A Pilot Study on Using Wearables to Identify Difficulties for Ensemble Musicians

Christian Brinkhaus
Osaka Metropolitan University
Graduate School of Informatics
Sakai, Osaka, Japan
sw25858f@st.omu.ac.jp

Tenka Naka
Osaka Metropolitan University
College of Engineering
Sakai, Osaka, Japan
sl23434c@st.omu.ac.jp

Andrew Vargo
Osaka Metropolitan University
Graduate School of Informatics
Sakai, Osaka, Japan
awv@omu.ac.jp

Ko Watanabe
DFKI GmbH
Smart Data & Knowledge Services
Kaiserslautern, Germany
ko.watanabe@dfki.de

Koichi Kise
Osaka Metropolitan University
Graduate School of Informatics
Sakai, Osaka, Japan
kise@omu.ac.jp

Abstract

Through a self-observation approach, this 24-day study investigates whether physiological signals from wearable sensors differ by session context and repertoire in orchestral settings, and whether self-rated piece difficulty can be predicted from these signals. The participant wore an EmbracePlus during the study period, collecting physiological data during home practice, group rehearsals, and two concerts. An analysis with generalized linear models found that the session type and the piece played could be used to explain the physiological changes captured by the EmbracePlus wristband. These findings inform a research arc that can enable us to identify demanding repertoire that needs more practice and could help with the identification and prevention of performance anxiety.

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Keywords

Stress; Music; Physiological Sensing; Wearables

ACM Reference Format:
Christian Brinkhaus, Tenka Naka, Andrew Vargo, Ko Watanabe, and Koichi Kise. 2026. A Pilot Study on Using Wearables to Identify Difficulties for Ensemble Musicians. In *Companion of the 2026 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '26), October 11–15, 2026, Shanghai, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3798063.3837215

1 Introduction

Musicians face stressors [15] leading to performance anxiety [14] and a higher load on working memory [1] influencing performance quality [6, 19]. Ensemble music requires a high degree of alignment between co-performers and conductor [5]. Higher cognitive load and stress can lead to less reliance on auditory feedback and poorer adaptation to tempo changes, reducing flexibility to adjust to music [10]. Traditional approaches to measuring performance stress rely on subjective self-report questionnaires such as the State-Trait Anxiety Inventory (STAI). However, these do not correlate with physiological measurements [13], highlighting the need for objective assessment methods. Wearable physiological sensors can offer continuous and objective measurement that overcomes these limitations [16]. Previous work has shown that the heart rate of orchestral musicians is significantly higher in concerts than in rehearsals [4], and the physiological stress markers of a professional pianist increase further when performing in front of an audience compared to without an audience [17]. However, none of these studies followed musicians over a longitudinal period of time.

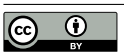
Table 1: Concert schedule and the first author's subjective difficulty labels.

Date	Piece	Difficulty
28 Dec. 2025	Grieg - Symphonic Dances	Easy
	Mahler - Symphony No. 1	Hard
12 Jan. 2026	Mendelssohn - The Hebrides Overture	Medium
	Chopin - Piano Concerto No. 2	Easy
	Stravinsky - Pulcinella Suite	Medium

This 24-day study follows the first author, an amateur violinist in two orchestras. Using the EmbracePlus [1], a validated wearable sensor [8, 11], we collected physiological data related to stress such as electrodermal activity (EDA) [12, 18] and skin temperature [3], and we collected audio recordings and timing labels during practice and performance. With these data points, we analyzed the relationship between physiological factors, the type of session (home practice, group rehearsal, or concert), and the piece being performed. The first author rated the subjective difficulty of each piece (Table 1).

