OPTICAL OR INERTIAL? EVALUATION OF TWO MOTION CAPTURE SYSTEMS FOR STUDIES OF DANCING TO ELECTRONIC DANCE MUSIC

Ragnhild Torvanger Solberg
University of Agder, Department of Popular Music
ragnhild.t.solberg@uia.no

Alexander Refsum Jensenius
University of Oslo, Department of Musicology
a.r.jensenius@imv.uio.no

ABSTRACT

What type of motion capture system is best suited for studying dancing to electronic dance music? The paper discusses positive and negative sides of using camera-based and sensor-based motion tracking systems for group studies of dancers. This is exemplified through experiments with a Qualisys infrared motion capture system being used alongside a set of small inertial trackers from Axivity and regular video recordings. The conclusion is that it is possible to fine-tune an infrared tracking system to work satisfactorily for group studies of complex body motion in a “club-like” environment. For ecological studies in a real club setting, however, inertial tracking is the most scalable and flexible solution.

1. INTRODUCTION

There has been a rapid growth in studies of music-related body motion over the last decades [1, 2], many of which have focused on musicians’ sound-producing actions [3, 4] or people’s spontaneous motion to music [5, 6]. Relatively few studies have been carried out on music–dance correspondences, and those have primarily focused on one or a few people dancing to music [7, 8].

We are interested in studying (larger) groups of people moving to music, and to look more closely at the intersubjective relationships found in such music–dance settings. More specifically, we are looking at the relationship between body motions and the musical sound, and dancers’ engagement with electronic dance music. Our long-term ambition is to carry out a large-scale experiment in a real club context. Due to ethical, practical and methodological challenges, however, we are currently running experiments in our controlled lab environment.

The aim of this paper is to present some of the challenges we have faced in setting up and running motion capture experiments with groups of 10–15 people dancing together. To our knowledge, few to none empirical studies have investigated how groups of people dance and relate to electronic dance music—even though this is a common and wide-spread form of engaging with this musical style. Therefore, we have realized the need to develop an ecologically valid motion capture research design that can be used to study such musical group behaviour in both a lab context and real-life settings.

We will start by presenting an overview of some relevant motion capture technologies. This is followed by brief presentations of the experiments we have conducted thus far and a discussion of solutions found to different technical and methodological challenges.

2. MOTION TRACKING TECHNOLOGIES

When it comes to systems for tracking human body motion, we may, very roughly, differentiate between two main types of technologies: camera-based and sensor-based systems. Each of these can further be subdivided into a number of categories. For the discussion here, however, we will consider three concrete solutions that in different ways could be used to capture dance motion:

- Video recordings, using a single, off-the-shelf video camera to record the entire dance space, followed by the application of computer vision techniques to extract relevant features
- Optical, infrared motion capture, using a setup with multiple infrared cameras to record the position of reflective markers on the body of the dancers
- Small inertial sensor devices with built-in accelerometers, gyroscopes and magnetometers, recording directly to an on-device memory storage

In the following we will briefly discuss benefits and possible challenges of each type of system.

2.1 Regular video recordings

There are both theoretical and practical limitations when it comes to using regular video recordings as the basis for tracking human motion. Even though there has been enormous progress in the field of computer vision in recent years [9–11], this method is still limited to primarily tracking motion in two dimensions. That means that the position of the camera is crucial for the final result, since only what can be seen, can be tracked. If one wants to track the position of a group of dancing people, the most sensible camera position would be in the ceiling. Such a position allows for
capturing people’s horizontal motion, but not much of the vertical displacement.

More specialized cameras, such as time-of-flight and stereo cameras, may allow for recording “pseudo-3D,” or at least get some depth information. But it is only with multiple cameras surrounding the capture space that it is possible to carry out true 3D motion tracking from cameras.

Another problem with using video recordings for our type of studies is the influence of changing lights. An important element of a club environment is that of light and laser effects, rapidly moving lights with changing colours. This problem can be overcome to some extent by using cameras with some kind of “night mode,” using a filter that only passes through the infrared light. The latter might entail the use of infrared light sources to function properly, which further complicates the setup.

