





# A method for the reproduction of cello bow kinematics using a robotic arm and motion capture

Montserrat Pàmies-Vilà\* , Alexander Mayer, Ewa Matusiak, and Vasileios Chatziioannou 

Department of Music Acoustics – Wiener Klangstil, mdw - University of Music and Performing Arts Vienna,  
Anton-von-Webern-Platz 1, 1030 Vienna, Austria

Received 25 March 2024, Accepted 10 July 2024

**Abstract** – Numerous experimental and theoretical methods have focused on the bow–string interaction in bowed string instruments, including several artificial bowing setups. The current research aims to present an experimental approach to reproduce bowing techniques using a robotic arm. First, optical motion capture is used to track the 3D kinematics of the bow. The cello bow and corpus are equipped with reflective markers. The cello is mounted on a playing platform. The recorded 3D trajectories of the bow markers are used to control the motion of the robotic arm. This process requires converting the 3D data between the coordinate frames of the two systems. This conversion is described in detail in this paper. To demonstrate the performance of the proposed method, an experienced cellist was asked to play an adapted piece on the cello, which was then repeated using the robotic arm. The robotic arm is capable of accurately reproducing the bow velocity, but even minimal variations in position can compromise proper bow–string contact. To illustrate this, the study compares two similar robotic situations and discusses the challenges of adapting the robot’s coordinates as a function of a given playing parameter or the sound produced.

**Keywords:** Bowed-string instruments, Bowing machine, Motion capture, Musical robotics, Musical acoustics

## 1 Introduction

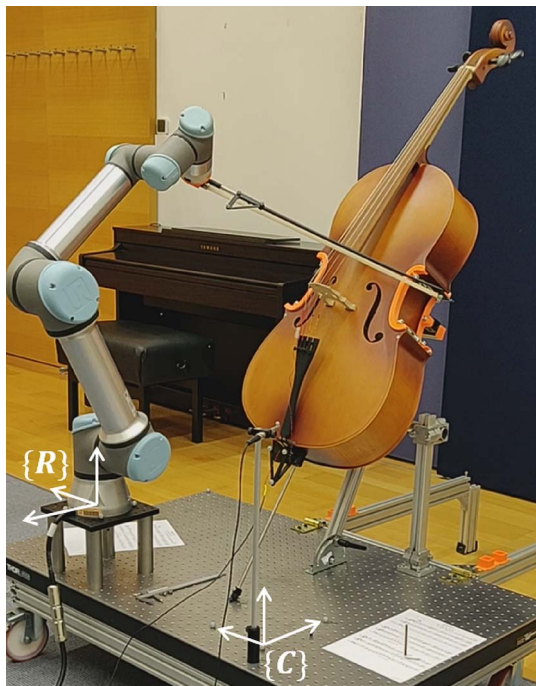
The acoustics of bowed-string instruments have been extensively studied to examine the non-linear interaction between the bow and the string [1–3] as well as to quantify the player’s actions on the instrument [4–8]. In pursuit of a better understanding of the underlying physics, numerous bowing devices have been developed to analyse the bow–string interaction under controlled laboratory conditions [1, 7, 9–13]. Since the first bowing devices, like Raman’s in 1918 [9] and later Schelleng’s in 1973 [10], artificial bowing approaches have aimed at analysing sound production and the properties of the so-called Helmholtz motion [1], originally only during steady-state [9–11] and more recently also considering transient states [7, 12, 13]. Following Schelleng’s contributions [10], which identified a narrow set of playing parameters where Helmholtz motion could possibly occur – as depicted in the Schelleng diagrams [14] –, many subsequent studies have focused on how these control parameters are used to create and sustain Helmholtz motion [1, 2]. It was shown that the bow velocity plays a crucial role in sound production, and that changes in timbre mainly occur depending on the bow force and the bow–string contact position [2, 15, 16]. Askenfelt and Guettler

[17] and later Schoonderwaldt [7] made significant contributions to the study of articulation styles using a bowing machine to describe, among other aspects, how articulation relates to changes in the attack transients.

These devices often strived to achieve controlled playing conditions, reducing the main playing parameters (bow velocity, bow force and bow position) to a set of constant values to analyse dependencies among them. Because of the nature of controlled experimentation, these devices considered the bow trajectory on the string exclusively in a linear motion, which was required in order to analyse playability in bowed-string instruments via Schelleng or Guettler diagrams [2, 12]. Although high precision has been achieved in previous bowing setups, a bow trajectory that mimics the three-dimensional motion of a bow as played by a musician was not considered.

The current study proposes a method designed to record the bow kinematics of a real player using motion capture and to replicate the bow motion with a robotic arm (Fig. 1). The ultimate purpose of such a motion-capture-aided robotic player is to enable a meticulous analysis of playing nuances, including articulation and realistic changes in bowing direction, while assuring a controlled laboratory environment. For instance, compared to human players, a mechanical playing setup would allow for a high number of exact repetitions of the bow motion, whereby the mechanical player

\*Corresponding author: [pamies-vila@mdw.ac.at](mailto:pamies-vila@mdw.ac.at)



**Figure 1.** Robotic arm situated on the playing platform with the cello secured on the supporting structure. The motion-capture coordinate frame is given as  $\{C\}$ , which establishes the ground plane, and the robot coordinate frame  $\{R\}$  defines the robot’s world frame.

would not introduce any bias based on their opinion on the materials used (bow, instrument, etc.), or their level of comfort or tiredness. For playability analysis, the robotic arm can also be set to play controlled, straight bow trajectories [16, 18], for which no motion-capture assistance is required.

Mechanical setups to play musical instruments have been used for centuries, with examples ranging from piano automata to sophisticated robotic performers. Within the field of musical robotics, a discipline found at the intersection between musical creativity and robotics, there have been many approaches to recreating a human performer with a mechanical setup. The reader can refer to reference [19] for an overview. To this date, though, the robotic approaches dealing with bowed strings, have not focused on the analysis of musical instruments or their playing techniques. They were rather seen as a robotics challenge on its own or as a tool for musical creativity. For example, the violin playing robot by Shibuya et al. [20], the humanoid robot performer by Jordà et al. [21], or other robotic setups taking part in artistic performances (such as “Empty Vessels” by David Gardener and Greg Debicki, or the work by composer Fredrik Gran on the cello). These approaches demonstrated the many possibilities of playing bowed-string instruments mechanically. We believe, though, that there have been no previous investigations of the bow-string interaction that include human-like bow trajectories.

Given the importance of bowing actions in sound production, motion capture technology [22] is well established as a method to record the 3D-trajectory of the bow, the instrument, and the player’s arms during playing. Many

studies, both for violin [15, 23] and less often for cello performances [24, 25], have shown the adequacy of using motion capture to capture musician’s and instrument movements. We build upon this existing technology because of its availability in our and many research institutions. However, the presented methodology could be adapted to use another technology.

The objective of the current publication is to showcase and evaluate the reproduction of bow kinematics with a robotic arm. The motion of the bow is first recorded in three dimensions using optical motion capture technology. Subsequently, these captured data are processed and used to instruct a robotic arm. The limitations of the proposed method are discussed, as we acknowledge that the challenging problem of reproducing player-like bow strokes is not yet completely solved. Note that the current approach relies on the reproduction of the bow kinematics, i.e. its position and velocity, but does not adapt or control for the bow force. The bow force on the string is one of the key parameters defining the timbre of the played sound [15, 26]. Moreover, the bow-string contact forces are also responsible for the vibrotactile and kinesthetic feedback perceived by the player on the bowing hand [27, 28] and used, together with the perceived sound, to meticulously introduce adjustments while playing. The current setup, however, is restricted to the direct control of the robot action, without providing any feedback loop to re-adjust the robot’s trajectory while playing, or accounting for a target bow force.

Following the previous conference paper, where we presented an analysis of different dynamic levels (*piano* vs. *forte*) on all open strings of a cello [29], we now would like to examine a recording of a more realistic playing situation using a cello piece adapted to be played on notes  $G_2$ ,  $D_3$  and  $A_3$  using only open strings.

The rest of the paper is structured as follows: The robotic arm and motion capture technology used in this study are presented, and the methodology to convert bow-motion data between these two systems is described in detail (Sect. 2). Section 3 suggests several metrics to evaluate the robot performance. Section 4 illustrates how a gentle change in bow position in the direction of the robot-tool influences the outcome. The manuscript concludes with a discussion of the capabilities and shortcomings of the presented approach (Sect. 5).

## 2 Methodology

This publication presents a methodology for converting motion-captured bow kinematics to the coordinates necessary to instruct a robotic arm. The presented method can be adapted for other types of motions that might be required to play other string instruments, as well as for other types of motion capture or robotic systems.

