No cover image available

The Oxford Handbook of Music Performance, Volume 2

Gary E. McPherson (ed.)

https://doi.org/10.1093/oxfordhb/9780190058869.001.0001

Published: 2022 **Online ISBN:** 9780190058890 **Print ISBN:** 9780190058869

CHAPTER

22 Motion Capture of Music Performances

Marcelo M. Wanderley

https://doi.org/10.1093/oxfordhb/9780190058869.013.30 Pages 465–C22.P182

Published: 18 March 2022

Abstract

Motion capture (mocap)—the recording of three-dimensional movement using high-accuracy systems—has become a standard research tool in the analysis of music performances in the last two decades. A variety of systems is currently available, ranging from optical, multi-camera (passive and/or active) infra-red systems and inertial systems (using orientation sensors) to electromagnetic trackers providing six degrees-of-freedom (DoF) measurement per marker/sensor. Recent advances in technology have made many of these systems more affordable, allowing access to a large research community. Music-related mocap applications include the tracking movements of solo or group, beginner, or expert performers and instruments for teaching performance skills, comparing movement strategies across performers, generating movement synthesis parameters in animation, and use in real-time music interaction. This chapter introduces the basic concepts behind motion capture, reviews the most common mocap technologies used in the study of music performance, and presents several examples of research, pedagogy, and artistic uses. Mocap of single acoustic instrument performances is reviewed, including violin, cello, piano, clarinet, timpani, and acoustic guitar, as well as examples of mocap of multiple instruments. Finally, we discuss the limitations of mocap and possible solutions to overcome them.

Keywords: motion capture, music performance, movement analysis, musical instruments, musical interaction

Subject: Applied Music, Music

Series: Oxford Handbooks

Introduction

MOVEMENT is an integral part of music. In fact, it can be difficult to conceive music without movement, at least in the case of music created by acoustic musical instruments. The assessment of performer movements is an essential aspect of music performance, an integral step in instrument pedagogy, but also an aid in preventing injuries and analyzing differences in style or creating novel digital musical instruments (Miranda & Wanderley, 2006).

Physical movements that involve or evoke meaning can be called *gestures*. The concept of gesture in music (Cadoz & Wanderley, 2000) is useful as it describes different movements and actions musicians do while performing. *Sound-producing gestures* are the movements made by musicians to effectively produce sounds, but many other gestures are also part of music performances, though they might not be directly involved in sound production (Davidson, 1994; Wanderley, 2002). They can help performers communicate with each other, facilitate the execution of sound-producing gestures, or be induced by music (Jensenius et al., 2010).

The study of performer movements can take various forms, and an excellent review of multiple perspectives is provided by Dahl et al. (2010), though it does not explicitly address movement-measuring devices and techniques. Two more recent works discuss several methodologies to study movement in music.

Goebl et al. (2014) take a holistic view, presenting techniques for the analysis of *movements*, *audio*, and *perceiver response*, addressing the physical world, the experiential world, and the world of thought and knowledge, respectively (p. 221). This is a very interesting approach, as movement, sound, and perception are all intertwined in music performance. An overview of various strategies to measure movement is
p. 466 provided, from ↳ video-based to sensor-based ones, with a section on the three-dimensional motion capture mostly focusing on systems based on infrared (IR) cameras (optical mocap).

Jensenius (2018), on the other hand, focuses specifically on methods for studying body motion in music. An essential contribution of this work is the list of *key challenges* to be considered when studying movement in music. An exhaustive list of methods is then presented, including movement annotation and video analysis using different software tools. A review of motion capture tools is introduced, from IR marker-based to biosignal interfaces, showing the wide range of solutions used to study performer movements.

The current chapter expands on these works by focusing on motion capture's practical details using mostly optical, electromagnetic, and inertial systems. It presents the basic notions underlying the technologies of such systems and several examples of mocap of music performances using them, as well as practical issues that researchers and performers face when studying movement in music using motion capture.

What Is Motion Capture?

Wikipedia defines motion capture (mocap) as "the process of recording the movement of objects or people" (Wikipedia, 2020). This recording is typically done today using digital technologies.

In music performance applications of *mocap*, we are interested in measuring performers' three-dimensional movements and (most of the time) their instruments. Though not necessarily a recent technique, with early examples going back a long time,[1] the use of mocap technologies has become widespread during the last two decades, allowing researchers to assess the intricacies of performers' technique accurately.

When carrying out motion capture, measuring performer movements typically involves defining one or more locations in the body that will be tracked. This tracking is done by placing artifacts in these body locations, commonly referred to as *markers*, though not all mocap systems require markers.

Markers, anyone?

Note that *marker* is a generic term that encompasses many technologies. In passive optical mocap systems, markers are artifacts built with light-reflecting tape that "shine" when IR light is present in the capture space. In some systems, markers are actually sensors measuring physical properties and producing an
p. 467 electrical signal proportional ↳ to the property measured; for instance, acceleration and angular momentum in inertial systems. In active optical mocap systems, markers are light-emitting diodes that send IR light to the cameras. Details of these systems will be discussed later in this chapter.

Marker sizes might vary substantially. This is an important factor to consider depending on the body part to be tracked. For instance, when measuring finger movements, small markers are needed to allow for unobtrusive measurements. In optical systems relying on passive markers, marker sizes range from 2.5 mm to around 2 cm. Electromagnetic systems also provide different marker sizes, typically larger and tethered (wired), which can be a limitation in some measurements.

No markers, please!

Some mocap systems, collectively known as *markerless mocap systems*, do not rely on markers (nor on sensors placed in the body) to record human body movements. Examples include depth cameras and systems that use several standard video cameras and computer vision techniques to track the human body. The advantage of such systems is the absence of artifacts to be tied to the body location to be tracked. Their disadvantage is that it might not be straightforward to track instruments and other devices.

The choice of a mocap system is an important issue, as no single technology fits all needs (Welch & Foxlin, 2002). Important issues to consider when choosing a mocap system will be discussed later. It is also important to point out that one does not necessarily need to purchase one's own mocap system. Several academic laboratories in the fields of kinesiology, rehabilitation, sport sciences, computer graphics, and music might already have mocap systems that can be used for one-off or not too frequent captures, as the investment to purchase and run one's own system can be substantial, from a few thousand to tens of thousands of dollars.

Why Use Mocap for the Analysis of Music Performances?

Imagine a musician learning to perform a piece on a given instrument, such as a specific violin bow movement sequence. Some of the actions required in the performance may not be obvious for their technical level, so studying a professional performer's gestures might provide a useful reference for what needs to be achieved. While watching the movements of a skilled performer with the naked eye, it might be challenging to grasp the intricacies of the motor control strategies at play. Video cameras can be of help here, as movements can be recorded and played back in slow motion to show them in detail. Nevertheless, depending on the movement executed, the camera angle, and light conditions, video recordings might not be sufficient to display intricate movement patterns.

p. 468 By placing one or more markers in the joints of the limbs to be studied, mocap of one or several performers can be made, and the full movement of the limb reconstructed in 3D. One can look at it from any perspective (e.g., frontal, rear, top, bottom, sideways). For instance, in a violin performance, the bow's exact position and orientation can be recorded and displayed with respect to the instrument. If a force sensor is available, and its output is synchronized with the mocap system, the force between the bow and the string can also be measured, or the bow force can be estimated from the motion capture data itself (Marchini et al., 2011). The sum of these measurements allows for a much more complete understanding of the interaction between the bow and the violin than what would be possible with the naked eye or a single video camera.

Schoonderwaldt and colleagues have developed such a system (Schoonderwaldt & Wanderley, 2007; Schoonderwaldt & Demoucron, 2009). Using an optical passive marker mocap system and force sensors, the following bowing parameters were obtained: *bow position*, *velocity*, and *acceleration*; *bow-bridge distance*; *bow force*; as well as *skewness*, *inclination*, and *tilt of the bow*. A screenshot of the visualization of the various parameters is shown in Figure 22.1.[2]

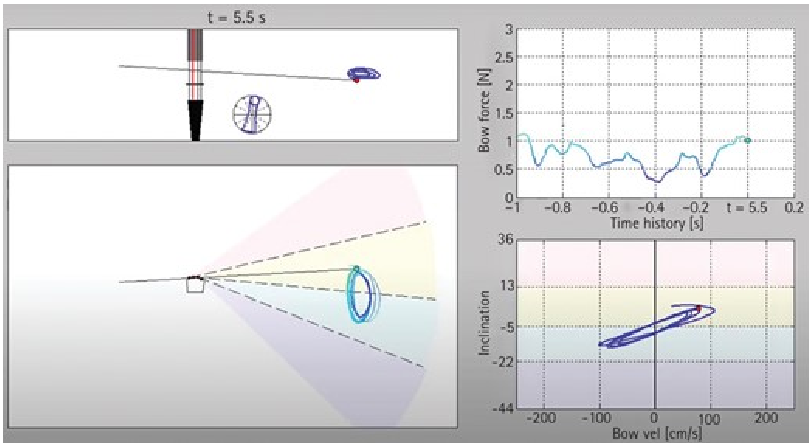
Related research has been carried out by Shan and Visentin (2003), Maestre and colleagues (2010), and in the context of the European project i-Maestro (Ng & Nesi 2008), among others.

How to Do Mocap in Practice

There are various ways of studying performer movements using mocap. For instance, when studying clarinettists, one can focus on the movement of the fingers (Palmer et al. 2009), the instrument (e.g., the clarinet bell) (Wanderley, 2002), the musician's head (Teixeira et al., 2015), the instrument and select locations in the body (Wanderley et al., 2005; Wanderley & Vines; 2006; Rodger et al., 2012), or the full body (Nusseck & Wanderley, 2009; Desmet et al., 2012; Nusseck et al. 2017; Weiss et al., 2018).

