Sketches vs Skeletons: video annotation can capture what motion capture cannot

Marco Gillies, Harry Brenton, Matthew Yee-King, Andreu Grimalt-Reynes and Mark d'Inverno
Department of Computing, Goldsmiths, University of London
New Cross, London SE14 6NW, UK
{m.gillies, h.brenton, m.yee-king, agrim025, dinverno}@gold.ac.uk

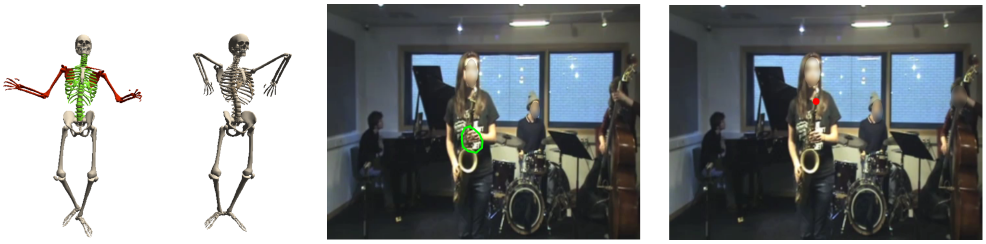


Figure 1. This paper investigates the use of digital representations of movement and posture to support music education. A skeleton based motion capture system (left) was found to have many problems, which a video based sketching system (right) may be able to address.

ABSTRACT

Good posture is vital to successful musical performance and music teachers spend a considerable amount of effort on improving their students' posture. This paper presents a user study to evaluate a skeletal motion capture system (based on the Microsoft Kinect^{TM}) for supporting teachers as they give feedback to learners about their posture and movement whilst playing an instrument. The study identified a number of problems with skeletal motion capture that are likely to make it unsuitable for this type of feedback: glitches in the capture reduce trust in the system, particularly as the motion data is removed from other contextual cues that could help judge whether it is correct or not; automated feedback can fail to account for the diversity of playing styles required by learners of different physical proportions, and most importantly, the skeleton representation leaves out many cues that are required to detect posture problems in all but the most elementary beginners. The study also included a participatory design stage which resulted in a radically redesigned prototype, which replaced skeletal motion capture with an interface that allows teachers and learners to sketch on video with the support of computer vision tracking.

Author Keywords

Motion Capture; Feedback; Education; Music

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Input devices and strategies

INTRODUCTION AND RELATED WORK

Posture is a key element of effective instrumental performance and poor posture also increases risks of long term injury[20]. For this reason teaching good posture is a key part of music pedagogy. This paper investigates technologies supporting music teachers in giving feedback on students' posture and in particular the effectiveness of motion capture in music teaching.

Recent advances in bodily and gestural interfaces (e.g. Bevilacqua *et al.*[2] and Fiebrink *et al.*[8]) have opened up the possibility of using movement tracking and other sensors to support learners in improving their posture. A number of researchers have investigated this possibility, for example the i-Maestro project[15] used motion capture as a means of generating both visual and auditory feedback for learners of bowed instruments, and Johnson *et al.*[12] used a gyroscope as a sensor in order to give vibrotactile feedback. Motion capture in particular seems a promising way of giving this feedback, it is able to track a person's movements and represent them as a virtual skeleton, with rigid bones that rotate relative to each other. The emergence of low cost consumer motion capture systems has made motion capture widely accessible. Motion capture has been used to give feedback on movement in domains other than music, for example Velloso *et al.*'s work on feedback for weight lifting exercises[19].

This paper presents the results of a study that aimed to evaluate how skeletal motion capture can support music teachers in giving feedback about learners' posture, and to understand how such a system could be improved. This paper will begin by briefly describing the prototype system. We will then describe the study and our participants feedback. This study discovered that there were serious problems with skeleton

MOCO'15, August 14–15, 2014, Vancouver, Canada.

motion capture in music teacher, particularly that it reduces the information so much that it is no longer useful for music teaching. We will end by describing the proposed new prototype and how it addresses the participants' concerns.

CONTEXT OF THE WORK

We evaluated a system developed within the context of a major European research project, PRAISE: Practice and peRformance Analysis Inspiring Social Education, that aims to create a social network for music learning which supports students and teachers in giving feedback about each other's performances, via comments on time based media such as audio and video. This work aimed to extend the platform to include feedback on posture and movement. The focus of the work is therefore not only on automatic feedback but on using technology to enhance human feedback.

MusicCircle is a website that allows you to upload and annotate music. Music can be uploaded in several ways: using a free iPhone App; uploading a file from a computer's hard drive; or recording directly into a browser. Once the recording has been uploaded and transcoded, it appears as a waveform with controls for playing, pausing and scrubbing the audio (figure 2). Users can select and comment upon a region of the waveform, which then appears as a coloured rectangle next to an avatar of the user. The coloured blocks represent sections of the waveform that have been highlighted and annotated. Pop-up comments are revealed when the cursor hovers over a block.

