VIPPSA: A WEB-BASED TOOL FOR VIDEO SCORE SYNCHRONIZATION AND ANNOTATION

Xavier Davenport
University of Illinois
xavierd2@illinois.edu

Craig Vear
University of Nottingham
craig.vear@nottingham.ac.uk

Christoph Guttandin
Media Codings
chrisguttandin@media-codings.com

ABSTRACT

Based on feedback from musicians who have rehearsed and performed digital scores, in this paper we present a novel technology for Video Playback Position Synchronization and Annotation (ViPPSA). We attempt to ameliorate some of the issues described by performers when rehearsing and interpreting video scores by providing a convenient platform for synchronizing and annotating pre-rendered scrolling-scores. We describe the annotation and synchronization features of ViPPSA and how they may be utilized in rehearsal and performance settings. We also briefly present a new musical work designed to take advantage of the unique features of this software, as well as possible future enhancements informed by early feedback from performers rehearsing using this technology.

1. INTRODUCTION AND PROBLEM STATEMENT

1.1 Digital Scores

The Digital Score is a 5-year, European Research Council-funded research project investigating the transformation of the music score through digital and computational technologies. It has been running since 2021, and currently in the final stage of evaluation and synthesis. The principle investigator is Prof. Dr. Craig Vear (University of Nottingham).

Vear defines a digital score as *a communications interface of musical ideas between musicians utilizing the creative potential of digital technology.* [1, 2] There are many types of digital scores that cover a broad range of approaches and technologies, but they all have one thing in common: they *communicate musical ideas using the creative potential of technology*. In this paper, we focus on a small subset of digital scores commonly referred to as *scrolling scores* or *video scores* where the score is pre-rendered as a video, then presented to the performer using a video-player.

1.2 Technical Difficulties

One of the findings that the Digital Score research project highlights is how performers have encountered difficulties when attempting to synchronize individual video score parts in rehearsal and performance. One such piece is the piano trio *Traveling Vagabond* by Davenport where synchronization of unique scrolling-score parts was a challenge that was never overcome before the premiere. Although a video score is often successful in conveying information when all the performers are able to look at a single screen, ensembles involving the piano often traditionally have performers facing radically different directions on stage. For the premiere performance of *Traveling Vagabond*, the strings players looked at an iPad, while the pianist looked at a Windows laptop. It was discovered in rehearsal that these different devices had radically different response times between clicking the play-button and the video timeline beginning to move. To make matters worse, the response times between clicking and playing were also inconsistent, likely depending on the background tasks each device was performing.

Several scrolling scores were commissioned and presented as part of the 2022 TENOR conference in Marseille, France. Musicians were open and willing to explore these new works, but offered critical feedback concerning the challenges around rehearsing and performing the works. [3] The performers specifically recommended that there should be functional rehearsal markers which enable rehearsal of individual sections in isolation.

More digital scores were presented at a DigiScore concert as part of *Encuentro Montrealia* in 2025. Performers at this event highlighted issues with technical coordination of the scores. [4] They noted that synchronization issues, multiple screen displays, and hardware constraints directly impacted the flow and coherence of their performance, though they attempted to alleviate some of these issues with the on-site sound engineers.

To address these issues reported by performers of digital works, we have developed a web-based environment for annotating and synchronizing video scores. The technology reduces the hardware barrier to entry by being accessible in any web-browser. Though it is currently optimized to work over a local area network, pieces can still be rehearsed and performed if necessary by accessing a publicly available website. As of submitting this paper, the annotation capabilities of ViPPSA can be accessed at https://vippsa2025.web.illinois.edu/. In the following, we describe the ways these issues of synchronization and rehearsability have been addressed by others before describing the unique features of ViPPSA, then finally present some possible future enhancements to the software based on feedback from performers using the web application.

