Time's up for the Myo? The smartwatch as a ubiquitous alternative for audio-gestural analyses.

William Wilson
DMT Lab
Birmingham City University
Birmingham, United Kingdom
william.wilson@bcu.ac.uk

Niccolo Granieri
DMT Lab
Birmingham City University
Birmingham, United Kingdom
granieriniccolo@gmail.com

Islah Ali-MacLachlan
DMT Lab
Birmingham City University
Birmingham, United Kingdom
islah.ali-maclachlan@bcu.ac.uk

ABSTRACT

The utility of gestural technologies in broadening analytical- and expressive-interface possibilities has been documented extensively; both within the sphere of NIME and beyond.

Wearable gestural sensors have proved integral components of many past NIMEs. Previous implementations have typically made use of specialist, IMU and EMG based gestural technologies. Few have proved, singularly, as popular as the Myo armband. An informal review of the NIME archives found that the Myo has featured in 21 NIME publications, since an initial declaration of the Myo's promise as "a new standard controller in the NIME community" by Nyomen et al. in 2015 [10]. Ten of those found were published after the Myo's discontinuation in 2018, including three as recently as 2022 [7, 12, 15].

This paper details an assessment of smartwatch-based IMU and audio logging as a ubiquitous, accessible alternative to the IMU capabilities of the Myo armband. Six violinists were recorded performing a number of exercises using VioLogger; a purpose-built application for the Apple Watch. Participants were simultaneously recorded using a Myo armband and a freestanding microphone. Initial testing upon this pilot dataset indicated promising results for the purposes of audio-gestural analysis; both implementations demonstrated similar efficacy for the purposes of MLP-based bow-stroke classification.

Author Keywords

NIME, MYO Armband, IMU, Gestural Analysis, Musicology, Violin, Apple Watch, Smart Watch

CCS Concepts

•**Human-centered computing** → **Ubiquitous and mobile computing;** •**Information systems** → *Multimedia information systems;*

NIME'23, 31 May–2 June, 2023, Mexico City, Mexico.

1. BACKGROUND

If considering a musical performance to be the culmination of a performer's gestural execution, two avenues for analysis emerge; these are the audible and gestural content. Just as audible content may be quantified through the use of a microphone, the gestural content of a performance may be similarly quantified through the use of gestural sensors. An abundance of such technologies exist, varying markedly in both form and function [2]. Countless studies have explored the efficacy of implementing such, to both creative and analytical ends.

1.1 Related Works

Produced by Thalmic labs between 2014 and 2018, the Myo is a consumer device comprising of 8 EMG Sensors and a 9-DoF IMU sensor. Despite it's short commercial life-span, the demonstrable utility of the Myo as an interface for gestural interaction has maintained its prevalence within the NIME community, continuing to average around three appearances per year in NIME proceedings.

While the Myo's suitability as a compositional and performance tool has been documented extensively [7, 6, 8], the device's use as an analytical tool for purposes both musical and otherwise has proved similarly effective.

Dalmazzo et al. [5] and Sarasúa et al. [13] both made use of the Myo during studies of gestural execution in violin performance; the authors reported high classification accuracies when using the Myo to identify a range of bow articulation conditions within performed material. During integration within an HMM-based classification system, Sarasúa et al. found that the inclusion of EMG data increased early gestural recognition rates, but decreased overall gestural recognition rates when compared to classification upon lone IMU data [13]. Dalmazzo et al. compared the utility the Myo with an optoelectric system; reporting respective classification accuracies of 99.847%, and 99.460% through the use of a J48 Decision tree algorithm; the authors noted a disparity in cost between the two implemented technologies, asserting that "this result shows that it is possible to develop music-gesture learning applications based on low-cost technology which can be used in home environments for self-learning practitioners" [5].

A similar methodology was employed by Auepanwiriyakul et al. [1]. while assessing the utility of the Apple Watch for the purposes of hospital inpatient monitoring. The authors compared these to a "gold standard" optoelectronic OptiTrack system, in addition to a number of specialist IMU sensors, concluding that "with relatively few drawbacks, consumer-grade smartwatches can be objectively used within a clinical- and research-grade setting"

2. METHODOLOGY

A methodology was devised based upon a prior study [17], wherein the utility of the Myo was previously investigated for the purposes of violin bowstroke classification using an existing dataset.

