

## Exploring physiology-based classification of flow during musical improvisation in mixed reality

Ruben Schlagowski, Silvan Mertes, Dominik Schiller, Yekta Said Can, Elisabeth André

### Angaben zur Veröffentlichung / Publication details:

Schlagowski, Ruben, Silvan Mertes, Dominik Schiller, Yekta Said Can, and Elisabeth André. 2024. "Exploring physiology-based classification of flow during musical improvisation in mixed reality." In *Deep Learning Theory and Applications: 5th International Conference, DeLTA 2024, Dijon, France, July 10–11, 2024, proceedings, part I*, edited by Ana Fred,ALLEL Hadjali, Oleg Gusikhin, and Carlo Sansone, 296–309. Cham: Springer. [https://doi.org/10.1007/978-3-031-66694-0\\_18](https://doi.org/10.1007/978-3-031-66694-0_18).

### Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

**Deutsches Urheberrecht**

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



# Exploring Physiology-Based Classification of Flow During Musical Improvisation in Mixed Reality

Ruben Schlagowski<sup>(✉)</sup>, Silvan Mertes, Dominik Schiller, Yekta Said Can,  
and Elisabeth André

Chair for Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany  
{ruben.schlagowski, silvan.mertes, dominik.schiller,  
yekta.can, elisabeth.andre}@uni-a.de  
<https://hcai.eu/>

**Abstract.** The flow state is desirable in many activities, e.g., while making music or being active in Virtual, Augmented, and Mixed Realities. With the long-term goal of creating affective systems that can consider the user’s flow state in real-time, we evaluated an approach for real-time flow classification during networked music performance using a deep neural network. We trained our classifier based on physiological signals (PPG and GSR) that we recorded and annotated in a laboratory study, including jamming musicians. The results that we present in this paper confirm the technical validity of this approach while also facing challenges that stem from inter-rater reliability and a heavily unbalanced dataset.

**Keywords:** Affective computing · Machine learning · Physiological signals · Flow state · Music collaboration · Networked music performance · Mixed reality · Augmented reality · Virtual reality

## 1 Introduction

Flow is the state of being fully immersed in an activity, losing track of time and space while being hyper-focused. It has been observed while engaging in sports, programming, and music. The state of flow is usually reached while being in a “sweet spot” regarding the perceived challenge level, and it was found to have positive impacts on productivity and well-being. However, measuring flow through means such as self-report questionnaires can be problematic since it is crucial not to interrupt or disturb the activity through the employed measurement method. As such, an optimal method for flow assessment would assess the user state purely from observation without deriving any attention from the test subject. Classifiers that predict the test subject’s flow state based on data gathered during activities could be highly advantageous for researchers and enable effective systems that consider the user flow state and adapt to it in real time.

To investigate the feasibility of physiology-based flow classification in real-time, we previously gathered data streams of Photoplethysmogram (PPG) and Glavanic Skin Response (GSR) sensors (see Fig. 2) from jamming musicians in a laboratory study

[38] (see Fig. 1), including mixed reality headsets for networked music performance. After the jam sessions, the participants retrospectively annotated their flow state for the recorded data streams. The study reported a substantially increased correlation between PPG-based heart rate (BPM) and self-reported flow when participants wore mixed reality headsets while making music and incentivized further research that explores real-time flow classification using physiological signals.

In this study, we follow up on this work by using the corpus of previous studies [38,39] to evaluate a first model architecture for flow classification using k-fold cross-validation, separately for data from jamming while wearing mixed reality headsets (condition B) and while only hearing each other in separated rooms, not wearing additional headsets (condition A). We also conducted a feature ablation to evaluate the impact of individual features on model performance. In this paper, we report our results and discuss limitations that need to be overcome to develop physiology-based flow classifiers for musicians.