The perhaps biggest challenge with regular video cameras, however, is that of actually tracking people or objects. Tracking an individual person can be hard enough if the background in the image is too noisy. Needless to say, it is quite a challenge to track individual people within a group of dancers in a dark setting and with changing lights. With this in mind, we never really tried to use regular video tracking for our current experiments, but rather used video recordings only for documentation and reference purposes.

2.2 Optical, infrared motion capture

While progress is being made for carrying out markerless, camera-based motion capture [12], the current state-of-the-art is still setups of multiple infrared cameras placed around the capture space, and with subjects wearing reflective markers. Using markers on the body allows for high spatiotemporal accuracy and precision of the joints being tracked. The use of infrared cameras with built-in light sources makes it possible to also track people in rooms with no, limited or changing light. One problem remains, though, the need for line-of-sight from cameras to markers.

Another challenge with infrared motion capture systems, is that they work best in a controlled laboratory setting. We have experienced numerous challenges when setting up the system outside of the lab, such as in a regular concert venue. First, carrying out the calibration process—moving a wand with reflective markers around the space—may be problematic if there are people present. This means that it would be necessary to calibrate the system before people arrive to the venue, which could possibly be several hours before an actual recording would take place.

A second challenge is that the calibration of the system could easily be ruined if any of the cameras move after the calibration has been performed. In our experience, it is a high risk for someone to bump into a camera stand or cable in a public space, which would result in the need for a new calibration to be performed. Even if we were to mount the cameras in the ceiling, the vibrations alone in a club space with loud music might very well be sufficient to require a re-calibration of the system.

Finally, infrared systems are very sensitive to reflections, everything from reflective materials on people’s clothes, to bottles and glasses. Such reflections would end up as tracked markers in the system, thus complicating the tracking of individuals. In the best of cases, many such “ghost” markers would require a very long post-processing process to identify individual markers. In a worst case scenario, too many reflections could possibly ruin an entire data set.

To conclude, it may very well be theoretically possible to use an infrared system in a real club context, but due to the many practical challenges we have for now decided to work in our controlled lab environment.

2.3 Inertial sensor-based systems

Many of the challenges presented above are non-existent for systems based on inertial sensors. The two main types of inertial sensors are accelerometers and gyroscopes, and both of these sensor types are based on measuring the displacement of a small “proof-mass.” Accelerometers measure the positional displacement of such a mass, while gyroscopes measure the rotational. By combining three accelerometers and three gyroscopes it is possible to capture both three-dimensional position and three-dimensional rotation in one small sensor unit.

One of the most compelling features of inertial sensors, is that they rely on physical laws (gravity), which are not affected by external factors, such as lighting. They can also be made into very small and self-contained units, with low power consumption and high sampling rates. These are probably some of the reasons why inertial sensors are now becoming integrated in a lot of technologies, further propelling down the cost of single units and securing even broader integration in all sorts of electronic devices.

The downside to inertial sensing is that accelerometers do not measure the position, but rather the rate of change of the subjects. It is possible to estimate the position through integration, and, combined with the data from gyroscopes and magnetometers, this can lead to satisfactory results [13]. However, while the relative position estimates may be good, such position data often suffer from a considerable amount of drift [14]. One way to overcome some of the drift problems in inertial systems, is by adding other sensor types and possibly also cameras [15]. This is common in more advanced inertial motion capture systems, but is not possible with smaller and cheaper integrated units.

3. DANCE EXPERIMENTS

Due to the many challenges of working in a real-life setting, we decided to carry out our current studies in a (motion capture) lab environment. Still we wanted to make the experiments as ecological as possible, so care was taken in transforming the lab into a “club-like” environment. The club setting is characterized by many people dancing relatively close to loud music in a darkened space with light effects. Therefore, we covered all the lab’s walls in black, turned off the lights, and added various changing light effects. The 60-channel sound system secured an immerse sound experience. Thus the final visual and audible appearance was comparable to that of a club setting (Figure 1).

We will in the following briefly describe the two experiments we have conducted so far. Our focus is on method-
3.1 Dance Experiment 1

The first experiment was carried out in June 2014, with 16 people participating in a 15-minute long dance session. In this experiment we used a high-quality infrared, marker-based system from Qualisys, with nine Oqus 300 cameras surrounding the capture space and running at 100 Hz. The system was for this experiment calibrated at the level of the floor. Each subject was equipped with two reflective markers: one positioned on the head and another on one of the wrists. The initial idea was to capture both general motion patterns (from the head) and more local activity (from the arm) of the subjects while dancing.