### 2.1 The robotic arm

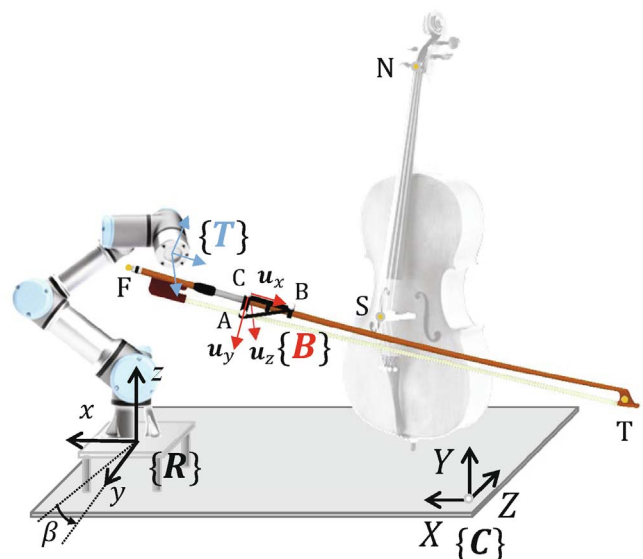
The robotic arm chosen for the current purpose is the model UR5e by Universal Robots, which can safely operate alongside humans in a shared workspace. It has a reach of 850 mm and a 5 kg object-holding capacity (Fig. 1). With

its six independent joints, the robotic arm can position an object in space, assuring three degrees of freedom for position and three for orientation (total of 6 degrees of freedom). We selected this type of robotic arm because of its reduced operation noise (below 65 dB, according to the technical specifications [30]), and a pose repeatability of  $\pm 0.03$  mm. To attach the bow to the robot, 3D-printed components are employed and screwed at the tool flange, which is the last flat surface at the mobile end of the robotic arm, where a certain tool can be added. Hereafter, we refer to the union of the last joint of the robot and the 3D-printed clamp as the tool<sup>1</sup>. The robot might be controlled directly via the angles of its joints or via the tool's position and orientation in the robot's Cartesian space, also called the robot frame. In the latter, an inverse kinematics algorithm calculates the corresponding joint angles [31]. The robot's tool position, or TCP (tool-center-point), is therefore defined by its 3D spatial coordinates relative to the robot base ( $T_x$ ,  $T_y$ ,  $T_z$  coordinates with respect to the robot's origin), along with the tool inclination given in axis-angle representation. The purpose of the presented methodology is to convert the recorded motion capture data of the bow kinematics to these six coordinates of the tool (three for position, three for orientation). In robotics, the establishment of three positions and three rotations is called the pose of the tool [31]. A given pose is always relative to the coordinate frame of the robot, centred at its origin. In Section 2.7 further details about the control and communication protocols with the robot are provided.

## 2.2 Optical motion capture system

The current approach uses optical motion capture with passive reflective markers and infrared cameras. The setup consists of 12 infrared cameras (OptiTrack Prime 13), distributed on a metallic grid suspended from the ceiling. Given that the cello consistently faces forward, a configuration with fewer cameras could also be sufficient. Typically, these systems involve the placement of a minimum of six cameras strategically around the capture space [32]. The Prime 13 cameras emit infrared light, which is reflected from the markers positioned on the bow, cello, playing platform, and, if needed, specific body parts of the player. The cameras, connected via LAN with a PC, capture the reflected infrared light. A commercially available software (Motive by OptiTrack) calculates the position of each marker in three dimensions in space at a rate of 240 frames per second (fps), with a typical precision of around  $\pm 0.3$  mm. Note that the error introduced due to imprecision of the camera recordings is about one order of magnitude larger than the error in controlling the robotic arm ( $\pm 0.03$  mm). This might introduce inaccuracies in the reproduction of the recorded geometrical data of the bow motion. A minimum of three reflective markers are necessary on the bow to calculate its position and orientation in space. For better visualisation, in the current setup, five reflective markers are affixed to the bow: one at the tip T, one at

<sup>1</sup> Other nomenclatures, such as end effector, holding joint, object, last link, or wrist, are omitted on purpose for simplicity



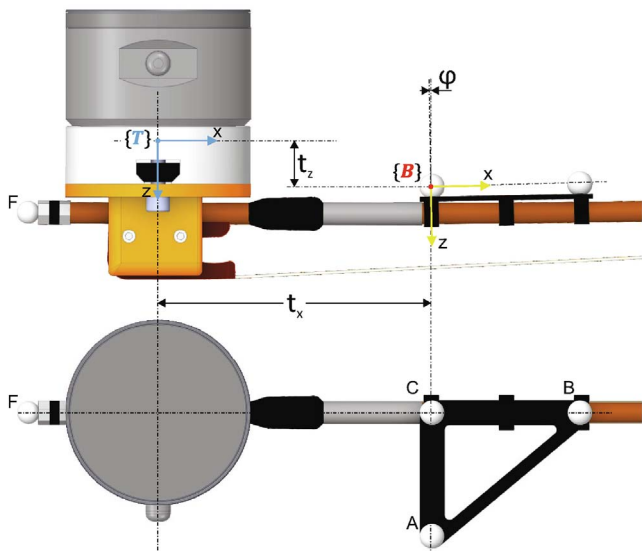
**Figure 2.** Indication of the motion-capture coordinate frame  $\{C\}$ , the robot coordinate frame  $\{R\}$ , the bow coordinate frame  $\{B\}$  and the tool coordinate frame  $\{T\}$ . Location of the bow-markers F, T, A, B, C (see Fig. 3) and cello-markers S, N.

the frog F, and three in between them, labelled as A, B, and C, as seen in Figures 2 and 3. A triangular 3D-printed mount is used to clamp these three markers on the bow-stick, right in front of the winding. The distance between markers C and B is 6 cm; between markers C and A, 5 cm, and between markers F and C, 17 cm.

## 2.3 Coordinate frames

Both the motion capture and the robotic systems function based on a Cartesian coordinate system or frame, where three orthogonal axes are chosen to form a right-handed coordinate frame. Motion capture data are recorded as the distance of all markers to the three axes (X, Y, and Z) of a user-defined camera coordinate frame  $\{C\}$ , i.e., the origin of the motion capture space. This camera coordinate frame  $\{C\}$  is situated on the surface of the platform, in front of the instrument and next to the robot (marked with  $\{C\}$  in Figs. 1 and 2). The motion capture system and software require a one-time recording of a normalised L-shape (3 reflective markers at  $90^\circ$ ), which defines the direction of the X and Z axes and, therefore, the ground floor of the recording space. In the present setup, the three markers forming the L-shape are screwed on the playing platform. To avoid introducing errors because of the positioning of the origin markers, the position of the L-shape is recorded at all times, and it is subsequently subtracted from the motion capture data collected.

The robot coordinate frame, denoted as  $\{R\}$ , serves as the world frame for the robot, establishing the reference origin for all data that are input or output from the robot. This frame represents the relation of the robot coordinate system to the camera frame  $\{R\}$  and is ultimately defined by the positioning of the robot arm on the platform. We decided to maintain the direction of the  $x$ -axis, so that



**Figure 3.** Location of the bow coordinate frame  $\{B\}$  and the tool coordinate frame  $\{T\}$ , with an indication of the distances  $t_x$ ,  $t_z$  and the angle  $\varphi$  that allows to align the two coordinate frames.

up-bow and down-bow motions correspond to growing or decreasing values in the  $x$ -axis in both camera and robot data. The same frame  $\{R\}$  is used in Section 4 as the reference origin of the presented data.

## 2.4 Post-processing of motion capture data

After the recordings, the motion capture software (Motive by OptiTrack) is used to assign every marker with a label, and to interpolate possible occlusions. Occlusions appear if a marker is hidden from the camera view for several frames, introducing a gap in the recorded data. Because the markers on the bow (markers F, A, B, C, and T) are facing the cameras all the time, no occlusions are found for these markers. The markers on the instrument are usually also not covered, but in cases where this might happen, the data for the cello-markers are linearly interpolated. Note that the cello is rigidly mounted to the playing platform for all tests, so the cello markers show nearly constant values. All markers of interest (bow and cello markers, playing-platform markers, and clapperboard markers) are exported without any applied filtering, and are used in the next steps of the algorithm.

## 2.5 Change of basis and conversion of recorded data

After exporting the motion capture data, all 3D data are first converted from the camera coordinate frame  $\{C\}$  to the robot coordinate frame  $\{R\}$  via a change of basis. The change of basis is given by three translations and a rotation matrix, and it is applied to all recorded data. The three translations are the positions of the robot coordinate frame  $\{R\}$  given in the camera frame  $\{C\}$  and the rotations are the orientations of the axes of  $\{R\}$  ( $x, y, z$ ) with respect to  $\{C\}$  ( $X, Y, Z$ ). The  $x$ -axis is therefore a result of a translation of  $X$ -axis by distance  $d_x$ ; the  $y$ -axis is a rotation by  $180^\circ$  of

the  $Z$ -axis and a translation by  $d_y$ , and the  $z$ -axis is a translation of  $Y$ -axis by distance  $d_z$ . To correct any misalignment of the horizontal axes of the robot ( $x, y$ ) that might occur when mounting the robot on the playing platform, we introduce an additional rotation around the  $z$ -axis by an angle  $\beta$  (Fig. 2). In other words, in the current configuration, the robot base is placed on the playing platform ( $z$ -axis points upwards), but it might not be exactly parallel to the camera frame. To put it briefly, the transformations of any point  $P = [P_x P_y P_z]^T$  with respect to the camera frame  $\{C\}$  to the same point  $p = [p_x p_y p_z]^T$  with respect to the robot frame  $\{R\}$  are first a translation by  $d_x$ ,  $d_y$  and  $d_z$ , then a change of orientation of the axis (rotation by an angle of  $-\pi/2$  radians about the  $X$ -axis), and a further rotation by an angle  $\beta$  about the  $z$ -axis. This can be mathematically written as:

$$\begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} \cos \beta & \sin \beta & 0 \\ -\sin \beta & \cos \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} P_x - d_x \\ P_y - d_y \\ P_z - d_z \end{bmatrix} \quad (1)$$

After this conversion, one can think of the converted recorded data as if the L-shape defining the origin of the motion capture data had been placed at the origin of the robot coordinate system.