Defining a motion capture setup for a given experiment is not necessarily a trivial task. It is worth spending time designing a mocap recording session rather than to quickly improvise it and likely have to redo part or all of the measurements. *Rule of thumb*: pilot experiments are fantastic! It is not unusual, at least for beginners, to plan for long, complicated, one-size-fits-all sessions aiming to capture all that is possible simultaneously. Though this can be a valid strategy, for instance, when looking at multiple levels of movements (e.g., gestures, respiration, and facial expressions) in a given performance, there is a high risk of failure when pursuing such a strategy. Even if successful, because of the massive amount of data
p. 469 generated and the time required to pre-process ↳ and analyze it, many times at least part of the capture will likely stay untouched in a hard drive for years, perhaps forever, as the analysis of the data might take much longer than initially envisaged. After capturing and analyzing data, one might realize the need to gather new captures not initially foreseen. It is an excellent strategy to design a short and focused session to answer one research question and then carry out follow-up studies if needed.

Figure 22.1.



Screenshot of a video displaying a violinist's movement captured using a six-camera passive IR optical mocap system (Schoonderwaldt & Wanderley, 2007). Four panels are shown. From the top left, clockwise: The violin is seen from above, showing the position and orientation of the bow (more specifically, of the bow-hair ribbon from the frog to the tip). A dot indicates the frog, and its past trajectory is shown. The keyhole-like shape inside the small circle indicates the bow tilt as if looking at the frog along the bow direction. The string being played is indicated in grey. The second panel shows the bow force in newtons, measured by calibrated strain gauge sensors placed close to the frog and the bow's tip. The third panel shows a phase diagram with bow inclination and velocity, a useful way to display frog movement patterns. The fourth and largest panel shows the violin seen from the angle of its bridge. Each string angular zone is associated with a color, though not visible in this black-and-white picture. The dashed lines show string crossing angles. This figure shows the performer playing the A and D strings.

Used with permission from Erwin Schoonderwaldt.

Many factors will influence the choice of a specific motion capture technology for a given experiment. These include whether one is tracking movements of a single musician or of a group of performers, the complexity of the movements to be captured, the instruments to be played, and the level of detail required in the capture, such as the "speed" of the acquisition (more accurately, the frequency of measurements) in frames per second or in Hz. Other issues, including the context where this capture will occur and the level of technological interference that can be tolerated, are also essential factors. For instance, is the experiment
p. 470 taking place in a research laboratory or during ↳ a live concert with an audience in a hall? What is the complexity of the movements to be captured (large arm gestures or small finger movements)? Do factors such as the obtrusiveness due to marker size, marker cables, or the need to wear a Lycra suit prevent a natural performance? Finally, cost, equipment and room size, and system portability can be decisive factors in choosing to use one technology instead of another for a given experiment.

Which acquisition frequency should one use?

Inertial and electromagnetic mocap systems have acquisition frequencies varying from dozens to a few thousand Hz (measurements per second), independently of the gestures to be captured. On the other hand, optical systems can have much higher acquisition frequencies, currently up to 10,000 frames per second, typically with reduced resolution. Choosing a good frame rate to capture a given movement is an important question when setting up an experiment (Song & Godøy, 2016).

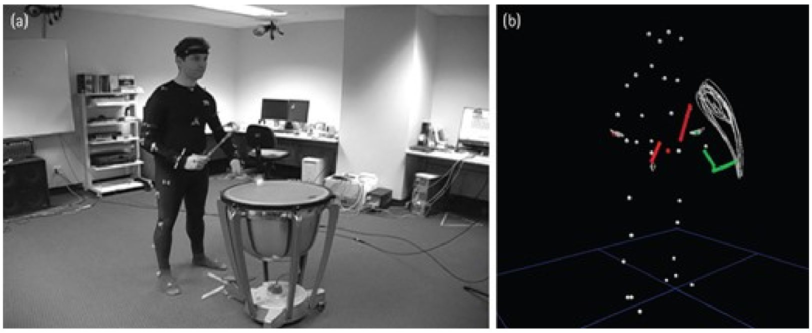
In these systems, the camera frame rate implies the maximum volume where movements can be captured: higher frequencies reduce the maximum possible volume where movements can be captured. This may not be a problem when studying the body sway movements of clarinettists, for instance, when one can choose a camera frame rate of 60 or 100 fps, largely enough to capture sway movements. However, it can become an issue when capturing percussionist's movements, when acquisition frequencies of 500 frames per second or more might be needed to capture in detail the impact movements of a drumstick hitting the drum membrane. For example, Bouënard et al. (2011), using a Vicon System 460, chose a frame rate of 250 fps as a compromise between a high frame rate and the capture of movements of the entire body of the performer (and drumsticks) using a mocap system with six IR cameras and passive markers in a medium-sized room (approximately 30 m^2), cf. Figure 22.2. Ideally, a frame rate of 500 fps would be needed, but at the expense of body movement coverage. Newer systems allow for much higher resolutions, for instance, a full field of view (FOV) resolution of 5M pixels at 700 fps (Qualisys Arqus A5 camera).

Portable and wireless, really?

Here, it is important to distinguish between *genuinely portable* and *transportable* mocap systems. Inertial and electromagnetic systems are mainly portable, in that they can relatively easily be moved from one environment to another (ideally, in one suitcase). On the other hand, most camera-based systems are *transportable*; that is, they *can* be moved but usually require the transportation of many items (boxes), lengthy setups, and multiple calibration steps.

Another distinction would be between *completely wireless* (or *wire-free*) and *wireless* mocap systems. Many p. 471 manufacturers advertise their systems as wireless—which would ↳ ideally mean that markers do not have any wires attached. Some wireless systems have wired (tethered) markers connected to a wireless measurement unit communicating with the central computer, so while markers still have wires, these systems allow for some degree of movement freedom if compared with wired mocap systems. Recent inertial mocap systems are typically truly portable and completely wireless, making them an excellent option when these characteristics are needed.

Figure 22.2.



Left: Percussionist Fabrice Marandola in a motion capture session at Input Devices and Music Interaction Laboratory (IDMIL), McGill University, around 2008, captured by Alexandre Bouënard. Note the number of markers placed on the Lycra suit, head, feet, and mallets. The mocap system used in this experiment was a six-camera Vicon System 460. Right: Data obtained from the capture. Note the points representing the markers on the body and the sticks connecting two markers for the mallets and the wrists (represented by the grey sticks). The spatial trajectory of the tip of the left-hand drumstick is shown. There are two distinct trajectories for the left mallet tip, depending on whether the percussionist performs an impact movement with it—the long trajectory, the actual strike—or with the right mallet. In the latter case, the left mallet's tip performs a shorter trajectory, which somehow simultaneously follows the strike performed with the right hand. This effect is not reversible, that is, the trajectory of the tip of the right mallet (dominant hand) does not follow the left one (non-dominant hand).

Number and location of markers

The number and location of markers may vary substantially with the mocap experiment's goal, which depends on the types of movements to be recorded (see summary of mocap studies in Table 22.1). For instance, to measure the orientation of a pianist's left wrist with an optical mocap system, a minimum of three markers are needed so that a plane can be defined in space instead of a point (one marker) or a line (two markers). If one is to measure the same pianist's arms and upper body movements, dozens of markers will likely be needed. Using an electromagnetic mocap system, one marker (sensor) would be needed to
p. 472 measure the wrist's orientation and perhaps around eight markers ↳ to measure the arms and upper body movements. For this example, an inertial mocap system would require a similar number of markers to the electromagnetic system. This variability happens because a marker measurement may yield, depending on
p. 473 the mocap technology used, a position (optical systems) or orientation (inertial systems) in three ↳ dimensions or a full six degree-of-freedom (DoF) output, that is, three positions: X, Y, Z, *and* three orientations: pitch, yaw, row (electromagnetic systems).

Table 22.1. A few examples of different numbers of markers used in the mocap of musical performances. Note the wide variability in the number of markers and systems used to carry out the various experiments.

Authors	Year	# Markers total	# Markers, body part	# Markers, instrument	System/Technology
Rodger et al.	2012	6	5 (head, 2 in elbows, and 2 in knees)	1 (clarinet)	Qualisys/Passive IR
Wanderley	2002	10	8 on one side of the body	2 (clarinet)	Optotrak/Active IR
Maes et al.	2015	10	4 (right hand, elbow, shoulder, and forehead)	6 (4 on the cello and 2 on the bow)	Qualisys/Passive IR
Palmer et al.	2009	12	8 (on fingertips)	4 (clarinet)	Optotrak/Active IR
Thompson & Luck	2012	28	26 (upper body)	2 (keyboard)	Qualisys/Passive IR
Thompson et al.	2017	31	26 (joints)	5 (3 on the violin and 2 on the bow)	Qualisys/Passive IR
Rozé et al.	2018	38	29	9 (cello and bow)	Vicon/Passive IR
Goebl & Palmer	2008	40	25 (fingernails, joints, and hand)	15 on piano keys	Vicon/Passive IR
Nusseck & Wanderley	2009	41	39 (full body)	2 (clarinet)	Vicon/Passive IR
Massie-Laberge	2019	61	57 (full body)	4 (piano)	Qualisys/Passive IR
Cossette et al.	2008	89	89 (chest, front and back) of flautists	–	BTS/Passive IR
Cossette, Palmer, & Wanderley	ca. 2010	119	119 (89 on chest, 14 on face, and 16 on arms and head) of a singer	–	BTS/Passive IR
Maestre et al.	2017	140	Upper body of performers	Instruments and bows (string quartet)	Qualisys/Passive IR
Maestre et al.	2017	8	–	Instruments and bows (string quartet)	Polhemus/EMF

Even when using optical mocap systems, rather small sets of markers might be sufficient in many studies. For measuring clarinettists' movements, Wanderley (2002) placed eight markers on one side of the performers (three on the head, one in the shoulder, elbow, wrist, hip and above the knee) and two markers on the instrument (clarinet bell and mouthpiece), tracked by an Optotrak 3020 active IR optical mocap system. Rodger et al. (2012) used six passive (reflective) markers (on the head, elbows, knees, and the clarinet bell) with a Qualisys passive IR optical mocap system. To study the movements of fingers in clarinet performances, Palmer et al. (2009) used an Optotrak 3020 active IR mocap system to track twelve markers: eight markers on performers' fingertips (one on each finger, except for the thumbs) and four attached to the clarinet (two on the bell, one above the bell, and one below the mouthpiece).

