




On the Beat: Analysing and Evaluating Synchronicity in Dance Performances

Malte Menzel¹ , Jan-Philipp Tauscher¹ , and Marcus Magnor¹ 

¹ Institut für Computergraphik, TU Braunschweig, Germany
malte.menzel@tu-bs.de, {tauscher, magnor}@cg.cs.tu-bs.de

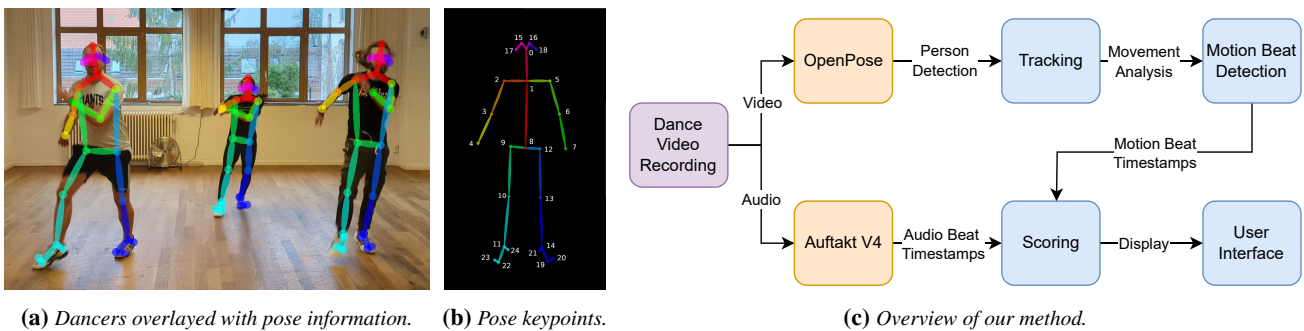


Figure 1: A typical application of our method: a smartphone video recording of a dance rehearsal session is analysed for pose information. Then, beat information is extracted from the accompanying audio stream and correlated to the pose information to evaluate synchronicity.

Abstract

This paper presents a method to analyse and evaluate synchronicity in dance performances automatically. Synchronisation of a dancer's movement and the accompanying music is a vital characteristic of dance performances. We propose a method that fuses computer vision-based extraction of dancers' body pose information and audio beat tracking to examine the alignment of the dance motions with the background music. Specifically, the motion of the dancer is analysed for rhythmic dance movements that are then subsequently correlated to the musical beats of the soundtrack played during the performance. Using a single mobile phone video recording of a dance performance only, our system is easily usable in dance rehearsal contexts. Our method evaluates accuracy for every motion beat of the performance on a timeline giving users detailed insight into their performance. We evaluated the accuracy of our method using a dataset containing 17 video recordings of real world dance performances. Our results closely match assessments by professional dancers, indicating correct analysis by our method.

CCS Concepts

• **Human-centered computing** → **Information visualization**; • **Applied computing** → **Performing arts**;

1. Introduction

An essential part of dance performances is accurately performing rhythmic movements to the beat of the music [Rep06]. For the audience, even slight deviations in the temporal alignment of music and motion can drastically change how they perceive a performance. Especially in group performances, it is vital for all dancers to be synchronised with the music. Otherwise, the performance will seem chaotic and disjointed. However, research has shown that keeping pace is a common issue for dancers [DWB09], especially during rehearsal. Combined with the fact that dance training can be complex and mentally demanding [BCMC*12], synchronicity

is a common problem in training both for dance groups and solo dancers. To solve this issue, a common technique dance groups utilise is to record videos of their training performances and have someone, usually their instructor or trainer, analyse this video regarding musical synchronicity for individual dancers. This analysis can, however, be tedious and time-consuming.

In this work, we present an approach to automatically analyse videos of dance performances and evaluate each dancer on their synchronicity. Our goal is to support dancers by making their training more effective by giving them an estimate of the accuracy of their dance performance over time. Specifically, we focus on Jump-

style dancing, which consists of well articulated and defined movements. To achieve our goal of assisting dancers in their training, our approach first separates dance performance videos into its audio and video. The video is then analysed using an existing pose recognition method *OpenPose* [CHS*19] to estimate the individual dancers' locations via machine learning and extract the position of body joints for each video frame. We developed a novel algorithmic method to consistently follow dancers through the video to add temporal tracking to OpenPose. Based on the movement of these joints, we design and implement a method to analytically detect the rhythmic beats the dancers are performing and accentuating in their movement. We then examine the audio stream using the beat tracking software *AUFTAKT* [zpl23], to find the tempo and rhythm of the musical track playing during the dance performance. Afterwards, our method matches the dancers' visual beats to the track's musical beats, thereby evaluating how synchronised each dancer is. An overview of our pipeline is shown in Fig. 1c.

Finally, we evaluate our proposed method in collaboration with dancers from the multiple German championship-winning dance group *Jump It* [Jum].

2. Music and Dance

2.1. Synchronicity and Dance

Dance and music are innately human activities. Researchers describe music and dance as "a necessary and integral dimension of human development" [CRO01]. Barring a single definition, researchers generally agree that music consists at least of some form of rhythm and pitch [HM81]. Dance is then described in its broadest sense as movement accompanying music [HM81].

Modern music features a consistent rhythm, with dance movements mirroring the rhythm of the music. Research suggests that humans can innately perceive the so-called *beat* of the music [Hon12]. A musical piece's beat, or measure, is its rhythmic pulse, usually established by percussion instruments, with its speed or tempo measured in beats per minute (BPM).