2.1 Data Capture Methodology

A pilot dataset was collected comprising of synchronous gestural and audio recordings; the development of two distinct recording mechanisms was necessitated for this purpose.

Recording via two Myos and a freestanding DPA4090 microphone was triggered via a Python script; this was developed through use of the PyoMyo api [16]. Recorded IMU data comprised of three-dimensional accelerometer and gyroscopic data, Euler angles and 4-unit quaternions, at a sample rate of 50Hz; EMG data was recorded at 200Hz. Due to hardware constraints, these are the highest sample rates at which the respective data types may be recorded using the Myo [11]. Audio was recorded at 44.1KHz.

An purpose-built application for the Apple Watch was developed for the logging of IMU and Audio data, based upon Logger7 by GitHub user Shakshi3104 [14]. In addition to three-dimensional accelerometer and gyroscopic data, derived Euler angles and 4-unit quaternions are also logged, paralleling the IMU data-types offered by the Myo, albeit at a sample rate of 100Hz. Audio is recorded simultaneously at 44.1KHz.

Data was timestamped extensively for the purposes of time alignment. An initial timestamp was taken as audio recording commenced, while a second timestamp was taken upon its termination. Individual audio timestamps were later interpolated between these. Gestural samples were timestamped individually upon receipt.

Six violinists were recorded performing two-octave G and D major scales. These scales were chosen with the intention to capture a comprehensive range of both the violin's typical performance register, and movement along the four strings. Participants were required to play each note twice, such that both an up-bow and a down-bow were captured on each note; each bow-stroke one beat in duration, at a tempo of 110BPM. Each scale was performed in two bow-articulation techniques; Spiccato and Legato. Three takes of each exercise were recorded. Participants were sourced from the student body of the Royal Birmingham conservatoire, comprising of both undergraduate and postgraduate students and alumni.

2.2 Analysis Methodology

Synchronous recorded signals were first trimmed to their maximum concurrent length, such that any samples preceding the latest initial sample of any one data type was discarded; similarly, any samples following the earliest final sample of any one data type were discarded.

Recorded concurrent signals were then trimmed further, to remove silence or unwanted noise at the start and end of each recording; this was automated as follows. An initial RMS envelope was first computed; subsequently a threshold value was calculated for each, equal to 0.6x the mean of each RMS envelope. Two low-pass filters were subsequently applied to the initial envelope, with respective cutoffs of 0.5 and 2.0 Hz. Trimming signals initially to the first and last intersections of the threshold with the prior filtered envelope was found to effectively remove transient sounds prior to any playing. Subsequently trimming signals similarly, to the threshold's intersection with the second filtered envelope, was found to effectively remove any remaining unwanted silence. The trimmed audio signals were then normalised.

A linear de-trend function was applied to each channel of IMU data to counteract drift - Kok et al. [9] note the presence of drift to be characteristic of IMU sensors, attributing this to accumulated error over time.

Proportional normalisation was applied to the IMU data such that the maximum magnitude of a signal was bounded by 1, while the proportional difference in maximum magnitude between concurrent channels of data (e.g. IMU Accelerometer signals X, Y, Z) was maintained.

A system was developed wherein each recording was segmented into a series of individual bow-strokes. Use of the Madmom bi-directional RNN onset detector [3] provided sample indices within the audio data at which point note-onsets were likely to have occurred. Of the interpolated audio timestamps, those corresponding to these sample indices were then identified. These audio timestamps were then used to split all data-types into individual series of inter-onset-intervals at the closest timestamped samples.

Following this, MIR features were derived from the audio data, intending to reduce the resolution of the data. For each segment of audio data, sequential arrays of MFCCs, Delta-MFCCs, Delta-Delta-MFCCs, and Chroma coefficients were calculated. These were intended to depict both timbral and pitch characteristics of each note, while preserving temporality.