For the study of cellist's movements, Maes et al. (2015) used ten passive markers (3 mm) placed on the limbs and forehead of cellists (right hand, elbow, shoulder, and forehead), the instrument (four markers), and the bow (two markers). These were tracked by a twelve-camera Qualisys system to study working memory's role in the temporal control of cellists' discrete (staccato) and continuous (legato) movements.

Nevertheless, extensive sets of markers might be needed for specific experiments. Goebl and Palmer (2008), when studying right-hand finger movements in piano performances, used a six-camera Vicon system 460 to track twenty-five 4 mm markers on the fingernails, joints, and the hand of pianists, plus fifteen markers on piano keys. Thompson and Luck (2012) used twelve Qualisys ProReflex cameras to track twenty-six markers placed on pianists' upper body, plus two markers at the keyboard's ends.

Thompson et al. (2017), when studying interpersonal coordination in dyadic violin performances using a Qualisys system, placed twenty-six markers on the joints of each performer and five markers on the instruments and bows (three markers on the body of the violin and two markers on the bow). Rozé et al. (2018), using a Vicon system, placed twenty-nine passive markers on the performer (cellist) and nine markers on the instrument and bow to study postural constraints on performance. Nusseck & Wanderley (2009) used a six-camera Vicon system to track thirty-nine markers on clarinetists following the marker placement indicated in the Vicon Plug-In Gait model,[3] plus two markers on their instrument. Massie-Laberge placed fifty-seven markers on pianists and four markers on a digital piano for the analysis of performances of Romantic-era piano pieces using a seventeen Oqus camera Qualisys system (Massie-Laberge 2019).

For the analysis of breathing patterns in flute performances, Cossette et al. (2008) used eighty-nine 6 mm passive markers on the chest (front and back) of flautists, tracked by seven IR cameras of an Elite System, from BTS Bioengineering,[4] five in front of and p. 474 ↳ two behind the performer. They also measured simultaneous electromyography (EMG) activity of the respiratory muscles, mouth pressure, and recorded sound.

In a pilot experiment around 2010 by Cossette, Palmer, and Wanderley at the Centre for Interdisciplinary Research in Music Media and Technology (CRIMMT), McGill University, 119 markers of different sizes were placed on a singer to simultaneously measure upper body and head movements, respiration, and facial expression. The acquisition system consisted of a nine-camera SMART OEP system from BTS bioengineering, with three cameras focusing on the face and six cameras devoted to the rest of the body. Of the 119 markers, eighty-nine markers were placed on the chest, fourteen on the face, and sixteen on the singer's arms and head. This pilot's goal was to verify the feasibility of using such a reduced camera set to measure the three regions of the performer body (yes, it worked!).

Capturing the movements of more than one performer at the same time naturally tends to increase the minimum number of markers needed. When measuring the movements of a string quartet, Maestre and colleagues placed 140 passive markers on the upper body of the performers and their instruments tracked by a 32 Oqus 400 IR camera Qualisys system, while simultaneously using a Polhemus Liberty eight-tracker at the International Laboratory for Brain, Music, and Sound Research (BRAMS) at the Université de Montréal (Maestre et al., 2017). Eight Polhemus markers were placed on the four instruments and respective bows.

The synchronization strategy used in this rather complex setup, which also included audio and video recording, is discussed in the subsection titled "Synchronization issues."

Nevertheless, elaborate group performances can be recorded with somewhat limited setups, depending on the study's goals. Hilt et al. (2019) used a seven-camera Qualisys system to study ensemble coordination. Each musician wore a cap with three passive markers, while additional markers were placed on the violin bow and on the conductors' baton.

When It Can Be Hard to Use Mocap

There are several limitations to the use of motion capture systems in musical performance contexts. As expected, motion capture systems excel at measuring, well, *motion*, so performer actions that mainly involve forces but not much movement are not well captured or not captured at all with mocap. For instance, consider Figure 22.2. The large arm, upper body, and mallet movements are well captured with a passive mocap system at 250 fps, though with some limitations on the fidelity of the mallet movement's tip. On the other hand, the subtleties of the performer grip, that is, the way the fingers and hand hold the mallet, are not captured. Expert performers developed advanced grip techniques to perform different strikes, which involve complex finger movements and forces applied on the stick that will not be measurable with a motion capture system. To p. 475 ↳ capture such actions, one could use EMG sensors and/or calibrated force sensors in the mallets while synchronously capturing movements using mocap.

Marker-based systems have further limitations. For instance, placing markers on body joints can be obtrusive, many times jeopardizing or negatively impacting the performance. This can be the case with electromagnetic mocap systems, where marker sizes and wires might limit their application. Most important, markers should only move with the limbs they are attached to, not with the performer's clothes. Therefore, it is usually the case that markers are attached to limbs using Velcro straps or many types by requiring performers to wear Lycra suits to avoid extraneous movements (cf. Figure 22.2, not the standard dress choice in concerts!). Though such elastic fabric allows for an acceptable solution in many cases, movements involving high acceleration and impacts might not be adequately captured due to *soft tissue artifacts* (STA) or measurement errors introduced because markers are attached to the skin and not directly to the bones. Advanced analysis techniques can mitigate this issue but are beyond the scope of this chapter.

Besides, many systems require *line-of-sight* between markers and the measurement devices. For instance, markers need to be seen by the cameras when using optical mocap systems. Several cameras will need to be placed around the performers and at each time "see" a marker so that 3D measurements can be made. Depending on the movements' complexity or the number of performers to be captured, twelve or more cameras might be needed. Similarly, body parts need to be visible by multiple cameras when using markerless mocap systems. Finally, the environment should be somewhat *controlled*; in other words, one needs to remove possible sources of light or reflective surfaces that might create spurious reflections when using optical mocap or place electromagnetic sources in positions as far as possible from metals in the room to avoid measurement distortion. These issues add constraints to the performance that might not be acceptable in many cases.

Choosing a System: Main Mocap Technologies

So which system to choose? There is no simple answer because, as mentioned, every system has advantages and drawbacks. To check the advantages of the various systems, one can check manufacturers' webpages, which tend to present products as the most advanced, most used, or most flexible. It is much harder, however, to learn about systems' drawbacks this way.

This section presents an overview of the most common technologies used in music performance mocap contexts (optical IR, electromagnetic, and inertial). Several works present detailed information on these and
p. 476 other mocap technologies and are an ↳ excellent resource for the interested reader, for instance, Burdea and Coiffet (2003) and Fuchs et al. (2011).

Infrared optical systems

Infrared (IR) optical systems are typically based on multiple cameras that track several markers. These mocap systems can use *active markers—light emitting diodes* (LEDs), which emit a signal to be detected by cameras—or *passive markers—objects that reflect light emitted* by LEDs placed around the lens of the cameras. Though widely referred to as IR optical mocap systems, such systems typically allow for a choice of light wavelength between *visible red*, *near infrared*, and *infrared*, making the light shed by the LEDs more or less visible. These systems are commonly referred to as *infrared optical mocap*.

In IR optical mocap systems, the three-dimensional position (X, Y, and Z) of each marker is obtained from the two-dimensional information provided by the multiple cameras. In a sense, one can think about this process as filtering the visible information in the picture (removing background, objects, etc.) and tracking the individual points from one frame to another (cf. Figure 22.3).

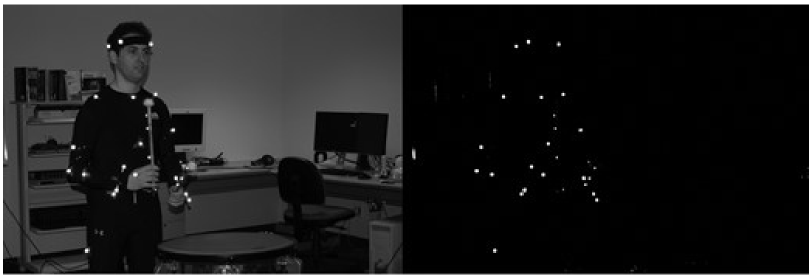
IR mocap systems using passive markers (aka *passive IR systems*) do not require wires (or batteries), can use markers of various sizes, and do not have limitations in the number of markers used at a given time. Passive IR systems are perhaps the most commonly used mocap systems in analyzing musical performances, with dozens of published works using this technology. Leading manufacturers of passive IR systems used in music performance studies include BTS Bioengineering, Motion Analysis Corporation, Natural Point, Qualisys, and Vicon.