The relationship between a song's beat and a dancer's rhythmic movement is called *synchronicity*. We call a dance performance *synchronous* or *on-beat* if the dancer temporally aligns their movement in a way that matches the beat of the music. In the other case, we call a performance *asynchronous* or *off-beat*. Especially in group performances, it is vital that all dancers are synchronous to the beat, which establishes group harmony.

2.2. Motion Beats

A dancer's movement should be temporally aligned with the beat of the music to be synchronous. As movement itself is a change over time, what exactly should be aligned with the occurrence of each musical beat? To answer this question, dance research has introduced the concept of a *motion beat* [DA18, HTLC13, ZXY21]. Motion beats are defined as a dancer's '*deliberate changes in movement*' during a dance performance. For example, a dancer raising their arm and then lowering it again. By periodically changing their movement, dancers accentuate certain moments in time, constituting a rhythm of motion. The motion beat then would be the exact

moment the dancer's arm reaches its highest point and its movement direction changes. So whilst the full motion of raising and lowering their arm takes an entire time period, the motion beat is just a specific time point during the course of the movement. These motion beats define synchronicity, as during a synchronous dance performance, they will occur in temporal alignment with the beat of the music, thereby uniting the rhythm of movement and the rhythm of the music to form a consistent ensemble.

2.3. Jumpstyle Dancing

This work focuses on Jumpstyle dancing, as it features pronounced and well defined movements to music with a distinct, salient beat, making it a good choice for this work. Members of the professional dance group *Jump It* agreed to test our work. Jumpstyle is a modern electronic dance style. It originated in the early 1990s in Belgium, where it appeared as an underground/street dance accompanying a new fusion of tech-trance, hardcore and techno music [Ive08]. Whilst it has remained a small scene and still holds the image of youth culture, Jumpstyle tournaments and championships are held worldwide [jum08]. There are both online tournaments in the form of video submissions and offline championships such as the National or European Championships. The main music accompanying Jumpstyle today is Hardstyle music, a fast dance music style. Hardstyle music features tracks at around 140-160 BPM, which are characterised by a prominent kick drum in a 4/4 beat pattern and melodic synthesizers [QD20]. Jumpstyle dancing itself matches this musical style by being an energetic dance form. In this dance style, performers jump on every beat whilst performing intricate twists and turns. As such, Jumpstyle performances come across as fast and energetic to viewers [Ege12].

3. Related Work

3.1. Dance and Music

Research has found that dance as an art form depends on the dancer's ability to coordinate rhythmic movement with rhythmic sounds in music [Rep06]. Dancers are interesting to watch for an audience if they are synchronised to the music [RJK*16]. Especially for competitive dance groups that participate in contests, all dancers must be synchronised to the music. Hadley found that viewers prefer such dance performances over off-beat ones [HTW12].

3.2. Music Beat Tracking

The first music beat tracker able to analyse entire songs instead single instruments was presented by Goto et al. [GM94] in 1994. Further beat tracking methods were developed, some still being used today [Dix01, DP07, Ell07]. Since then, beat tracking saw significant improvements in recent years incorporating machine learning like AUFTAKT V4 by zPlane that we also use in this work.

3.3. Pose Estimation

Accurate human pose estimation from images and videos is an active research area in computer vision. In this work, we use OpenPose [CHS*19] that predicts the location of 18 human keypoints,

such as the face, hands and feet, using a convolutional neural network (CNN). Alternatives similar to OpenPose include AlphaPose [FLT*22] and BlazePose [BGR*20].

3.4. Motion Beat Tracking

Motion beats, i.e. when a dancer changes their movement, can be detected using depth cameras or inertia measurement units (IMUs) [EAT*12, AD14, HTLC13]. Purely video based studies track the movement of feature points throughout the video [CT11], or split each video frame into motion and background and then analyse the motion part for periodic movement [AII6]. [Gue06] transforms the luminance change in each pixel into audio-like signals and then uses spectral analysis algorithms and [BKCO18] use an image-space method to detect frames at which there is a significant change in motion direction.

3.5. Dance Practice Systems

The first group of dance practice systems records a master performance by a teacher to then compare the performance to the students. These systems usually use depth cameras [TSS18, KK18, AD14, LT16], smartphone sensors [WYB*14] and IMUs [EAT*12]. The second group of dance practice systems uses videos of group dancers to evaluate how synchronised these dancers are to each other. One such system is SyncUp [ZXY21] that uses pose estimation to compare how similar each dancer's pose is and motion beat extraction to compare the dancers' rhythm of movement. However, these practice systems only evaluate dancers against other dancers. Our system, on the other hand, checks if dancers are synchronised with the music. To our best knowledge, no video-based dance practice system such as ours exists.

4. Processing

4.1. Audio Processing

To extract the time position of audio beats from the background track of the video, we first need to separate the audio stream from the video. We then use *beat tracking* to find the exact occurrence of each audio beat. Since we aim to analyse synchronicity between audio and motion beats, it is paramount that a beat tracker finds the beat positions as accurately as possible. Therefore we chose AUFTAKT V4 from zPlane, a modern, state-of-the-art proprietary beat tracker that is widely used in the music industry. It employs a deep neural network-based approach and can handle a multitude of different musical styles. AUFTAKT V4 outputs the tempo and time signature of the musical piece as well as the timing of each individual beat and its location within its measure. To further increase its accuracy we also constrain it to a tempo range between 120 – 170 BPM, including any song from the music genres Jumpstyle dance is generally performed to.