Our results are promising, as they show that the type of session and the piece played do have a modest explanatory effect on each physiological measure. Our prediction results for piece difficulty show that a K-Nearest Neighbors classifier achieves a cross-validation (CV) accuracy of 44.6% (±28.8%) using the TimeSeriesSplit algorithm from scikit-learn [9]. These results show that it is feasible to develop more complex models that account for many

[1] https://www.empatica.com/embraceplus/



This work is licensed under a Creative Commons Attribution 4.0 International License.
UbiComp Companion '26, Shanghai, China
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2533-3/2026/10
https://doi.org/10.1145/3798063.3837215

Table 2: GLM Results for EDA.

Variable	Coefficient	Std. Err.	p-value
Intercept	2.52461	0.21160	< 2e-16***
Home Practice	−1.10215	0.13611	1.44e-15***
Group Rehearsal	−0.02885	0.04540	0.525
Grieg Symphonic Dances	−2.40997	0.20924	< 2e-16***
Mahler Symphony No. 1	−2.34435	0.20918	< 2e-16***
Mendelssohn The Hebrides Overture	−1.37430	0.24558	2.75e-08 ***
Stravinsky Pulcinella	−1.17106	0.24276	1.60e-06***
			$R^2_{\text{McFadden}} = 0.168$

Pairwise (Tukey)	Coefficient	Std. Err.	p-value
Concert - Home Practice	1.1021	0.1360	<0.0001***
Concert - Group Rehearsal	0.0288	0.0454	0.8006
Home Practice - Group Rehearsal	−1.0733	0.1300	<0.0001***
Chopin Piano Concerto No. 2 - Grieg Symphonic Dances	2.4100	0.2090	<0.0001***
Chopin Piano Concerto No. 2 - Mahler Symphony No. 1	2.3443	0.2090	<0.0001***
Chopin Piano Concerto No. 2 - Mendelssohn The Hebrides Overture	1.3743	0.2460	<0.0001***
Chopin Piano Concerto No. 2 - Stravinsky Pulcinella Suite	1.1711	0.2430	<0.0001***
Grieg Symphonic Dances - Mahler Symphony No. 1	−0.0656	0.0326	0.2584
Grieg Symphonic Dances - Mendelssohn The Hebrides Overture	−1.0357	0.1330	<0.0001***
Grieg Symphonic Dances - Stravinsky Pulcinella Suite	−1.2389	0.1270	<0.0001***
Mahler Symphony No. 1 - Mendelssohn The Hebrides Overture	−0.9700	0.1330	<0.0001***
Mahler Symphony No. 1 - Stravinsky Pulcinella Suite	−1.1733	0.1270	<0.0001***
Mendelssohn The Hebrides Overture - Stravinsky Pulcinella Suite	−0.2032	0.0656	0.0167*
			$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

physiological features and move towards building support systems to manage practice sessions and performance anxiety.

2 Experiment Setup

This study follows the first author, an amateur violinist who has played since childhood. Having experience in diverse orchestral settings, the first author is familiar with the structure and conventions of music ensembles across cultures while maintaining the perspective of a non-professional musician. The orchestra setting provides a structured performance environment for this experiment. Unlike sports, which depend on decision making [2] and complex interactions between athletes [7], orchestral performances follow a predetermined structure, reducing external variability.

The data collection took place between 20 December 2025 and 12 January 2026, during which the first author was active in two amateur orchestras. This period captured the late stages of the practice cycles, including two public concerts, three home practice sessions, and seven group rehearsals. The study covered a repertoire of five orchestral works spanning various styles and difficulty (Table 1). The participant had no prior experience with these pieces.

We recorded physiological measurements with the EmbracePlus at its highest sampling rate in milliseconds. The device was continuously worn on the left wrist. The participant did not have access to physiological data during the study period to prevent bias. Audio recordings provided by the orchestras and by the first author were used for post-hoc labeling. After the data collection period, we analyzed the EmbracePlus data which includes all biomarkers calculated at a 1-minute resolution using Empatica's proprietary algorithm. In this analysis, we focus on EDA and skin temperature. In addition, we tested several machine learning algorithms to predict piece difficulty from physiological responses.