Systems using active markers (henceforth, *active IR systems*) are less sensitive to spurious reflections, typically require less time for calibration, and are more immune to data loss during acquisition due to their greater tolerance to ambient light. Active IR systems can thus be used in brighter light conditions. For example, outdoor captures that can be challenging for passive systems, though recent developments have presented passive systems as usable outdoors. Many works on motion capture of solo and group performances used active IR systems (Winold et al., 1994; Wanderley, 2002; Palmer et al., 2009; Keller & Appel, 2010; Mota et al., 2017). Leading manufacturers of active IR systems used in music performance studies include MetaMotion, Northern Digital (NDI), and Phoenix Technologies Inc (PTI).

A few manufacturers propose systems that can work with passive or active markers, one type of marker at a time, for instance, Natural Point, Northern Digital (NDI), and Qualisys.

Although some active IR systems provide wireless markers (using batteries attached to the marker), the size and weight of these markers are usually more significant than the smallest passive markers available, sometimes limiting their use in experiments that require the use of many markers measuring small movements, for example, in capturing facial expressions.

p. 477 **Figure 22.3.**



Obtaining marker information from 2D images: Starting with a color image of the performer (and background), color is removed (left picture), then the contrast and luminosity are adjusted to remove less visible features until only the bright dots associated with markers are left (right picture). Note that, in this case, spurious light reflections not coming from mocap markers are also present in the image; for instance, the reflection from the metallic edge of a whiteboard on the left side. In practice, optical mocap systems use optical filters and various processing techniques to obtain marker positions.

A preliminary calibration[5] is typically necessary for passive IR systems, consisting of a static calibration (static reference) and a dynamic calibration (known as the "dance with the wand," needed to define the space where the movement will be captured). These two calibrations can be done simultaneously or in sequence, depending on the model and make of the mocap system. This step is needed to establish a known relationship between the multiple cameras placed in a setup and is done at the beginning of each capture session, or anytime cameras are moved. Active IR systems usually are sold pre-calibrated, but calibration may also be required if more than one *tracker*[6] is used.

The number of cameras available in a passive IR system is an important variable to consider. If, theoretically, two cameras can provide a 3D position measurement, in practice, three or more cameras are needed due to marker occlusion. As a rule of thumb, the more complex the measurement, the more cameras would be needed to provide a complete measurement of the markers at any moment. Mocap recordings of complex movements with few cameras typically yield many marker losses (occlusions), making it difficult or even impossible to reconstruct the missing trajectories of occluded markers.

For active or passive IR systems, the complexity of a mocap measurement is a function of the *types of movements to be captured* (e.g., small movements such as finger movements or facial expressions, hands
p. 478 crossing and feet movements in piano ↳ performances), *the number of people captured at any given time* (e.g., solo, duets, or larger ensembles), and *the obtrusiveness of the setup* (e.g., an acoustic upright piano would block the view of markers for cameras placed behind it, as would an acoustic guitar if one wants to capture right-hand finger movements; Pérez, 2019), among other issues.

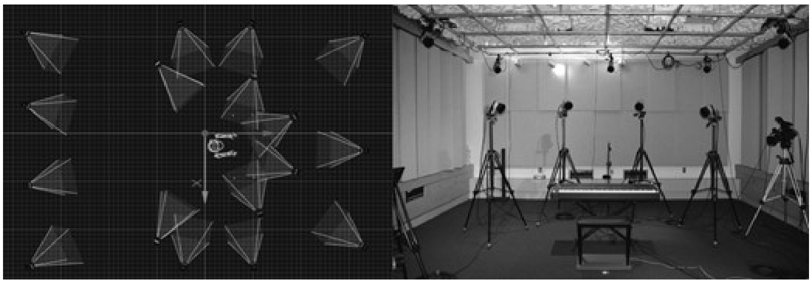
Passive IR systems, for instance, those made by Qualisys, Vicon, and Natural Point, have de facto become a reference in the field of motion capture in music performance contexts, although active IR systems, for instance, those from NDI or PTI, are also common.

How many cameras does one need?

Defining the right number of cameras is not a trivial task. Ideally, the more cameras, the better. This is because the more cameras are available, the higher the chance that three or more cameras see each marker at any time. However, given that the cost of an optical mocap system is directly related to the number of cameras/trackers purchased, significant budgets of several dozen to a few hundred thousand dollars might be at stake when setting up a mocap laboratory, not an option for all researchers (and musicians!). A minimum of three to six passive cameras or one tracker for active mocap systems would be required to allow for various mocap tasks. Systems of thirty, fifty, or many more cameras can be assembled these days, but for a considerably higher cost.

Figure 22.4 shows a recent motion capture experiment carried out by C. Massie-Laberge using seventeen passive IR cameras (Qualisys) to measure pianists' movements and seating information force using a Bertec force plate under the piano stool (Massie-Laberge et al. 2019).

Figure 22.4.

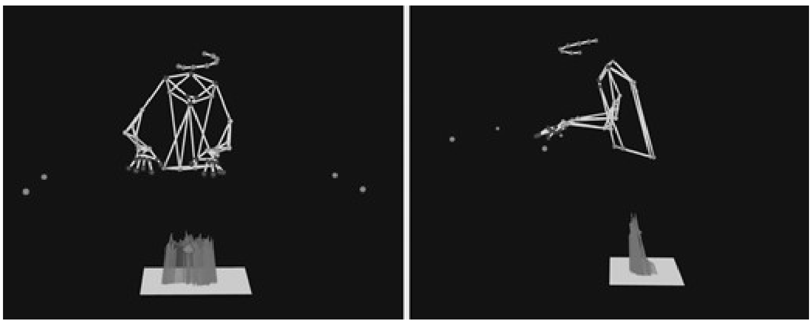


Left: Motion capture camera placement in Qualisys Track Manager (QTM). An example of the mocap-skeleton of one pianist is shown, and the 3D coordinate axes. Right: Laboratory setup at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), McGill University, showing motion capture cameras, MIDI keyboard, force plate, and a video camera.

Reprinted with permission from Catherine Massie-Laberge.

p. 479 Figure 22.5 shows the reconstructed upper body of a pianist and the ground forces produced while playing. A short description of force plates and their use with mocap will be presented later in the chapter.

Figure 22.5.



Reconstruction of a pianist skeleton performing an excerpt of the Medtner Sonata Reminiscenza (front and side view). Mocap and force data are captured and synchronized using a Qualisys optical mocap system and a Bertec force plate.

Reprinted with permission from Catherine Massie-Laberge.

Electromagnetic systems

Electromagnetic motion capture systems (aka *magnetic trackers*) use *sensors* to measure the intensity and orientation of an electromagnetic field (EMF) generated by a *source*. In the basic architecture, sensors (markers) are positioned on the objects to be tracked inside a measurement range, a function of the source size. Typical electromagnetic systems have sources allowing for measurement radiuses roughly between 1 and 3 meters, using up to sixteen (*wired*) sensors of various sizes (e.g., Polhemus Liberty).

Wireless magnetic trackers exist, such as the Polhemus Liberty LATUS (Large Area Tracking Untethered System), where up to twelve markers are placed on objects to be tracked, and one or more *receptors*[7] are wired to the central processing unit. Up to sixteen receptors can be used simultaneously to increase the measuring range, each one covering an area of approximately 2.5 m in diameter. In this model, the wireless markers are much larger than those used in the wired magnetic trackers.

One of the main advantages of magnetic trackers is that one marker can completely define the position *and* orientation of a body segment, that is, only one marker is needed to give the three positions and three orientations of an object, while at least three active or passive markers are needed in IR mocap systems to do the same. Magnetic trackers also have the advantage of not suffering from *line-of-sight* issues inherent to optical p. 480 ↳ systems, meaning that sensors can still be tracked when not "seen" by the electromagnetic field source (as when placed behind certain objects), which can be a significant advantage in many situations. Polhemus has recently released a video of a mocap session with classical guitarist Julian Gray where fourteen small (micro) sensors are used to track the fingers of his left hand.[8] Though possible with optical trackers, occlusion of markers would be expected when capturing advanced playing techniques.

When studying musical performances, the main disadvantage of magnetic trackers is the need for larger and/or tethered markers if compared to the markers used with optical systems. Furthermore, metals present in the environment can affect the accuracy of magnetic systems, as *eddy currents*[9] may cause distortion in the measurement. Such distortion increases with the electromagnetic field source's size and range (Vigliensoni & Wanderley, 2012). Though current systems allow for the correction of such distortions, care should nevertheless be taken when a mocap experiment will be set to avoid measurement errors. This issue might also be less critical in devices using a variation of the mainstream technology in magnetic trackers, as was the case in the discontinued tracker *Flock of Birds* by Ascension Technologies, now part of NDI.

In practice, magnetic trackers can be very useful in measuring performer limb movements (e.g., arms) when in a more restrained position, for instance, when standing up or sitting in a (wooden) chair. In 2009, we carried out experiments using a Polhemus Liberty 8 to track the arms and head movements of percussionists playing several acoustic and electronic instruments: a snare drum, a Roland V-Drum, the Mathews Radio Baton, and the Buchla Lightning 2 (Collicutt et al., 2009) (Figure 22.6). In this case, eight markers were placed on each hand's back, the forearms, biceps, the performers' back, and top of their head on a baseball cap. Because each sensor yielded a 6 DoF measurement, only eight sensors allowed to measure the performers' upper body and head movements, which would require many more markers if using an IR mocap system.