4.2. Video Processing

To extract motion beats from the video, we first need to find the position of each dancer in each frame using OpenPose. OpenPose detects possible positions of so-called human *keypoints* in images

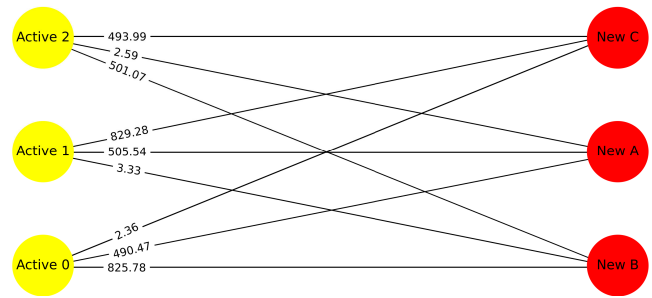


Figure 2: Bipartite graph of matching 3 persons between 2 frames.

and classifies them based on an internal confidence value. Depending on which version of OpenPose is used, detected keypoints differ. In this work, we use the *BODY_25B* model, available from the OpenPose training data repository. The output of this model consists of 25 keypoints representing the human body, as can be seen in Fig. 1b. *BODY_25B* is accurate enough compared to larger models and fast enough compared to smaller models for our purpose. For each person in each frame, OpenPose returns pixel coordinates of each keypoint detected, as visualised in Fig. 1a.

For each frame, it returns a unordered list of all persons found, each containing all their respective keypoints. Since we seek to analyse movement that spans multiple frames, we require temporal consistency for tracking persons across frames. We call this process *tracking* the individual dancers throughout the video.

4.2.1. Temporal Tracking

To consistently follow the individual dancers throughout the video, we develop our own tracking solution which parses OpenPose's frame-by-frame output. We maintain a list of 'active persons' to keep record of each person currently detected. We then match each person in the new frame to an active person detected in the previous frames. As we are processing videos at standard modern frame rates, it is reasonable to assume that each person will only move slightly and change their pose minimally from frame to frame. Therefore we need to find the likelihood for dancers being the same person in two consecutive frames.

To do so, we calculate a score for each pairing, with lower values indicating a higher chance for a person to be the same:

$$\text{score} = \frac{\sum_{\text{keypoints}}^{\text{for all detected}} |x_{\text{frame 1}} - x_{\text{frame 2}}| + |y_{\text{frame 1}} - y_{\text{frame 2}}|}{\text{amount of detected keypoints}} \quad (1)$$

with x and y corresponding to the position of each keypoint. In other words, the score is defined as the distance that each detected keypoint moved from one frame to the next. It is then normalised by the total number of keypoints detected to correct for persons with fewer detected keypoints. It also rewards pose similarity since the pair-wise distance of each keypoint is considered, not just the distance to the centre of each person.

After computing scores for all persons in two consecutive frames, we match them based on this score. We interpret this as a *bipartite matching problem* and solve it using a graph-based ap-

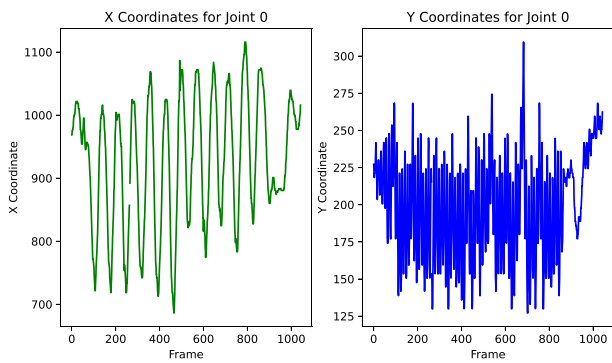


Figure 3: Typical movement of a single keypoint during dancing.

proach. In Fig. 2, three active person are matched between two adjacent frames using Karps algorithm [Kar80] to find a minimum weight matching. The number of detected persons can change between frames, i.e. when a new person enters the recorded area. In this case, we still perform our matching algorithm but now from the side with fewer participants. Therefore, if a new person enters the frame, we create an additional active person entry for the new person that did not receive a partner. Conversely, we do the opposite when a person leaves the recorded area. All active persons who received a partner continue to exist for the next frame, whilst those who did not will be removed. We now construct the matching from the current frame perspective, as this one contains fewer persons. In the end, we have a list of persons, each being active from a specific frame to another specific frame.

Although OpenPose is quite robust against false positives, we still regularly encountered instances of OpenPose detecting persons where there actually weren't any. However, OpenPose can be constrained to a specific maximum number of people to be seen in a video to significantly remove false positives. To find this number, we first run OpenPose unconstrained. We then filter out persons that got assigned a low confidence value (<0.6) and persons that are active for less than one second. After filtering the detected persons by these criteria, we calculate the maximum number of active people for any frame. We then rerun OpenPose and our tracking algorithm, now constrained to the newly calculated maximum number of people. This result is then used for the rest of our methodology.

5. Dance Motion Analysis

5.1. Finding Movement Changes

To detect motion beats, we first identify when a dancer changes their movement by looking at each keypoint separately. A common pattern in Jumpstyle dancing is keypoints moving up and down in y-direction as the dancer consistently jumps. In Fig. 3, the dancer is also moving continuously from left to right and back, as can be seen in the x-coordinate.