The development of the platform was informed by an extensive study of music teaching at university level [citation removed for anonymity] including 23 lesson observations with music teachers and many interviews with teachers and students. The key focus of this research was the ways in which teachers gave feedback, and the content of this feedback. This work identified movement and posture as a key topic of feedback during instrumental teaching. For example, teachers gave feedback on a number of aspects of movement: ease, independence, muscle co-ordination and / or support, dexterity, strength and overall posture. However, these are high level, human understandable concepts. It is not clear to what degree computers can support teachers and learners in giving and receiving feedback on these terms which may well be difficult to detect automatically or even semi-automatically.

For this reason we sought to better understand how music teachers and learners can use a technologically mediated system to give and receive feedback on movement and posture. To do this we used what Hutchinson et al. [11] call a *technology probe*, an example technology used as a means of studying participants interactions and eliciting their needs. We implemented a rapid prototype of a feedback system using the Kinect commercial motion capture system. The aim of this prototype was to have users try a mediating technology for movement feedback and engage them in informed discussion about the merits and problems with such an approach. To support this the prototype was deliberately rough and unpolished to encourage critique and feedback in the study, acting as what Buxton[4] call a "sketch of user experience". The prototype is described in the next section and the user study in the following section.

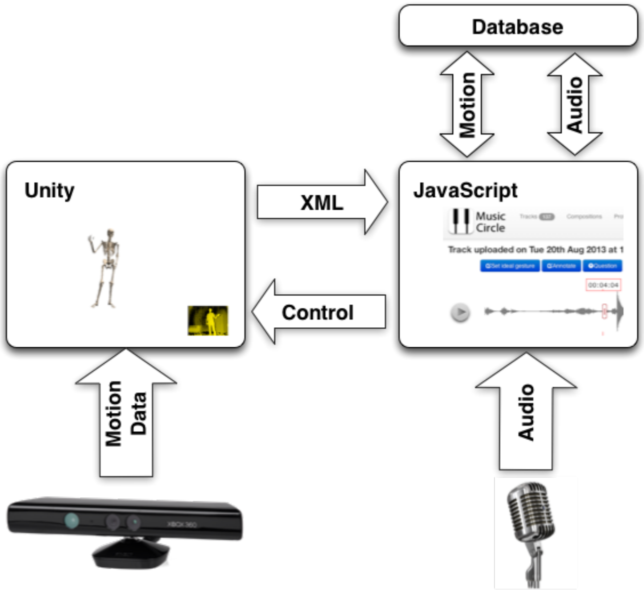


Figure 3. Overview of The motion capture system.