Other researchers have used mocap to study percussion sound-producing gestures. Dahl (2000) used active optical mocap systems to capture the strikes of three professional and one amateur drummer's right arm and stick performing on a rubber pad. Bouënard et al. 2011 captured movements of timpanists (cf. Figure 22.2) using a Vicon and later a Qualisys passive mocap system to create animations of virtual percussion avatars using a hybrid motion synthesis method controlled by the recorded trajectory of the mallets. Finally, it is worth mentioning a study using a single off-the-shelf video camera to track stick movement in controlled snare drum performances (Van Rooyen & Tzanetakis, 2015), allowing for a simple, albeit affordable and non-invasive method to accurately track the sticks.

p. 481 **Figure 22.6.**

Four snapshots of percussionist Fernando Rocha performing the same exercise in acoustic and electronic instruments, captured by Carmine Casciato. From left to right: a snare drum, a Roland V-Drum, the Mathews Radio Baton, and the Buchla Lightning 2. The exercise's goal was to study how performance techniques would be changed to perform with real and virtual instruments, that is, electronic instruments with no resonating membrane or even a surface to be hit.

Inertial and markerless systems

Inertial systems use combinations of sensors (typically, accelerometers, gyroscopes, and magnetometers) to obtain the orientation of an object (or part of the body).

Inertial mocap systems' main advantage is that they do not need an artificial source of physical phenomenon (e.g., IR light or electromagnetic field) for the measurement. The constant cost reduction of inertial sensor technologies makes it possible to obtain accessible, high-performance systems in the order of a few thousand dollars for a complete mocap solution.

An early example would be XSens, which allows measurement of orientation and position by using inertial sensors with navigation data. Several commercial systems have been proposed in recent years, including systems manufactured by Nansense, Noitom (Perception Neuron), Noraxon, Shadow, and STT Systems, among many others. Also, open-source inertial mocap projects exist (e.g., Chordata Motion).

*Markerless mocap system*s are based on standard video cameras where "thousands of natural points are captured, as in (for instance), a real-time three-dimensional (3-D) scan" (Bregler, 2007). This is an application of computer vision, where an algorithm analyzes the images from each camera and looks for points that are part of a model of the human body after a preliminary calibration step.

Depth cameras, associating a 2D (X, Y) image sensor and an IR depth sensor for the Z dimension, allow object tracking information in 3D. The Microsoft Kinect is a well-known example of such a device that has been extensively used in interactive musical performances and for the analysis of performer movements, for instance, for tracking the head of a duo of violinists (Hadjakos et al., 2013).

p. 482

Other mocap technologies

Several other motion capture technologies exist and are worth discussing, though they have been somewhat less used in music performances so far.

Mechanical mocap systems are those where a mechanical contact exists between the point whose position will be measured using metal structures with joints whose angle is measured using accurate potentiometers. An example would be exoskeletons, systems worn by a subject and measure their limbs' movements.

Ultrasound systems measure the propagation time of an inaudible sound with a frequency usually around 40 kHz. A sound source sends a signal, and the propagation time to the receiver is measured, which is proportional to the distance between the source and the receiver.

Other measurement devices and systems used in conjunction with mocap

Several other recording devices and systems are commonly used with mocap to simultaneously record different aspects of music performances.

The most obvious one, given the music most likely involves sound, is *audio recording.* Music performance audio can be simply recorded by one microphone in a standalone audio recording device or as an audio channel recorded by a video camera, but also as more complicated setups for stereo, binaural, or multiple speaker systems. MIDI (Musical Instrument Digital Interface) information is handy in many musical mocap applications, mostly involving piano performances. *Video recording* has become ubiquitous thanks to portable phones and is an essential complement to mocap data.

Force plates are devices that measure ground reacting forces, for instance, forces reacting to the weight of a person stepping onto the plate. Advanced force plates typically provide three or six measurements: the forces in X, Y, and Z or these forces plus the moments (rotation forces) about the three axes. Specialized force plate manufacturers include AMTI, Bertec, and Kistler, and their products can cost several thousand dollars, though cheaper options such as the Nintendo Wii balance board can be a viable solution in several applications. *Pressure mats* yield measurements distributed across a surface, providing a map of force distribution. They can be made in the form of individual platforms or as insole products, allowing for force measurements while someone is walking.

Eye trackers are devices that measure eye movements, pupil dilation, and blinking activity and are becoming popular in music studies (Bishop et al., 2019, Fink et al., 2019; Marandola, 2019). They can be handy, for instance, in the study of music score reading or of the gaze of musicians (and/or audience) during ensemble performances. In such situations, eye tracking can complement the movement information obtained by mocap to obtain a more comprehensive picture of the performance.

p. 483 *Biosignal interfaces* such as EMG (*electromyography*, the measurement of electrical activity in muscles) and EEG (*electroencephalography*, the measurement of electrical activity in the brain) are also common in the analysis of musical performances and can be used in conjunction with mocap. EMG provides a measurement of muscle activation complementary to the kinematics of the movement captured with mocap. EEG systems are becoming more portable and user-friendly, allowing their use with motion capture systems (Maidhof et al., 2014).

More recently, several manufacturers started offering *hybrid systems* that use more than one of the technologies here described. Novel combinations of multiple mocap technologies allow for more comprehensive performance measurements mitigating the limitations of individual systems. Such combinations include *marker* and *markerless* optical systems, inertial and IR systems, motion capture, and different types of biosignal interfaces.

For example, Verdugo and colleagues measure the movements and muscle activation patterns of several pianists using an eighteen-camera Vicon system and a wireless Delsys Trigno EMG system. In this study, isolated keystrokes were captured to investigate features of touch and articulation and the impact of trunk motion on these features (Verdugo et al., 2020). Similarly, a twelve-camera Qualisys mocap system and a

Trigno EMG system were used by Gonzales-Sanchez et al. (2019) to characterize the fluency in performances on snare drum and cello.

Other sensor systems can also be useful for the recording of performer movements. An interesting alternative direction for the recording of pianist movements by MacRitchie and McPherson (2015) used one high-speed (117 fps) video camera for tracking painted markers placed on the hands and capacitive sensors (TouchKeys) to measure the location of finger contact with the piano keys on a Yamaha Disklavier grand piano. By doing so, they could simultaneously investigate large-scale hand movements using video and small-scale finger movements using position sensors. A multi-step strategy for data alignment of the video camera, sensor, and MIDI (generated by the Disklavier) data was devised to compare the various data sources captured.

Synchronization issues

As illustrated in the last paragraph, when using several acquisition devices together, for instance, mocap systems, audio, MIDI and video recording, force plates, eye trackers, or biosignal interfaces, aligning the data streams from the various systems is an important step. This process is commonly known as synchronization.

Synchronization between mocap and other devices is an essential step for the analysis of musical performances. If multiple recording devices run independently (*asynchronously*), how can one be sure that what seems to be the same event in multiple data streams is actually happening at the same time? The exact timings of multiple data streams can become pretty hard to figure out when the capture involves fast-paced excerpts.

p. 484 At the most superficial level, synchronization can be done by measuring a short event with the multiple recording systems and identifying the event's time in each system. A clapboard can also be useful. Assuming the clocks of the various systems do not drift over time with respect to each other, this is an acceptable solution that avoids the use of rather costly specialized hardware. For instance, quickly nodding at the beginning of the recording can be used for aligning optical mocap and eye-tracking data (Burger et al., 2018).

Some systems accept an external control signal generated by another system to start recording data, that is, they can be triggered by another system. Other systems work on their own but have a synchronization input that takes a signal (e.g., a pulse) indicating an event and integrates it with the captured data for further reference (for instance, the Polhemus Liberty tracker that has a *sync_in* pin). Once data are analyzed, one can retrieve the reference time by identifying the pulse in the recorded data.

Some systems can be controlled by an external clock, that is, they can run according to an external clock signal generated by a specific unit or by another mocap system. This is the case of the Qualisys cameras, which can either generate a clock for other devices or be controlled by an external clock. Having a common clock controlling multiple devices is a common strategy in music studios (known as *word clock*), which guarantees that multiple devices run synchronously.

A more complete synchronization strategy for a variety of devices would involve data stamping using SMPTE (Society of Motion Pictures and Television Engineers) timecode[10] and a common clock to control all systems in use (Maestre et al., 2017). SMPTE timecode is usually streamed in the form of an LTC (*linear timecode*) signal. It can then be recorded as an analog signal (e.g., audio) and decoded to provide synchronization and timestamp information. SMPTE timecode provides a reference time information containing data on the hour, minute, second, and frame, which is sent to all devices for synchronization. MIDI timecode (MTC) has a similar data format. Network synchronization protocols might also be supported by the different manufacturers, for instance, Labstreaminglayer (LSL),[11] which allows synchronizing data from different sources including mocap, eye tracker and brain measurement devices.

Mocap data analysis

Motion capture can quickly generate large amounts of data, typically in the form of multiple time series (one time series per direction per marker) and associated data such as videos, audio, MIDI, and force plate data.

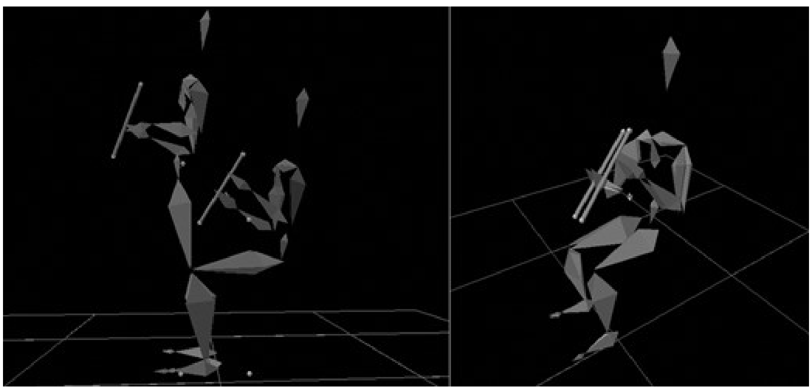
Mocap data typically need to be processed, for instance, to correct eventual missing information due to occlusion or recording conditions (Tits et al., 2018). A useful step when using optical mocap systems is p. 485 marker labeling, that is, associating markers to ↳ pre-determined locations in the body or instrument. The use of marker placement models such as Vicon's plug-in gait can facilitate data capture, as spurious reflections can be eliminated as they are not present in the model. Marker labeling can be done either manually or automatically, for instance, using tools such as Qualisys AIM (Automatic Identification of Markers).