**Fig. 1.** A study participant jamming in mixed reality.

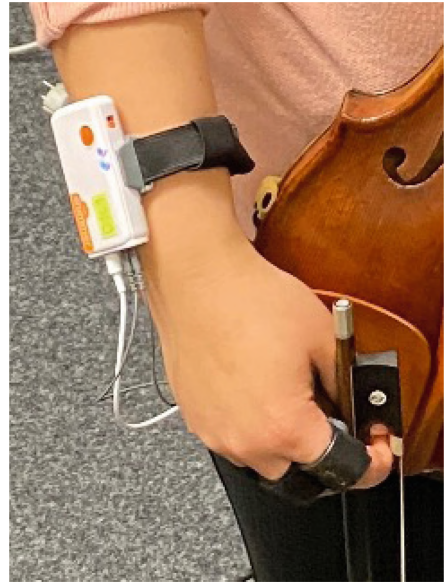


**Fig. 2.** A shimmer sensor used for data collection.

## 2 Related Work

### 2.1 Emotional State and Musical Collaboration

Numerous research endeavors have illustrated the advantageous impacts of engaging in music-related pursuits across various age groups, cultural contexts, and degrees of musical inclination [44]. These benefits encompass a spectrum ranging from bolstering self-esteem momentarily [10] to enhancing an individual's emotional state [36] and

cognitive faculties [37]. Keeler et al. determined that communal singing diminishes stress and arousal levels, noting an elevation in plasma oxytocin concentrations among musicians following improvisational singing sessions, which they attributed to the communal aspect of musical improvisation [19].

Moreover, individuals engaged in musical endeavors may enter a state of flow [6, 11], a concept initially introduced by Csikszentmihalyi et al. in 1975 [3]. In a musical context, this state is characterized by an immersive joy experienced during musical creation [4]. While direct induction or instruction of flow is improbable, enhancing conducive conditions may heighten its occurrence [18]. For musicians, fostering optimal performance conditions, such as interpersonal and emotional connectedness, or suitable environmental settings, may facilitate flow [12]. Recent years have witnessed a surge in flow research within musical contexts, with commonly utilized evaluation methods including quantitative self-report surveys and qualitative interviews [41]. Fritz et al. noted in 2007 that musicians frequently encounter flow during diverse musical activities, and its presence correlates with overall well-being [14]. Additionally, Wrigley et al. discovered in 2023 that the experience of flow during musical practice remains relatively stable across different age groups, genders, and instrumental families [45]. This aligns with observations of flow among improvisational jazz musicians [11] and pianists [6]. Loepthien et al. identified a relationship between flow and subjective well-being among individuals possessing a highly adaptable self-concept, along with evidence linking flow to prior musical engagement and experience [23].

It is worth noting that the term “well-being” is frequently synonymous with positive affect, which various studies indicate can be augmented through musical engagement [9, 13, 15, 26]. However, subjective well-being comprises both positive and negative affect [7], with a nuanced interplay influencing long-term well-being and happiness [7].

## 2.2 Music Making, Flow and Physiological Signals

Prior investigations conducted by Morgan et al. have delved into affective and behavioral sensors [28] as well as physiological signals [27] during musical improvisation. In the former study, the researchers noted synchronicities between shifts in rhythmic patterns (defined as changes in rhythmic structures) and the heart rate of musicians engaged in jam sessions. In the latter work, correlations between Electroencephalography (EEG) and Galvanic Skin Response (GSR) metrics and self-reported levels of Energy, Positivity, and Boredom were identified [27]. However, it should be noted that these self-reported measures lacked psychometric validation.

A psychological state frequently observed in musical practice and extensively explored in psychology literature [6, 11, 45] is the phenomenon of flow (refer to Subsect. 2.1). De Manzano et al. proposed that flow represents the subjective experience of a harmonious interplay between positive affect and heightened attention during performance [6]. Empirical findings suggest that the underlying physiological mechanisms of the flow state are associated with the Autonomic Nervous System [30], cortical activity [8], and dopamine release [5]. Peifer et al. posited that the relationship between flow and sympathetic arousal, as well as hypothalamic-pituitary-adrenal (HPA) axis activation, follows an inverted U-shaped curve, where moderate physiological arousal facilitates

flow, while excessive arousal impedes it [30]. Harmat et al. demonstrated an association between the flow state and respiratory depth, as well as Low-Frequency Heart Rate Variability (LF-HRV) [17].

In a study investigating flow experiences within virtual reality (VR) gaming contexts, Tozman et al. observed that flow was linked to reduced LF-HRV activity during tasks requiring a balanced skill set [42].