Whenever the motion direction of a keypoint changes, there are local extrema, i.e. minima and maxima, in at least one of the axes. We use a simple peak finding algorithm to locate these extrema, which identifies every occurrence of a value being greater

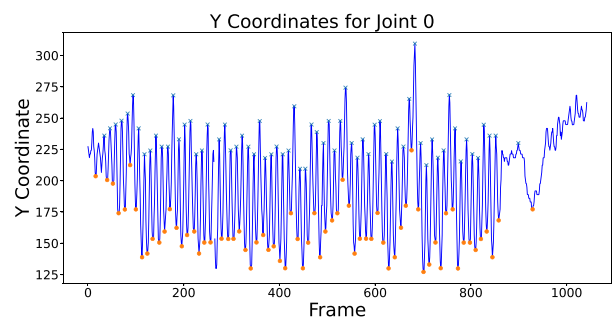


Figure 4: Keypoint movement annotated with filtered local extrema. Maxima are marked by an 'x' and minima by a circle.

or smaller than both of its neighbouring values. Not all of these occurrences should be classified as complete changes in movement, as multiple reasons could have led to small changes in position, such as camera shake or OpenPose shifting the location of a keypoint by a few pixels in a new frame. Considering this, we need to filter our detected peaks. One useful property of extrema is their prominence, which measures how much a peak stands out due to its intrinsic height relative to neighbouring peaks. The prominence of a peak is defined as the distance in value to the closest valley that needs to be crossed to reach an even larger peak. Therefore, prominence allows us to measure how much a peak differs from its neighbouring extrema. Finding a suitable prominence value for filtering is essential, as we neither want to exclude actual movement changes or to include noise in our detected extrema. The correct value differs for each person as we are working with absolute values in pixels: Both the size of a person and their distance to the camera influence the prominences of each extremum. Thus, we find this value by calculating the average size of each person throughout the video. For each frame, we calculate the bounding box around this person and measure the width and height of this box. We then average the width and height over the entire duration of the video that the person is active in. The prominence filter is then calculated as $\frac{1}{30} * \text{avg}(\text{width})$ for the x-axis and $\frac{1}{30} * \text{avg}(\text{height})$ for the y-axis. This constant was empirically determined from our data set of 17 videos (see Sec. 7 for details). The result of filtering all extrema by prominence can be seen in Fig. 4.

5.2. Jumpstyle Specific Filtering

For each joint we now have four data categories: The maxima and minima in x-direction and the maxima and minima in y-direction. As discussed in Sec. 2, a Jumpstyle dancer jumps and lands on every beat. Therefore we only need to identify the moment of full ground contact to determine synchronicity to the music. As the origin of video coordinate systems are usually located in the upper left image corner, we are interested in a dancer's vertical extremum. Therefore, we focus on the timestamps of motion beats in the positive y-direction, or the y-maxima. The other three data categories are not relevant for Jumpstyle dancing specifically.

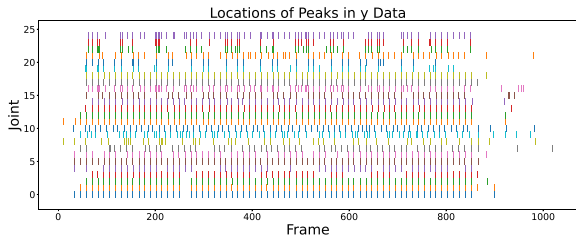


Figure 5: Peak time positions of each of the 25 keypoints, indicating a change in motion.

5.3. Selecting Motion Beats from all Movement Changes

Now we analyse the movement changes of all keypoints to derive the actual motion beats. We begin by calculating how many of the 25 keypoints are peaking for each frame in y-direction (see Fig. 5). As the entire person constantly jumps, most keypoints peak at similar timings. We sum these peaks per frame, where the peak amount is 0 if no keypoints are peaking, and the maximum being 25, if all keypoints are peaking simultaneously (see Fig. 6 (left)).

To find the actual motion beats for the entire person, we now perform peak finding once again, this time on the cumulative data. However, before that, we must consider that a person is not entirely rigid, as the movement of each keypoint influences the movement of other keypoints. If, for example, a person lands from a jump their body compresses like a suspension and their lower body parts reach their lowest point a few frames earlier than their upper body parts. The actual motion beat should then be somewhere between those two frames. The effect can be seen in Fig. 6 (left), as most large peaks of about ten keypoints peaking simultaneously are surrounded by a few keypoints peaking a few frames earlier or later.

We account for the the context of neighbouring frames by applying kernel smoothing using a sliding window. We use Epanechnikov’s kernel [Epa69] with a window size of $[-4, 4]$, i.e. four frames before and after the current frame are taken into consideration when its smoothed value is calculated. This window size was chosen after manual testing, with smaller values omitting important context and larger values only slightly increasing accuracy at a performance cost. We chose Epanechnikov’s kernel, as it leads to the optimal global accuracy [WJ94] and decays to 0 at its edges. The impact of kernel smoothing on our data can be seen in Fig. 6 (middle). Most peaks are now wider, including the context of a few keypoints peaking earlier or later.