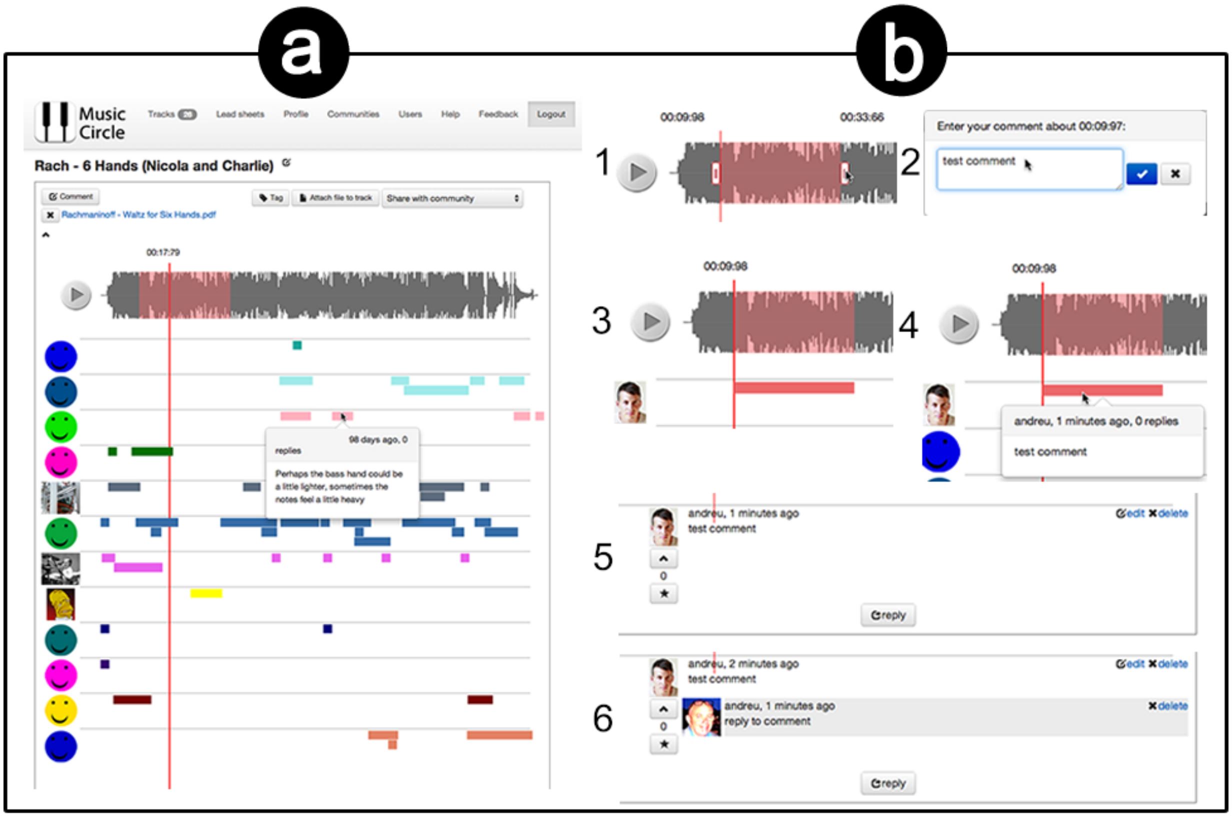
A MOTION CAPTURE BASED FEEDBACK SYSTEM

The study used a prototype of a motion capture based feedback system integrated within a larger software platform for collaborative feedback. This platform uses the Social Timeline model proposed by Brenton et al.[3], in which users are able to attach comments to particular moments or periods of time within time based media such as audio or video, using an interface similar to the timelines commonly found in audio and video editing software. Musicians' movements are recorded using a first generation Microsoft Kinect^{TM}for XBox 360 motion capture camera, using the OpenNI 1 drivers distributed by Zigfu. The movement data is synchronised with audio of the performance.

Figure 3 shows an overview of the system. Motion data is recorded by the Unity 3D engine using a Microsoft Kinect device. The resulting data is saved as XML and transferred to the javascript web interface. This is synchronised with audio recording and the two sets of data are transferred to an online database. The data can then be downloaded again to be played back.

A Recording interface allows users to record both audio and movement simultaneously in the browser. The user must have a Microsoft Kinect motion tracking device attached to their computer. If they do, their movements will be tracked and displayed on a virtual character in the browser. This character is a representation of a human skeleton that abstracts away gender and any individual features of a person.

When the user presses the record button, audio recording will begin and recording of the motion data will begin simultaneously. When the stop button is pressed, the recording will finish resulting in an audio and a motion data set, which are



Rach - 6 Hands (Nicola and Charlie)

98 days ago, 0 replies

Perhaps the bass hand could be a little lighter, sometimes the notes feel a little heavy

Enter your comment about 00:09:97:

test comment

andreu, 1 minutes ago, 0 replies

test comment

andreu, 1 minutes ago
test comment

andreu, 2 minutes ago
test comment

andreu, 1 minutes ago
reply to comment

Figure 2. The social timeline: showing learner's comments as coloured rectangles).

synchronised together. When the upload button is pressed, a new track is created on the database and the motion data is uploaded as additional data to that track.

The playback interface (figure 4) is similar to the recording interface with movement played on a skeleton avatar. The real time motion tracking features of the recording interface are disabled, so it will work without a kinect device. Users can select performances from the database, and if there is movement data associated with that track they can choose to view and discuss the movement. When the audio track is loaded, the motion data associated with that track is loaded into the gesture module and displayed on the skeleton. Initially the skeleton displays the first frame of the motion, but when the user plays the audio data the motion is played back in synchrony with the motion data. The playback can be paused and users can scrub through the performance using a timeline interface to find particular moments. Users can add textual comments to particular moments in the performance or to extended periods of time.

The playback module contains a second skeleton avatar, which is used to display a comparison pose. Users can select a particular frame of the motion as the pose to compare the playing motion with. This would typically be a frame which displays a good playing posture. As the motion is played back the user can view how the current playing movement compares with that comparison posture. In addition, the system calculates the differences in joint rotation between the current pose and the comparison pose.

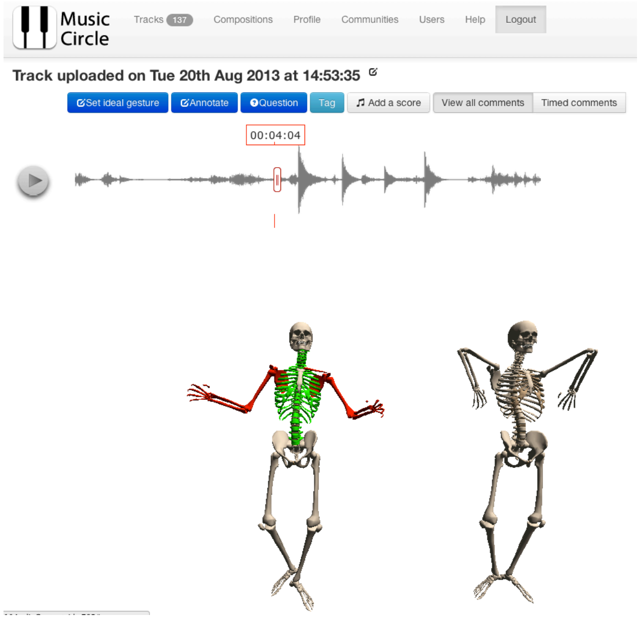
This difference is used to change the color of that joint. Small values are colored green and large differences are colored red, to show a warning for bad posture. Currently this analysis is done on the lower spine and shoulder, but the gesture module supports comparing any joints. Users an select a comparison pose from any frame in the current playing motion, but the gesture module supports comparison with poses in different motions. This will be implemented in the interface at a future date.