The session type was determined from the practice schedule. We manually labeled the audio data for piece (5 works, Table 1) and activity (actively playing vs. resting). The labels were converted from relative timestamps in milliseconds to absolute time using the recordings' end times. The data were summarized into 1-minute segments; minutes containing mixed activities were labeled based on the majority duration and then synchronized with the EmbracePlus data. The participant provided a subjective difficulty rating (easy, medium, hard; Table 1) for each piece after the study, without access to physiological data. We filtered for active minutes, yielding a final dataset of 1136 active minutes.

3 Analysis

We used a series of generalized linear models (GLMs) with families to match the distribution of two physiological measurements: EDA and skin temperature. Each best-fitting model has the same independent variables and excludes self-labeled difficulty, as it is a less informative confounder with the piece. In Table 2 and Table 3 we can see the results for EDA and skin temperature, respectively. The GLM for EDA produced a pseudo $R^2 = 0.168$, while the GLM for skin temperature produced a pseudo $R^2 = 0.268$. Although neither of these results explains the majority of each respective physiological

Table 3: GLM Results for skin temperature.

Variable	Coefficient	Std. Err.	p-value
Intercept	31.69546	0.08844	< 2e-16***
Home Practice	0.12676	0.11517	0.271266
Group Rehearsal	−0.33162	0.07871	2.72e-05***
Grieg Symphonic Dances	−0.66025	0.09557	< 8.15e-12***
Mahler Symphony No. 1	0.26481	0.07555	<0.000475***
Mendelssohn The Hebrides Overture	−1.36377	0.10438	< 2e-16***
Stravinsky Pulcinella Suite	−0.86326	0.08434	< 2e-16***
			$R^2_{\text{McFadden}} = 0.268$

Pairwise (Tukey)	Coefficient	Std. Err.	p-value
Concert - Home Practice	−0.127	0.1150	0.5138
Concert - Group Rehearsal	0.332	0.0787	<0.0001***
Home Practice - Group Rehearsal	0.458	0.0937	<0.0001***
Chopin Piano Concerto No. 2 - Grieg Symphonic Dances	0.660	0.0956	0.0001***
Chopin Piano Concerto No. 2 - Mahler Symphony No. 1	−0.265	0.0756	0.0043**
Chopin Piano Concerto No. 2 - Mendelssohn The Hebrides Overture	1.364	0.1040	0.0001***
Chopin Piano Concerto No. 2 - Stravinsky Pulcinella Suite	0.863	0.0843	0.0001***
Grieg Symphonic Dances - Mahler Symphony No. 1	−0.925	0.0914	0.0001***
Grieg Symphonic Dances - Mendelssohn The Hebrides Overture	0.704	0.1170	0.0001***
Grieg Symphonic Dances - Stravinsky Pulcinella Suite	0.203	0.0990	0.2429
Mahler Symphony No. 1 - Mendelssohn The Hebrides Overture	1.629	0.1010	0.0001***
Mahler Symphony No. 1 - Stravinsky Pulcinella Suite	1.128	0.0799	0.0001***
Mendelssohn The Hebrides Overture - Stravinsky Pulcinella Suite	−0.501	0.0942	0.0001***
			* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

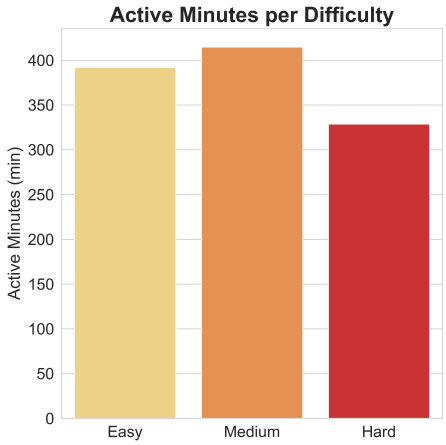


Figure 1: Distribution of active minutes per difficulty label.