Lastly, some noise remains in the data from dancers not performing absolutely perfectly, that we filter by prominence again. Here the prominence value indicates how many keypoints were peaking around this frame. Since this differs across dance performances, we calculate an individual prominence limit using kernel density estimation (KDE). In essence, KDE applies kernel smoothing to find the probability density estimation, i.e. assigns a probability to each of the prominence values. The KDE result is shown in Fig. 6 (right), with a local maximum at 12.5 keypoints peaking simultaneously for the actual motion beats and a second local maximum at 1.5 keypoints peaking simultaneously for the noise (red vertical bars). To find our prominence limit, we discard all peaks with a promi-

nence lower than the local minimum between those two groups. Here this value is 5.5 (green vertical bar), but it can shift based on how cleanly a dancer is performing.

Finally, we now select all peaks in the smoothed data with a higher prominence than the local minima from the KDE. These peaks represent the timestamps of a dancer landing on the floor that are our motion beats.

6. Synchronicity Metrics

So far we have extracted each motion beat position of a dancer and each audio beat position from the background music track. With this information we now analyse how synchronised a dance is to the music.

6.1. Data Rate Mismatch

We need to consider the different sampling frequencies of music and video. Motion data is extracted from video that is bound to its frame rate. Standard smartphone frame rates include 30 and 60 FPS, producing new data every ~ 33.3 milliseconds (ms) or ~ 16.6 ms. Audio data is usually sampled at 44 kHz or 48 kHz, recording a new sample every ~ 0.0227 ms or every ~ 0.0208 ms. Therefore, even when a video is shot at 60 FPS, we receive audio information at about 800 times the rate of new video information. The musical tempo and the video frame rate do not necessarily align (Fig. 7). Thus, an audio beat can occur while the video still shows a slightly older image.

6.2. Scoring

We use a scoring algorithm that compensates for the data rate mismatch. The temporal accuracy of every motion beat is evaluated separately to obtain a continuous assessment of dance performance over time:

$$\text{score} = \frac{\text{distance} - \text{frame time}}{\text{beat length}} \quad (2)$$

$$\text{distance} = \left| \frac{\text{motion beat pos. in frames}}{\text{FPS}} - \text{closest audio beat pos.} \right| \quad (3)$$

where all timings are measured in seconds if not stated otherwise. First, we calculate the distance of each motion beat to the closest audio beat. This distance cannot surpass half a beat length since in the worst case a motion beat sits right in the middle between two audio beats. To compensate for the difference in information frequency, we subtract the duration of a video frame from these distances, as we do not know precisely when the actual motion beat happened during the frame’s duration. This is in favour of the user as we overcompensate when the actual motion beat lies directly on a frame time stamp. Finally, the score is computed by dividing the distance by a duration of an audio beat, which depends on the song’s tempo. In musical terms, this scoring describes ‘*how many beats is a dancer off-beat?*’. The best score here is 0, where a dancer is perfectly synchronous. The worst possible score is 0.5, when the dancer is perfectly off-beat or asynchronous, as their motion beats sit exactly between two audio beats. These scores allow the user to infer information on their synchronicity and the choreography parts to revisit.

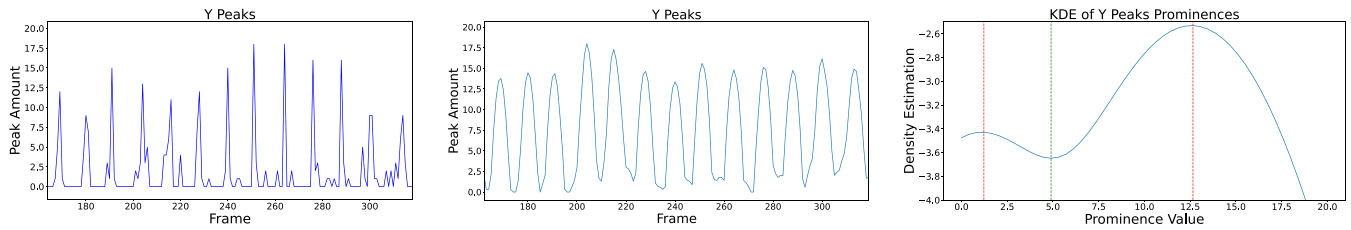


Figure 6: Sum of keypoints peaking each frame in the y-direction (left), and after applying kernel smoothing (middle). On the right, the result of kernel density estimation (KDE) of a typical dance performance.

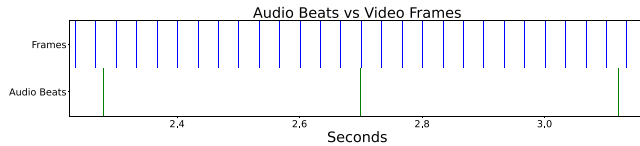


Figure 7: Temporal alignment of video frames and audio beats.

7. Results

7.1. Individual Component Test

We created a custom data set to verify the individual components of our pipeline for correct operation by using 22 unlabelled videos of Jumpstyle dance performances. Here, the videos are unlabelled, as the actual dance performance content is irrelevant for this kind of test. We begin by testing the removal of false positives in our test videos, by manually checking the rendered OpenPose output. We find no false positives in any of the 22 videos.

We test our tracking algorithm by manually checking the IDs of every person overlaid on every detected keypoint of that person. Of the 22 videos, 20 show no problems in tracking. In two videos, one person entry suddenly represents two actual dancers, as it is not closed correctly and instead reassigned to another dancer entirely. As this however is only a problem when OpenPose malfunctions, we do not attribute this error to our tracking method.

To verify our motion beat detection, we generate a new video for each test video, where we mark every keypoint currently peaking each frame for manual inspection. This works correctly for all videos, except in two instances where the tracking fails, as mentioned above.