Even though we had done several smaller pilot studies prior to the actual experiment, we ended up with a lot of tracking problems. The biggest challenge was the large dropout rate of the wrist markers, since the subjects danced so close to each other that the markers were covered up most of the time. We also experienced challenges with reliably tracking the head markers due to people raising their arms and shifting positions while dancing. The raised arms covered many of the head markers, and when the wrist markers came close to the head markers, it also confused the proximity-based trajectory detection. Other markers disappeared for some time when some dancers bent down and danced close to the floor, and others when they moved around in the space.

All in all, the tracking percentage of the head markers was on average quite good, even though there were too many broken trajectories to reliably track individual subjects throughout an entire recording. Thus the final data set could not be used for the individual analysis that we had originally hoped for, but it still presented a solid and useful data set with the possibility to estimate the general “quantity of motion” of all the subjects.

3.2 Dance Experiment 2

The knowledge gained from the first experiment was vital when planning our second experiment, which was carried out in August 2015. Here we did several adjustments and updates to the research design. First, we decided to put a limit on 10 participants at a time, so the 29 recruited participants were distributed in three groups. Even though this is a somewhat smaller group than in the first experiment, it can still be qualified as a sufficient amount of people to simulate an ecologically valid dance setting. Also, by using smaller, but several groups, we had the added benefit of looking at differences and similarities between groups.

To ensure the recording of at least one good data set, we decided to do parallel recording with all the three techniques mentioned in Section 2: infrared optical tracking, inertial sensors and regular video recording.

Several measures were taken to improve the quality of the infrared tracking. First, each subject was equipped with only one reflective marker, positioned on the head, to reduce the problem of marker occlusion and confusion. Nine Qualisys Oqus 300 cameras were hanging around the walls of the room, in the same configuration as for the first experiment. This time four additional cameras (Oqus 400) were positioned in the ceiling above the capture space. These additional cameras greatly improved the tracking percentage to nearly 100% for each tracked subject. To reduce any possible measurement errors, the system was calibrated at head’s height, approximately 1.6 m above the floor. We also increased the frame rate of the Qualisys system to 200 Hz, since it has recently been shown that the minimum frame rate needed to capture motion should not be chosen based on the Nyquist-Shannon theorem, but rather according to the ratio between the maximum speed and the minimum spacing between markers [18].

Since it had proven difficult to capture arm motion using the infrared system, we opted for using inertial sensing of the activity of the arms. Here the AX3 armband sensor unit from Axivity was chosen. These sensors are made for long-term motion recording (up to one month continuously) and are running as standalone and individual modules. They each have an internal clock that is updated when connecting to a computer and this clock is used to record time-stamps to the data file. From our initial testing, we found that the clocks’ time-stamps deviated too much to be used for synchronization. So the solution was to do a manual synchronization routine with a set of repeated non-periodic spikes with all devices at the same time after starting the devices and before stopping them (Figure 2).
To further fail-proof the setup, the sessions were recorded using four regular Canon prosumer cameras. They were placed in each of the labs corners so as to cover the scene from different angles. The cameras’ “night-mode” setting was used to remove the changing light effects. The four video streams were run into a quad picture-in-picture video mixer, so that we could record a combined full HD video stream in the QTM software. This stream would carry audio as well, ensuring synchronization between audio, video and motion capture data.

It turned out that all of the three recording types worked well. The visual quality of the video recordings was quite poor, as expected when using the “night-mode” setting, but they were still useful for visual inspection and also worked well for some basic quantitative video analysis techniques. The data files from the Axivity sensors were flawless as well for some basic quantitative video analysis techniques. The AX3 sensors is that each device has a unique ID, which makes it possible to track individuals over time, even the ones that move around a lot in the space. We could also pair the inertial sensors to each of the subject’s infrared marker. The downside to inertial systems, however, is that the devices measure relative motion (based on the gravitational pull) and not the exact location in space.

