Chapter 1

Methods for studying music-related body motion

Alexander Refsum Jensenius, Department of Musicology, University of Oslo

Abstract

This chapter presents an overview of some methodological approaches and technologies that can be used in the study of music-related body motion. The aim is not to cover all possible approaches, but rather highlight some of the ones that are more relevant from a musicological point of view. This includes methods for video-based and sensor-based motion analyses, both qualitative and quantitative. It also includes discussions of the strengths and weaknesses of the different methods, and reflections on how the methods can be used in connection to other data in question, such as physiological or neurological data, symbolic notation, sound recordings and contextual data.

1.1 Some key challenges

The last decades have seen a rapid growth of interest in studying *music-related body motion*, that is, all types of human motion that appear in a musical context (Wanderley and Battier, 2000; Gritten and King, 2006, 2011; Godoy and Leman, 2010). This includes the motion carried out by *performers*, such as musicians, conductors and dancers, and the motion of *perceivers*, such as that of audience members during concerts, people dancing at clubs, or people's spontaneous motion to music in everyday life. As such, music-related motion is a functionally diverse category, ranging from describing purely instrumental motion (such as hitting a piano key with a finger) to purely communicative motion (such as gesticulating in the air with the arms). Furthermore, music-related motion may occur in any type of location, for example in a concert hall, at home, or in the street.

The challenge for musicologists interested in studying motion as part of their empirical material, is to choose methods that allow for studying such motion in a systematic manner. However, before deciding on a methodological approach, it is important to properly evaluate the content and context in which the motion is to be studied. Some questions to consider are:

- Aim: why is music-related motion interesting in this study? What kind of interaction is planned (sound–human, human–sound, human–human)?
- Subjects: how many subjects will be studied? What is their demography (gender, age, music/motor abilities) and personal context (familiar/unfamiliar with the task)? Will they move individually or in groups? What is the social context of the study?
- Motion: what type of motion is expected, and in which parts of the body? Are they large or small? Are they slow or fast? Will the subjects be stationary, or will they move about? Is it necessary to find the absolute

position in space, or is relative motion information (such as acceleration) sufficient?

- Environment: will the study be carried out in a "controlled" environment (such as a lab) or in an "ecological" setting (such as a concert hall)? Is power available? How much time is there to set up equipment? What are the lighting conditions?

- Artefacts: will there be any instruments, tools or other types of technologies used in the setup, and how will they be captured and synchronised?

- Audio: what type of sound recording is needed? How many channels? What sampling frequency and bit rate? What is the necessary level of synchronisation between motion and sound data?

- Video: what type of video recording is needed? How many cameras? What resolution and frame rate?

- Data handling: how will the different types of data be synchronised? What software and data formats will be used? What type of storage, backup, and sharing solutions are planned?

- Analysis: what type of analysis is planned? What types of visualisations are needed? What types of features will be extracted?

When it comes to the latter—analysis—there are numerous approaches to choose from, but they can be broadly categorised into two main categories:

- Descriptive analysis: the motion is described through *kinematics* (such as velocity or acceleration), *spatial features* (such as size and position in the room) or *temporal features* (such as frequency) components.

- Functional analysis: the motion is described through functional properties, such as whether the actions are sound-producing or sound-accompanying (see (Jensenius et al., 2010) for an overview of functional categories).

The former may often be associated with quantitative analysis approaches, such as using statistical methods on numerical data, while the latter may be based on qualitative analysis approaches. In most cases, however, one would typically need to carry out both descriptive and functional analyses, and utilise both qualitative and quantitative methods. As such, the methods should be seen as complimentary rather than competing.

This chapter will not focus on the analytical methods per se, but rather on methods that prepare the ground for such analyses to be carried out. We will begin by presenting a few methods for qualitative motion analysis, before moving to some quantitative approaches.

1.2 Qualitative motion analysis

There exist numerous systems for systematic notation, analysis and exploration of body motion from a qualitative and observational point of view, including the Alexander technique (Barlow, 1975), Rolfing (Feitis, 1978), Expressive motion (Pierce and Pierce, 1989), Dalcroze (Findlay, 1971) and Benesh (Parker, 1996). Several of these systems were developed in parallel to, or influenced by, the work of Rudolf Laban (1879–1954). Here we will look more closely at two of Laban's methods: the *Labanotation* and the *Laban Movement Analysis* (LMA).

1.2.1 Labanotation

Rudolf Laban worked as a dancer and choreographer before he became interested in the description and analysis of human motion at large, with the aim of developing a universal system for motion analysis. For that reason he spent time studying all sorts of human motion, including that of factory workers. In 1928 he presented a system called *Schrifttanz* (Guest, 2004), a method for the notation of motion as symbols along a vertical axis. This system was later to become *Labanotation*.

When writing a motion score through Labanotation, it is common to start with a description of *motifs*. These motifs are reduced versions of a full score, and they provide a rapid approach to writing down the main elements of a motion sequence. This first sketch is later used to write a more detailed Labanotation based on a structured set of symbols allowing for the notation of any type of motion of any body part. Figure 1.1 shows examples of Labanotation, with time running vertically, from bottom to top. The vertical line in the centre of each notation system marks the centre of the body, and motion is notated through symbols on the left and right sides of the centre line. Each of the body parts has their own symbol, and the duration of motion and position are given through different types of symbols inside the shapes of the body parts.

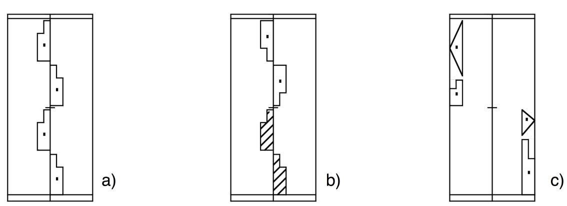

Figure 1.1: Three short motion sequences written in Labanotation, with time running vertically from bottom to top: (a) normal walking forwards with right and left foot; (b) right and left foot moving forwards on their toes, followed by walking backwards; (c) right side of the body moves forwards and then to the right, followed by the left side of the body leaning forwards and then to the left.