Three bow-stroke classification tasks were completed through use of an MLP neural network trained upon various data-type combinations. Participant identification was approached as a multi-class classification problem, while bow-articulation and scale-belonging were approached as binary classification problems given the availability of only two classes.

3. RESULTS

Through inspection of the classification accuracies depicted in Table 1, a number of trends can be observed. Classification accuracies were highest in the "Bow Articulation" task, wherein segmented bowstrokes were classified as being either Spiccato or Legato, with a mean accuracy of 97.89% achieved across all data-types.

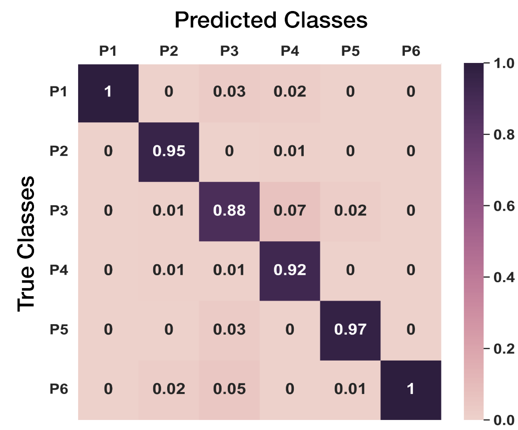

Figure 1: Apple Watch IMU+MIR Features TPR Confusion Matrix - Participant Identification

Of the three tasks, the lowest accuracies were achieved in the "Scale Belonging" task, wherein segmented bow-strokes were classified as having occurred in either of the recorded scales. In this instance, lone gestural data-types performed poorly, exhibiting accuracies approximating that of random classification. The inclusion of audio-derived MIR features

Table 1: MLP Classification Accuracies

	Classification Task Test-Accuracies		
Training Data Types	**Participant Identification**	**Bow Articulation**	**Scale Belonging**
Watch IMU	94.79%	97.30%	49.65%
Watch MIR Features	82.63%	100.0%	96.96%
Watch IMU + Watch MIR Features	95.54%	100.0%	98.48%
Myo IMU	88.81%	88.61%	50.68%
Myo EMG	90.07%	95.18%	50.98%
MIR Features	89.45%	100.0%	98.48%
Myo IMU + Myo EMG	91.19%	100.0%	50.99%
Myo IMU + MIR Features	91.49%	100.0%	95.28%
Myo IMU + Myo EMG + MIR Features	91.60%	100.0%	99.39%

increased classification accuracies markedly, however; datatype combinations wherein MIR features were included averaged 97.71%.

Of the two recording mechanisms, differences between the observed classification accuracies proved negligible. During participant identification, slightly higher classification accuracies were achieved through the use of the watch's combined audio and IMU datatypes. Using these same combined data types, classification accuracies in the "Bow Articulation" task proved consistently high between both devices. Although further incorporation of EMG data in the "Scale Belonging" task saw the Myo outperform the Watch by 0.91%, Myo classification accuracies were 3.2% lower through the use of IMU and audio data types alone.

4. DISCUSSION

The previously detailed results demonstrate the prospective utility of the smartwatch as an alternative to the IMU capabilities of the Myo Armband. Considering the Apple Watch's lack of EMG sensing functionality, it is arguable that a true parallel cannot be drawn between devices; while wearable IMU technologies are well suited towards the analysis of bowing - a gross motor skill - it is likely that such utility would be diminished while attempting analysis of fine motor skills such as fingering. For such analyses, EMG data would be better suited, given the efficacy demonstrated by Dalmazzo et al. [4].

Further considering the native functionalities of each device, the lack of an integrated microphone within the Myo armband necessitates the use of additional hardware for the purposes of audio-gestural analyses.

While the affordability of the Apple Watch may prove debatable, since discontinuation the Myo may be considered a specialist device; significant technical knowledge was required to facilitate recording, through the circumvention of now-outdated drivers. As the smartwatch may be considered to be comparatively ubiquitous, its utility could help to democratise both the conduction of audio-gestural research and the products thereof.