We verify the audio processing by visually overlaying the waveform with the timestamps of each audio beat. We also created click sound overlays for the audio beat positions for auditory evaluation. In 18 of 22 videos, AUFTAKT V4 correctly detects every single audio beat. In four videos, AUFTAKT V4 fails for short parts of the audio ($\sim 10\%$) and identifies the remainder correctly.

7.2. Accuracy Test

Now we analyse the accuracy of our proposed dance evaluation system. We modified our previous data set to include a wider variety of dance scenarios and performance qualities. We annotated our data set in collaboration with the dancers of the Jump It formation to classify whether the performance in the videos is synchronous

Table 1: Results of analysing 19 dancers from 12 on-beat videos. SD is standard deviation, Dur is video duration in minutes.

Video	Dancer	Dur	Video Info		Score	
			FPS	BPM	Average	SD
1	1	0:21	30	155	0.0357	0.0631
	2	0:21	30	155	0.0326	0.0488
2	3	0:51	60	155	0.0382	0.0632
	4	0:30	30	150	0.0163	0.0318
3	5	0:30	30	150	0.0257	0.0444
	6	0:34	30	155	0.0461	0.0551
4	7	0:20	59.98	155	0.0252	0.0355
	8	0:38	60	155	0.0201	0.0591
7	9	0:19	59.98	150	0.1249	0.0856
	10	0:40	29.98	150	0.0175	0.035
8	11	0:40	29.98	150	0.0242	0.046
	12	0:40	29.98	150	0.0294	0.0634
9	13	0:54	30.07	155	0.0951	0.0888
	14	0:54	30.07	155	0.0835	0.1047
	15	0:54	30.07	155	0.1091	0.1253
	16	0:54	30.07	155	0.1137	0.1182
10	17	0:18	59.98	150	0.0169	0.0239
11	18	0:22	59.98	160	0.0460	0.0339
12	19	3:18	30.12	150 - 200	0.0504	0.0496
Average					0.0505	0.0618

or asynchronous. This labelled data set contains 17 videos of real-world dance performances (12 on-beat, 5 off-beat). Our verification generalises to many different scenarios: The videos contain 16 different dancers, where between 1 and 12 dancers perform simultaneously. They were shot at multiple different indoor and outdoor locations with varying lighting conditions from bright sunlight to barely lit night scenes. Multiple smartphones and camera settings were used to account for different camera chips, encoding formats, and video parameters.

7.2.1. Analysis of On-Beat Videos

We begin by analysing the subset of our data set labelled on-beat (results in Tab. 1). This set includes 19 dancers who are spread over 12 total videos. The average score of each dancer is between 0.0163 and 0.1249 beats. This value can be interpreted as 'how many beats is a dancer off-beat?' (Sec. 6.2). The average of all 19 dancers' average score is ~ 0.0505 beats, or only $\frac{1}{20}$ th of a beat off-beat. This is in line with the expert classifications rating these performances as on the beat. Even the worst value detected in dancer 9,

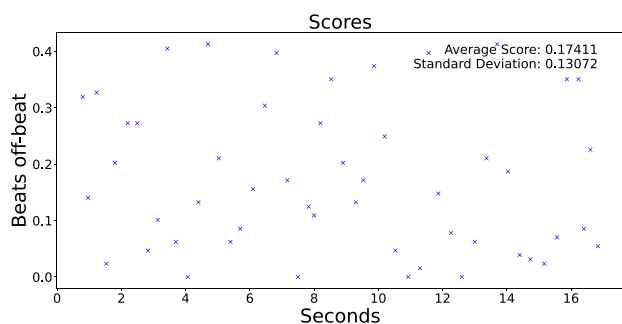
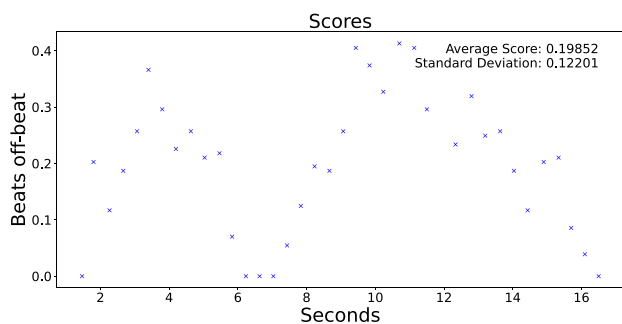
Table 2: Analysis results for the five off-beat videos.

Video		Score	
ID	Issue	Average	SD
20	Phase Shift	0.2024	0.0574
21	Phase Shift	0.2316	0.0665
22	Wrong Tempo	0.1741	0.1307
23	Short Term Misstep	0.1985	0.1220
24	No Dancing	0.1739	0.1447
Average		0.19026	0.1043

with an average score of 0.1249 beats off-beat, is still much closer to being perfectly on-beat at a score of 0.0 than being perfectly off-beat at a score of 0.5. The standard deviation values are similar to the average score values and, therefore, relatively small. On average, the standard deviation of all 19 dancers is only ~ 0.0618 beats showing they are consistently holding the correct tempo.

7.2.2. Analysis of Off-Beat Videos

The remaining five videos of our data set are labelled as off-beat by Jump It. In two videos, dancers are constantly off-beat, which means they hold the correct tempo but at a permanent delay, akin to a phase shift. In another video a dancer performs at a varying wrong tempo, thereby constantly cycling the alignment of their motion beats, like in a polyrhythm. Furthermore, a video includes real world examples of dancers losing track of the beat for a short time. Finally, one video includes a section of a dancer not dancing at all but instead moving randomly.