Although not very widespread, Labanotation is still in active use today, particularly in larger ballet and dance companies. There are also examples of the scholarly use of Labanotation, with an increasing interest from researchers within human-computer interaction and machine-assisted motion analysis (Choensawat et al., 2014). However, the complexity of the system means that quite few people really master Labanotation at a level with which they can read and write scores fluently. So, as opposed to music notation, which one can easily assume that most musicians and music researchers master well, it is hard to find people that can comfortably read and write Labanotation outside specialised circles.

1.2.2 Laban Movement Analysis

The theoretical basis for the Labanotation system was developed within a framework that has later been called *Laban Movement Analysis* (LMA). Even though Labanotation and LMA coexist, they can be used independently of each other. In fact, the LMA has received more widespread usage than Labanotation, also within the study of music.

As opposed to Labanotation's focus on writing motion structures in time, the LMA system is based on describing motion qualities. Fundamental to using the LMA, is that the analysis should always start from the observer's point of view, and by asking the question "how does this motion feel from my own body?" (Schrader, 2004). As such, the LMA is by default subjective, even though it aims at being a general method for the observation of motion.

The LMA system is based on four main categories: *body*, *space*, *shape* and *effort*, each of which are subdivided into categories describing different motion qualities, such as outlined in Figure 1.2. In this context we will mainly focus on one of the four main categories, the *effort* element (Laban and Lawrence, 1947), which has proven to be particularly relevant for musicological studies.

The effort category can be further subdivided into four subcategories, each with a descriptive axis:

- Space (direct–indirect). Space describes how one moves through the physical space, and

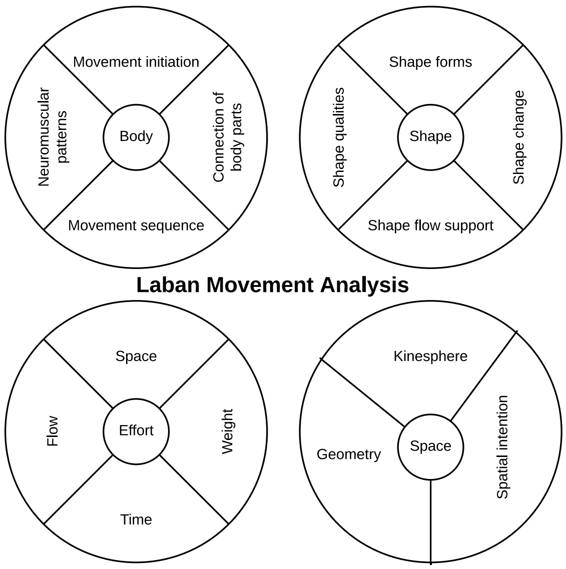

Figure 1.2: Overview of all the categories in the Laban Movement Analysis (LMA).

how one relates to the body's *kinesphere*, the maximal volume we can reach around our body when standing fixed on the floor.

- Time (quick–sustained). Time is used to describe the rhythmic character of motion. Laban was concerned that time and rhythm should not be split. We experience rhythmical patterns all the time, and we have our own bodily rhythms defined by our pulse and breathing.
- Weight (strong–light). Weight is related to gravity, and the fact that we need to use muscular activity to work against gravity when moving upwards. We also use gravity to help us when moving downwards, and it is this constant interplay between our body and the earth's gravitational pull that help shape our motion.
- Flow (bound–free). Flow describes how motion unfolds in time and space. Free flow motion is both relaxed and continuous at

the same time, while bound flow is when the motion is hindered.

Even though the four effort parameters are but one small part of the full LMA system, they have been used separately in several musicological studies, including sound–motion correlations in free dance to music (Haga, 2008), the phrasing motion of clarinettists (Campbell and Wanderley, 2005), and relationships between music, emotion and dance (Van Dyck et al., 2013).

1.3 Video-based analyses

If thinking about an axis from purely qualitative methods to purely quantitative, video-based analyses can be argued to cover quite a large part of the axis. For example, having a video recording that can be played back multiple times, and at various speeds, is very useful for carrying out LMA analysis. And, as we shall see in the next section, even a regular video recording can be used to extract meaningful quantitative motion data. Furthermore, it is also common to use video recordings as a reference when recording motion with sensor-based motion tracking technologies. In such cases, the video recording can be used to help the qualitative interpretation of numerical results.

For many applications, a regular video recording is often the easiest, fastest and cheapest solution to start working systematically with the study of music-related motion. Nowadays, everyone has access to high quality video cameras even in their mobile phones, and the cost of professional-quality video cameras is also within the reach for many. The main challenge for musicologists, then, is to record the video in a manner suitable for later analysis.

1.3.1 Recording video for analysis

One thing to bear in mind is that a video recording meant for analytical purposes is quite different from a video recording shot for documentary or artistic purposes. The latter type of video is usually based on the idea of creating an aesthetically pleasing result, which often includes continuous variation in the shots through changes in the lighting, background, zooming, panning, etc. A video recording for analysis, on the other hand, is quite the opposite: it is best to record it in a controlled studio or lab setting with as few camera changes as possible. This is to ensure that it is the *content* of the recording, that is, the human motion, which is in focus, not the motion of the camera or the environment.

Even though a controlled environment may be the best choice from a purely scientific point of view, it is possible to obtain useful recordings for analytical purposes also out in the field. This, however, requires some planning and attention to detail. Here are a few things to consider:

- Foreground/background: place the subject in front of a background that is as plain as possible, so it is possible to easily discern between the important and non-important elements in the image. For computer vision recordings it is particularly important to avoid backgrounds with moving objects, since these may influence the analysis.

- Lighting: avoid changing lights, as they will influence the final video. In dark locations, or if the lights are changing rapidly (such as in a disco or club concert), it may be worth recording with an infrared camera. Some consumer cameras come with a "night mode" that serves the same purpose. Even though the visual result of such recordings may be unsatisfactory, they can still work well for computer-based motion analysis.

- Camera placement: place the camera on a tripod, and avoid moving the camera while recording. Both panning and zooming makes it more difficult to analyse the content of the recordings later. If both overview images and close-ups are needed, it is better to use two (or more) cameras to capture different parts of the scene in question.

- Image quality: it is always best to record at the highest possible spatial (number of pixels), temporal (frames per second) and compression (format and ratio) settings the camera allows for. However, the most important is to find a balance between image quality, file size and processing time.