## 2.6 Conversion from the bow markers to the tool-centre-point

The second part of the conversion algorithm considers the position of the bow markers with regard to the position of the robot tool-center-point or TCP. The TCP is a coordinate system given with respect to the robot coordinate frame, which is used to instruct the robot to move. It is by default located at the last flat surface of the robot arm, called the tool flange, as marked with  $\{T\}$  in Figures 2 and 3. For simplicity, the current approach also keeps the TCP in its default setting, but one could relocate it somewhere else on the tool (via the robot software). Since the proposed method is purely geometrical, the bow and the last joint of the robotic arm need to be rigidly clamped. This is achieved by combining two custom-made 3D printed parts to which, on one side, the bow is attached (orange part in Fig. 3) and, on the other side, the 3D-printed clamp is screwed to the tool flange (white part in Fig. 3). In order to make the robot move according to the recording, the recorded bow data (bow coordinate frame  $\{B\}$ ) have to be translated to the TCP (tool coordinate frame  $\{T\}$ ). For that, we must assume that the bow is a rigid body. Attaching the considered bow markers as close as possible to the robot-bow junction reduces the error introduced by this assumption (the marker at the tip of the bow is therefore not used in the conversion, but later for the analysis). The next steps describe how to implement the conversion from the bow markers to the TCP and adapt the motion capture data to be ready for communication with the robot.

### 2.6.1 Determine the bow coordinate frame

To obtain the bow location in the 3D space, one needs its position and its orientation. The position of an object is given by the position of one point of the object. The orientation can be represented using a rotation matrix  $\mathbf{B}$ , which is constructed using the unit vectors defining its coordinate frame  $\{\mathbf{B}\}$  as columns. To determine the unit vectors  $\mathbf{u}_x$ ,  $\mathbf{u}_y$  and  $\mathbf{u}_z$ , the bow markers are employed.

The origin of the bow coordinate frame  $\{\mathbf{B}\}$  is placed on the marker C (see Figs. 2 and 3). Several sets of bow markers have been tested to calculate the unit vectors. The logical use of markers ABC to define the unit vectors as the normalised version of vectors  $\overline{CB}$  and  $\overline{CA}$  and their cross-product led to increased errors. This is because the stick of the bow might slightly turn along its axis, thus forcing marker A to rotate as well. This gentle rotation of the stick is due to the bow adjuster, a mechanism of the bow to tighten the bow hair through a screw-system at the frog.

We therefore use the frog F-marker instead of the A-marker. Note that the vectors  $\overline{CB}$  and  $\overline{FC}$  are linearly independent since C and B are above the bow stick and F is screwed on the bow adjuster (Fig. 3). Using markers F, C and B, one can calculate the rotation matrix of the bow  $\mathbf{B}$  which has unit vectors  $u_x$ ,  $u_y$ ,  $u_z$  as columns, as:

$$B = [u_x \quad u_y \quad u_z], \text{ with} \quad (2)$$

$$u_x = \frac{\overline{CB}}{|\overline{CB}|}, u_y = u_x \times u_{FC} \quad \text{and} \quad u_z = u_x \times u_y$$

where  $u_{FC} = \overline{FC}/|\overline{FC}|$  is an auxiliary unit vector used to calculate  $u_y$ ;  $u_x$  is the longitudinal unit vector of the bow;  $u_y$  is the transversal unit vector, and  $u_z$  is perpendicular to both and pointing to the hair of the bow (Fig. 2).

At this step, one might want to consider a certain rotation  $\varphi$  along the transverse axis  $u_y$  that would allow introducing a given angle between the longitudinal axis of the bow and the corresponding axis of the tool, as showcased in Figure 3. To do that,  $u_y$  is maintained,  $u_x$  is rotated in its 3D coordinates by an angle  $\varphi$  and  $u_z$  is recalculated as the cross product of  $u_y$  and the new  $u_x$ . In the current algorithm, we allow a rotation of  $\varphi = -1^\circ$ . This value can be adjusted if another bow-tool clamp system is used.

### 2.6.2 Alignment of the bow and tool coordinate frames

For robotic arms, it is common to use a coordinate frame called tool-center-point (TCP) that is located either at the flat surface of the last robot joint (tool flange) or at a chosen point related to the tool in use [31]. By convention, this coordinate frame is set so that its  $z$ -axis points out of the tool flange. The  $z$ -axis is sometimes called the *approach vector*, since it moves the tool closer or further away from the object it interacts with. In the current approach, the  $z$ -axis is crucial to approaching the string with sufficient force to satisfactorily bow on it. The next step involves converting the data describing the motion of the bow coordinate frame  $\{\mathbf{B}\}$  to that of the tool coordinate frame, denoted as  $\{\mathbf{T}\}$  (Figs. 2 and 3). The alignment between these two coordinate frames is achieved through a change of basis from  $\{\mathbf{B}\}$

**Table 1.** Values for the conversion parameters as used in this study.

Sampling frequency	Motion capture: 240 fps Robot control: 250 Hz Audio recording: 44100 Hz
Robot center respect to camera origin	$d_x = 576.62$ mm $d_y = 163.00$ mm $d_z = 25.84$ mm $\beta = -0.5157^\circ$
Bow marker C to TCP alignment	$t_x = 115.45$ mm $t_y = 0$ $t_z = 16$ mm (case A) $t_z = 15$ mm (case B) $\varphi = -1^\circ$

(centred at the marker  $C = [C_x \ C_y \ C_z]^T$ ) to  $\{\mathbf{T}\}$  (centred at the origin of the TCP  $T = [T_x \ T_y \ T_z]^T$ ). Since the orientation of the bow has already been calculated as the orientation of the TCP in the previous step (with the  $z$ -axis pointing to the bow hair), and the bow and the tool are rigidly clamped, the change of basis implies just a translation from  $C$  to  $T$ , given by:

$$\begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = \begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix} - B \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (3)$$

where  $t_x$ ,  $t_y$  and  $t_z$  are the geometrical distances between the C-marker and the tool-center-point  $T$  (Fig. 3). Table 1 shows the geometrical parameters used for the conversion in the present study. The values  $T = [T_x \ T_y \ T_z]^T$  are the first three inputs that define the position of the robot tool. The next three are its orientation, given in axis-angle representation.

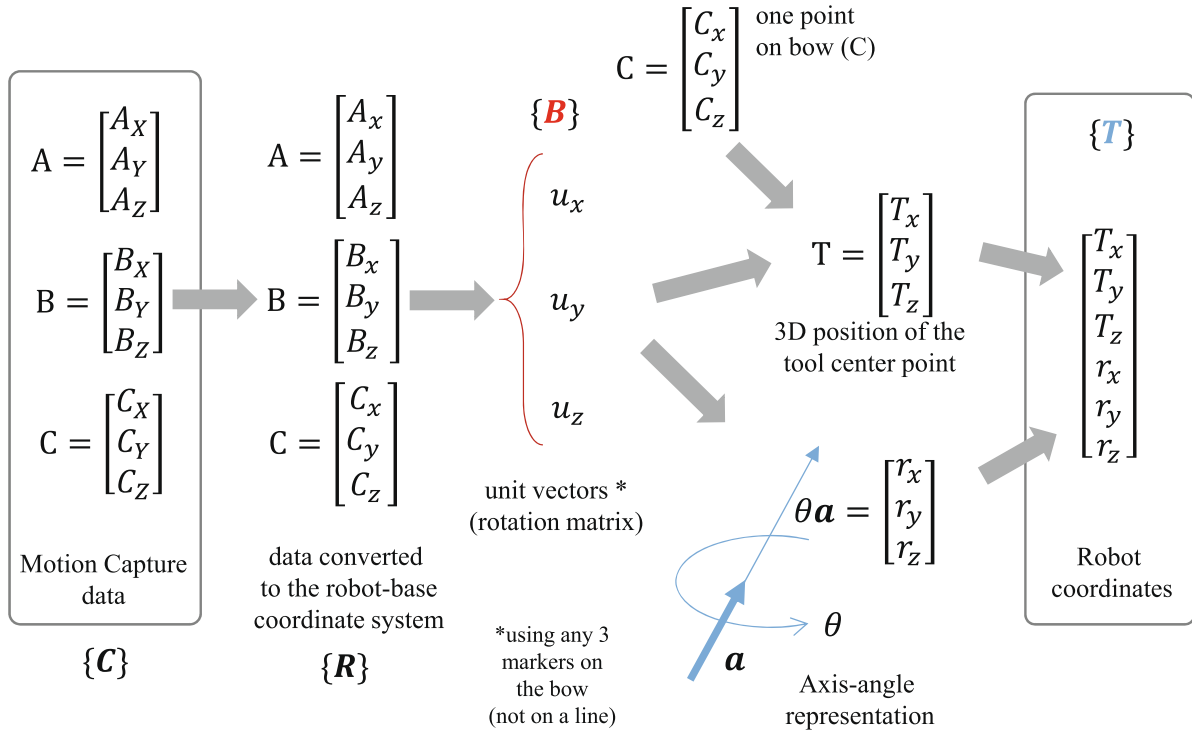
### 2.6.3 The axis-angle representation

The orientation in 3D of the robot tool is given by three values corresponding to its axis-angle representation. The rotation matrix of the bow  $\mathbf{B}$  (Eq. (2)) is used to compute the axis-angle representation, which is required by the robot software to read out the orientation of the object, and it is expressed as  $[r_x \ r_y \ r_z]^T = \theta \mathbf{a}$ . Here,  $\theta$  represents the angle (in radians) and  $\mathbf{a}$  the vector characterising the orientation of the bow, as:

$$\theta = \arccos \frac{\text{Tr}(\mathbf{B}) - 1}{2} \quad (4)$$

$$\mathbf{a} = \frac{1}{2 \sin \theta} \begin{bmatrix} b_{32} - b_{23} \\ b_{13} - b_{31} \\ b_{21} - b_{12} \end{bmatrix} \quad (5)$$

where  $\text{Tr}(\mathbf{B})$  is the trace of the rotation matrix  $\mathbf{B}$  and  $b_{ij}$  are the cell values at the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of the rotation matrix. As the rotation matrix is formed with the unit vectors, the values  $b_{ij}$  are obtained in the previous step. We remark that the columns of the matrix must be of norm 1. The above formula is valid only when the matrix  $\mathbf{B}$  is not



**Figure 4.** Schematic of the process to convert data from the motion-captured 3D position of three markers on an object to the robot coordinates defining its pose.

symmetric, that is, when the rotation angle is not 0 nor  $\pm\pi$ . Such singularities can be detected by checking if the trace  $\text{Tr}(\mathbf{B})$  is equal to 1 (case  $\theta = \pm\pi$ ) or 3 (case  $\theta = 0$ ). If the trace equals 3,  $\mathbf{B}$  is an identity matrix, and any vector can be chosen for  $\mathbf{a}$ . In the case of a trace equal to 1, known as Gimbal lock, a different approach to finding the vector  $\mathbf{a}$  is required (the reader can refer to a general robotics reference for more details [31]).