**Figure 8:** Scores of a dancer performing at a wrong tempo.**Figure 9:** Scores of a dancer making two short mistakes.

The results of our analysis of these videos can be seen in Tab. 2. The average scores here range from ~ 0.1741 to ~ 0.2316 beats off-beat and are much higher than for on-beat videos. The average over all off-beat videos is ~ 0.19026 beats off-beat, compared to the ~ 0.0505 of the on-beat videos. This is a quite significant difference and allows us to confidently classify these videos as off-beat. Apart from a general evaluation however, we can also gather more information on what specifically was wrong with each performance by looking at the scores of each individual motion beat. First, the standard deviation is a good indicator if a dancer is at least keeping the correct pace as visible in Tab. 2. While the average scores of videos 20 and 21 are definitely off-beat, the standard deviations of ~ 0.0574 and ~ 0.0665 are similar to the average standard deviation of the on-beat videos at ~ 0.0618 . As these are the two phase shift videos, their motion beat scores are all similarly off-beat. If the dancers however perform at varying incorrect speeds (video 22, 23 and 24) the scores will continuously shift like in a polyrhythm, leading to a high standard deviation (see Fig. 8). Second, viewing the score timeline allows dancers to identify moments in a performance, where they were off-beat. For example in Fig. 9, the dancer loses track of the beat twice around 4 and 12 seconds.

8. Conclusion and Discussion

In this work, we present a method to automatically analyse and evaluate dance performances regarding their synchronicity using only a single video recording. Our method successfully and correctly analyses and evaluates videos of people performing Jumpstyle dance. During our entire testing process, we encounter not a single video where the analysis results do not match the content of the videos. Our approach enables classification of dance performances as either on-beat or off-beat. It tracks each dancer throughout the video, then evaluates every motion beat for each dancer and scores temporal alignment of motion and audio beats. The average score of all detected motion beats is a suitable measure for overall performance synchronicity. We find performances with an average score of ~ 0.1 beats off-beat or less are generally considered synchronous by professional dancers. Larger scores, especially from 0.2 beats off-beat upwards, strongly indicate an asynchronous performance. Furthermore, the standard deviation of the scores is a good indicator of a dancer keeping the correct tempo (Sec. 7.2.2). Lastly, tracking the accuracy scores over time throughout the video allows to gather even more context on the performance, e.g. identify when a dancer temporarily loses the beat.

Although we consider our approach successful, there are a few limitations. First, our simple algorithm for adding person tracking to OpenPose features no persistence over occlusion. This leads to performances with frequent occlusion producing many person entries which can be impractical for a user to interpret. Second, our approach can only ever be as accurate in its evaluation as the frame rate of the video. For video data, there is only new data when the camera records a new frame, making it the upper limit for accuracy. Third, our approach depends on two third-party techniques, OpenPose and AUFTAKT V4. Although we try to give the best input possible to both, we can do nothing if they fail for some reason.

In the future, an improved tracking method would improve analysis experience for videos with many occlusions. Furthermore, our

approach could be extended to other dance styles by generalising motion beat selection and adapting audio beat tracking to the respective style. Finally, we could extend our method for real-time feedback during training.

In conclusion, we consider our approach successful. It requires no other input than a dance performance video, making it easy to use in training scenarios. The analysis results it produces for dance performances match the assessments made by professional dancers. Ultimately, our method can support dancers in their training and help them improve their synchronicity.

Acknowledgments

We would like to thank the dance formation Jump It (Dance Company Braunschweig) and Tim Flohrer and Martin Schwerdtfeger (zPlane GmbH) for their kind support. The authors gratefully acknowledge funding by the German Science Foundation (DFG MA2555/15-1 “Immersive Digital Reality”) and the L3S Research Center, Hanover, Germany.