As mentioned earlier, a video recording can be used as the starting point for both qualitative and quantitative analysis. We will here look at a couple of different possibilities, moving from more qualitative visualisation methods to advanced motion capture techniques.

1.3.2 Video visualisation

Videos can be watched as they are, but they can also be used to develop new visualisations to be used for analysis. The aim of creating such alternate displays from video recordings is to uncover features, structures and similarities within the material itself, and in relation to, for example, score material. Three useful visualisation techniques here are *motion images*, *motion history images* and *motiongrams*.

Motion images

One of the most common techniques when working with motion analysis from video files is to create a *motion image* by calculating the absolute pixel difference between subsequent frames in the video file (Figure 1.3). The end result is an image where only the pixels that have changed between the frames are displayed. This can be interesting in itself, but motion images are also the starting point for many other video visualisation and analysis techniques.

Motion history images

Motion images only display the motion happening between two frames in a video file, but of-

Figure 1.3: A *motion image* is created by subtracting subsequent frames in a video file (Frame(2) - Frame(1)).

ten it is desirable to visualise motion over a period of time, say, a few seconds. This can be done through *motion history images* that display the temporal development of a motion sequence. There are numerous ways of creating such displays (Ahad et al., 2012), but many of the most common techniques are based on adding several motion image frames together. The usefulness of motion history images depend on carefully selecting a time window that fits the content of the motion, as can be seen in the examples of percussion strokes in Figure 1.4.

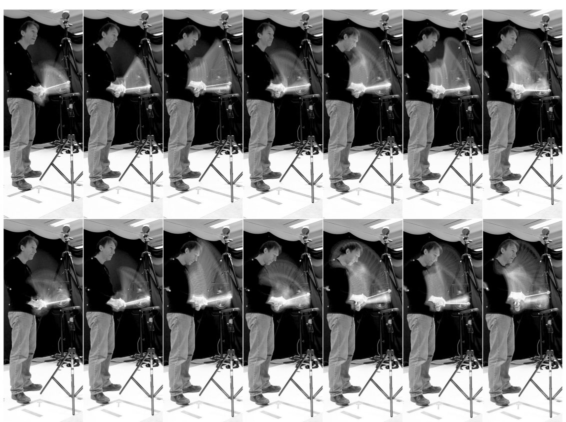

Figure 1.4: Individual motion history images of 14 separate percussion strokes allow for studying stroke heights and patterns. The images have been made by adding a motion history image on top of a picture of the scene, thus showing both motion features and the contextual information.

Motion history images have been used in various types of music analysis, such as in the study of music and dance (Camurri et al., 2003), and are also often to be seen in visual arts and creative practice.

Motiongrams

While a motion history image may reveal information about the spatial aspects of a motion sequence over a fairly short period of time, it is possible to use a *motiongram* to display longer sequences (Jensenius, 2013). Figure 1.5 shows a motiongram created from a dance improvisation recording, and this display is created by plotting the normalised mean values of the rows of a series of motion images. The motiongram makes it possible to see both the location and quantity of motion of a video sequence over time, and is thus an efficient way of visualising longer motion sequences.

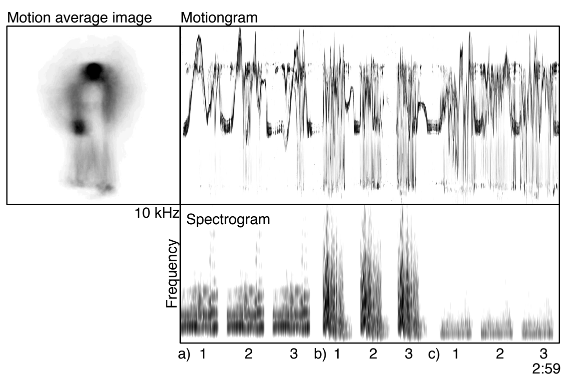

Figure 1.5: A motion average image and a motiongram of a 3-minute free-dance sequence to music make it possible to study both spatial and temporal features of a performance in connection to the spectrograom of the audio.

A motiongram is only a reduced display of a series of motion images, with no analysis being done. It might help to think of the motiongram as a display of a collapsed series of pictures, or "stripes," where each "stripe" summarises the content of a whole motion image.

Dependent on the frame rate of the video file, motiongrams can be created from recordings as short as a few seconds to several hours.

For short recordings it is possible to follow detailed parts of a body, particularly if there are relevant colours in the image, while motiongrams of longer recordings will mainly reveal larger sections of motion. Motiongrams work well together with audio spectrograms, and other types of temporal displays such as graphs of motion or sound features.

1.3.3 Computer vision

The broad field of *computer vision* (CV) is concerned with extracting useful information from video recordings. There is a lot of progress in the field, as summarised in (Moeslund and Granum, 2001; Moeslund et al., 2006; Rautaray and Agrawal, 2015), and we will here only look at a few possible methods.

Some basic motion features that are commonly used in music research are derived directly from the motion image. Since the motion image only shows pixels that have changed between the two last frames in a video sequence, the sum of the values of all these individual pixels will give an estimate of the *quantity of motion* (QoM). Calculating the QoM for each frame will give a numeric series that can be plotted and used as an indicator of the activity, such as illustrated in the graph of a dance sequence in Figure 1.6. Here it is possible to see where the dancer moved or stood still.

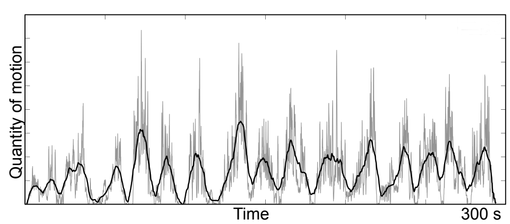

Figure 1.6: A plot of the quantity of motion for a 5-minute long dance sequence. The grey line is a plot of the tracked data, and the black line is a filtered version of the same data set.

The *centroid of motion* (CoM) and *area of motion* (AoM) are other basic features that can easily be extracted from a motion image, and the differences between them are illustrated in Figure 1.7. The CoM and AoM features can be used to illustrate *where* in an image the motion occurs as well the spatial displacement of motion over time.