Finally, at each time instance, the six robot coordinates are determined, i.e. the pose of the tool, comprising three positions and three angles, denoted as  $[T_x, T_y, T_z, r_x, r_y, r_z]$ . The conversion process is schematised in Figure 4. To achieve the motion of the bow with the robotic system, a sequence of positions and orientations is continuously transferred to the robotic arm at a rate of 250 Hz.

## 2.7 Data transfer between systems

The original motion capture data are exported in a CSV file containing the X, Y, Z positions of the required markers in the  $\{C\}$  frame. These data were used without any filtering for the purpose of data transfer to the robot. After the aforementioned geometrical conversion (implemented in MATLAB, by MathWorks), the frame rate is interpolated from 240 Hz (motion-capture rate) to 250 Hz (robot data-exchange rate). A series of tool coordinates at every time-step  $n$  are exported to a txt file as  $[T_x^n, T_y^n, T_z^n, r_x^n, r_y^n, r_z^n]$ . This file is then read line-by-line in LabVIEW (by National Instruments) to transfer the data to the robot.

The robotic arm might be controlled via a user-friendly environment that allows for a set of custom-made instructions [16, 33]. However, if the requested motion requires

reading a long list of poses (such as in the present case), the robot should be controlled externally, so that it can be given one pose at a time. Using the Real-Time Data Exchange (RTDE) method [34], both the TCP pose (Cartesian space) or the joint angles of the robot can be driven externally. In the present approach, the PC-software communicating with the robot is written in NI – LabVIEW [35]. A graphical user interface (GUI) is used to visualise and control the data transfer and the recordings. As data transfer to the robot happens at a rate of 250 Hz, the robot coordinates are updated every 0.004 s. The receiving robot program performs the inverse kinematics conversion in real time on the internal computer of the robotic arm (from the positions and angles of the TCP to the angles of the robot joints).

## 3 Evaluation

For the present publication, several recordings were made so that a comparison could be made between the human performance (with a professional cellist) and two performances by the robotic arm (instructed with the recording by the human player). We introduce several metrics to evaluate the errors in the robot performances.

### 3.1 Experimental procedure

A cello of size 4/4 (adult size) was attached to a stable support structure on a playing platform, to remain immobile during the tests (see Fig. 1). The position and inclination of the cello were predetermined based on a study of six

cellists, taking into account their height, body proportions and comfort [36]. The recordings consisted of synchronised measurements of motion capture and sound using a microphone at 45 cm in front of the instrument (ROGA Instruments RG50 – also affixed to the playing platform), as well as a piezoelectric sensor on the bridge, with a sampling rate of 44100 Hz. Supplementary video images were taken to facilitate data analysis.

First, an experienced cellist was invited to perform various open-string exercises on the prepared instrument and bow (i.e. only using the bow, without fingering any notes with the left hand), as well as a real piece adapted to be played on three open strings (cello part of “Spring” 1<sup>st</sup> movement of The Four Seasons by Antonio Vivaldi). After data conversion (see Sect. 2), the robot was instructed to replicate the bow motion. Using the same recording setup as for the human player, the 3D trajectories of the bow were again recorded for comparison (motion capture, audio and video data). The data collection using the robotic-arm took place two weeks after the session with the human player. The protocol was approved by the Ethics Committee of the mdw – University of Music and Performing Arts Vienna.

This paper showcases two very similar situations of bow-motion replication with the robotic arm. They differ only by 1 mm in the  $t_z$  value (Fig. 3). Changes in parameter  $t_z$  allow adapting the whole data set in the direction of the longitudinal axis of the tool ( $z$ ). Since the plane of the bow hair approximately faces the instrument during playing (so that the bow hair touches the strings), a decrease in  $t_z$  moves all recorded data in the direction of the  $z$ -axis of the tool, and thus brings the bow closer to the instrument. In case A, the parameter  $t_z$  equals 16 mm, and thus the bow is slightly further away from the instrument. And in case B, the parameter  $t_z$  equals 15 mm, and the bow is closer to the instrument. After observing some inaccuracy in the bow-string contact in case A, case B is expected to provide a more stable contact.

### 3.2 Data processing and analysis

The motion capture data shown in this paper regard the motion of selected markers to compare human playing with the two robot-playing conditions (cases A and B). The data were low-pass filtered using a Butterworth 3<sup>rd</sup> order filter at 12 Hz to reduce measurement noise. The markers’ positions  $x$ ,  $y$ ,  $z$  are given with respect to the robot coordinate frame  $\{\mathbf{R}\}$ . This implies that when the player bows “to their right” (down bow),  $x$  increases, and when the player bows “to their left” (up bow),  $x$  decreases. The  $z$  coordinate is vertical (positive upwards), and the  $y$  coordinate grows to the front and diminishes to the back with respect to the instrument.

In the previous conference publication [29], we compared human vs. robot performances by first considering only one marker on the bow (i.e. the frog marker), and later considering two reflective markers on the bow and two on the cello. On the bow, the frog and the tip markers (F, T in Fig. 2) create a line  $\overline{FT}$ ; on the cello, the bridge and the nut markers (S and N in Fig. 2) create the line  $\overline{SN}$ .

The parameter  $\sigma_{\text{contact}}$  is then calculated as the minimum geometrical distance between these two lines in the 3D space. In an ideal scenario,  $\sigma_{\text{contact}}$  should exhibit similar behaviour for both human and robot performances. Any deviations in  $\sigma_{\text{contact}}$  are an indication of the inaccuracies that might arise at the bow-string contact. Therefore, it serves as a measure to evaluate the robot’s performance, specifically at the contact point of the bow on the string. We showed that the accuracy of  $\sigma_{\text{contact}}$  seemed to vary depending on the string being played, with the best results for string  $D_3$  [29]. Yet, a systematic analysis that compares the motion capture data to the resulting sound was missing.

For the present study, two more parameters related to the produced sound are introduced. The first parameter compares the root-mean-square (RMS) value per note among the three playing situations (human, robot case A, robot case B), calculated as

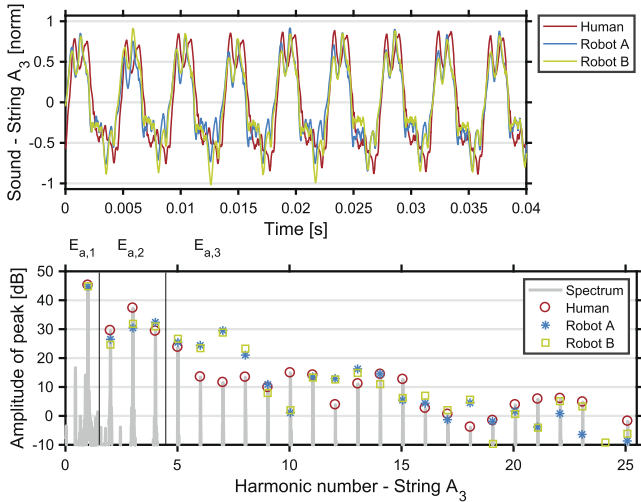
$$\text{RMS}_n = \sqrt{\frac{1}{N_n} \sum_{i=1}^{N_n} s_i^2} \quad (6)$$

where  $s_i$  ( $i = 1, 2, \dots, N_n$ ) are the excerpts of the sound files with a length of  $N_n$  per note  $n$ . The RMS-values are then subtracted as  $E_{\text{RMS}} = |\text{RMS}_{\text{human}} - \text{RMS}_{\text{robot}}|$  to obtain the absolute error per note. The mean absolute errors are calculated and classified according to the played string. The error  $E_{\text{RMS}}$  indicates whether the sound level produced in each of the two robot cases is comparable to that of the human playing situation. It can then be compared to the errors appearing in the parameter  $\sigma_{\text{contact}}$ , calculated as  $E_{\sigma} = |\sigma_{\text{human}} - \sigma_{\text{robot}}|$  at each note, and averaged and classified at each string (see Tab. 2).

A second sound-related measure is introduced in order to consider not only the sound level but also the spectrum of the produced sounds. This measure aims to indicate differences in the amplitude of the spectral harmonics. Figure 5 shows the similarity between the human and robot playing, both in the waveform and on the frequency peaks (shown for string  $A_3$ ).