References

- [AD14] ALEXIADIS D. S., DARAS P.: Quaternionic signal processing techniques for automatic evaluation of dance performances from mocap data. *IEEE Trans. Multimedia* (2014), 1391–1406. 3
- [AII16] ARGÜELLO C., IREGUI M.: Exploring rhythmic patterns in dance movements by video analysis. In *DHM* (2016), pp. 123–131. 3
- [BCMC*12] BLÄSING B., CALVO-MERINO B., CROSS E. S., JOLA C., HONISCH J., STEVENS C. J.: Neurocognitive control in dance perception and performance. *Acta psychol.* (2012), 300–308. 1
- [BGR*20] BAZAREVSKY V., GRISHCHENKO I., RAVEENDRAN K., ZHU T., ZHANG F., GRUNDMANN M.: BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204* (2020). 3
- [BKCO18] BELLINI R., KLEIMAN Y., COHEN-OR D.: Dance to the beat: Synchronizing motion to audio. *Computational Visual Media* (2018), 197–208. 3
- [CHS*19] CAO Z., HIDALGO MARTINEZ G., SIMON T., WEI S., SHEIKH Y. A.: OpenPose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.* (2019). 2
- [CRO01] CROSS I.: Music, cognition, culture, and evolution. *Ann. N. Y. Acad. Sci.* (2001), 28–42. 2
- [CT11] CHU W.-T., TSAI S.-Y.: Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos. *IEEE Trans. Multimedia* (2011), 129–141. 3
- [DA18] DAVIS A., AGRAWALA M.: Visual rhythm and beat. *ACM Trans. Graph.* (2018), 1–11. 2
- [Dix01] DIXON S.: Automatic extraction of tempo and beat from expressive performances. *J. New Music Res.* (2001), 39–58. 2
- [DP07] DAVIES M. E., PLUMBLEY M. D.: Context-dependent beat tracking of musical audio. *IEEE/ACM Trans. Audio Speech Lang. Process.* (2007), 1009–1020. 2
- [DWB09] DROBNY D., WEISS M., BORCHERS J.: Saltate! a sensor-based system to support dance beginners. In *CHI EA* (2009), ACM, pp. 3943–3948. 1
- [EAT*12] ESSID S., ALEXIADIS D., TOURNEMENNE R., GOWING M., KELLY P., MONAGHAN D., DARAS P., DRÉMEAU A., O’CONNOR N. E.: An advanced virtual dance performance evaluator. In *IEEE Int. Conf. Acoust. Speech Signal Process.* (2012), pp. 2269–2272. 3
- [Ege12] EGE A.: Schnelle hüpfen im jumpstyle, 2012. max.de, accessed 22/08/2023. 2
- [Ell07] ELLIS D. P. W.: Beat tracking by dynamic programming. *J. New Music Res.* (2007), 51–60. 2
- [Epa69] EPANECHNIKOV V. A.: Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications* (1969), 153–158. 5
- [FLT*22] FANG H.-S., LI J., TANG H., XU C., ZHU H., XIU Y., LI Y.-L., LU C.: Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Trans. Pattern Anal. Mach. Intell.* (2022). 3
- [GM94] GOTO M., MURAOKA Y.: A beat tracking system for acoustic signals of music. In *ACM MM* (1994), pp. 365–372. 2
- [Gue06] GUEDES C.: Extracting musically-relevant rhythmic information from dance movement by applying pitch tracking techniques to a video signal. In *Sound Music Computing* (2006). 3
- [HM81] HERNDON M., MCLEOD N.: *Music as culture*. Norwood, 1981. 2
- [Hon12] HONING H.: Without it no music: beat induction as a fundamental musical trait. *Ann. N. Y. Acad. Sci.* (2012), 85–91. 2
- [HTLC13] HO C., TSAI W.-T., LIN K.-S., CHEN H. H.: Extraction and alignment evaluation of motion beats for street dance. In *IEEE Int. Conf. Acoust. Speech Signal Process.* (2013), pp. 2429–2433. 2, 3
- [HTW12] HADLEY L., TIDHAR D., WOOLHOUSE M.: Effects of observed music-gesture synchronicity on gaze and memory. In *ICMPC* (2012), pp. 384–388. 2
- [Ive08] IVERS B.: What Is It? Jumpstyle, 2008. <https://xlr8r.com/features/what-is-it-jumpstyle>, accessed 22/08/2023. 2
- [Jum] JUMP IT: *Jumpstyle Dance Formation, Braunschweig, Germany*. <http://www.youtube.com/@formationjumpit2863>. 2
- [jum08] JUMPSTHESTYLE.COM: Over jumpen (about jump), 2008. <http://jumpisthestyle.com/jumpstyle/overjump>, accessed 22/08/2023. 2
- [Kar80] KARP R. M.: An algorithm to solve the $m \times n$ assignment problem in expected time $O(mn \log n)$. *Networks* (1980), 143–152. 4
- [KK18] KIM Y., KIM D.: Real-time dance evaluation by markerless human pose estimation. *Multimed. Tools Appl.* (2018), 31199–31220. 3
- [LT16] LARABA S., TILMANNE J.: Dance performance evaluation using hidden markov models. *Comput. Animat. Virtual Worlds* (2016), 321–329. 3
- [QD20] Q-DANCE: What is hardstyle music?, 2020. <https://www.q-dance.com/en/static/hardstyle>, accessed 22/08/2023. 2
- [Rep06] REPP B. H.: Musical synchronization. *Music, motor control, and the brain* (2006), 55–76. 1, 2
- [RJK*16] REASON M., JOLA C., KAY R., REYNOLDS D., KAUPPI J.-P., GROBRAS M.-H., TOHKA J., POLLICK F. E.: Spectators’ aesthetic experience of sound and movement in dance performance: A transdisciplinary investigation. *Psychol. Aesthet. Creat. Arts* (2016), 42. 2
- [TSS18] TONGPAENG Y., SRIBUNTHANKUL P., SUREEPHONG P.: Evaluating real-time thai dance using thai dance training tool. In *ECTI DAMT* (2018), pp. 185–189. 3
- [WJ94] WAND M. P., JONES M. C.: *Kernel smoothing*. CRC press, 1994. 5
- [WYB*14] WEI Y., YAN H., BIE R., WANG S. T. M. D., SUN L.: Performance monitoring and evaluation in dance teaching with mobile sensing technology. *Pers. Ubiquitous Comput.* (2014), 1929–1939. 3
- [zpl23] ZPLANE.DEVELOPMENT: *AUFTAKT V4*. [zplane.development GmbH & Co KG Grunewaldstr. 83 10823 Berlin, Germany, 2023. https://licensing.zplane.de/technology#auftakt](https://licensing.zplane.de/technology#auftakt). 2
- [ZXY21] ZHOU Z., XU A., YATANI K.: Syncup: Vision-based practice support for synchronized dancing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2021). 2, 3