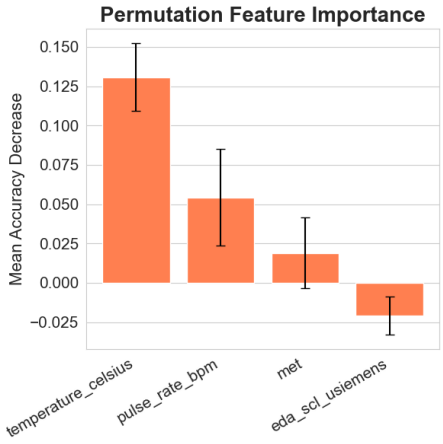


Figure 2: Permutation feature importance for KNN-based piece difficulty prediction.

factor, that the models can explain some of the response is notable given the noise in the data.

It is important to note that the results of the two GLMs do not agree. For example, with EDA, concert sessions showed a significant increase in response compared to home practice, while with skin temperature there was no significant difference. This indicates that the interplay between the physiological features and the session/piece labels is complicated and requires a detailed analysis of the possible interaction. In addition, it is notable that the participant's difficulty labels do not align with the physiological responses. For example, while Grieg and Mahler were identified as *Easy* and *Hard*, we found no difference in the EDA response. This could indicate that merely being in the same concert provided too

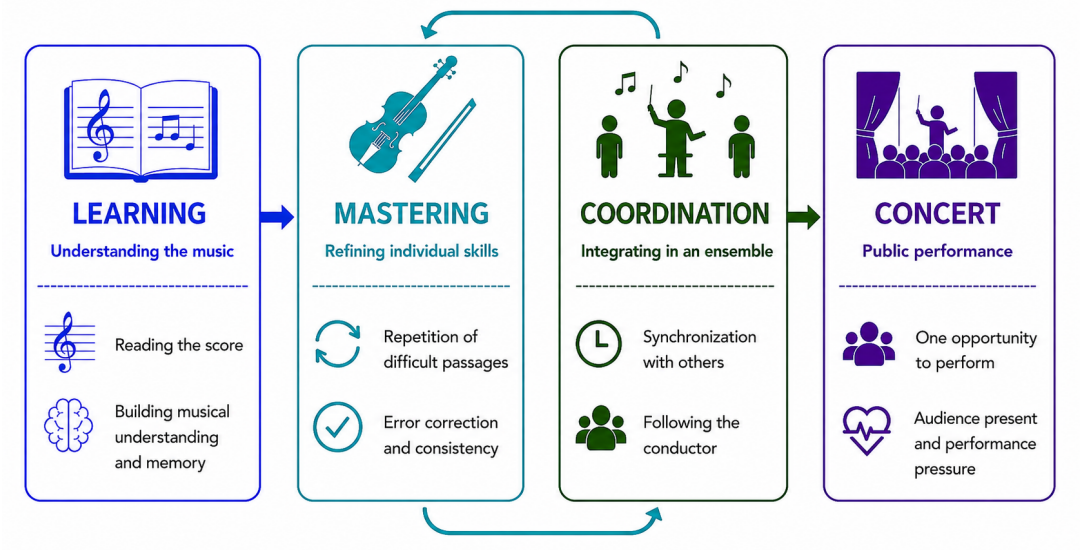


Figure 3: Stages of preparing for a musical performance; learning, mastering, coordination, and concert. Mastering and coordination iterate across rehearsals to finally end in a performance.

much noise for differentiation or that the participant was impacted by the difficulty outside of the periods for the labeled data.

To investigate whether physiological signals can predict the self-evaluated difficulty rating (Figure 1), we trained seven classifiers: Random Forest, Gradient Boosting, Logistic Regression, SVM, K-Nearest Neighbors, Naive Bayes, and Decision Tree. For prediction, we used four features: EDA, MET, pulse rate, and skin temperature. The best model was selected using 5-fold TimeSeriesSplit CV from scikit-learn [9] to respect the temporal structure. Before CV, we divided the dataset into an 80/20 train-test split, using the last 20% as the test set to avoid data leakage.