To further explore this comparison for the whole recording and quantify it, three parameters inspired by the tris-timulus timbre model [37] are calculated. They aim to compare separately the errors in amplitude of the fundamental frequency ( $E_{a,1}$ ), the average of the second, third, and fourth harmonics ( $E_{a,2}$ ), and the averaged amplitude of the higher harmonics ( $E_{a,3}$ ), as indicated with vertical lines in Figure 5. To do that, the errors in amplitude are first calculated for all considered harmonics as  $e_{a_h} = |a_{h,\text{human}} - a_{h,\text{robot}}|$ , where  $a_h$  is the amplitude of the  $h^{\text{th}}$  harmonic, for  $h = 1, 2, 3, \dots, H$ . Then the three parameters are computed as:

$$\begin{aligned} E_{a,1} &= e_{a_1} \\ E_{a,2} &= \frac{1}{3} \sum_{i=2,3,4} e_{a_i} \\ E_{a,3} &= \frac{1}{H-4} \sum_{i=5}^H e_{a_i}. \end{aligned} \quad (7)$$



**Figure 5.** Waveform (top) and harmonic peaks (bottom) comparing human and robotic playing at string  $A_3$ . The spectrum frequency-axis is given as normalised by the fundamental frequency (so that the peaks align with each harmonic number). The spectrum of the human-played tone is plotted in gray for better visualisation. The vertical lines separate the harmonic peaks used for the errors  $E_{a,1}$ ,  $E_{a,2}$  and  $E_{a,3}$ .

In the current study, the number of harmonics considered is  $H = 30$ . Note that the first parameter directly depicts the error in the amplitude of the fundamental frequency. To calculate the spectra of the signals, time-windows of 0.4 s with 0.2 s overlap are used, which are windowed with a Hamming function, and their fast-Fourier transformation (FFT) is computed. After that, the mean absolute errors are calculated at each of the three strings for comparison (Table 3).

## 4 Results

We present in this paper the comparison of three playing situations: a real cellist playing a piece of music adapted for 3 open strings, and its robotic reproduction with the proposed method in two cases (A, B). Cases A and B differ by a change of 1 mm in the direction of the robot tool, i.e., in case B the bow is slightly closer to the instrument than in case A. If there is bow-string contact in both cases, the change from A to B results in increased bow force on the string. If there is no contact in case A, then a change in  $t_z$  (towards the instrument) may lead to bow-string contact in case B. For data regarding all four open strings of the cello in an exercise concerning dynamic changes (from soft to loud dynamic range), the reader can refer to the proceedings of Forum Acusticum 2023 [29].

### 4.1 Position and velocity

Figure 6 shows an example of the kinematic data gathered in the tests. On the left-side, the position in three axes of the frog marker F ( $x_F$ ,  $y_F$ ,  $z_F$  - given according to the robot world frame  $\{\mathbf{R}\}$ ) demonstrates how similarly the

robot (robot case A and B) mimics the human playing. Even though one would expect that the note-onsets (vertical lines) happen at the bow change (changes from down-bow -  $x$  increases, to up-bow -  $x$  decreases), as it was observed in the previous publication ([29], Fig. 4), it appears not to be the case in the present recording. Here, due to the presence of playing nuances (e.g. accents or notes that are shortened due to a certain articulation style), the note onsets do not always directly coincide with the bow change. As it was observed in Ref. [29], the biggest errors in the positioning of the bow after the conversion from motion capture data to robotic coordinates appear at the bow change. In the following lines we give the errors of the recorded kinematics of marker F as  $E_x = x_{F,\text{human}} - x_{F,\text{robot}}$  for coordinate  $x$  as well as for coordinates  $y$ ,  $z$  and their first derivatives.

Considering the whole recording, the mean errors for case A and case B, respectively, report: mean ( $E_x$ ) = {2.3, 1.9} mm, mean ( $E_y$ ) = {-0.1, 0.5} mm, mean ( $E_z$ ) = {-2.9, -2.1} mm. If the robot data showed a consistent offset in any of the axes, one could consider adapting the whole setup by manually fine-tuning the distances  $d_x$ ,  $d_y$  and  $d_z$ , respectively. These distances are an approximation of the distance between the centre of the camera coordinate frame and the centre of the robot coordinate frame. An analysis of the mean error of one marker on each axis can already assist into the evaluation of the setup, for example at the beginning of the recording session, to check that the outcome conversion-methodology is as expected. Such a procedure was performed at the beginning of the recording session, distances  $d_x$ ,  $d_y$  and  $d_z$  were then fixed (values given in Table 1). After that, only a change in the bow-tool distance  $t_z$  was tested for the experiment (cases A and B).

Errors in the position along the  $x$ -axis (main bowing axis) are less problematic than errors in the  $z$ -axis (horizontal - approaching the instrument from the front) and the  $y$ -axis (vertical - approaching the instrument from above): inaccuracies in  $y$  and  $z$  axes can result in a loose bow-string contact, while inaccuracies in the  $x$ -axis mainly translate the bow along its longitudinal axis. Although the reported mean errors are fairly small, maximum errors can reach values of around 1 cm at each axis. The maximum errors are reported for the showcased passage (Fig. 6), for case A and case B, respectively, as  $\max(E_x) = \{9.9, 9.6\}$  mm,  $\max(E_y) = \{4.7, 5.4\}$  mm,  $\max(E_z) = \{2.8, 3.8\}$  mm. And for the whole recording as  $\max(E_x) = \{17.3, 16.7\}$  mm,  $\max(E_y) = \{7.9, 8.4\}$  mm,  $\max(E_z) = \{7.1, 8.0\}$  mm. These maximum errors are found at the instances where extreme direction changes take place. Since the robotic arm is much more massive than the person's arm, the robotic arm reduces its velocity around these fast changes, acting as a low-pass filter effect on the recorded motion capture data. This effect becomes more apparent when regarding the velocity curves (right-side of Fig. 6). Note that the reported mean errors and maximum errors also depend on the precision of the motion capture camera setup used to compare the performances.

The mean errors regarding velocity (right-side of Fig. 6) are a good indicator of the precision of the robotic system:



mean ( $E_{v_x}$ ) = {0.84, 0.86} mm/s, mean ( $E_{v_y}$ ) = {0.20, 0.21} mm/s, mean ( $E_{v_z}$ ) = {-0.19, -0.17} mm/s, for case A and B, respectively. This means that although there might be an offset in all three axes in terms of position (resulting from geometry imprecisions), the velocities present no offsets. Since most of the control of dynamic level in bowed-string instruments happens because of the velocity of the bow [7], it is crucial that the system achieves the required velocity. Moreover, the almost identical reproduction of the velocity curves between A and B (Fig. 6) shows the high repeatability of the robotic setup. Although the peaks of velocity of the recorded motion are below the limits for typical TCP speed specified by the robot’s manufacturer (1000 mm/s [30]), the tested motions are already reaching the limits of the setup. The difficulties of the current requirements are the fast change of direction of the bow (reduce velocity and suddenly increase velocity in the opposite direction), and the short peak in velocity happening in between bow changes. Maximum errors in velocity for this passage report:  $\max(E_{v_x}) = \{-95.1, -94.6.1\}$  mm/s,  $\max(E_{v_y}) = \{-104.9, -106.3\}$  mm/s,  $\max(E_{v_z}) = \{-140.7, -141.0\}$  mm/s. Notice that less expressive music (as the exercises presented in [29]) would show a velocity plot that resembles a squared wave. Here instead, it resembles a sinus curve, with higher velocity peaks. For comparison, the passage shown in Figure 6 shows values of  $v_x$  that double those found for the exercises in reference ([29], Fig. 4). This suggests that expressive music challenges the requested reproduction task even more than controlled bowing exercises.

## 4.2 Closer contact: better sound?

Since the analysis was so far centred on one marker (frog marker), one would need to consider more markers if a comparison with the produced sound is needed. Therefore, the parameter  $\sigma_{\text{contact}}$  is proposed as a measure of the distance between the 3D-lines formed by two extreme bow markers (frog F and tip T markers) and two extreme string markers (bridge S and nut N markers in Fig. 2). Figure 7 shows the same passage of the recorded piece as in Figure 6, comparing the sound (waveforms and RMS-envelopes) and the parameter  $\sigma_{\text{contact}}$ .

Although it was shown in Figure 6 that differences between case A and B are difficult to analyse out of the raw motion capture data (left-side of Fig. 6), parameter  $\sigma_{\text{contact}}$  amplifies these differences and facilitates the analysis. At the bottom of Figure 7, parameter  $\sigma_{\text{contact}}$  shows that neither case A nor case B exactly follow the human performance, yet case B appears to be closer to the desired result in  $\sigma_{\text{contact}}$  than case A. Observations of an offset in the parameter  $\sigma_{\text{contact}}$  could therefore be used to adjust the system – as it is done between cases A and B – by applying a slight modification in the parameter  $t_z$  when converting the data. Similarly, differences in RMS-envelope (middle plot), are in this example wider for case A than case B. RMS-envelopes in this plot are computed using a running RMS envelope with a window of 4000 samples (about 10 periods of the lowest note). An analysis of the errors

**Table 2.** Absolute mean errors  $E_\sigma$  and  $E_{\text{RMS}}$  comparing the human and robot performances (robot case A and robot case B) as a function of the played cello open string ( $A_3$ ,  $D_3$ ,  $G_2$ , from high to low). Column # indicates on how many notes case B presented a smaller error than case A.