2. RELATED RESEARCH

2.1 Score Synchronization

The idea of synchronizing distinct animated part-scores between devices is not new – the Decibel Score Player does exactly that as well as providing an environment for bespoke score creation. [5] While the Decibel Score Player provides an immense amount of flexibility for the composer, the required collection of iPads may be prohibitively expensive for an ensemble to adopt. The creators of the Decibel Score Player cite synchronization issues due to "different operating systems, networking protocols and personal settings" as the motivation for restricting the software to only iPads. [6] Other frameworks for cross-device communication exist to fulfil specific compositional needs such as INScore and DRAWSOCKET which rely on OSC messages for synchronization, as well as browser-based systems for distributed notation like PITL, A&M, and John. [7, 8, 9] SmartVox is a web app very similar in concept to ViPPSA in that it synchronizes mp4 video files between devices using socket.io. [10] Another web-based app was developed a few years after SmartVox in 2021 called Autoconductor which can handle multi-channel audio as well as synchronizing scores, though the method of synchronization used was susceptible to fluctuations in the WiFi network. [11] The primary differences between these two synchronization web apps and our new software are the underlying algorithm used for synchronization between devices and the ability to annotate the score.

With the exception of SmartVox, the software listed in the previous paragraph was designed with interactivity in mind. SmartVox and our new software ViPPSA are concerned with a smaller subset of digital scores where the score has been pre-rendered into one or more videos, several of which can be found at www.animatednotation.com. We have opted to focus on scrolling video scores because of their portability and reliability – it is significantly easier for performers to put together a performance when they do not need to manage a bespoke system for algorithmic score generation.

2.2 The Timing Object

The video elements in the browser are synchronized across multiple devices with the TimingObject as specified by the Multi-Device Timing Community Group. [12, 13] The final version of the Timing Object specification was published December 3rd, 2024, and is designed to integrate with the Web Audio API, among other Web APIs. [14] We use an implementation of the TimingObject provided by Suenc which allows for precise timing between devices on a local area network without the need to reference a clock hosted on an external server. A demonstration of this technology is hosted at suenc.io. The TimingObject is essentially a vector (called the TimingStateVector) with position, velocity, and acceleration imformation at a certain point in time. It allows for precise and reliable synchronization between any number of devices. [15] This contrasts with what is used in SmartVox, which is dependent only on position with an additional algorithm layered on top to correct for timing drift between devices. [16] Correcting the playback position when it has sufficiently drifted from the desired timestamp can lead to an irregular playback velocity between devices. The TimingObject implementation constantly verifies the timing offset between devices and then adjusts the timing information of the synchronized TimingStateVector transparently.

3. DESIGNING VIPPSA

There are four significant difficulties performers have expressed when performing using video scores: [17] (1) It is challenging to rehearse small sections of animated scores in a repeatable way. (2) Scores should be flexible enough so that they can be hidden from the audience. Many scores need to be seen by the entire ensemble on a single screen, so they are often projected somewhere the audience can also see the score. (3) Bespoke technology is unstable. Latency, poor internet signal, unstable connectivity, and more can cause unnecessary stress to performers. (4) The performers should be able to engage with the piece before the group rehearsal. ViPPSA is intended to ameliorate each of these issues.

When performers go to the ViPPSA website, they will first need to press the "Load Video" button and select the video they wish to play on their device. So long as the *synchronize* checkbox is selected, any manipulation of the video playback position or playback rate is synchronized between all connected clients. If a performer wishes to navigate throughout the score with influencing or receiving influence from others, they may deselect the *synchronize* checkbox.

3.1 Lower Controls

Many elements in the video area will look familiar to anyone who frequently consumes content on YouTube or other streaming services, as shown in *Figure 1*. On the bottom left of the screen is the play/pause button as well as a mute button with volume slider. On the bottom right of the screen is a fullscreen button, as well as a number indicating the current playback rate, for example 1x, 0.2x, 0.8x, etc. All the other buttons across the bottom of the screen are tools used for annotating the score. First there are two horizontal sliders which appear when the user hovers over the corresponding icon, the first one used for pen-size and the second one used for pen-opacity. After that there are five buttons, the first of which allows the user to select a color for their next annotation. After that are four tools: the pointer, the text tool, the pen tool, and the eraser. The pointer is the default tool and does not allow the user to make annotations. After selecting the text tool, the user may click or tap somewhere on the video to make a text annotation. With the pen tool selected, the user can drag their mouse or finger across the screen to draw an annotation. Finally, the eraser can be used to delete both text and pen annotations.

After drawing an annotation using the pen tool and releasing the click or finger from the screen, a dialog appears asking for the annotation start time and end time. In the in-

Figure 1. Screenshot of the lower video controls container. From left to right, there is a play/pause button, a mute/volume button, the current video time, the video duration, a pen/text size slider, an annotation opacity slider, a color picker, a default pointer button, a text-tool button, a pen-tool button, an eraser button, a playback-rate button, and a fullscreen button.