The analysis of the captured data can be a complicated task. Once data has been processed, one needs to choose methods to analyze the obtained data. These can vary widely, from simply visualizing each time series individually to applying more advanced techniques. A simple step in analyzing mocap data is to reconstruct the body movements with stick figures (connecting the individual markers) and then inspect the resulting data from several angles before delving into each marker's time series. This can also be a good strategy when comparing multiple versions of performances, as seen in Figure 22.7. Segmenting gesture mocap time series can help identify basic movement patterns and shed light on performance strategies (Caramiaux et al., 2012).

A variety of calculations can be carried out to describe the data. Low-level features describing the kinematics of movements (e.g., velocity, acceleration, or jerk, the derivative of acceleration) are calculated on one or on a few markers. Higher-level features describing the structure of the data are done on multiple markers using tools such as PCA (Principal Components Analysis), "a method that transforms a large group of variables into a reduced group of uncorrelated variables called principal components (PCs), which are linear combinations of the original variables." (Toiviainen et al., 2010).

Nevertheless, no single, universal technique can be applied to all cases, and the choice of tool to use highly depends on the research questions and experiment setup.

Figure 22.7.



Screenshots of two clarinet performances superposed using Vicon's Polygon software. Left: lateral view of the performer standing up and seated. Right: angle view of two seated performances.

p. 486 Mocap system manufacturers and third-party companies offer software that can provide advanced options for data analysis. A well-known example is Visual3D by C-Motion. The cost of such tools can be substantial, though, sometimes in the range of thousands of dollars per license. One can also design one's own analysis techniques in various programming environments, which is often done in kinesiology and computer science laboratories, but this takes expertise and time.

Fortunately, several freely available toolboxes created to explore mocap data sets, though they might require a certain level of computer literacy. Some of these toolboxes have been designed with musical mocap data in mind.

The most comprehensive and widely used option is the University of Jyväskylä's MoCaP Toolbox (Burger & Toiviainen, 2013), a Matlab[12] toolbox providing dozens of functions to carry out simple as well as complex data analysis tasks. Though it requires commercial software, it is relatively accessible for musicians with some familiarity with programming languages.

The Topos toolkit was designed in the context of dance performances, focusing on dance-music interactions. Based on PureData,[13] it aims "to present and explore realtime features of choreographic gesture in space for computer music applications" (Naveda & Santana, 2014, p. 470).

A more recent example is Modosc, "a set of Max abstractions for computing motion descriptors from raw motion capture data in real time" (Dahl & Visi 2018). The focus of Modosc is on real-time analysis, therefore requiring a tool like Max.[14] Though also a commercial software product, Max is a well-known visual programming language for musical and interactive applications developed for music interaction and is widely used by musicians and interactive artists.

Another useful platform is EyesWeb[15] (Volpe et al. 2013), "an open platform to support the design and development of real-time multimodal systems and interfaces" that allows for mocap data input. However, it has much more to offer and could be an excellent tool for the simultaneous analysis of mocap, video, and sensor data.

One can also use data *sonification*[16] to explore mocap data (Dahl et al. 2019). Mocap data sonification eliminates the need to inspect data visually and might help, for instance, to identify movements with rhythmic characteristics. It has been used for studying ancillary movements of clarinettists where a set of four repetitive movements were sonified (clarinet bell circular movements, body weight transfer, body curvature, and knee bending), providing a sonic illustration of the various movements performed (Verfaille et al. 2006).

p. 487

Other Mocap Applications in Music

Apart from its use in music performance analysis and music pedagogy (Caramiaux et al. 2018), motion capture can be an interesting tool in other contexts as well. For instance, it can provide data for the animation of musician characters (Peinado et al., 2004; Bouënard et al., 2011), as it's done in video games and movies, or it can be used as a real-time input control to generate live music.

Early examples by Bevilacqua et al. (2001) and by Dobrian and Bevilacqua (2003) use motion capture data obtained with a Vicon System 8 as the input control for sound synthesis, to trigger sampled sounds and to control visual rendering in real time. Eckel et al. (2009) describe several live musical performances using real-time motion capture of up to three musicians using a Vicon system with passive markers placed on their instruments, arms, or heads. Nymoen et al. (2011) describe the motion capture of four passive markers on the SoundSaber, a 120 cm long, 4 cm diameter rod, using either an Optitrack or a Qualisys passive mocap system. Mamedes et al. (2013) use a similar Vicon system in the performance of Mamedes' piece *Entoa*, simulating the interaction with a virtual sphere (Figure 22.8). A version of the piece was created with a Kinect replacing the original Vicon system.

Figure 22.8.

Snapshot of the video with excerpts of the piece *Entoa*, composed and played by Clayton Mamedes (Mamedes et al., 2013). Markers on the hands and shoulders are captured by a six-camera Vicon 460 and used to control and process sounds in real time. A video describing the process is available at https://www.youtube.com/watch?v=H4W71MgBhsc. A version of the piece using the Kinect instead of the Vicon is available at https://www.youtube.com/watch?v=yQt1Xomw5iQ.

Used with permission from Clayton Mamedes.

p. 488 Real-time use of electromagnetic and inertial mocap systems in music and artistic applications is common due to lower cost, higher portability, and simplified setup. For instance, an experimental musical performance called *Dance Jockey* was proposed to perform electronic music while dancing, using an Xsens inertial mocap system (de Quay et al. 2011).

Conclusion

Motion capture technologies are becoming standard tools for the analysis of musical performances. Several works in the last twenty-five years or so have addressed multiple uses of mocap in musical contexts, and this list will surely grow with constant reduction of price and increase of quality of newer technologies.

Though still dominated by optical IR systems, mostly those tracking passive markers, electromagnetic, inertial, and more recently markerless mocap systems are commonly used.

Several toolboxes for main programming languages such as Matlab, Max, or Pure Data have been proposed, and the options of tools for mocap data analysis should increase with the introduction of machine learning techniques in such toolboxes.

Key Resources

General resources

General information on motion capture can be obtained from the various manufacturer websites, though it might sometimes be hard to disentangle important system features from rather useless marketing claims (at least for researchers!).

That said, an excellent resource for general information and detailed explanations on motion capture using passive IR systems is Qualisys' QAcademy, containing dozens of videos that clearly explain basic and advanced steps into setting up mocap experiments and analyzing data.

Resources on technological aspects of mocap

An excellent technical source of information on various technologies, including motion capture, is G. Burdea and P. Coiffet, *Virtual Reality Technology*, 2nd ed. It describes several mocap technologies in extensive detail and considers their limitations and possible applications, though not in musical performance.

p. 489 Similarly, the 2014 edited book by P. Fuchs, G. Moreau, and P. Guitton, *Virtual Reality: Concepts and Technologies*, has many details about various technologies and a section on motion capture, also focusing on virtual reality, not music.

Finally, the 2002 paper by Welch and Foxlin is still an excellent source for understanding the most common motion capture technologies at the time. Though around twenty years old, it is still a useful reference for most common technologies.

Resources on the musical application of mocap

The papers by Goebl et al. (2014) and by Jensenius (2018) are excellent sources about assessing musical performances and include discussions on motion capture technologies. They are a very good starting point to contextualize and complement the present chapter.

Reflective Questions

1. Think about an opportunity to study the movements of one or more performers. What questions might you seek to answer? How would you go about addressing these questions? In what ways might technology help?
2. Which mocap technology would you choose to capture a given performance with your instrument? Which other data capture would be necessary? How would you synchronize multiple recording devices? How would you set up a research space to use this system?
3. Imagine a specific musical task to be captured in a mocap experiment. If it involves the performance of existing musical pieces, which musical excerpts would you chose and why? How long do you imagine each capture should take? How many markers would you need, and where would you place them on the performers and their instruments?
4. Suppose you have already defined a mocap experiment using your instrument. How would you describe the process to analyze the data you will gather? Which tools do you feel comfortable to use for that?