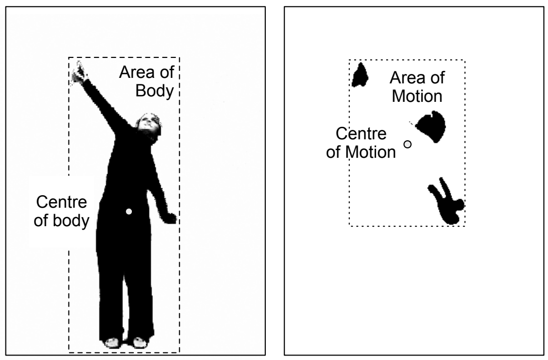

Figure 1.7: Illustrations of the area and centroid of body and motion.

The field of computer vision has diverged into a number of different directions over the years. Most notably, there are now numerous methods available for tracking bodies and body parts in space, many of which are also being used in music interaction and analysis through tools such as EyesWeb (Camurri et al., 2004b), GEM for PureData (Zmölnig, 2004) and Jitter for Max (Levin, 2006). In addition to using regular video cameras for such analyses, there are several different types of specialised cameras that are made particularly for computer vision methods, including:

- infrared cameras capturing only light in the infrared (non-visible) range. Such cameras can be very useful in musical applications, since they work well also in dark spaces (such as a concert hall) or in locations with changing lights (for example club or stage lights).

- time-of-flight cameras emitting a modulated acoustic signal and receive the reflected signal from which it is possible to measure the time it took for the signal to return to the camera.

- stereo-cameras employing the same strategy as human vision, using two cameras next to each other (like our eyes), and then using the differences between the two images to estimate motion and rotation.

There are also examples of the combination of these three capturing methods, as well as the use of multiple cameras around the capture space. By combining multiple cameras distributed in space it is possible to create true three-dimensional recordings. This can now be done without markers on the body (Sigal et al., 2010), but still the state-of-the-art when it comes to camera-based motion tracking are the marker-based systems using infrared cameras.

1.3.4 Infrared, marker-based motion capture

What many people refer to as "motion capture", can more precisely be described as optical, infrared, marker-based systems. Such systems usually consist of at least six cameras positioned around the capture space (Figure 1.8). Each of the cameras contains a ring of infrared light sources, and this infrared light is reflected on small markers and captured by the cameras. The system then calculates the exact position in space based on triangulating all the marker positions from each individual camera. The end result is a three-dimensional tracking of the markers in space, often captured at high speeds (more than 100 Hz) and at a high spatial resolution (in the range of millimetres). The captured points can be visualised directly or used as the basis for further analysis.

An advantage of marker-based motion capture systems is that they allow for reliable tracking of the position (and sometimes orientation) of individual body joints. Furthermore, together with force plates and physiological sensors such motion capture systems may provide very pre-

Figure 1.8: Setup for motion capture of pianist Christina Kobb. In addition to the cameras placed in front, there are also cameras placed behind the pianist to capture the motion in three dimensions.

cise and accurate data about the kinematics and kinetics of human motion.

There are, however, also some drawbacks of infrared, marker-based motion capture systems. The price tag is one, although such systems have become more affordable in recent years. More problematic from a user's perspective are the constraints enforced by the need of a controlled recording environment. Also, even though the markers used in such systems are small and lightweight, a recording session necessarily feels somewhat unnatural for the subjects. So when deciding on whether to use such a system for a study, it is important to weigh the benefits of a high-quality motion recording against the obvious limitations of such an "unecological" setup.

1.4 Sensor-based Motion Capture

Sensor-based systems are very different technologically to the camera-based solutions mentioned above. One main difference is that sensor systems are often modular, allowing for more flexibility when it comes to the types and numbers of *sensors* being connected to a sensor *interface*.

1.4.1 Sensor interfaces

The sensor interface is the unit that digitises the electrical signals coming from the sensors through an analog-to-digital converter (ADC), and which also contains the electronics needed for recording and/or connecting to a computer. Many sensors are analog and therefore have a theoretically infinite resolution, so the sensor interface is often the limiting factor of a digital sensor-based system. It is therefore important to choose a suitable sensor interface for the task (Wanderley et al., 2006), and some elements to consider are:

- Sampling rate: the speed at which the sensors are read, often ranging from a few to several thousands of samples per second.

- Bit rate: the "resolution" of each recorded sample, often ranging from 7-bit (like MIDI) to 10-bit or even higher.
- Connectivity: whether the interface is cabled or wireless, and the types of wired (USB, ethernet, etc.) or wireless (Wi-Fi, Bluetooth, etc.) connection being used.
- Power: whether the system needs external power or can run on batteries.
- Expandability: how many sensors that can be connected at the same time, with what types of connectors, etc.
- Size: anything from the smallest imaginable to fairly large-scale devices, typically dependent on all of the above together with the price.

Fortunately, sensor interfaces are constantly becoming smaller, faster, cheaper and more reliable, so there are numerous solutions to choose from, both ready-made and modular systems (Wanderley et al., 2006).

When it comes to the sensors themselves, there are many ways of classifying and evaluating these (Begg and Palaniswami, 2006; Zhou and Hu, 2008; Richard, 1987; Patel et al., 2012). The presentation below follows the classification of tracking technologies used by Bishop et al. (Bishop et al., 2001), in which the different systems are sorted according to the physical medium of the technology: *acoustic*, *mechanical*, *magnetic*, *inertial* and *electrical*.

1.4.2 Acoustic tracking

Motion capture systems based on acoustic sensing use the principle of time-of-flight mentioned above, which measures the round-trip an acoustic signal takes from being transmitted, reflected on an object and received back at the sensor. In the same way as for infrared, marker-based systems, it is necessary to have multiple transmitter/receiver pairs to be able to estimate the three-dimensional position of an object through triangulation of the individual points.

Acoustic sensor systems usually work in the ultrasonic range, and can therefore be used unobtrusively for music-related applications. A challenge, however, is that such systems are often less accurate and precise for the spatial range of music-related motion. Thus acoustic tracking has received relatively little attention in music research, although there are some examples of its application in new electronic instruments (Vogt et al., 2002; Ciglar, 2010).