A USER STUDY

Two viola players (trainee and teacher), two drummers (trainee and teacher) and two conductors (trainee and teacher) recorded performances of their choosing using the gesture capture interface. During these performances the teacher gave advice to the trainee (directly in person, not via the system) and referred to a projected image of the system. After the performances the participants annotated their performances using the social timeline and took part in a group discussion where they described their experiences and evaluated the strengths and weaknesses of the system.

Reliability

There were some technical problems during the session. Recording quality was inconsistent between participants, for example the Kinect sensed the movement of the Viola teacher



Track uploaded on Tue 20th Aug 2013 at 14:53:35

Figure 4. The interface for viewing and annotating motion capture. Comparisons between the two skeletons are shown in green for similar postures (e.g. the torso) and red for dissimilar ones (e.g. the arms).

much better than the Viola trainee. The capture of the drummers was particularly problematic with only very large arm movements recorded, this was notably worse than a previous capture session. This variability suggests that factors such as lighting conditions, camera angle, participant body shape, clothing and stance have a noticeable impact upon the quality of the recording. These types of 'Glitches' are common in motion capture causing legs or arms to twitch rapidly into violently bent angles. These glitches and other issues created a problem of trust: *"There was an issue of lag, you can't quite tell at the moment, you can't quite trust . . . you know, what you are seeing"* (while the quote speaks of "lag", it is clear from the surrounding discussion that the participant was referring to a combination of multiple issues). The lack of trust seems to be due to the uncertainty of whether a movement was really made by the musician or whether it was was an artifact due to a glitch.

Diversity

The viola teacher said that players can have very different physical dimensions *"some people are more flexible than others and some people are less flexible, people have different body types"* which can result in different playing postures: *"There are a lot of great players that do play in a lot of different ways"*. This causes problems for a one-size-fits all pedagogy of posture, which is a known issue in current teaching: *"Some people say you should all play like Heifitz and have a really flat instrument like this and that doesn't actually work for most people, most people are comfortable with something more angled . . . in an ideal world you would get master players of very diverse physical dimensions as archetypal models"*. This suggests that any postural feedback needs to be tailored to the needs of a particular student, a view that is consistent with Johnson et al.'s[12] use of sensor calibration as a means of achieving personalized feedback.

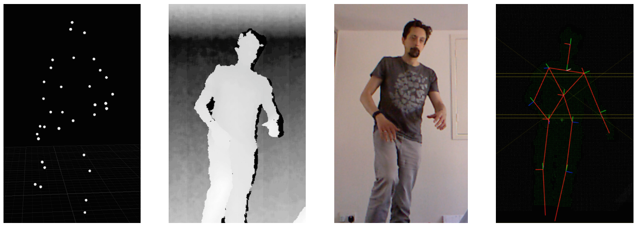


Figure 5. The information available in a number of forms of recording: optical motion capture markers, kinect depth map, video, skeleton reconstruction.

Nature of Movement

The most important criticism was the nature of the movements that the system picked up. Like most motion capture systems, it records gross movements of the skeleton. However, the participants thought that gross postural problems were general resolved very early in learning an instrument and would not be relevant to most learners. The viola teachers commented that the system could be useful *"when you are working with a really beginner, beginner and just working on these really gross movements"*. Similarly, the conducting teacher commented *"if you are dealing with a skeleton, you are de-facto reducing the area of analysis to posture .. within conducting, yes you can come up with exercises that have postural constraints and that could possibly be useful, but they are very few, they would really be in the first couple of lessons where you are looking at general stance"*. Later study would involve a more complex, subtle and diverse aspects of movement. The conducting teacher continued by saying: *"the rest of it, the rest of conducting is so much more expressive than that, it's about face, its about eyes . . . "*. The issues are very diverse: *"you're looking at left hand gestures, eye contact, your looking at the facial message, you're looking at the whole body message really right from how the come up to the podium and how the present themselves, through to the first down beat, how they are breathing, how they are connecting to the music and expressing the music, . . . it's very holistic"*. For the viola teacher small movements were very important *"What you really need to be able to see . . . is . . . these microscopic movements . . . so if you're doing something and maybe you're shifting, and you are just bringing your shoulder up just a little bit"*. For example: *"you can watch people's neck muscle. . . some people play with this neck muscle tensed all the time . . . you can also watch how the motion is happening under the collar bone up here . . . if some one is wearing a shirt with no sleeves . . . A lot of it is kinesthetic . . . You need some one who will hold onto your arm and they will feel if you're flexing something randomly"*. This shows that our prototype was not able to pick up the majority of signals that instrumental teachers look to in order to judge students' posture.