References

Bevilacqua, F., Naugle, L., & Valverde, I. (2001). Virtual dance and music environment using motion capture. *Proceedings of the IEEE Multimedia Technology and Applications Conference* (pp. 1–4). Irvine, CA.
Google Scholar WorldCat

Bishop, L., Cancino-Chacón, C., & Goebl, W. (2019). Eye gaze as a means of giving and seeking information during musical interaction. *Consciousness and Cognition*, *68*, 73–96. https://doi.org/10.1016/j.concog.2019.01.002.
Google Scholar WorldCat

p. 490 Bouënard, A., Wanderley, M. M., Gibet, S., & Marandola, F. (2011). Virtual gesture control and synthesis of music performances: Qualitative evaluation of synthesized timpani exercises. *Computer Music Journal*, *35*(3), 57–72. https://doi.org/10.1162/COMJ_a_00069.
Google Scholar WorldCat

Bregler, C. (2007). Motion capture technology for entertainment [in the spotlight]. *IEEE Signal Processing Magazine*, *24*(6), pp. 156, 158, 160. https://doi.org/10.1109/MSP.2007.906023.
Google Scholar WorldCat

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology.* (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc.
Google Scholar Google Preview WorldCat COPAC

Burger, B., Puupponen, A., & Jantunen, T. (2018). Synchronizing eye tracking and optical motion capture: How to bring them together. *Journal of Eye Movement Research*, *11*(2). https://doi.org/10.16910/jemr.11.2.5.
Google Scholar WorldCat

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox: A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the 10th Sound and Music Computing Conference* (pp. 172–178). Stockholm, Sweden.
Google Scholar Google Preview WorldCat COPAC

Cadoz, C., & Wanderley, M. M. (2000). Gesture-music. In M. Wanderley & M. Battier (Eds), *Trends in Gestural Control of Music*. Paris: IRCAM–Centre Pompidou.
Google Scholar Google Preview WorldCat COPAC

Caramiaux, B., Bevilacqua, F., Wanderley, M. M., & Palmer, C. (2018). Dissociable effects of practice variability on learning motor and timing skills. *PLoS ONE*, *13*(3), e0193580. https://doi.org/10.1371/journal.pone.0193580.
Google Scholar WorldCat

Caramiaux, B., Wanderley, M. M., & Bevilacqua, F. (2012). Segmenting and parsing instrumentalist's gestures. *Journal of New Music Research*, *41*(1), 13–29. https://doi.org/10.1080/09298215.2011.643314.
Google Scholar WorldCat

Collicutt, M., Casciato, C., & Wanderley, M. M. (2009). From real to virtual: A comparison of input devices for percussion tasks. *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 1–6), Pittsburgh, PA.

Cossette, I., Monaco, P., Aliverti, A., & Macklem, P. T. (2008). Chest wall dynamics and muscle recruitment during professional flute playing. *Respiratory Physiology & Neurobio*logy, *160*, 187–195. https://doi.org/10.1016/j.resp.2007.09.009.
WorldCat

Dahl, L., Knowlton, C., & Zaferiou, A. (2019) Developing real-time sonification with optical motion capture to convey balance-related metrics to dancers. *Proceedings of the 6th International Conference on Movement and Computing* (pp. 1–6). https://doi.org/10.1145/3347122.3359600.

Dahl, L., & Visi, F. (2018). Modosc: A library of real-time movement descriptors for marker-based motion capture. *Proceedings of* *the 5th International Conference on Movement and Computing*. https://doi.org/10.1145/3212721.3212842.

Dahl, S. (2000). The playing of an accent: Preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, *29*(3), 225–233.
Google Scholar WorldCat

Dahl, S., Bevilacqua, F., Bressin, R., Clayton, M., Leante, L., Poggi, I., & Rasamimanana, N. (2010). Gestures in performance. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 36–68). New York and Oxford: Routledge.
Google Scholar Google Preview WorldCat COPAC

Davidson, J. W. (1994). Which areas of a pianist's body convey information about expressive intention to an audience? *Journal of Human Movement Studies*, *26*, 279–301.
Google Scholar WorldCat

de Quay, J., Skogstad, S., & Jensenius, A. (2011). Dance jockey: Performing electronic music by dancing. *Leonardo Music Journal*, *21*, 11–12. https://doi.org/10.1162/LMJ_a_00052.
Google Scholar WorldCat

Desmet, F., Nijs, L., Demey, M., Lesaffre, M., Martens, J. P. & Leman, M. (2012). Assessing a clarinet player's performer gestures in relation to locally intended musical targets. *Journal of New Music Research*, *41*(1), 31–48. https://doi.org/10.1080/09298215.2011.649769.
Google Scholar WorldCat

Dobrian, C., & Bevilacqua, F. (2003). Gestural control of music using the Vicon 8 motion capture system. *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp.161–163). Montreal, Canada.

Eckel, G., Pirró, D., & Sharma, G. K. (2009). Motion-enabled live electronics. *Proceedings of the Sound and Music Computing Conference*. Porto, Portugal.

Fink, L. K., Lange, E. B., & Groner, R. (2019). The application of eye-tracking in music research. *Journal of Eye Movement Research*, *11*(2). https://doi.org/10.16910/jemr.11.2.1.
Google Scholar WorldCat

Fuchs, P., & Mathieu, H. (2011). Location sensors. In P. Fuchs, G. Moreau, & P. Guitton (Eds.), *Virtual reality: Concepts and technologies* (pp. 105–121). Leiden: CRC Press/Balkema.
Google Scholar Google Preview WorldCat COPAC

Fuchs, P., Moreau, G., & Guitton, P. (Eds.) (2011). *Virtual reality: Concepts and technologies* (pp. 105–121). Leiden: CRC Press/Balkema.
Google Scholar Google Preview WorldCat COPAC

Goebl, W., Dixon, S., & Schubert, E. (2014). Quantitative methods: Motion analysis, audio analysis, and continuous response techniques. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 221–239). Oxford: Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

Goebl, W., & Palmer, C. (2008). Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, *186*(3), 471–479. https://doi.org/10.1007/s00221-007-1252-1.
Google Scholar WorldCat

Gonzalez-Sanchez, V., Dahl, S., Hatfield, J. L., & Godøy, R. I. (2019). Characterizing movement fluency in musical performance: Toward a generic measure for technology enhanced learning. *Frontiers in Psychology*, *10*, 84. https://doi.org/10.3389/fpsyg.2019.00084.
WorldCat

Hadjakos, A., Großhauser, T., & Goebl, W. (2013). Motion analysis of music ensembles with the Kinect. *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 106–110). Daejeon, Korea. https://doi.org/10.5281/zenodo.1178540.
Google Scholar Google Preview WorldCat COPAC

Hilt, P. M., Badino, L., D'Ausilio, A., Volpe, G., Tokay, S., Fadiga, L., & Camurri, A. (2019). Multi-layer adaptation of group coordination in musical ensembles. *Scientific Reports*, *9*, 5854. https://doi.org/10.1038/s41598-019-42395-4.
Google Scholar WorldCat

Jensenius, A. R. (2018). Methods for studying music-related body motion. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 805–818). Berlin and Heidelberg: Springer-Verlag.
Google Scholar Google Preview WorldCat COPAC

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gestures: Concepts and methods in research. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12–35). New York and Oxford: Routledge.
Google Scholar Google Preview WorldCat COPAC

Keller, P., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, *28*(1), 27–46. https://doi.org/10.1525/mp.2010.28.1.27.
Google Scholar WorldCat

MacRitchie, J., & McPherson, A. P. (2015). Integrating optical finger motion tracking with surface touch events. *Frontiers in Psychology*, *6*, 702. https://doi.org/10.3389/fpsyg.2015.00702.
Google Scholar WorldCat

Maes, P.-J., Wanderley, M. M., & Palmer, C. (2015). The role of working memory in the temporal control of discrete and continuous movements. *Experimental Brain Research*, *233*(1), 263–273. https://doi.org/10.1007/s00221-014-4108-5.
Google Scholar WorldCat

Maestre, E., Blaauw M., Bonada J., Guaus E., & Pérez A. (2010). Statistical modeling of bowing control applied to violin sound synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(4), 855–871. https://doi.org/10.1109/TASL.2010.2040783.
Google Scholar WorldCat

Maestre, E., Papiotis, P., Marchini, M., Llimona, Q., Mayor, O., Pérez, A., & Wanderley, M. M. (2017). Enriched multimodal representations of music performances: Online access and visualization. *IEEE Multimedia*, *24*(1), 24–34. https://doi.org/10.1109/MMUL.2017.3.
Google Scholar WorldCat

p. 492 Maidhof, C., Kästner, T., & Makkonen, T. (2014). Combining EEG, MIDI, and motion capture techniques for investigating musical performance. *Behavioral Research Methods*, *46*(1), 185–95. https://doi.org/10.3758/s13428-013-0363-9.
Google Scholar WorldCat

Mamedes, C. R., Wanderley, M. M., Manzolli, J., & Garcia, D. H. L. (2013). Strategies for mapping control in interactive audiovisual installations. *10th International Symposium on Computer Music Multidisciplinary Research* (pp. 766–778). Marseille, France.
Google Scholar Google Preview WorldCat COPAC

Marandola, F. (2019). Eye-hand synchronization in xylophone performance: Two case-studies with African and Western percussionists. *Journal of Eye Movement Research*, *11*(2), 7. https://doi.org/10.16910/jemr.11.2.7.
Google Scholar WorldCat

Marchini, M., Papiotis, P., Pérez, A., & Maestre, E. (2011). A hair ribbon deflection model for low-intrusiveness measurement of bow force in violin performance. *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 481–486). https://doi.org/10.5281/zenodo.1178097.
Google Scholar Google Preview WorldCat COPAC

Massie-Laberge, C. (2019). *Kinematic, kinetic and perceptual analyses of piano performances*. Unpublished PhD thesis, McGill University.
Google Scholar Google Preview WorldCat COPAC

Massie-Laberge, C., Cossette, I., & Wanderley, M. M. (2019). Kinematic analysis of pianists' expressive performances of Romantic excerpts: Applications for enhanced pedagogical approaches. *Frontiers in Psychol*ogy, *9*, 2725. https://doi.org/10.3389/fpsyg.2018.02725.
WorldCat

Miranda, E. R., & Wanderley, M. M. (2006). *New digital musical instruments: Control and interaction beyond the keyboard*. Middleton, WI: A-R Editions.
Google Scholar Google Preview WorldCat COPAC

Mota, D., Loureiro, M., & Labossière, R. (2017). Gestural interaction in ensemble performance. In M. Lessafre, P. J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 177–185). New York: Routledge.
Google Scholar Google Preview WorldCat COPAC

Naveda, L., & Santana, I. (2014). "Topos" toolkit for pure data: Exploring the spatial features of dance gestures for interactive musical applications. *Proceedings of the 40th International Computer Music Conference (ICMC) | 11th Sound & Music Computing Conference (SMC)*. Athens, Greece.