The K-Nearest Neighbors classifier ($k = 7$) achieved the highest cross-validation accuracy of 44.6% (±28.8%), above the proportional random baseline of 31.8%, though the large standard deviation indicates that the estimate is unstable across folds. On the held-out test set it reached an accuracy of 51.8% (Precision = 0.4632, Recall = 0.3448, F1-macro = 0.3948), which shows some predictability. Looking at the weighted metrics, we obtain Precision = 0.69, Recall = 0.52, and F1-Score = 0.59. However, these values are skewed as our test set did not include "hard" labeled pieces due to the temporal split. The model could thus be assessed only on easy and medium pieces, but not on the hardest class. We therefore treat the prediction results as exploratory evidence that physiological features carry some difficulty-related signal, and more data are needed to build a full classifier.

The permutation feature importance (Figure 2) revealed that skin temperature was the strongest predictor, followed by pulse rate and MET. EDA had a negligible predictive value (−0.021), suggesting that arousal-related signals alone are insufficient for predicting difficulty.

4 Conclusion and Outlook

This study shows that in the context of an ensemble musical performance, the sessions and music played can account for part of stress-related physiological responses. This is promising as it shows that a wearable could be used to inform performers about unknown stressors and performance issues.

This limited single-participant study highlights both the potential and challenges of collecting physiological data in musical performances. Music presents an interesting process in the lead-up to a concert: pieces are first individually learned, gradually mastered through repeated practice, while also requiring group coordination. In addition, it is not a linear process, as group coordination often informs what needs to be mastered (Figure 3). These building blocks make music a rich area for HCI research. At the same time, data acquisition is challenging, as the participant cohort is limited and devices are costly.

To expand beyond the limitations of our single-participant study and improve the generalizability of our findings, we are extending this work into a larger study. A key direction is to rework the labeling process. Currently, labels are derived manually from audio recordings, focusing only on sections where the participant is playing. However, the labeling process is time-consuming and raises privacy concerns when extended to more participants or group practices with no orchestral audio available. Instead, we plan to collect self-reported data after each rehearsal: the practice schedule (which pieces were practiced, when, and in which context), the perceived difficulty per piece for each session, and an overall difficulty rating per piece. Beyond this, our goal is to expand the participant pool and examine the interplay between physiological indicators.

Generative AI Disclosure

ChatGPT GPT-5.5 was utilized to generate Figure 3 from author-written prompts with iterative revision instructions from the author.

References

[1] Mário Cardoso, Levi Leonido, Antonino Pereira, and Elsa Morgado. 2025. Mental health challenges in professional musicians: A systematic review of stress, anxiety, and depression. *International Journal of Innovative Research and Scientific Studies* 8, 2 (April 2025), 3611–3621. doi:10.53894/ijirss.v8i2.6065

[2] Silke De Waelle, Griet Warlop, Matthieu Lenoir, Simon J. Bennett, and Frederik J.A. Deconinck. 2021. The development of perceptual-cognitive skills in youth volleyball players. *Journal of Sports Sciences* 39, 17 (Sept. 2021), 1911–1925. doi:10.1080/02640414.2021.1907903

[3] Katherine A. Herborn, James L. Graves, Paul Jerem, Neil P. Evans, Ruedi Nager, Dominic J. McCafferty, and Dorothy E.F. McKeegan. 2015. Skin temperature reveals the intensity of acute stress. *Physiology & Behavior* 152 (Dec. 2015), 225–230. doi:10.1016/j.physbeh.2015.09.032

[4] Claudia Iñesta, Nicolás Terrados, Daniel García, and José A. Pérez. 2008. Heart rate in professional musicians. *Journal of Occupational Medicine and Toxicology* 3, 1 (July 2008), 16. doi:10.1186/1745-6673-3-16

[5] Peter E Keller. 2014. Ensemble performance: Interpersonal alignment of musical expression. In *Expressiveness in music performance: empirical approaches across styles and cultures*, Dorottya Fabian, Emery Schubert, and Renee Timmers (Eds.). Vol. 1. Oxford University Press, Oxford, 260–282.