Figure 2. Screenshot of the upper video controls container. From left to right, there is a synchronization checkbox, a selectable list of warp points, and a settings button. Selecting the synchronization check box allows the user to both receive timeline updates from others as well as influence the timeline of other users. Selecting an item in the list of warp points immediately moves the timeline to the corresponding timestamp.

terim between the start and end times, the annotation will appear on the video, but will not otherwise. The same is true of the text tool, however there in an additional input field for the text the user wants to appear.

3.2 Upper Controls

There is an additional option in the text annotation dialog asking whether the annotation should be added as a warp point. This feature was added in response to performers wishing it was easier to rehearse specific parts in a piece of music; adding a warp point adds the text annotation to a list of timestamps accessible in the "Warp to Annotation" dropdown in the center of the upper video controls container shown in *Figure 2*. When one of the timestamps is selected, the video warps all synchronized videos to that position. At any time, the user may save all the annotations they made. The annotations are saved in a plain-text json file and can be loaded again next time the website is accessed.

The feature-set detailed above streamlines the rehearsal process and aims to ease the four difficulties described by performers: (1) Small sections of a piece can be more easily rehearsed by setting warp points using text annotations. (2) Since scores can be so easily and precisely synchronized between devices, each performer can look at their own device and there is no need to show the score on a large screen. (3) This technology is using time-tested browser-based technology. Even if performers are not on a dedicated local area network, the TimingObject underlying ViPPSA allows for consistent synchronization between many devices over the internet, even when the WiFi is somewhat unreliable. (4) ViPPSA is hosted on a website accessible to performers at any time. Either as an individual or as an ensemble, performers can rehearse and annotate a piece without the composer or director needing to be in the room.

4. USAGE IN REHEARSAL

In the following, we outline a hypothetical large ensemble rehearsal using this technology. We assume that everyone has access to a personal device with a screen sufficiently large to view the score.

4.1 The First Run

Before rehearsal, each performer must download a video file of their part. At the beginning of the rehearsal, the performers should navigate to the ViPPSA webpage, click the Load Video button, and select their score. At this point, they can freely navigate and annotate the score on their own and should wait to select the "Synchronize" checkbox until everyone is ready to navigate the score together.

When the ensemble decides to begin playing the piece as a group, each member should ensure that the "Synchronize" checkbox is selected. Once everyone is synchronized, any member of the ensemble can navigate the video and influence the video playback position of every other performer's score.

4.2 Annotations

During a pause in rehearsal, performers may wish to desynchronize from the ensemble to navigate the score and add annotations. For instance, a performer could set the pen tool to be yellow with a large pen size and low opacity to highlight an easy-to-miss moment in the score. They might also use the text tool to add supplemental information to the score such as expressive text or a technical detail.

Using the text-tool, users can also add warp points to the score as a way to quickly revisit a specific moment in the score. Users can individually de-synchronize and add personal rehearsal markers. Alternatively, the ensemble could synchronize before individuals add rehearsal markers to ensure the markers are synchronized. There is currently not a simple way to transfer rehearsal markers between performers within ViPPSA, so every performer will need to add the markers to their own score. In either case, only one performer needs to select a warp point for all synchronized users to be influenced.

4.3 Example Score: *Stormbrush*

As a way to test the capabilities of ViPPSA, Davenport composed a piece for five instruments where each performer receives a unique video score. In the piece, each performer is playing at a different tempo by referencing what Kocher calls a *virtual conductor* on the screen. [18] This polytempic texture requires a high level of temporal synchronization to achieve the desired effect. In certain moments of the piece, it is also suggested that the performer annotate reminders for themselves on how to interpret the score depending on their instrument, as shown in one of the parts in *Figure 3*. Details about this piece can be found in another submission to this TENOR conference titled SONIC WORK SUBMISSION: STORMBRUSH.