It is important to note that this is not simply a shortcoming of our implementation that could be fixed with some adjustments, or of a particular capture technology like the kinect. It is a fundamental problem of most current motion capture systems that they capture only the movement of the skeleton. The issue is with the skeleton representation that is shared by most motion capture systems including the kinect but also Vicon, Optitrack, Animazoo and others (though it is not the case that this representation is used in all systems which we will discuss below). In fact in some ways the problem is more severe with high end motion capture systems such as the Vicon or Optitrack. While they are highly accurate, they reduce all human movement data to a relatively small number of marker points (figure 5, far left). The kinect on the other hand makes use of a depth map that is richer in information, in theory it would be possible to extract muscle tension from the depth map (figure 5 center left), though in practice the resolution and accuracy are not sufficient (though it maybe be possible with a Kinect 2). Both of these representations are considerably less rich, to a human view than video (figure 5 center right), though video is much less interpretable by computer. However, the focus of the study is the skeleton representation that is inferred from the marker points or depth map (figure 5 far right). This has much less information than video, a point that was explicitly mentioned by our participants *The Problem is that skeletal motion capture reduces the information about body movement so much that it is not useful for the needs of teaching gesture in playing of musical instruments*. This is because the movement is reduced to the positions and orientations of a small number of joints (about 12 for the kinect, an optitrack system has 17), or possibly the position of a slightly larger number of marker points (34-38 in the case of the optitrack). This is a key benefit of motion capture, that is simplifies data and makes it explicit, however, it also loses data which can be an issue for many applications, including ours. Our participants mentioned it explicitly: *"if you are dealing with a skeleton, you are de-facto reducing the area of analysis to posture . . . conducting is so much more expressive than that, it's about face, its about eyes . . . "*.

An easy way to see this loss of information is to hold your arm up (maybe pretend to hold a violin) and relax your arm muscles without moving your arm. You should then be able to tense your muscles without otherwise moving them. If you do this in front of a mirror with a sleeveless shirt, the muscle tension should be clearly visible, but the movement in terms of joint rotations would be minimal, certainly not distinguishable from other small movements or identifiable as muscle tension. One of our participants makes a similar point: *"you can watch people's neck muscle. . . some people play with this neck muscle tensed all the time . . . you can also watch how the motion is happen under the collar bone up here . . . if some one is wearing a shirt with no sleeves"*.

Given the diversity of factors and of students' body types these criticisms may even apply to any automated feedback systems that relied on a single modality (though participants were enthusiastic about the possibility of measuring muscle tension).

Comparison to video

Another theme of the feedback was that video could capture much of what is possible with the motion capture system and also give a richer view *"What's the real benefit of the skeleton vs actual straightforward video for analyzing beat patterns? The beat patterns are less for me about posture . . . essentially it is about clarity of the beat, about where the stick is falling, so many other things really." "[experimenter: do you think video can capture that effectively?] yes, and that is how it is done in any conducting school"*. However, this applies to feedback given by a teacher, automated feedback from raw video would be challenging.

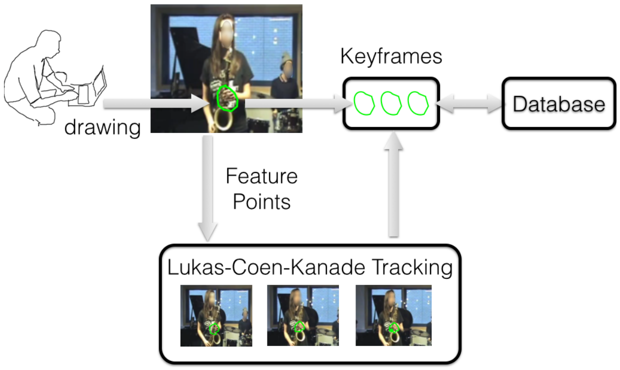
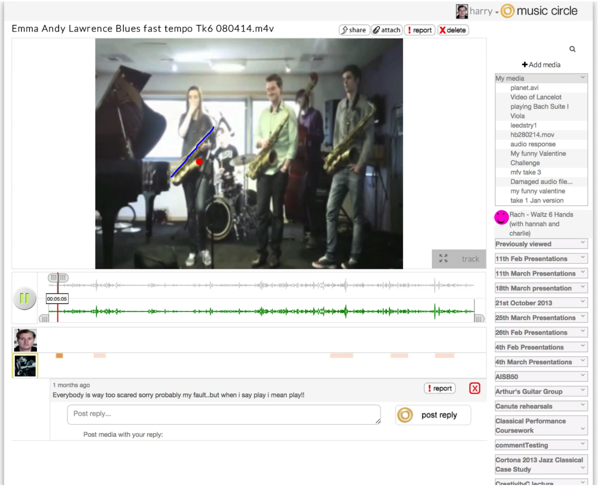


Figure 6. An overview of the digital ink system.