	$E_\sigma$ [mm]			$E_{\text{RMS}}$ [norm.]		
	Case		#	Case		#
	A	B		A	B	
String $A_3$	1.80	0.96	25/26	0.068	0.044	24/26
String $D_3$	1.98	0.37	30/36	0.019	0.012	32/36
String $G_2$	0.37	0.58	4/10	0.039	0.028	7/10

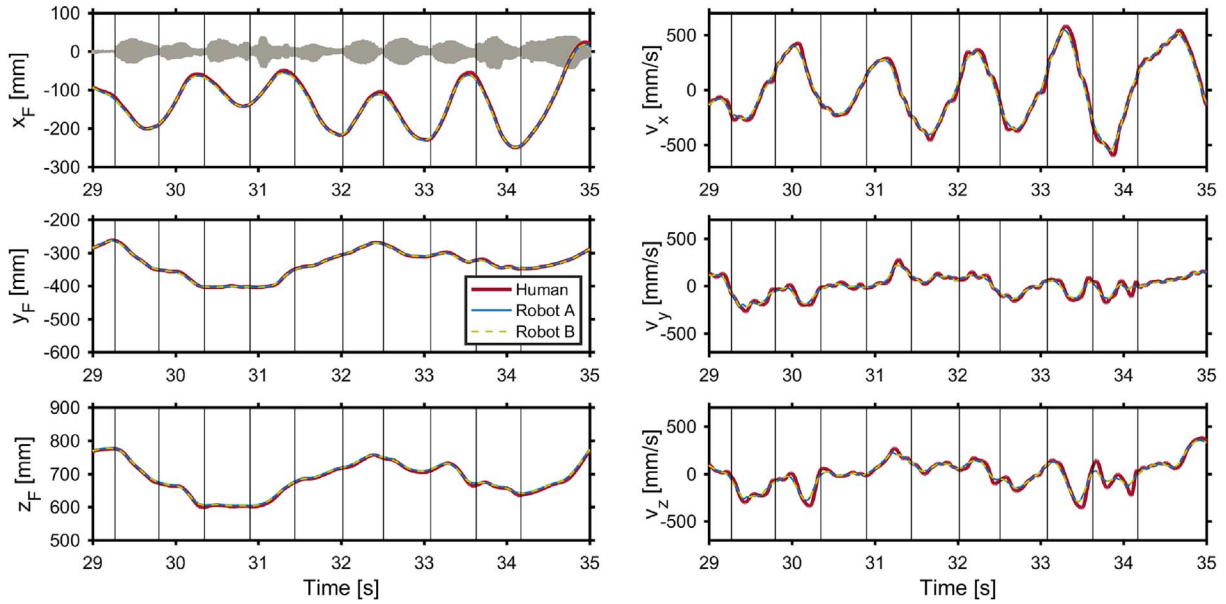
**Table 3.** Absolute mean errors in the harmonic-amplitude-parameters  $E_{a,i}$  comparing the human and robot performances (robot case A and robot case B) as a function of the played cello open string ( $A_3$ ,  $D_3$ ,  $G_2$ , from high to low).

	$E_{a,1}$ [dB]		$E_{a,2}$ [dB]		$E_{a,3}$ [dB]	
	Case		Case		Case	
	A	B	A	B	A	B
String $A_3$	7.5	5.7	6.4	5.3	7.8	8.3
String $D_3$	7.1	4.7	7.6	5.8	9.0	7.6
String $G_2$	2.5	1.8	5.9	4.8	6.9	6.8

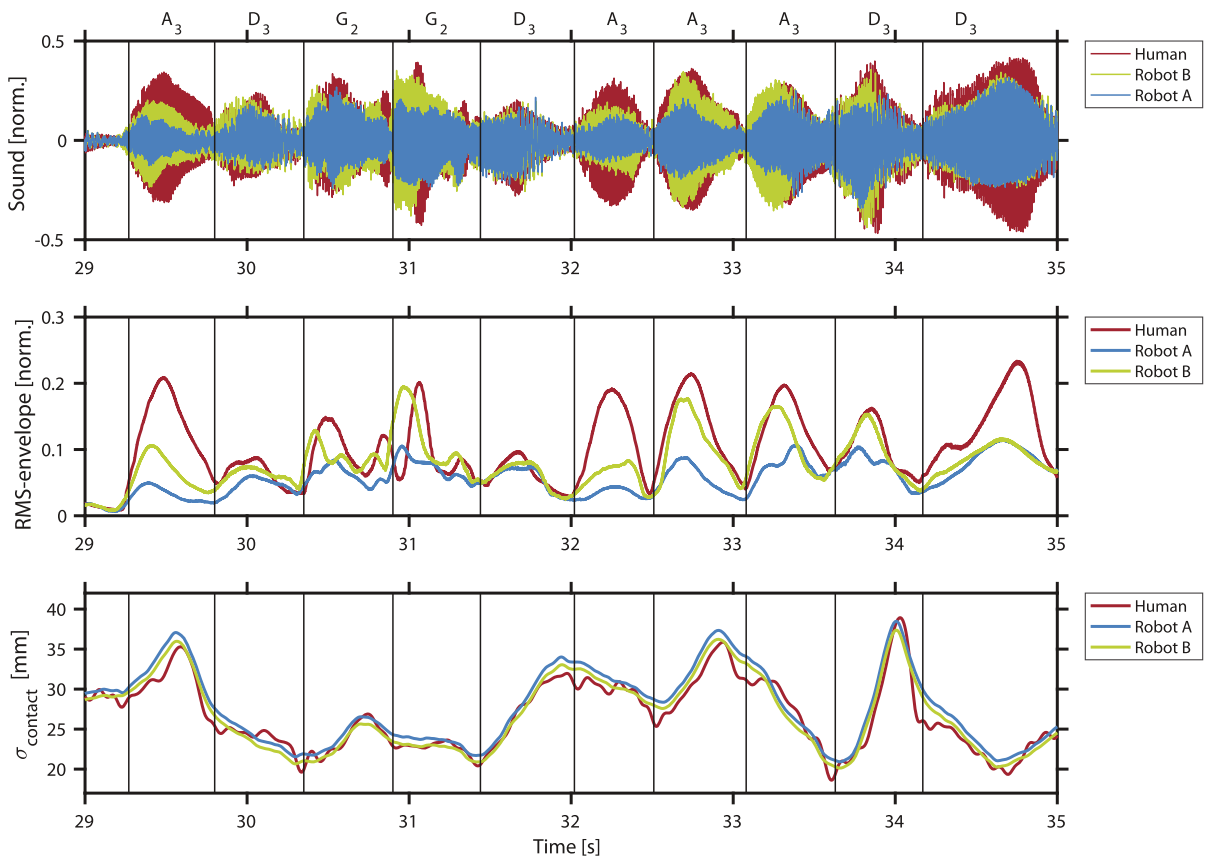
throughout the recorded signals using the selected parameters follows.

Table 2 summarises the average per played string ( $A_3$ ,  $D_3$ ,  $G_2$ ) of the two parameters  $E_\sigma$  and  $E_{\text{RMS}}$ . It shows how much error the robot performances (case A and case B) have with respect to the human performance at each string, and it also indicates on how many notes the case B resulted in a smaller error than case A (marked with #). The notes are repeated 26, 36 and 10 times at strings  $A_3$ ,  $D_3$ ,  $G_2$ , respectively. From Table 2, it becomes clear that case B performs significantly better than case A for strings  $A_3$  and  $D_3$  (highest and middle-high string of the cello), but not for string  $G_2$  (middle-low string of the cello). This observation is in agreement with reference [29].

The errors  $E_{a,1}$ ,  $E_{a,2}$ ,  $E_{a,3}$  that express the three tristimulus-inspired parameters regarding the amplitude of the spectral harmonics are calculated according to equation (7) for robot case A and case B with respect to the human performance. Table 3 shows the three errors classified according to the string played. In this case, for the fundamental frequency,  $E_{a,1}$  always shows an improvement of case B compared to case A. The same applies to the middle frequencies:  $E_{a,2}$  shows a smaller error in B than in A. For the high frequencies represented with  $E_{a,3}$ , a positive observation can only be made for string  $D_3$  but not for strings  $A_3$  and  $G_2$ . Again, string  $D_3$  is the one that shows the greatest advantage of a change from case A to case B. The differences observed in the high frequencies are also dependent on the pitch played, since the parameter  $E_{a,3}$  considers the harmonic peaks starting at the 5<sup>th</sup> peak, and these frequencies are the highest for string  $A_3$ ; that could explain the increased errors at the highest string.



**Figure 6.** Example of the raw motion capture data for all axes on the frog marker ( $x_F$ ,  $y_F$ ,  $z_F$ ) and their first derivatives ( $v_x$ ,  $v_y$ ,  $v_z$ ). The plot top is synchronised with the sound waveform of this passage for comparison (arbitrary units). Vertical lines indicate the note onsets.



**Figure 7.** Passage of the recorded data showing the sound envelopes (waveform on the top and RMS-envelope in the middle) and the parameter  $\sigma_{\text{contact}}$ , comparing human and robot cases A and B. Vertical lines correspond to the note onsets (played notes  $A_3$ ,  $D_3$ ,  $G_2$  indicated at the top).

**Table 4.** Improvement rate given as the percentage of the occurrences where case B shows a smaller error than case A in the recorded data (if improvement rate  $> 50\%$ ). Errors in the parameter  $\sigma_{\text{contact}}$ , the running RMS-envelope (as in Fig. 7) and the harmonic-amplitude errors  $E_{a,i}$  comparing the human and robot performances (robot case A and robot case B). First without string-classification (global) and then classified at each string.

	Improvement rate				
	$E_{\sigma}$	$E_{\text{RMS}}$	$E_{a,1}$	$E_{a,2}$	$E_{a,3}$
Global	73.6%	67.3%	64.8%	75.2%	64.8%
A <sub>3</sub>	87.3%	74.4%	48.8%	71.6%	36.4%
D <sub>3</sub>	73.5%	66.1%	72.6%	78.4%	79.4%
G <sub>2</sub>	55.6%	55.0%	62.5%	65.6%	56.2%

Tables 2 and 3 show the mean errors per string played, after averaging the values obtained at each played note repetition. If the whole recording is now considered, without separating the files according to the note played, one could calculate the errors of the selected parameters at a more global scale. This is reported at the top of Table 4, which regards whether case B is an improvement with respect to case A (improvement rate  $> 50\%$ ) or not (improvement rate  $< 50\%$ ). This improvement rate quantifies the time instances where case B presents a smaller error than case A. This is calculated as the amount of samples where B is closer to the human ground-truth than A divided by the total length of the recording. If the same procedure is repeated on the segments of the recording that correspond to each played string, the improvement rate can be calculated per string (A<sub>3</sub>, D<sub>3</sub>, G<sub>2</sub>), as also shown in Table 4.