### 4.2 Spatiotemporal Resolution

It has been shown that the spatiotemporal resolution of a Qualisys system is much higher than what is needed for studying human body motion [19]. This is the case, even though the spatial accuracy and precision is uneven throughout the space [20]. It was the new knowledge about the proportionality between the speed and spacing of markers [18] that made us increase the frame rate from Experiment 1 to Experiment 2 (from 100 Hz to 200 Hz). So while such a frame rate is not necessary to capture the motion observed in the dancers, it clearly reduced the number of marker dropouts in the recordings.

It has been shown that the AX3 does not provide the same spatiotemporal accuracy and precision as the Qualisys system [21]. Still, the spatial resolution and date rate is more than sufficient for capturing the large-scale body motion seen in dance studies. An added benefit of inertial sensors is that they provide an even spatial accuracy and precision all over the recording space, as opposed to the infrared markers.

### 4.3 Synchronization to Audio

One of the most challenging parts when it comes to working with motion capture systems in a musical context, is the need for synchronizing motion data to related audio and video files. One of the positive sides of using a complete motion capture solution like that provided by Qualisys, is that it allows for SMPTE-based synchronization of cameras to an audio interface. This makes it possible to record high-quality audio and video with frame-based synchronization to the motion capture data.

The AX3 sensors, on the other hand, are standalone devices with no proper synchronization mechanism. The easiest solution is to use the built-in clocks for synchronization, but our tests have shown that they drift apart for longer recordings. This is negligible in many cases, but they are not accurate enough when we want to synchronize several hour-long recordings to rapid, beat-based music. Fortunately, the AX3s sample evenly, and can hence be synchronized based on reference points at the beginning and end of recordings.

Our solution was to carry out a manual synchronization routine (Figure 2) consisting of five non-periodic spikes created over a period of about 30–40 seconds. This synchronization routine was performed at the beginning and end of the experiment, with several hours in between. A simple cross-correlation algorithm in Matlab aligned the data sets based on the spikes at the beginning and end of each of the 20 data files. As the plots in Figure 4 show, the result was near-perfect time alignment of all the AX3 data. Since the routine was carried out in front of the video cameras, which also recorded audio, it was possible to use the spikes to time-align the AX3 data sets, which could then also easily be synchronized with the audio and video files.

**Figure 3.** XY-plots from the infrared motion capture data, showing the horizontal motion patterns of subjects in each of the three groups.
4.5 The subject’s perspective

Most people would find dancing together with others in a lab environment somewhat unnatural and awkward—at least in the beginning. The “clubification” of the lab certainly helped in creating a relaxed and natural atmosphere for the subjects. We received a lot of positive feedback from the people participating in the experiments about the layout of the lab. It also helped having sofas, music, food and non-alcoholic drinks outside the lab space, thus creating a social atmosphere surrounding the experiments.

The use of a very limited sensor/marker setup also helped in reducing the feeling of being part of an experiment. The reflective marker on the head is lightweight and barely noticeable when put on, and the AX3 sensor feels like a regular watch. It probably also helped that the participants could easily put the equipment on themselves. None of the participants commented that the sensors had invaded their personal space, and they quickly forgot about them as soon as they had put them on.

This shows that even though the participants are not dancing in an actual club space, it is, indeed, possible to carry out such group studies with a certain level of ecological validity in a mocap lab.

5. CONCLUSION

As far as we know, there have been few experiments using advanced motion capture systems with such a number of people and in such a noisy environment that we have attempted. After having experienced several tracking problems in Experiment 1, we obtained near-perfect infrared motion tracking results for all three groups dancing together in Experiment 2. We also managed to successfully beat-synchronize hour-long recordings of 20 inertial trackers through a simple calibration routine.

As expected, regular video recordings do not work very well for the experiments in question. Video recordings, even in full HD quality, have limited spatiotemporal resolution, and it is difficult to adequately track individuals in a larger group of people. That said, video recordings with “night-mode” turned on, are of high value for documentation purposes and for assisting in the post-processing of sensor and marker data of a large group of people.

Based on the knowledge gained from these experiments, we are currently planning new lab-based recordings using the infrared motion capture system to further investigate musical group behaviour. We are also one step closer to realizing a larger study with inertial sensors in a real club context.

Acknowledgments

Thanks to all the dancers that participated in the study, to Diana Kayser, Minho Song and Mari Romarheim Haugen for their assistance in running the experiments, and to Kristian Nymoen for help with creating the AX3 synchronization routine in Matlab.
REFERENCES


