, A. (2016). Analysis of Intrapersonal Synchronization in Full-Body Movements Displaying Different Expressive Qualities. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, volume 6455533, pages 136–143, New York, New York, USA. ACM Press.

Aristidou, A., Charalambous, P., and Chrysanthou, Y. (2015). Emotion Analysis and Classification: Understanding the Performers' Emotions Using the LMA Entities. *Computer Graphics Forum*, 34(6):262–276.

Aristidou, A., Stavrakis, E., and Chrysanthou, Y. (2014). LMA-Based Motion Retrieval for Folk Dance Cultural Heritage. *Euromed*, pages 207–216.

Cuykendall, S., Soutar-Rau, E., and Schiphorst, T. (2016a). POEME: A Poetry Engine Powered by Your Movement. *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 635–640.

Cuykendall, S., Soutar-Rau, E., Schiphorst, T., and Dipaola, S. (2016b). If Words Could Dance: Moving from Body to Data through Kinesthetic Evaluation. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, pages 234–238.

DanceVirtual (2018).

dos Santos, A., Yacef, K., and 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*, pages 183–191, New York, New York, USA. ACM Press.

Drumond, R. R., Marques, B. A. D., Vasconcelos, C. N., and Clua, E. (2018). Peek. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: GRAPP, (VISIGRAPP 2018)*, pages 215–222. INSTICC, SciTePress.

Hicks, J. D., Myers, G., Stoyen, A., and Zhu, Q. (2004). Bayesian-game modeling of c2 decision making in submarine battle-space situation awareness. Technical report, NEBRASKA UNIV AT OMAHA DEPT OF COMPUTER SCIENCE.

Merom, D., Cumming, R., Mathieu, E., Anstey, K. J., Rissel, C., Simpson, J. M., Morton, R. L., Cerin, E., Sherrington, C., and Lord, S. R. (2013). Can social dancing prevent falls in older adults? a protocol of the Dance, Aging, Cognition, Economics (DAnCE) fall prevention randomised controlled trial. *BMC Public Health*, 13(1):477.

Merom, D., Grunseit, A., Eramudugolla, R., Jefferis, B., Mcneill, J., and Anstey, K. J. (2016a). Cognitive Benefits of Social Dancing and Walking in Old Age: The Dancing Mind Randomized Controlled Trial. *Frontiers in aging neuroscience*, 8:26.

Merom, D., Mathieu, E., Cerin, E., Morton, R. L., Simpson, J. M., Rissel, C., Anstey, K. J., Sherrington, C., Lord, S. R., and Cumming, R. G. (2016b). Social Dancing and Incidence of Falls in Older Adults: A Cluster Randomised Controlled Trial. *PLOS Medicine*, 13(8):e1002112.

Ozcimder, K., Dey, B., Lazier, R. J., Trueman, D., and Leonard, N. E. (2016). Investigating group behavior in dance: An evolutionary dynamics approach. In *Proceedings of the American Control Conference*, volume 2016-July, pages 6465–6470. IEEE.

Paez Granados, D. F., Kinugawa, J., Hirata, Y., and Kosuge, K. (2016). Guiding Human Motions in Physical Human Robot Interaction through COM Motion Control of a Dance Teaching Robot. In *IEEE International Conference on Humanoid Robots (Humanoids)*, pages 279–285. IEEE.

Piana, S. (2016). Movement Fluidity Analysis Based on Performance and Perception. *CHI Extended Abstracts on Human Factors in Computing Systems*, pages 1629–1636.

Powers, R. S. U. Brief Histories of Social Dance.

Senecal, S., A. Nijdam, N., and Thalmann, N. (2018). Motion analysis and classification of salsa dance using music-related motion features. pages 1–10.

Senecal, S., Cuel, L., Aristidou, A., and Magnenat-Thalmann, N. (2016). Continuous body emotion recognition system during theater performances. In *Computer Animation and Virtual Worlds*, volume 27, pages 311–320. Wiley.

Shum, H. P., Komura, T., Shiraishi, M., and Yamazaki, S. (2008). Interaction patches for multi-character animation. *ACM Transactions on Graphics (TOG)*, 27(5):114.

Theodorou, L., Healey, P. G. T., and Smeraldi, F. (2016). Exploring Audience Behaviour During Contemporary Dance Performances. In *Proceedings of the 3rd International Symposium on Movement and Computing - MOCO '16*, pages 1–7, New York, New York, USA. ACM Press.

Whyatt, C. P. and Torres, E. B. (2017). The social-dance. In *Proceedings of the 4th International Conference on Movement Computing - MOCO '17*, pages 1–8, New York, New York, USA. ACM Press.

Won, J., Lee, K., O'Sullivan, C., Hodgins, J. K., and Lee, J. (2014). Generating and ranking diverse multi-character interactions. *ACM Transactions on Graphics (TOG)*, 33(6):219.