If a more strict threshold of 60% improvement rate is set, one can determine whether case B is really an amelioration compared to case A. As a general observation, one can confirm that case B performs better than case A, showing improvement rates above 60% in the “global” row for all considered parameters (top of Tab. 4). Yet when considering strings separately, the analysis is not as straightforward. Regarding errors  $E_{\sigma}$  and  $E_{\text{RMS}}$ , a similar behaviour is observed. Namely, for string G<sub>2</sub> the adjustment from case A to case B does not significantly affect the outcome ( $< 60\%$ ), whereas for strings A<sub>3</sub> and D<sub>3</sub>, case B significantly improves both parameters. The best improvement appears for string A<sub>3</sub>, for which case B shows a smaller error than case A at 87.3% of the time instances for  $\sigma_{\text{contact}}$  and at 74.4% for the RMS-envelope.

Regarding the spectral parameters  $E_{a,i}$ , it is shown again that the best improvement rate is found at string D<sub>3</sub> ( $> 70\%$  for all  $E_{a,i}$ ), but string A<sub>3</sub> does not experience an improvement from case A to case B at the fundamental frequency and at high frequencies (only the middle frequencies improve  $E_{a,2}$ ). String G<sub>2</sub> shows a light improvement at all frequency bands. Therefore, the previously observed trend (better match of case B at higher strings than at lower strings [29]) is not found here. Instead, only the central strings D<sub>3</sub> and G<sub>2</sub> show a certain improvement. In general, these results confirm that robot case B not only

achieved similar sound-level values to the human performance, but it also resulted in a closer sound spectrum.

## 5 Discussion

This study presents an approach to replicate the bow movements on bowed-string instruments using a robotic arm, and it describes the methodology to use motion capture data to obtain the bow motion. Through the integration of optical motion capture technology and the control of the robotic arm, we have demonstrated the feasibility of reproducing the bow kinematics in a controlled setting. The proposed methodology does not use any active regulation, instead, it uses a set of coordinates that depict the motion of the bow to control the motion of the robotic arm. Thanks to this direct control of the robotic arm, the motion of the bow as moved by the robot is fairly close to that of the cello player. Yet, given the complexity of the bow-string interaction, minimal changes in the bow motion and force on the string are known to result in great changes in the sound production. This poses an extra challenge when aiming at reproducing playing nuances such as articulation techniques and precise dynamic-level changes. The current manuscript exemplifies this challenge by presenting the recording of a piece played on open strings on the cello, and comparing the robot outcome in two similar playing situations.

Inaccuracies in the positioning of the bow on the string lead to imprecisions at the bow-string contact. Note that optical motion capture was not developed for the purpose of precise measurements of position but for motion animation in the film and video games industry and, subsequently, in performance science studies [22]. The main source of position error in the current approach is therefore found in the acquisition and conversion of motion capture data between systems. To minimise inaccuracies in the position of the bow, one could adapt the setup by bringing the whole trajectory closer to the strings. To do that, an axis orthogonal to the tool flange is used and all recorded 3D data points are moved on that axis (so-called approach vector). This means that the data are not adapted according to the fixed coordinate frame of the robot base, but depending on the position of the robot tool, the bow trajectory can be adapted forward and backward from the position of the robot tool. This is achieved by a change in the geometrical parameter  $t_z$ . Comparing two similar robot-playing situations, which differ only by 1 mm in  $t_z$ , it was shown that, as expected, small changes in the robot control result in great changes in the produced sound.

The recorded motion capture data and sound signals analysed in this study are available under the repository <https://doi.org/10.5281/zenodo.10696680>, as well as the music score. When listening to the three provided sounds, it is noteworthy to mention how similar the three conditions appear to the lay listener (despite changes in intonation, and the directivity of the sound that might have been influenced by the room and the lack of a person behind the instrument). This is particularly remarkable for case B at loud dynamic

level. This resemblance between the audio recordings again shows the problematic nature of the current approach to aim either at a certain level of accuracy in the bow motion, the evaluation of the resulting sound, or some perceptual evaluation of the performance. The current study suggests that misalignments at the bow positioning are caused by the limitations of the robotic system (limits of velocity), as well as by the inaccuracies introduced at the geometrical conversion of the motion capture data (imprecision of the motion capture system and difficulty of precise measurement of the distances between coordinate frames), yet the instrument can still be played following the bow velocity and trajectories. If the misalignment persists and the bow-string contact is not guaranteed, the bow strokes might produce no string oscillation, or a very soft one. This is most noticeable at some attack transients, where the robotic performance fails to achieve a clear attack (for example, refer to the beginning of the recording in case A).

Several parameters to evaluate the result of the reproduction were scrutinised (position, velocity, distance between bow and strings, sound RMS values and spectral parameters). The parameter  $\sigma_{\text{contant}}$ , which quantifies the distance from the bow markers to the cello markers, was shown as the most reliable indicator of the proposed method's accuracy in positioning the bow on the string. This parameter together with the RMS-values per note allowed to compare two robotic performances. We showed that a slight modification in the geometric data conversion reduced the error in  $\sigma_{\text{contant}}$  significantly. It was also made clear that adjustments in the system might react differently on each played string, improving the outcome only at certain bow inclinations.

When considering how to further adjust or continue developing the system with the help of the presented sensors or additional ones, we take into account three scenarios: a motion-capture evaluation, a sound measurement, and a force sensor feedback loop.

- (I) If live data transfer of a motion capture system is available, one could consider visualising the parameter  $\sigma_{\text{contant}}$  to assist in the system adjustments before the recordings. Even if motion-capture real-time data transfer is possible, we do not consider it safe to directly use it in combination with the robot, because of the in-between calculations and the damage that could be caused to the robot or to the instrument if an unforeseen error appears. As an example of this risk, traditionally, robotics would rather use pre-designed robot motions and avoid direct robot control, so that after recognising a certain human motion, a given robot motion is activated that avoids exceeding the limitations of the robotic system. In the present methodology, though, the exact bow motions were aimed, and the limits of the robotic setup were challenged, so that the reproduction would be as close as possible to the bow movement by a real player.
- (II) If a motion capture system is not available when the robot-performance happens, the recorded sound would be the only parameter to evaluate the process

in the current configuration. In that case, the RMS value per note appeared as a fair indicator for a post-evaluation of the recordings. Yet, we believe that the RMS-level would not be suitable for a real-time adjustment in the form of a feedback loop. Since the system can be adjusted in many ways (positioning of the bow on the string, its velocity and force, among others) resulting in a variety of sounds (from no sound to raucous sound, passing through the many nuances of a Helmholtz sound) a simple feedback loop would probably fail at controlling the system. In other words, two very different sounds could have the same sound level. Yet, future work could consider more advanced classification strategies (also including spectral aspects of the sound) to decide how to adapt the system as a function of the recorded sound, while respecting the mechanical limitations of the system.

- (III) If the adjustment of the feedback loop were made considering the measurement of the bow force or the force at the bridge [15, 16, 38], one would still need to decide which parameters would be allowed to change when the loop condition is not matched. Although this might be the only feasible feedback-loop solution, it is an utmost difficult task that might surpass the possibilities of the system (calculation time, reaction time, security to protect the instrument) or might result in adjustments that differ even more from the ground truth (original recording with a cellist).

To address the current limitations, future work could involve measuring and assessing the dynamics of the system (bow force, or force at the bridge), as well as analysing which parameters would need to be manipulated to achieve the desired force values [38]. Measuring the bow force or the force at the bridge would also contribute to the ultimate objective of the setup, which is to replicate and analyse human performance under artificial excitation conditions. Future research could also evaluate the adjustments required to replicate more advanced playing techniques such as spiccato and other off-the-string bowing techniques [17], as well as contemporary techniques with wide variations in the playing parameters [39].

## 6 Conclusions

The present study demonstrates that combining optical motion capture technology with the control of a robotic arm can be used to replicate bow movements in a controlled environment. Despite this achievement, the analysis shows that slight variations in the playing situation, such as changing one parameter during the data conversion, have a great influence in the outcome. Towards an evaluation of the robotic performance, this paper emphasises the importance of a detailed analysis of the motion of the bow and the resulting sound. To that aim, several parameters were calculated regarding the distance from the bow to the strings, the sound amplitude and the spectral content of the sound. It was expected that, since case A seemed to

not provide enough contact between the bow and the string, after approaching the whole bow trajectory to the instrument by 1 mm (case B), an amelioration would be observed. This amelioration happened significantly yet only partially, since it implied a positive improvement rate only for some strings or only for some of the parameters selected for the evaluation.

This paper shows that, although the presented methodology is unable to replicate the exact sound produced by a human player without errors (particularly at the attack transients), it does provide excellent reproduction of the movement of the bow. This feature contrasts with previous bowing machines employing linear programming. The reproduction of human-like bow strokes and their corresponding sound, though, remains an open challenge. The limitations of the current approach are explained by i) the inaccuracies in the acquisition and geometrical conversion of the motion-capture data and ii) a lack of active control during playing. Because of these limitations, the accumulation of small errors can result in poor bow-string contact or even losing contact. The advantage of the presented robotic setup is that it can effortlessly repeat desired sequences at the touch of a button. Since the proposed method departs from traditional robotics and directly transmits almost unaltered motion capture data to the robot, the limitations of the robotic setup are challenged. This unconventional approach to instructing the robotic arm at such a high rate of change is, to the best of our understanding, essential in achieving the desired human-like outcome. As advances in technology continue to provide further hardware and software possibilities, future research could continue exploring ways to assist the analysis of musical expression during bowed-string musical performance using mechanical devices.