Remote distribution of VioLogger may allow participants to engage with future works globally, facilitating the development of a larger dataset for the purposes of more comprehensive analyses. While its offline recording mechanism may limit VioLogger's current utility, largely, to analytical ends, the introduction of a real-time data-transfer functionality may prove to facilitate its use as a tool for augmented performance or practice feedback tool.

5. CONCLUSION

Our findings indicate that the smartwatch could prove to be a perhaps surprisingly capable interface for musical audio-gestural analyses and interaction. Previously favoured technologies have presented obstacles in practicality, usability and accessibility; consequently, research artefacts developed for use with such systems have often lacked transferability due to their reliance upon inherently specialist hardware. It is hoped that through the use of such relatively ubiquitous hardware, replicability, engagement, and collaboration may be accommodated further during future works.

6. ETHICAL STANDARDS

This study was subject to approval by the ethics review board at Birmingham City University. Informed consent was sought and received from all participants prior to their participation.

7. REFERENCES

[1] C. Auepanwiriyakul, S. Waibel, J. Songa, P. Bentley, and A. A. Faisal. Accuracy and Acceptability of Wearable Motion Tracking for Inpatient Monitoring Using Smartwatches. *Sensors*, 20(24):7313, Dec. 2020.

[2] S. Berman and H. Stern. Sensors for Gesture Recognition Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(3):277–290, May 2012.

[3] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs, and G. Widmer. madmom: a new Python Audio and Music Signal Processing Library, May 2016. arXiv:1605.07008 [cs].

[4] D. Dalmazzo and R. Ramirez. Air violin: A Machine Learning Approach to Fingering Gesture Recognition. In *Proceedings of the 1st ACM International Workshop on Multimodal Interaction for Education*, pages 63–66, Glasgow UK, Nov. 2017. ACM.

[5] D. Dalmazzo, S. Tassani, and R. Ramírez. A Machine Learning Approach to Violin Bow Technique Classification: a Comparison Between IMU and MOCAP systems. In *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*, pages 1–8, Berlin Germany, Sept. 2018. ACM.

[6] J. Dooley. Colligation. In *NIME 2017*, Porto Alegre, Brazil, 2019.

[7] C. Erdem and A. Refsum Jensenius. Me & My Musical AI "Toddler". In *NIME 2022*, The University of Auckland, New Zealand, 2022.

[8] A. R. Jensenius, G. Sanchez, A. Zelechowska, and K. A. V. Bjerkestrand. Exploring the Myo Controller for Sonic Microinteraction. In *NIME 2017*, 2017.

[9] M. Kok, J. D. Hol, and T. B. Schön. Using Inertial Sensors for Position and Orientation Estimation. *Foundations and Trends® in Signal Processing*, 11(1-2):1–153, 2017. arXiv: 1704.06053.

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

[11] A. Phinyomark and E. Scheme. A feature extraction issue for myoelectric control based on wearable EMG sensors. In *2018 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, Seoul, Mar. 2018. IEEE.

[12] C. Rhodes. Membrana Neopermeable. In *NIME 2022*, The University of Auckland, New Zealand, June 2022. PubPub.

[13] Sarasúa, B. Caramiaux, A. Tanaka, and M. Ortiz. Datasets for the Analysis of Expressive Musical Gestures. In *Proceedings of the 4th International Conference on Movement Computing*, pages 1–4, London United Kingdom, 2017. ACM.

[14] Shakshi3104. Logger7, 2021.

[15] F. Verdugo, A. Ceglia, C. Frisson, A. Burton, M. Begon, S. Gibet, and M. M. Wanderley. Feeling the Effort of Classical Musicians - A Pipeline from Electromyography to Smartphone Vibration for Live Music Performance. In *NIME 2022*, The University of Auckland, New Zealand, June 2022. PubPub.

[16] P. Walkington. PyoMyo, 2021.

[17] W. Wilson, I. Ali-MacLachlan, and N. Granieri. Combining Gestural and Audio Approaches to the Classification of Violin Bow Strokes. In *Proceedings of the 10th International Workshop on Folk Music Analysis (FMA'22))*, Sheffield, United Kingdom, 2022. University of Sheffield.