Ng, K., & Nesi, P. (2008). i-Maestro: Technology-enhanced learning and teaching for music. *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 225–228).

Nusseck, M., & Wanderley, M. M. (2009). Music and motion: How music-related ancillary body movements contribute to the experience of music. *Music Perception*, *26*(4), 335–353. https://doi.org/10.1525/mp.2009.26.4.335.
Google Scholar WorldCat

Nusseck, M., Wanderley, M. M., & Spahn, C. (2017). Body movements in music performances: The example of clarinet players. In B. Müller & S. Wolf (Eds.), *Handbook of human motion*. Springer-Verlag.
Google Scholar Google Preview WorldCat COPAC

Nymoen, K., Skogstad, S. A., & Jensenius, A. R. (2011). SoundSaber: A motion capture instrument. *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 312–315). Oslo, Norway.
Google Scholar Google Preview WorldCat COPAC

Palmer, C., Koopmans, E., Loehr, J. D., & Carter, C. (2009). Movement-related feedback and temporal accuracy in clarinet performance. *Music Perception*, *26*(5), 439–449. https://doi.org/10.1525/MP.2009.26.5.439.
Google Scholar WorldCat

Peinado, M., Herbelin, B., Wanderley, M. M., Le Callennec, B., Boulic, R., & Thalmann, D. (2004). Towards configurable motion capture with prioritized inverse kinematics. *Proceedings of the Third International Workshop on Virtual Rehabilitation (IWVR2004)*, EPFL, Lausanne, Switzerland.

Pérez, A. (2019). Finger-string interaction analysis in guitar playing with optical motion capture. *Frontiers in Computer Science*, *1*, 8. https://doi.org/10.3389/fcomp.2019.00008.
Google Scholar WorldCat

Rodger, M. W., Craig, C. M., & O'Modhrain, S. (2012). Expertise is perceived from both sound and body movement in musical performance. *Human Movement Science*, *31*(5), 1137–1150. https://doi.org/10.1016/j.humov.2012.02.012.
Google Scholar WorldCat

Rozé, J., Aramaki, M., Kronland-Martinet, R., & Ystad, S. (2018). Assessing the effects of a primary control impairment on the cellists' bowing gesture inducing harsh sounds. *IEEE Access*, *6*, 43683–43695. https://doi.org/10.1109/ACCESS.2018.2856178.
Google Scholar WorldCat

Schoonderwaldt, E., & Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *Journal of the Acoustical Society of America*, *126*(5), 2695–2708. https://doi.org/10.1121/1.3227640.
Google Scholar WorldCat

Schoonderwaldt, E., & Wanderley, M. M. (2007). Visualization of bowing gestures for feedback: The Hodgson plot. *Proceedings of the i-Maestro 3rd Workshop*. Barcelona, Spain.

Shan, G. B., & Visentin, P. (2003). A quantitative three-dimensional analysis of arm kinematics in violin performance. *Medical Problems of Performing Artists*, *18*(1), 3–10. https://doi.org/10.7717/peerj.1299.
Google Scholar WorldCat

Song, M.-H., & Godøy, R. I. (2016). How fast is your body motion? Determining a sufficient frame rate for an optical motion tracking system using passive markers. *PLoS ONE*, *11*(3), e0150993. https://doi.org/10.1371/journal.pone.0150993.
Google Scholar WorldCat

Teixeira, E. C., Loureiro, M. A., Wanderley, M. M., & Yehia, H. C. (2015). Motion analysis of clarinet performers. *Journal of New Music Research*, *44*(2), 97–111. https://doi.org/10.1080/09298215.2014.925939.
Google Scholar WorldCat

Thompson, M. R., Diapoulis, G., Himberg, T., & Toiviainen, P. (2017). Interpersonal coordination in dyadic performance. In M. Lessafre, P. J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 186–194). New York: Routledge.
Google Scholar Google Preview WorldCat COPAC

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, *16*(1), 19–40. https://doi.org/10.1177/1029864911423457.
Google Scholar WorldCat

Tits, M., Tilmanne, J., & Dutoit, T. (2018) Robust and automatic motion-capture data recovery using soft skeleton constraints and model averaging. *PLoS ONE*, *13*(7), e0199744. https://doi.org/10.1371/journal.pone.0199744.
Google Scholar WorldCat

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception*, *28*(1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59.
Google Scholar WorldCat

Van Rooyen, R., & Tzanetakis, G. (2015). Pragmatic Drum Motion Capture System. *Proceedings of the International Conference on New Interfaces for Musical Expression*. Baton Rouge, LA.
Google Scholar Google Preview WorldCat COPAC

Verdugo F., Pelletier J., Michaud B., Traube C., & Begon M. (2020). Effects of trunk motion, touch, and articulation on upper-limb velocities and on joint contribution to endpoint velocities during the production of loud piano tones. *Frontiers in Psychology*, *11*, 1159. https://doi.org/10.3389/fpsyg.2020.01159.
Google Scholar WorldCat

Verfaille, V., Quek, O., & Wanderley, M. M. (2006). Sonification of musician's ancillary gestures. *Proceedings of the 2006 International Conference on Auditory Display (ICAD06)*, London, England.

Vigliensoni, G., & Wanderley, M. M. (2012). A quantitative comparison of position trackers for the development of a touch-less musical interface. *Proceedings of the International Conference on New Interfaces for Musical Expression*. Ann Arbor, MI.

p. 494 VIM (2008). *International vocabulary of metrology: Basic and general concepts and associated terms (VIM).* Joint Committee for Guides in Metrology (JCGM).
Google Scholar Google Preview WorldCat COPAC

Volpe, G., Coletta, P., Ghisio, S., & Camurri, A. (2013). EyesWeb XMI: A platform for recording and real-time analysis of multimodal data streams. *Proceedings of Joint 40th Italian Annual Conference on Acoustics (AIA) and 39th German Annual Conference on Acoustics (DAGA)* . Merano, Italy.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and sign language in human-computer interaction* (pp. 241–253). Berlin: Springer-Verlag.
Google Scholar Google Preview WorldCat COPAC

Wanderley, M. M., & Vines, B. W. (2006). Origins and functions of clarinettists' ancillary gestures. In A. Gritten & E. King (Eds.), *Music and gesture*. New York: Routledge/Ashgate Publishing.
Google Scholar Google Preview WorldCat COPAC

Wanderley, M. M., Vines, B., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97–113. https://doi.org/10.1080/09298210500124208.
Google Scholar WorldCat

Weiss, A. E., Nusseck, M., & Spahn, C. (2018). Motion types of ancillary gestures in clarinet playing and their influence on the perception of musical performance. *Journal of New Music Research*, *47*(2), 129–142. https://doi.org/10.1080/09298215.2017.1413119.
Google Scholar WorldCat

Welch, G., & Foxlin, E. (2002). Motion tracking: No silver bullet but a respectable arsenal. *IEEE Computer Graphics and Applications*, *22*(6), 24–38. https://doi.org/10.1109/MCG.2002.1046626.
Google Scholar WorldCat

Winold, H., Thelen, E., & Ulrich, B. D. (1994). Coordination and control in the bow arm movements of highly skilled cellists. *Ecological Psychology*, *6*(1), 1–31.
Google Scholar WorldCat

Notes

1 See, for instance, a very interesting overview of the history of motion capture in the video entitled "What Is Motion Capture?" from the excellent Qualisys QAcademy website: https://www.qualisys.com/my/qacademy/#!/tutorials/what-is-motion-capture.

2 The video with the full visualization of Lambert Chen's performance of the Preludio of the Partita No. 3 by J. S. Bach for solo violin is available at https://www.youtube.com/watch?v=W69LxKA0BdQ.

3 https://c-motion.com/v3dwiki/index.php/Tutorial:_Plug-In_Gait_Full-Body.

4 BTS Bioengineering currently offers an OEP (OptoElectronic Plethysmography) system to non-invasively measure lung volume variations consisting of eight IR cameras: https://www.btsbioengineering.com/products/bts-oep-system/.

5 Calibration is "the comparison of measurement values delivered by a device under test with those of a calibration standard of known accuracy" (Wikipedia, 2020). See also the formal definition provided by the International Vocabulary of Metrology (VIM 2008).

6 Groups of two or three cameras prearranged and pre-calibrated in a single device are called *trackers* by some manufacturers (e.g., PTI). Multiple trackers can be used in one mocap setup. Note that in virtual reality (VR), the whole mocap system might be called a *tracker* (Fuchs & Mathieu 2011). Burdea and Coiffet (2003) define a tracker as "a special purpose hardware used in VR to measure the real-time change in a 3D object position and orientation" (p. 17).

7 Note the change of term from *source* to *receptor*, reflecting the different technologies used in each system.

8 https://polhemus.com/motion-tracking/videos.

9 Wikipedia defines *eddy currents* as "loops of electrical current induced within conductors by a changing magnetic field in the conductor according to Faraday's law of induction" (Wikipedia, 2020, https://en.wikipedia.org/wiki/Eddy_current).

10 https://en.wikipedia.org/wiki/SMPTE_timecode.

11 https://labstreaminglayer.readthedocs.io/info/time_synchronization.html.

12 http://www.mathworks.com/products/matlab.html.

13 http://puredata.info/.

14 http://cycling74.com/products/max.

15 http://www.infomus.org/eyesweb.

16 Sonification is “the use of non-speech audio to convey information or perceptualize data” (Wikipedia, 2020) https://en.wikipedia.org/wiki/Sonification.