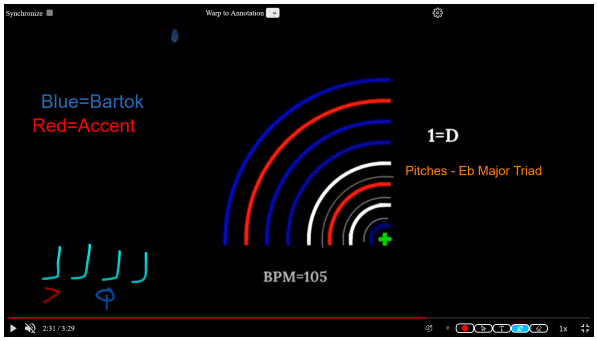

Figure 3. Screenshot of Davenport's score *Stormbrush* with a few annotations added. A user has added colored text as a reminder for what each color means on the screen, as well as a reminder of what pitch set they wish to use and a simple outline of a recurring rhythm in this section of the piece.

5. FUTURE ENHANCEMENTS

The goal in creating this software is to make digital scores more approachable and rehearsable for musicians. The future direction of ViPPSA design must prioritize an intuitive user interface while balancing a set of flexible features. Despite the need for flexibility, ViPPSA should not aim to be a comprehensive drawing program or a critical annotation tool. Instead, development should be driven by feedback from performers who desire a more efficient means of achieving their musical goals.

Based on early feedback from performers and colleagues, there are several features that could be reasonably added to ViPPSA which can be divided into three broad categories: synchronization, annotation, and distribution.

5.1 Distribution Features

Right now, the score is treated much as a paper score would be in an orchestral setting; the score must be "passed out" before or during the first rehearsal such that each performer has their own score loaded into their personal device. Depending on the size of the video file, this could be done via email, Box, Google Drive, or any number of other file sharing services. Some users expressed interest in a file-sharing system being integrated into ViPPSA such that videos can be downloaded and shared between users. This would make updating and sending new versions of a score much simpler for a composer.

Rather than relying on a locally stored video for rehearsal, some users express interest in playing videos hosted on YouTube or Vimeo embedded in ViPPSA while still being able to use the annotation tools. This addition to ViPPSA is not technically challenging, but requires duplicating a significant amount of code to handle both the HTML5 video element used for local video playback as well as the iframe element for embedding external videos. This feature would allow for the convenient rehearsal of a number of scrolling scores which are already posted on public platforms.

Performers who regularly work in larger ensembles saw the potential for ViPPSA to act as a system for the distribution of not only scores but also annotations. They conceived a scenario where an ensemble director may wish to add some general notes or rehearsal marks to everyone's score simultaneously, taking advantage of the digital interconnectivity provided by ViPPSA.

5.2 Synchronization Features

In addition to synchronizing videos between devices, the TimingObject can be mapped to other communication standards in audio. For example, timing object could be used to feed a string of OSC messages with timing information to another piece of software on one or more machines. This also provides a simple way to connect ViPPSA to a system for audio generation and other real-time elements.

Users also expressed an interest in even more precise temporal navigation in the score. Users can navigate by clicking on the timeline at the bottom of the video above the controls or by incrementing time using the left and right arrow keys, but typing in a specific time provides a much desired level of precision not currently in ViPPSA.

5.3 Annotation Features

The process of utilizing the annotation tools is designed to be as similar as possible to how a performer might annotate a paper score; a user draws something with a pen tool, and they could erase it later if desired. The annotation tools do not use any of the advantages of digital systems found in modern drawing programs. Some simple comforts that should be added to ViPPSA are functions like copy, cut, paste, undo, and redo. There is also presently no way to move an annotation after creating it; it must first be erased, then drawn again.

One more advanced computer drawing feature worth adding to ViPPSA is the idea of layering. Rather than every annotation being drawn to the same canvas, annotating on different layers would allow users to group annotations and choose whether to hide or lock them in place.

Perhaps the most desired feature for performers is the ability for annotations to move along with the video score. Before recent developments in AI technology, this would have been a feature exceptionally tedious to use: it would require the user to manually add key-frames to a video in order to attach an annotation to some feature on screen as it moves about. Now however, tools like the motion tracking video editors from Kapwing, Flexclip, and several other websites exist and are continuously improving. It is possible that this AI technology could be used to identify and track score elements in future versions of ViPPSA.

6. CONCLUSION

ViPPSA addresses challenges expressed by performers by allowing precise playback position synchronization between devices and for annotation of the score. Utilizing the timing-object standard allows latencies of less than a millisecond between an arbitrary number of web browsers. Our new annotation tool is simple, but gives room for personalized

note-taking within a video score. By prioritizing a user-focused design, we hope to broaden the appeal and accessibility of digital scores.