### Acknowledgments

We would like to acknowledge Anna Scheiblauber for playing the cello in this study, as well as Alessio Lampis, Werner Goebel and Sara D'Amario for their help and feedback during the preparation of the experimental setup.

### Funding

This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/{P34852 & T1295}]. For open access purposes, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

### Conflicts of interest

The authors declare no conflict of interest.

### Data availability statement

The research data associated with this article are available in Zenodo with a Creative Commons Attribution 4.0 International license, under the reference [40].

### Author contribution statement

Conceptualization and methodology, M.P.-V., A.M.; algorithm implementation, M.P.-V., A.M., E.M., V.C.; hardware and

software A.M.; data collection, curation and visualisation, M.P.-V.; writing – original draft, M.P.-V.; writing – review and editing, M.P.-V., A.M., E.M., V.C.; funding acquisition, V.C., M.P.-V. All authors have read and agreed to the published version of the manuscript.

### Statement concerning the originality of the work

The authors declare that the work is original. A preliminary study was presented in Forum Acusticum 2023 (Torino, Italy), after which we were invited to submit an extended version to the special issue in Acta Acustica.

### References

1. B. Lawergren: On the motion of bowed violin strings, *Acta Acustica united with Acustica* 44, 3 (1980) 194–206.
2. K. Guettler: On the creation of the Helmholtz motion in bowed strings, *Acta Acustica united with Acustica* 88, 6 (2002) 970–985.
3. J. Woodhouse, P. Galluzzo: The bowed string as we know it today, *Acta Acustica united with Acustica* 90, 4 (2004) 579–589.
4. A. Askenfelt: Measurement of bow motion and bow force in violin playing, *Journal of the Acoustical Society of America* 80, 4 (1986) 1007–1015.
5. A. Askenfelt: Measurement of the bowing parameters in violin playing. II: bow–bridge distance, dynamic range, and limits of bow force, *Journal of the Acoustical Society of America* 86, 2 (1989) 503–516.
6. M. Demoucron: On the control of virtual violins – physical modelling and control of bowed string instruments, PhD thesis, Université Pierre et Marie Curie – Paris VI; Royal Institute of Technology, Stockholm, 2008.
7. E. Schoonderwaldt: The violinist's sound palette: spectral centroid, pitch flattening and anomalous low frequencies, *Acta Acustica united with Acustica* 95, 5 (2009) 901–914.
8. E. Maestre: Modeling instrumental gestures: an analysis/synthesis framework for violin bowing, PhD thesis, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, 2009.
9. C.V. Raman: On the mechanical theory of the vibrations of bowed strings and of musical instruments of the violin family, with experimental verification of the results, part I, *Indian Association for the Cultivation of Science* 15 (1918) 243–276.
10. J.C. Schelleng: The bowed string and the player, *Journal of the Acoustical Society of America* 53, 1 (1973) 26–41.
11. A. Cronhjort: A computer-controlled bowing machine (mums), *Speech, Music and Hearing Quarterly Progress and Status Report* 33, 2–3 (1992) 61–66.
12. P. Galluzzo, J. Woodhouse: High-performance bowing machine tests of bowed-string transients, *Acta Acustica united with Acustica* 100, 1 (2014) 139–153.
13. R. Mores: Precise cello bowing pendulum, in: *Proceedings of Third Vienna Talk on Music Acoustics*, vol. 106, Vienna, Austria, 16–19 Sept 2015.
14. T. Rossing: *The Science of String Instruments*, Springer, New York, 2010.
15. E. Schoonderwaldt, M. Demoucron: Extraction of bowing parameters from violin performance combining motion capture and sensors, *Journal of the Acoustical Society of America* 126, 5 (2009) 2695–2708.
16. A. Lampis, A. Mayer, V. Chatziioannou: Assessing playability limits of bowed-string transients using experimental measurements, *Acta Acustica* (2024). <https://doi.org/10.1051/aacus/2024034>.

17. A. Askenfelt, K. Guettler: The bouncing bow: An experimental study, *Journal of the Acoustical Society of America* 3 (1998) 3–8.
18. A. Mayer, A. Lampis: A versatile monochord setup: An industrial robotic arm as bowing and plucking device, *IWK Tech Report 1-2024*, University of Music and Performing Arts Vienna, 2024. <https://doi.org/10.21939/iwk-tech-report-1-2024>.
19. S. Kemper: Locating creativity in differing approaches to musical robotics, *Frontiers in Robotics and AI* 8 (2021) 647028.
20. K. Shibuya, S. Matsuda, A. Takahara: Toward developing a violin playing robot-bowing by anthropomorphic robot arm and sound analysis, in: *Proceedings of 16th IEEE Int. Symposium on Robot and Human Interactive Communication*, Jeju, Korea (South), 26–29 August, 2007, IEEE, pp. 763–768.
21. S. Jordà: Afasia: the ultimate homeric oneman-multimedia-band, in: *Proceedings of New Interfaces for Musical Expression, NIME-02*, Dublin, Ireland, May 24–26, Media Lab Europe, 2002, pp. 132–137.
22. M.M. Wanderley: Motion capture of music performances, in: G.E. McPherson (Ed.), *The Oxford handbook of music performance*, Oxford University Press, 2022, pp. 465–494.
23. J. Van der Linden, E. Schoonderwaldt, J. Bird, R. Johnson: Musicjacket-combining motion capture and vibrotactile feedback to teach violin bowing, *IEEE Transactions on Instrumentation and Measurement* 60, 1 (2010) 104–113.
24. J. Rozé, M. Aramaki, R. Kronland-Martinet, S. Ystad: Exploring the perceived harshness of cello sounds by morphing and synthesis techniques, *Journal of the Acoustical Society of America* 141, 3 (2017) 2121–2136.
25. T. Wofford: Study of the interaction between the musician and the instrument. Application to the playability of the cello, PhD thesis, Sorbonne Université, Paris, 2018.
26. J. Rozé, M. Aramaki, R. Kronland-Martinet, S. Ystad: A technological platform for analyzing and improving musiciansâ sound-gesture interactions, in: *e-Forum Acusticum 2020*, Lyon, France, 7–11 December, 2020, pp. 439–447.
27. I. Wollman, C. Fritz, J. Poitevineau: Influence of vibrotactile feedback on some perceptual features of violins, *Journal of the Acoustical Society of America* 136, 2 (2014) 910–921.
28. A. Askenfelt, E.V. Jansson: On vibration sensation and finger touch in stringed instrument playing, *Music Perception* 9, 3 (1992) 311–349.
29. M. Pàmies-Vilà, E. Matusiak, V. Chatziioannou, A. Mayer: A cello bowing playback device? Motion capture meets robotic arm, in: *10th Convention of the European Acoustics Association*, Turin, Italy, 11–15 September, 2023.
30. Universal Robots: File: UR5e Technical Specifications, UR5e Product Fact Sheet – July 2021, 2021. Available at <https://www.universal-robots.com/products/ur5-robot/> (accessed 06 February, 2024).
31. P.I. Corke, W. Jachimczyk, R. Pillat: Robotics, vision and control: fundamental algorithms in MATLAB, vol. 73, Springer, 2011.
32. A.R. Jensenius: Methods for studying music-related body motion, in: R. Bader (Ed.), *Springer handbook of systematic musicology*, Springer, Berlin, Heidelberg, 2018, pp. 805–818.
33. M. Pàmies-Vilà, A. Scheiblaue, A. Mayer, V. Chatziioannou: A framework for the analysis of bowing actions with increased realisticness, in: *Proceedings of the 24th International Congress on Acoustics*, Gyeongju, Korea, 24–28, October, 2022.
34. Universal robots support: Real-time data exchange (RTDE) guide, 2022. Available at <https://www.universal-robots.com/articles/ur/interface-communication/real-time-data-exchange-rtde-guide/> (accessed 22 January, 2024).
35. A. Mayer, M. Pàmies-Vilà, V. Chatziioannou: The universal robots real-time data exchange (RTDE) and LabVIEW, Technical report, Department of Music Acoustics – Wiener Klangstil (IWK), 2022. <https://doi.org/10.13140/RG.2.2.14177.99684>.
36. A. Scheiblaue, A. Mayer, M. Pàmies-Vilà: Investigating the cello position, bow motion and cellist posture using motion capture, *Proceedings of Meetings on Acoustics*, Fourth Vienna Talk on Music Acoustics, 49 (2022) 035013.
37. H.F. Pollard, E.V. Jansson: A tristimulus method for the specification of musical timbre, *Acta Acustica united with Acustica* 51, 3 (1982) 162–171.
38. A. Mayer, M. Pàmies-Vilà, V. Chatziioannou: Ein Roboterarm spielt Cello, *OCG Journal* 48, 4 (2023) 16–19.
39. E. Fallowfield: Cello map: a handbook of cello technique for performers and composers, PhD thesis, University of Birmingham, UK, 2010.
40. M. Pàmies-Vilà, A. Mayer, E. Matusiak, V. Chatziioannou: Dataset to: A method for the reproduction of cello bow kinematics using a robotic arm and motion capture, Zenodo. <https://doi.org/10.5281/zenodo.10696680>.

**Cite this article as:** Pàmies-Vilà M. Mayer A. Matusiak E. & Chatziioannou V. 2024. A method for the reproduction of cello bow kinematics using a robotic arm and motion capture. *Acta Acustica*, 8, 45.