There are also some examples of the use of audible sound for motion tracking. One example is the walking experiment presented in (Styns et al., 2007), in which microphones were placed in the socks of participants, and periodicity peaks from the waveforms' amplitudes were used to measure the frequency of walking in relation to music.

1.4.3 Mechanical tracking

Mechanical tracking systems are based on measuring angles, distances or forces, using potentiometers or sensors measuring flexing, stretching or force (Figure 1.9). One advantage with mechanical tracking is that the sensors can easily be added to or embedded in suits, shoes and gloves, thus allowing for fairly unobtrusive tracking of body motion. Due to this versatility, mechanical systems have been popular for a lot of music applications, both for analysis and particularly for various types of electronic instruments and in interactive performance systems (Miranda and Wanderley, 2006).

1.4.4 Magnetic sensors

Motion capture systems using magnetic sensing are based on the idea of measuring disturbances in the magnetic field around the sensor. The perhaps simplest and cheapest of such sensors, *magnetometers*, is similar to a compass and measure the direction and strength of the Earth's magnetic field. Magnetometers are widely used in

Figure 1.9: Examples of small and large force and bend sensors (left) and a custom-built glove with similar sensors built-in (right).

combination with inertial sensors (see Section 1.4.5).

Active magnetic systems, on the other hand, are based on setting up a magnetic field around an electromagnetic source. Then one or more sensors can be used to measure the position of an object through sensing the induced electric current at a point in space (Figure 1.10).

One advantage with magnetic tracking is that such systems can often track the three-dimensional orientation of an object (with each of the rotation axes often being referred to as *yaw*, *pitch* and *roll*) in addition to the three-dimensional position (with the axes being called X, Y, Z). Furthermore, magnetic tracking is generally quite accurate and precise, and allow for high sampling rates. Since such systems are cable-based, there are no problems related to occlusion of markers such as for camera-based systems.

Having to deal with cables is also one of the main negative aspects of magnetic systems, since the cables severely limit the physical range of the tracking. The perhaps biggest challenge, still, is that the systems are suspect to interference from ferromagnetic objects within the magnetic field (Vigliensoni and Wanderley, 2012). This can be highly problematic in a musical context, since many musical instruments, including pianos, are constructed fully or partly with metal components.

Magnetic sensing has been used for a number of musical applications, both performance (Marrin and Picard, 1998; Lin and Wu, 2000; Marshall et al., 2007) and analysis (Marshall et al., 2002; Maestre et al., 2007; Jensenius et al., 2008). However, with advancements in other tracking types, and particularly that of optical systems, the use of magnetic sensing has been in decline in recent years.

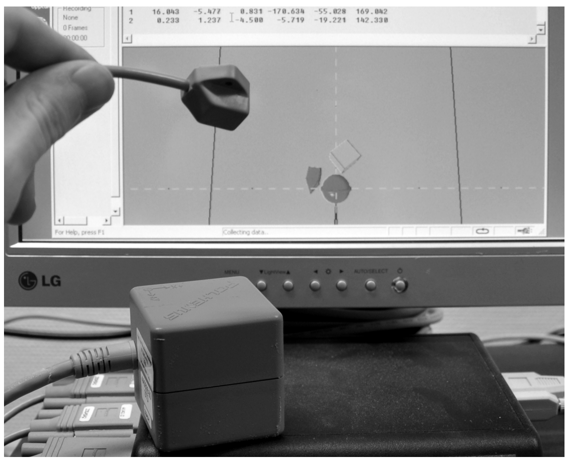

Figure 1.10: One six-dimensional sensor (held in the hand) and the electromagnetic transmitter (on table) from a Polhemus Liberty magnetic tracking system. The image on the screen displays the position and orientation of the object in realtime.

1.4.5 Inertial sensors

Besides camera-based systems, *inertial* sensing is currently one of the most popular types of motion tracking. 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" inside the sensor. Accelerometers measure the positional displacement of an object, while gyroscopes the rotational. By combining three accelerometers and three gyroscopes it is possible to capture both three-dimensional position and three-dimensional rotation in one sensor unit.

One of the most compelling features of inertial sensors is that they rely on physical laws (gravity), which is not affected by external factors such as ferromagnetic objects or lighting. Since they do not contain any transmitters (like infrared cameras and acoustic and magnetic sensors) they can also be made into very small and self-contained units (Figure 1.11), 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.

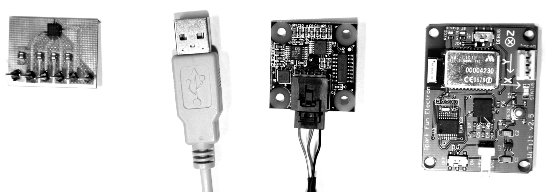

Figure 1.11: Different types of homemade and commercial inertial sensing systems.

The downside to inertial sensing is that accelerometers do not measure the *position* of objects, but rather the rate of change of the objects. It is possible to estimate the position through integration, and, combined with the data from gyroscopes and magnetometers, this can lead to satisfactory results (Wilmers, 2009). However, while the relative position estimates may be good, such position data often suffer from a considerable amount of drift (Skogstad et al., 2011). One way to overcome some of the drift problems is by combining inertial sensors with one or more of the other sensing solutions mentioned above (Welch and Foxlin, 2002). One example of this can be found in the commercially available motion capture suits from Xsens, that combine a number of accelerometers, gyroscopes and magnetometers with a kinematic model of the body (Figure 1.12).

Figure 1.12: Yago de Quay performing with the Xsens MVN suit in Oslo, Norway (de Quay et al., 2011).

1.4.6 Electrical sensors

Sensors measuring the electrical current of *biosignals* have become increasingly popular for musical applications over the last decades (Pérez and Knapp, 2008). Many such sensors, often also called *physiological* sensors, share the same sensing principle but are optimised to detect different types of biosignals:

- *Electromyograms* (EMG) are used to measure muscle activity, and are particularly effective on the arms of musicians to pick up information about hand and finger motion (Tanaka, 1993). Figure 1.13 shows an example of the commercially available MYO armband, with integrated EMG sensors, used for musical interaction (Nymoen et al., 2015).