PROPOSED REDESIGN

At this stage of the study it was clear that the assumption that skeletal motion capture data was suitable for giving feedback on instrumentalists posture was flawed and a radical rethinking of the system would be required. Participants had expressed that video was a more useful tool for teachers. The researchers explored ideas with the participants about augmenting video for feedback. The first suggestion was drawing on video which was enthusiastically received: *"that would be far more useful than the skeleton". "In conducting . . . I can imagine it would be useful to draw a guideline of where the beats should be falling and maybe track where they are actually falling"*. One of the participants suggested the possibility of tracking particular points in the video: *"in bowing . . . if you could have a line . . . that is at the end of the point or on the screw of the bow so that if somebody is doing a string crossing pattern you can see if it's a nice round circle or if it is . . . messy and unclear"*. Another participant made it clear that this would only be useful if it were overlaid on video otherwise: *"I could draw [a circular movement] that and I could be holding the batton like this [posture with arms very close to the body] it's actually how I'm generating movement"*.

Based on these proposals we developed a prototype that allows users to annotate video by drawing directly on it (figure 8). This type of interface is commonly referred to as Digital Ink, and has been used in several video annotation systems. For example, Ramos and Balakrishnan[16] integrate digital ink within a video editing workflow. Singh *et al.*[18] present a system that allows dancers and choreographers to draw annoations on videos of rehearsals. Cattelan *et al.* [5] use digital ink annotations as part of their "Watch and Comment" sys-


Emma Andy Lawrence Blues fast tempo Tk6 080414.m4v

1 months ago

Everybody is way too scared sorry probably my fault..but when i say play i mean play!!

Figure 7. The full digital ink interface.

tem for end user video editing and annotation. Kipp's Anvil system[13] for coding video for qualitative research also allows this form of annotation, called spatiotemporal annotations possible. Digital ink is now entering mainstream applications like Google Hangouts. Most examples simply allow static annotations, which are not interpreted and do not move with the video. However, recently, some researchers have used automated techniques to interpret or track the annotations. Goularte et al. [10] use character recognition techniques to recognise specific symbols in an annotation. Our work builds particularly on the work of Goldman et al.[9], who use automated tracking to support annotations that follow objects in a video or trace their paths. Silva et al. [17] use tracking to provide dynamic annotations for live video, this throws up a number of interaction design challenges that we do not need to consider in our work on recorded video.

Our work, and these other examples, leverege the richness of a now ubiquitous technology: video. After describing our prototype we will discuss how, based on our participants' comments, we can conclude that a very commonly available technology, augmented video, can be a more powerful tool for collaborative learning than more complex motion capture.

Figure 6 shows an overview of the redesigned prototype and figure 7 shows user interface. The system records video of a performance rather than motion capture data (with the audio of the performance recorded as the audio track of the video). Users are able to draw directly on the video. These annotations are saved to the database and are synchronised to the time line. Users must select a particular temporal region in which to do the drawing (on the social timeline as shown in figure 2. A drawing is done at a particular time in the video and stored at that time. Each drawing is displayed from the frame at which it is drawn until the end of the temporal region. However, the drawing does not have to be static. Users may choose to edit the region at later frames in the region. These edits become keyframes in an animation, allowing the drawings to move over time.

These keyframes allow users to manually create the movement of the drawing. It is also possible to make the drawings move automatically with the video, through video tracking. The drawings consist of a number of points, each of which is tracked on the video using a Lucas-Kanade optical flow algorithm [14]. If the user selects automatic tracking, the optical flow is calculated for each frame and the resulting points are used to automatically create new keyframes.

Our prototype allows two interactions. In the first (figure 8, top) users can do line drawings on the video, for example a circle to highlight a particular point on the body. These lines follow that point as the body moves. The second interaction is show in figure 8, bottom. Users can select a point in the image. The trajectory of that point is drawn as the video plays, as suggested by our participant.

Automatic tracking and keyframing can be used together. While automatic tracking can be used to quickly create annotations, the tracking is often lost if movement is too fast, or certain visual features are occluded. Manual keyframing can allow users to correct these errors and restart the tracking at a better position (figure 8, top right).

We will discuss how this prototype relates to the themes identified in our study. This method has the potential to show more diverse and multimodal information as it shows full video (though it will not show anything that is now shown in the video). This interface is potentially very general, it is simply drawing on video and can show anything that can be drawn and those drawing can be interpreted in the context of anything that can be shown in video. In this sense it is a very open interface as Dix [6] uses the term: an interface that makes possible multiple interpretations and therefore allows users to appropriate it for many different purposes. This can be true of any technology, as people appropriate all kinds of technologies despite the plans of the designers, to use Dourish's phrase "Users, not designers, create and communicate meaning"[7], however, deliberately making systems open encourages a diverse range of appropriation. This means that it is open for teachers to adapt their feedback to very diverse needs of a different students.