Acknowledgments

The Digital Score project (*DigiScore*) is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. ERC-2020-COG – 101002086).

7. REFERENCES

[1] C. Vear, *The Digital Score: Musicianship, Creativity and Innovation*. New York: Routledge, 2019.

[2] "The digital score," 2025. [Online]. Available: https://digiscore.github.io/

[3] "Digiscore tenor22 commissions," 2022. [Online]. Available: https://digiscore.github.io/2022-10-05-DigiScore_TENOR22_commissions/

[4] "Digiscore concert as part of encuentro montrealia, mexico-canada," 2025. [Online]. Available: https://digiscore.github.io/2025-06-05-CMMAS_concert/

[5] A. Wyatt and C. Hope, "The decibel scoreplayer – learning from and for usability," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'2022*, V. Tiffon, J. Bell, and C. de Paiva Santana, Eds. Marseille, France: PRISM Laboratory, 2022, pp. 143–148.

[6] C. Hope, L. Vickery, A. Wyatt, and S. James, "The decibel scoreplayer - a digital tool for reading graphic notation," in *Proceedings of the First International Conference on Technologies for Music Notation and Representation – TENOR'15*, M. Battier, J. Bresson, P. Couprie, C. Davy-Rigaux, D. Fober, Y. Geslin, H. Genevois, F. Picard, and A. Tacaille, Eds., Paris, France, 2015, pp. 58–69.

[7] J. Bell, "Distributed notation in the browser, an overview," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'20/21*, R. Gottfried, G. Hajdu, J. Sello, A. Anatrini, and J. MacCallum, Eds. Hamburg, Germany: Hamburg University for Music and Theater, 2021, pp. 251–259.

[8] A. Pirchner, "Ergodic and emergent qualities of real-time scores. anna & marie and gamified audiovisual compositions," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'20/21*, R. Gottfried, G. Hajdu, J. Sello, A. Anatrini, and J. MacCallum, Eds. Hamburg, Germany: Hamburg University for Music and Theater, 2020, pp. 189–197.

[9] V. Goudard, "John, the semi-conductor: A tool for comprovisation," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'18*, S. Bhagwati and J. Bresson, Eds. Montreal, Canada: Concordia University, 2018, pp. 43–49.

[10] J. Bell and B. Matuszewski, "Smartvox. a web-based distributed media player as notation tool for choral practices," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'17*, H. L. Palma, M. Solomon, E. Tucci, and C. Lage, Eds. A Coru na, Spain: Universidade da Coru na, 2017, pp. 99–104.

[11] L. Felten and C. Klinkenberg, "Autoconductor: Synchronizing graphic scores, multi-channel sound and video files," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'20/21*, R. Gottfried, G. Hajdu, J. Sello, A. Anatrini, and J. MacCallum, Eds. Hamburg, Germany: Hamburg University for Music and Theater, 2021, pp. 265–269.

[12] "Multi-device timing community group," 2025. [Online]. Available: https://www.w3.org/community/webtiming/

[13] *Data-independent sequencing with the timing object: a JavaScript sequencer for single-device and multi-device web media. In Proceedings of the 7th International Conference on Multimedia Systems (MMSys '16)*, 2016, pre SFI. [Online]. Available: https://www.w3.org/community/webtiming/files/2016/05/mmsys2016slides.pdf

[14] "Timing object," 2024. [Online]. Available: https://www.w3.org/community/reports/webtiming/CG-FINAL-timingobject-20241203/

[15] C. Guttandin, "The timing object - a pacemaker for the web," 2018. [Online]. Available: https://media-codings.com/articles/the-timing-object-a-pacemaker-for-the-web

[16] J. Bell, "Audiovisual scores and parts synchronized over the web," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'18*, S. Bhagwati and J. Bresson, Eds. Montreal, Canada: Concordia University, 2018, pp. 17–23.

[17] C. Vear and S. Moroz, "The digital score: a review of research spanning 2022-23," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'2023*, A. P. D. Ritis, V. Zappi, J. V. Buskirk, and J. Mallia, Eds. Boston, Massachusetts, USA: Northeastern University, 2023, pp. 154 – 160.

[18] P. Kocher, "Polytempo network: A system for technology-assisted conducting," in *Proceedings ICMCISMCI2014*, 09 2014.