- *Electrocardiograms* (EKG) measure the electrical pulses from the heart, and can be used to extract information about heart rate and heart rate variability. The latter has been shown to correlate with emotional state (Lee et al., 2005).

- *Galvanic skin response* (GSR) refers to changes in skin conductance, and is often used on the fingers. The GSR signal is highly correlated with emotional changes, and such sensors have been used to some extent in music research (Zimny and Weidenfeller, 1963; Craig, 2005) as well as in music performance (Ojanen et al., 2007; Pérez and Knapp, 2008). However, the signals may not be entirely straightforward to interpret, and elements like sweat may become an issue when worn for longer periods of time.

- *Electroencephalograms* (EEG) are used to measure electrical pulses from the brain, using either a few sensors placed on the forehead, or hats with numerous sensors included. Due to the weak brain signals, such sensors need to have strong amplifiers and are therefore also suspect to a lot of interference and noise. Nevertheless, such sensors have also been applied in both music analysis and performance (Miranda and Boskamp, 2005).

Figure 1.13: An example of music interaction with the commercially available MYO armband, which contains eight EMG sensors, accelerometers and gyroscopes (Photo: Kristian Nymoen.

1.5 Synchronisation and storage

As mentioned in the introduction, some important challenges for music researchers working with any of the above mentioned motion tracking solutions are to ensure synchronisation with other types of data and to store the data in a structured way (Jensenius et al., 2007). We will here briefly look into some possible solutions to these problems.

1.5.1 Motion data formats and protocols

A number of formats exist for storing motion data, most of which were designed to solve the needs of a specific hardware system and/or research problem. The BRD format, for example, is used with electromagnetic trackers from Flock of Birds, while the C3D format is used for infrared motion capture systems from Vicon (Motion Lab Systems, 2008). Several formats have also emerged for using motion capture data in animation tools, such as the BVA and BVH formats from Biovision, and the ASF and AMC formats from Acclaim, as well as formats used by animation software, such as the CSM format used by 3D Studio Max. There have been several attempts at creating XML-based standards for motion capture and animation data, such as the Motion Capture Markup Language (MCML) (Chung and Lee, 2004) and Multimodal Presentation Markup Language (MPML) (Tsutsui et al., 2000). But none of these nor the structures included in the MPEG-4 (Hartman et al., 2008) and MPEG-7 (Manjunath et al., 2002) formats seem to have achieved widespread usage. Of all these, the C3D format seems to be the one that is supported by most software, although its implementation and number of features varies somewhat between different software solutions.

Since none of the other formats available solve the problems of music researchers, there have been several initiatives to create new formats and standards specifically targeted at music applications. The Gesture Motion Signal (GMS) format (Evrard et al., 2006) is a binary format based on the Interchange File Format (IFF) standard (Morrison, 1985), and was mainly developed for storing raw motion data. The Gesture Description Interchange File Format (GDIF) was proposed as a structure for handling everything from raw motion data to higher-level descriptors within other data formats (Jensenius et al., 2006), and has been successfully used within namespaces for the Open Sound Control (OSC) protocol (Wright et al., 2003) and as an extension to the Sound Description Interchange Format (SDIF) (Burred et al., 2008). The Performance Markup Language (PML) was developed as an extension to the Music Encoding Initiative (MEI) (Roland, 2002) and focused on creating a structured approach to annotate performance data in relation to musical notation.

All of these formats and protocols solve some problems, even though none of them have emerged as de facto standards within the music research community.

1.5.2 Structuring multimodal data

Besides choosing a format for handling data, there are numerous conceptual and practical issues to deal with when working with music-related motion data. Different hardware devices use different protocols, formats and standards. They also work at different sampling rates, bit rates, with different numbers of sensors, different degrees of freedom, etc. Adding to this is the need to synchronise and store such data together with other relevant data, including MIDI data, audio and video files, as well as qualitative descriptors and various layers of analysis.

Fortunately, several software solutions have been developed for multimodal data acquisition and analysis, including the EyesWeb platform (Camurri et al., 2004a, 2007), the MoCap Toolbox for Matlab (Burger and Toiviainen, 2013), and various toolboxes for the graphical programming environment Max (Jaimovich and Knapp, 2010; Jensenius et al., 2008). There are also initiatives for creating online repositories for multimodal data storage and analysis, such as Repovizz (Mayor et al., 2011). In addition to helping in structuring and storing data, such systems also allow for carrying out collaborative and comparative studies on the same material.

1.6 Conclusion

In this chapter we have looked at a number of analytical approaches and technologies available for studying music-related motion, ranging from qualitative to quantitative, cheap to expensive, small to large, and simple to advanced. To simplify, we may differentiate between two main types of technologies used for motion tracking:

- Camera-based systems: this includes cameras of any sorts, recording in grey-scale, colour or infrared, and recording with or without markers on the body.
- Sensor-based systems: this includes all sorts of sensors, including acoustic, mechanical, magnetic, inertial and electrical.

It is impossible to give one answer to what type of method or technology to use; they all have their strengths and weaknesses. For that reason it is important to decide on the right analysis method for the research question at hand. The perhaps most decisive factor is whether to work in a laboratory setting or in a more ecological setting, say, a concert hall, as this will to a large extent guide which methods and tools to use.

More specialised motion capture solutions, like infrared, marker-based systems, electromagnetic systems and various types of inertial sensors, often provide high recording speeds, and high spatial accuracy and precision, but they also come with several drawbacks. Price is one, although such systems have quickly become more affordable. More problematic is that the person being studied has to wear markers/sensors on the body, something that may be both alienating to the performer and obtrusive to the motion being performed. This is particularly problematic when using electromagnetic and mechanical systems with fairly large and heavy sensors and cables. But even the lightweight, reflective markers typically used in optical infrared systems are noticed by performers, albeit to a lesser degree.

The perhaps biggest challenge with the larger systems is the non-ecological setting they require. Small inertial systems, on the other hand, can be mobile, wireless, and more or less invisible, although the physical sensors need to be placed on the body. So if the aim is to study musicians in a real-world concert situation, plain video recordings may be the only realistic solution.

Apart from selecting the right technology, however, is to have a clear plan for what to record, have a structured approach to synchronising the recordings with other data and media and to store the data using formats and protocols that maximise the potential for a diverse range of analytical approaches.