[6] Dianna T. Kenny. 2011. *The Psychology of Music Performance Anxiety*. Oxford University Press, Incorporated, Oxford.

[7] Tim McGarry, David I. Anderson, Stephen A. Wallace, Mike D. Hughes, and Ian M. Franks. 2002. Sport competition as a dynamical self-organizing system. *Journal of Sports Sciences* 20, 10 (Jan. 2002), 771–781. doi:10.1080/02640410232067562O

[8] Nir Milstein and Ilanit Gordon. 2020. Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Frontiers in Behavioral Neuroscience* 14 (Aug. 2020), 148. doi:10.3389/fnbeh.2020.00148

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[10] Bruno H. Repp and Yi-Huang Su. 2013. Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review* 20, 3 (June 2013), 403–452. doi:10.3758/s13423-012-0371-2

[11] Angela A. T. Schuurmans, Peter De Looff, Karin S. Nijhof, Catarina Rosada, Ron H. J. Scholte, Arne Popma, and Roy Otten. 2020. Validity of the Empatica E4 Wristband to Measure Heart Rate Variability (HRV) Parameters: a Comparison to Electrocardiography (ECG). *Journal of Medical Systems* 44, 11 (Nov. 2020), 190. doi:10.1007/s10916-020-01648-w

[12] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert. 2010. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (March 2010), 410–417. doi:10.1109/TITB.2009.2036164

[13] Esther M. van Fenema, Pim Gal, Maxime V. van de Griend, Gabriel E. Jacobs, and Adam F. Cohen. 2018. A Pilot Study Evaluating the Physiological Parameters of Performance-Induced Stress in Undergraduate Music Students. *Digital Biomarkers* 1, 2 (Jan. 2018), 118–125. doi:10.1159/000485469

[14] Johannes F. L. M. Van Kemenade, Maarten J. M. Van Son, and Nicolette C. A. Van Heesch. 1995. Performance Anxiety among Professional Musicians in Symphonic Orchestras: A Self-Report Study. *Psychological Reports* 77, 2 (Oct. 1995), 555–562. doi:10.2466/pr0.1995.77.2.555

[15] A Vervainioti and Ec Alexopoulos. 2015. Job-Related Stressors of Classical Instrumental Musicians: A Systematic Qualitative Review. *Medical Problems of Performing Artists* 30, 4 (Dec. 2015), 197–202. doi:10.21091/mppa.2015.4037

[16] Erica Volta and Nicola Di Stefano. 2024. Using Wearable Sensors to Study Musical Experience: A Systematic Review. *Sensors* 24, 17 (Sept. 2024), 5783. doi:10.3390/s24175783

[17] Aaron Williamon, Lisa Aufegger, David Wasley, David Looney, and Danilo P. Mandic. 2013. Complexity of physiological responses decreases in high-stress musical performance. *Journal of the Royal Society Interface* 10, 89 (Dec. 2013), 20130719. doi:10.1098/rsif.2013.0719

[18] Pamela Zontone, Antonio Affanni, Riccardo Bernardini, Alessandro Piras, and Roberto Rinaldo. 2019. Stress Detection Through Electrodermal Activity (EDA) and Electrocardiogram (ECG) Analysis in Car Drivers. In *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, A Coruna, Spain, 1–5. doi:10.23919/EUSIPCO.2019.8902631

[19] Muzaffer Çorlu, Pieter-Jan Maes, Chris Muller, Katty Kochman, and Marc Leman. 2015. The impact of cognitive load on operatic singers' timing performance. *Frontiers in Psychology* 6 (April 2015). doi:10.3389/fpsyg.2015.00429