The final issue is reliability and trust. This new system is likely to be significantly less reliable than skeleton tracking, for example the final image in figure 8 shows the system loosing tracking. However, there are reasons to believe that there will be less problems of trust. Part of the problem of trust for the first prototype is that the tracking was decontextualised from the original movement. Not only could the system make errors, but there was very little context to allow people to see whether a particular movement was a errors or not, because the could not see the original. This made it very difficult to trust the system as any movement made could be a glitch. The second prototype on the other hand keeps all annotations in the context of the video, making it easier to see if a particular movement is correct or not. We have also provided mechanisms for users to correct errors in tracking. Figure 8, top right, shows a point on the drawing being selected and moved. This allows users to manually correct errors in the tracking, and so keeps control with the user. Having said all

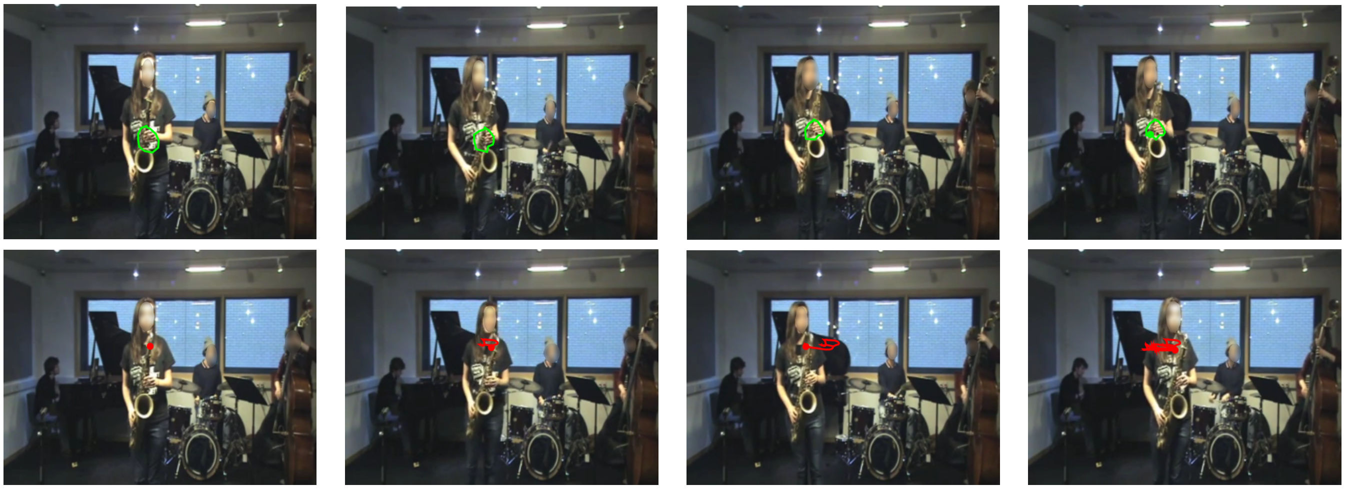


Figure 8. The new prototype: top: drawing on an image to emphasise neck posture. The top right image shows a point on the curve being selected end edited (the large green circle shows the point being moved). bottom: tracking a drummers hand, the point (large red circle) is tracked successfully for a period and a thin red trail shows the past trajectory of the point but tracking is lost in the last image

of this, the reliability of the tracking is likely to be a limiting factor in this type of interface, in particular certain visual features are easier to track than others, thus limiting in practice the systems ability to annotate the complex and diverse postural information that our participants discussed. These issues can only be understood by many more iterations of the type of prototyping and user testing that we have described.

CONCLUSION

This paper has presented a user study of a system to support feedback on music students' posture based on skeletal motion capture. HCI studies are a good way of determining: 1) the type of information that is appropriate for a particular group of learners; 2) the type of representation and interactions which are suited to unlocking and communicating that information. This study revealed significant issues not only with the system but with the underlying assumption of using skeleton poses. Firstly, glitches in the motion capture resulted in a lack of trust in the system, most likely compounded by the fact that the motion capture was presented out of context without the original video which could be used to judge whether a particular movement was made by the musician or was a glitch. The second issue was that musicians are very diverse in their physical dimensions and have different postural requirements, making it hard to develop a single form of automated feedback. Finally, and most importantly, the problems that learners show in their posture are more complex than gross pose and so are unlikely to show up in a skeleton representation. As non-musicians, we had high hopes for motion capture because to our untrained eyes it recorded movements and behaviours that would give musicians valuable feedback about their performances. However, feedback from subject matter experts revealed that gross movements failed to capture important aspects of performance such as muscle tension, finger movement and facial expression. Participants agreed that a video would be more informative than motion capture. This necessitated a fundamental redesign away from skeleton motion capture, and a new prototype was developed based on annotating video.

These insights and change of direction were enabled by a methodology of rapid prototyping and early piloting. The prototype was shown to participants in an early and incomplete state, where it had known flaws. It was what Buxton calls a sketch of user experience[4] (a second meaning of our title), in which the very roughness of the prototype is an invitation for participants to critique it. It reduces the researchers investment in the prototype and therefore increases their openness to change direction.