Bibliography

Ahad, M. A. R., Tan, J. K., Kim, H., and Ishikawa, S. (2012). Motion history image: its variants and applications. *Machine Vision and Applications*, 23(2):255–281.

Barlow, W. (1975). *Alexander-princippet.* Borgen forlag, Copenhagen.

Begg, R. and Palaniswami, M. (2006). *Computational intelligence for movement sciences: neural networks and other emerging techniques*. IGI Global, Hershey, PA.

Bishop, G., Welch, G., and Allen, B. D. (2001). Tracking: Beyond 15 minutes of thought. *SIGGRAPH Course 11*, pages 6–11.

Burger, B. and Toiviainen, P. (2013). MoCap Toolbox - A Matlab toolbox for computational analysis of movement data. In *Proceedings of the Sound and Music Computing Conference*, pages 172–178.

Burred, J. J., Cella, C. E., Peeters, G., Roebel, A., and Schwarz, D. (2008). Using the SDIF Sound Description Interchange Format for Audio Features. In *Proceedings of the International Conference on Music Information Retrieval*, pages 427–432.

Campbell, L. and Wanderley, M. (2005). The Observation of Movement. MUMT 609 report, McGill University.

Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., Ricci, A., and Volpe, G. (2004a). Toward real-time multimodal processing: EyesWeb 4.0. In *Proceedings of the Artificial Intelligence and the Simulation of Behaviour Convention*, pages 22–26, Leeds, UK.

Camurri, A., Coletta, P., Varni, G., and Ghisio, S. (2007). Developing Multimodal Interactive Systems with EyesWeb XMI. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 305–308, New York, NY.

Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *Applications of Affective Computing in Human-Computer Interaction*, 59(1-2):213–225.

Camurri, A., Mazzarino, B., and Volpe, G. (2004b). Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. In *Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, pages 460–467. Berlin Heidelberg: Springer-Verlag.

Choensawat, W., Nakamura, M., and Hachimura, K. (2014). GenLaban: A tool for generating Labanotation from motion capture data. *Multimedia Tools and Applications*, 74(23):10823–10846.

Chung, H. and Lee, Y. (2004). MCML: motion capture markup language for integration of heterogeneous motion capture data. *Computer Standards & Interfaces*, 26(2):113–130.

Ciglar, M. (2010). An Ultrasound Based Instrument Generating Audible and Tactile Sound. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 19–22.

Craig, D. G. (2005). An Exploratory Study of Physiological Changes during "Chills" Induced by Music. *Musicae Scientiae*, 9(2):273–287.

de Quay, Y., Skogstad, S., and Jensenius, A. R. (2011). Dance Jockey: Performing Electronic Music by Dancing. *Leonardo Music Journal*, 21:11–12.

Evrard, M., Couroussé, D., Castagné, N., Cadoz, C., Florens, J.-L., and Luciani, A. (2006). The GMS File Format: Specifications of the version 0.1 of the format. Technical report, INPG, ACROE/ICA, Grenoble, France.

Feitis, R. (1978). *Ida Rolf Talks about Rolfing and Physical Reality*. Harper and Row, New York.

Findlay, E. (1971). *Rhythm and Movement – Applications of Dalcroze Eurhythmics*. Summy-Birchard Inc., Miami, FL.

Godoy, R. I. and Leman, M., editors (2010). *Musical Gestures: Sound, Movement, and Meaning*. Routledge, New York.

Gritten, A. and King, E., editors (2006). *Music and gesture*. Ashgate, Hampshire.

Gritten, A. and King, E., editors (2011). *New Perspectives on Music and Gesture*. Ashgate, Hampshire.

Guest, A. H. (2004). *Labanotation*. Routledge, New York.

Haga, E. (2008). *Correspondences between music and body movement*. PhD thesis, University of Oslo, Oslo.

Hartman, E., Cooper, J., and Spratt, K. (2008). Swing Set : Musical Controllers with Inherent Physical Dynamics. In Camurri, A., Volpe, G., and Serafin, S., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 356–357.

Jaimovich, J. and Knapp, B. (2010). Synchronization of Multimodal Recordings for Musical Performance Research. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 372–374, Sydney.

Jensenius, A. R. (2013). Some video abstraction techniques for displaying body movement in analysis and performance. *Leonardo*, 46(1):53–60.

Jensenius, A. R., Camurri, A., Castagne, N., Maestre, E., Malloch, J., McGilvray, D., Schwarz, D., and Wright, M. (2007). Panel: the Need of Formats for Streaming and Storing Music-Related Movement and Gesture Data. In *Proceedings of the International Computer Music Conference*, pages 13–16.

Jensenius, A. R., Kvifte, T., and Godøy, R. I. (2006). Towards a Gesture Description Interchange Format. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 176–179.

Jensenius, A. R., Nymoen, K., and Godøy, R. I. (2008). A Multilayered GDIF-Based Setup for Studying Coarticulation in the Movements of Musicians. In *Proceedings of the International Computer Music Conference*, pages 743–746, Belfast, UK.

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., and Leman, M. (2010). Musical gestures: Concepts and methods in research. In Godøy, R. I. and Leman, M., editors, *Musical gestures: Sound, movement, and meaning*, pages 12–35. Routledge, New York.

Laban, R. and Lawrence, F. C. (1947). *Effort.* Macdonald & Evans.

Lee, C., Yoo, S.K., Park, Y.J., Kim, N.H., Jeong, K.S., and Lee, B.C. (2005). Using Neural Network to Recognize Human Emotions from Heart Rate Variability and Skin Resistance. In *Proceedings of the IEEE Engineering in Medicine and Biology*, pages 5523–5525, Shanghai, China.

Levin, G. (2006). Computer vision for artists and designers: pedagogic tools and techniques for novice programmers. *AI & Society*, 20(4):462–482.

Lin, E. and Wu, P. (2000). Jam Master, a Music Composing Interface. In *Proceedings of Human Interface Technologies*, pages 21–28, Vancouver, Canada.

Maestre, E., Janer, J., Blaauw, M., Pérez, A., and Guaus, E. (2007). Acquisition of Violin Instrumental Gestures Using a Commercial EMF Tracking Device. In *Proceedings of the International Computer Music Conference*, Copenhagen, Denmark.