We do not and cannot claim that our prototype is the "right" way. Many other approaches are possible. For example, sensing provides useful "hidden" information, that sketching cannot. This is particularly true of muscle sensing of the type used by Johnson[12] to give feedback to violin students. Our participants did comment on the particular usefulness of this type of interface. Using sensors such as Electromyogram (muscle sensing) runs the risk of creating a one size fits all form of feedback, though Johnson's use of calibration can mitigate this. However, video and motion capture are not mutually exclusive and an integrated representation which combines mocap and video annotation may ultimately be preferable as it may leverage the cognitive benefits of the different representations [1]. Another limitation of the first prototype is that it only compares static poses, rather movement over time, . This is an issue that we should address in future work, the point tracking shown in figure 8 (bottom). The other, more straightforward reason that we cannot make claims for our new prototype is that it is yet to be evaluated. It is currently a very rough and rapidly developed prototype that needs a lot of work, however, we will try it with users before doing that work to get early feedback, before we are committed to it, and, we hope, discover many interesting flaws in its design.

ACKNOWLEDGEMENTS

This research was supported by Praise: Practice and peRformance Analysis Inspiring Social Education a STREP project funded by the European Commission under the FP7 programme (grant number FP7-318770).

REFERENCES

1. Ainsworth, S. Deft: A conceptual framework for considering learning with multiple representations. *Learning and Instruction 16*, 3 (2006), 183–198.
2. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., and Rasamimanana, N. Continuous realtime gesture following and recognition. *Gesture in Embodied Communication and Human-Computer Interaction* (2010).
3. Brenton, H., Yee-King, M., Grimalt-Reynes, A., Gillies, M., Krivenski, M., and d'Inverno, M. A social timeline for exchanging feedback about musical performances. In *Proceedings of the 28th British HCI Group Annual Conference on People and Computers*, BCS (2014).
4. Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., 2007.
5. Cattelan, R. G., Teixeira, C., Goularte, R., and Pimentel, M. D. G. C. Watch-and-comment as a paradigm toward ubiquitous interactive video editing. *ACM Trans. Multimedia Comput. Commun. Appl. 4*, 4 (Nov. 2008), 28:1–28:24.
6. Dix, A. Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers*, BCS (2007).
7. Dourish, P. *Where The Action Is: The Foundations Of Embodied Interaction*. MIT Press, 2001.
8. Fiebrink, R., Cook, P. R., and Trueman, D. Human model evaluation in interactive supervised learning. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, ACM (2011).
9. Goldman, D. B., Gonterman, C., Curless, B., Salesin, D., and Seitz, S. M. Video object annotation, navigation, and composition. In *Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08*, ACM Press (New York, New York, USA, Oct. 2008), 3.
10. Goularte, R., Cattelan, R. G., Camacho-Guerrero, J. A., Incio, V. R., and da Graa C. Pimentel, M. Interactive multimedia annotations. In *Proceedings of the 2004 ACM symposium on Document engineering - DocEng '04*, ACM Press (New York, New York, USA, Oct. 2004), 84.
11. Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., and Eiderbäck, B. Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, ACM (New York, NY, USA, 2003), 17–24.
12. Johnson, R., Bianchi-Berthouze, N., Rogers, Y., and van der Linden, J. Embracing calibration in body sensing: Using self-tweaking to enhance ownership and performance. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, ACM (2013).
13. Kipp, M., et al. Spatiotemporal coding in anvil. In *Proceedings of the 6th international conference on Language Resources and Evaluation (LREC-08)* (2008).
14. Lucas, B. D., and Kanade, T. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, Morgan Kaufmann Publishers Inc. (1981).
15. Ng, K. C., Weyde, T., Larkin, O., Neubarth, K., Koerselman, T., and Ong, B. 3d augmented mirror: A multimodal interface for string instrument learning and teaching with gesture support. In *Proceedings of the 9th International Conference on Multimodal Interfaces*, ICMI '07, ACM (2007).
16. Ramos, G., and Balakrishnan, R. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th annual ACM symposium on User interface software and technology - UIST '03*, ACM Press (New York, New York, USA, Nov. 2003), 105–114.
17. Silva, J. a., Cabral, D., Fernandes, C., and Correia, N. Real-time annotation of video objects on tablet computers. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '12*, ACM Press (New York, New York, USA, Dec. 2012), 1.
18. Singh, V., Latulipe, C., Carroll, E., and Lottridge, D. The choreographer's notebook: A video annotation system for dancers and choreographers. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*, C&#38;C '11, ACM (New York, NY, USA, 2011), 197–206.
19. Velloso, E., Bulling, A., and Gellersen, H. Motionma: Motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (2013).
20. Zeuner, K. E., Shill, H. A., Sohn, Y. H., Molloy, F. M., Thornton, B. C., Dambrosia, J. M., and Hallett, M. Motor training as treatment in focal hand dystonia. *Movement disorders 20*, 3 (2005), 335–341.