Manjunath, B., Salembier, P., and Sikora, T. (2002). *Introduction to MPEG-7: Multimedia Content Description Interface*. New York: John Wiley and Sons.

Marrin, T. and Picard, R. (1998). The 'Conductor's Jacket': A device for recording expressive musical gestures. In *Proceedings of the International Computer Music Conference*, pages 215–219.

Marshall, M. T., Malloch, J., and Wanderley, M. M. (2007). Gesture Control of Spatialization. In *7th International Workshop on Gesture in Human-Computer Interaction and Simulation*, Lisbon, Portugal.

Marshall, M. T., Rath, M., and Moynihan, B. (2002). The Virtual Bodhran – The Vodhran. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 118–119, Dublin, Ireland.

Mayor, O., Llop, J., and Maestre, E. (2011). RepoVizz: A multimodal on-line database and browsing tool for music performance research. In *International Society for Music Information Retrieval Conference*, Miami, FL.

Miranda, E. R. and Boskamp, B. (2005). Steering generative rules with the EEG: An approach to brain-computer music interfacing. In *Proceedings of Sound and Music Computing*, Salerno, Italy.

Miranda, E. R. and Wanderley, M. M. (2006). *New digital musical instruments: control and* *interaction beyond the keyboard*. A-R Editions, Middleton, WI.

Moeslund, T. B. and Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3):231–268.

Moeslund, T. B., Hilton, A., and Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126.

Morrison, J. (1985). EA IFF 85: Standard for Interchange Format Files. Technical report, Electronic Arts.

Motion Lab Systems (2008). *The C3D File Format: User Guide*.

Nymoen, K., Haugen, M. R., and Jensenius, A. R. (2015). MuMYO - Evaluating and Exploring the MYO Armband for Musical Interaction. In *Proceedings of the International Conference on New Interfaces For Musical Expression*, Baton Rouge, LA.

Ojanen, M., Suominen, J., Kallio, T., and Lassfolk, K. (2007). Design Principles and User Interfaces of Erkki Kurenniemi's Electronic Musical Instruments of the 1960's and 1970's. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 88–93, New York, NY.

Parker, M. (1996). *Benesh Movement Notation for Ballet*. Royal Academy of Dance, London.

Patel, S., Park, H., Bonato, P., Chan, L., and Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 9(1):21.

Pierce, A. and Pierce, R. (1989). *Expressive Movement: Posture and Action in Daily Life, Sports, and the Performing Arts*. Perseus Publishing, Cambridge, MA.

Pérez, M. A. O. and Knapp, R. B. (2008). BioTools: A Biosignal Toolbox for Composers and Performers. In Kronland-Martinet, R., Ystad, S., and Jensen, K., editors, *Computer Music Modeling and Retrieval. Sense of Sounds*, number 4969 in Lecture Notes in Computer Science, pages 441–452. Springer Berlin Heidelberg.

Rautaray, S. S. and Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54.

Richard, W. M. (1987). A sensor classification scheme. *IEEE Transactons on Ultrasonic, Ferroelectrics, Frequency and Control*, UFFC-34(2):124–126.

Roland, P. (2002). The Music Encoding Initiative (MEI). In *Proceedings of the First International Conference on Musical Applications Using XML*, pages 55–59.

Schrader, C. A. (2004). *A Sense of Dance: Exploring Your Movement Potential*. Human Kinetics, Champaign, IL.

Sigal, L., Balan, A., and Black, M. (2010). Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision*, 87(1):4–27.

Skogstad, S., Nymoen, K., and Høvin, M. E. (2011). Comparing Inertial and Optical MoCap Technologies for Synthesis Control. In *Proceedings of Sound and Music Computing*, pages 421–426, Padova, Italy.

Styns, F., van Noorden, L., Moelants, D., and Leman, M. (2007). Walking on music. *Human Movement Science*, 26(5):769–785.

Tanaka, A. (1993). Musical technical issues in using interactive instrument technology with application to the BioMuse. In *Proceedings of the International Computer Music Conference*, pages 124–124, Waseda, Japan.

Tsutsui, T., Saeyor, S., and Ishizuka, M. (2000). MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions. In *Proc.(CD-ROM) WebNet*, pages 30–37.

Van Dyck, E., Maes, P.-J., Hargreaves, J., Lesaffre, M., and Leman, M. (2013). Expressing induced emotions through free dance movement. *Journal of Nonverbal Behavior*, 37(3):175–190.

Vigliensoni, G. and Wanderley, M. M. (2012). A Quantitative Comparison of Position Trackers for the Development of a Touch-less Musical Interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Ann Arbor, MI.

Vogt, F., Mccaig, G., Ali, M. A., and Fels, S. S. (2002). Tongue 'n' Groove: An Ultrasound based Music Controller. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 181–185, Dublin.

Wanderley, M. M. and Battier, M., editors (2000). *Trends in Gestural Control of Music*. IRCAM - Centre Pompidou, Paris.

Wanderley, M. M., Birnbaum, D., Malloch, J., Sinyor, E., and Boissinot, J. (2006). SensorWiki.org: A Collaborative Resource for Researchers and Interface Designers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 180–183, Paris, France.

Welch, G. and Foxlin, E. (2002). Motion tracking: no silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 22(6):24–38.

Wilmers, H. (2009). Bowsense – an Open Wireless Motion Sensing Platform. In *Proceedings of the International Computer Music Conference*, pages 287–290, Montreal, Canada.

Wright, M., Freed, A., and Momeni, A. (2003). OpenSound Control: State of the Art 2003. In

Proceedings of the International Conference on New Interfaces for Musical Expression, pages 153–159, Montreal.

Zhou, H. and Hu, H. (2008). Human motion tracking for rehabilitation—A survey. *Biomedical Signal Processing and Control*, 3(1):1–18.

Zimny, G. H. and Weidenfeller, E. W. (1963). Effects of music upon GSR and heart-rate. *The American journal of psychology*, pages 311–314.

Zmölnig, J. M. (2004). Gem for pd - Recent progress. In *Proceedings of the International Computer Music Conference*, Miami, FL.