Skin Conductance or Electrodermal Activity (EDA) serves as an indicator of overall arousal and attention, providing a reliable measure of sympathetic activation. Kivikan-gas et al. demonstrated that flow could be examined through EDA measurements, as flow states typically coincide with physiological arousal, which is accurately reflected in EDA readings [20]. Elevated levels of flow tend to correspond to moderate EDA levels, often depicted by an inverted U-shaped function. Bian et al. presented a comprehensive assessment model for quantifying flow experiences in VR gaming utilizing physiological signals such as Heart Rate Variability, Respiration, and facial Electromyography (EMG) [2]. These prior investigations informed the selection of EDA and HRV as primary physiological indicators for measuring and assessing the flow state. Specifically, Galvanic Skin Response (GSR) was utilized as the raw signal for EDA measurement, while Photoplethysmogram (PPG) was employed for HRV assessment.

### 2.3 Musical Collaboration in XR

Several endeavors have integrated Mixed Reality (MR) or Extended Reality (XR), encompassing Augmented Reality (AR) and Virtual Reality (VR), into the development of collaborative music systems. An overview of this domain is provided by Turchet et al. [43], categorizing collaborative Musical XR systems into “networked systems” or “multi-user experiences”. The former denotes systems capable of interfacing with a variety of devices, facilitating multi-user interactions, while the latter offer opportunities for creating shared social experiences [43]. Both types of systems can support collaboration either on-site or remotely [43].

Early collaborative musical XR systems commonly employed tangibles like Fiducial markers or cards for on-site collaborative manipulation [1, 21, 33]. Subsequent developments introduced features such as minigames [32] or gesture-controlled virtual instruments [16] for collective sound creation and manipulation. Other concepts include usage of virtual step sequencers [25] or loopers [29, 40], enabling holographic composition in virtual or augmented spaces. However, in the mentioned works, mixed reality primarily revolves around either comprising virtual instruments or being an integral part of them. To our knowledge, the emphasis on fostering co-presence through mixed reality systems, such as real-time reconstruction of a remote musician in the user’s environment, remains largely unexplored in a musical context.

Crucial design prerequisites for a system facilitating real-time remote music collaboration in MR include (1) enabling musicians to perceive their own musical instruments while (2) concurrently visualizing their remote partner with their chosen musical instrument in real-time. Mixed reality solutions capable of meeting these requirements may involve MR solutions incorporating 3D avatars [46], perspective-corrected life-size projections [31], or solutions integrating 3D data capturing and reconstruction [22, 35]. While the first approach heavily relies on pre-installed assets like 3D avatars (and for

music collaboration, 3D instrument proxies), the latter approaches rely solely on capturing and reconstructing sensor data, thereby offering instrument flexibility while maintaining high fidelity for posture and gesture transmission.

Reconstruction of sensor data in the remote user environment can be accomplished by rendering voxels on Head-Mounted Displays (HMDs), as advocated by Regenbrecht et al., who demonstrated real-time applications [35]. This approach has also been employed for immersive 3D video playback [34] or interacting with reconstructed environments [22]. Schlagowski et al. proposed a similar reconstruction of 3D sensor data as an overlay for pass-through AR based on their findings from focus group interviews with musicians [38] and observed increased levels of copresence and unhindered flow in a subsequent study [39].

### 3 Methodology

#### 3.1 Model Architecture

The models we trained and evaluated with k-fold cross-validation utilized three fully connected hidden layers (sizes 32, 16, and 8). The models were trained using Adam optimizer with a learning rate of 0.0002 and a batch size of 8. Further, we utilized BCE loss with sigmoid output activation. For the BCE Loss, the flow class was weighted with a factor of 5, as the corpus was unbalanced (in favor of no-flow).

For evaluating our approach, we used a nested k-fold cross-validation procedure. Here, one k-fold cross-validation was used to find the optimal number of epochs, while an additional, superordinate k-fold was employed for evaluating the performance.

#### 3.2 Corpus

Our dataset was gathered during a laboratory study by Schlagowski and colleagues in which pairs of jamming musicians were wearing Shimmer sensors, measuring and recording GSR and PPG signals during musical improvisation [39]. In total, 13 pairs of musicians were making music with each other in 13 individual *test sessions*, which were comprised of two jam sessions each. In one jam session, musicians were jamming while solely hearing each other (Condition A); in the other, they simultaneously saw each other's 3D representation in a video pass-through augmented reality mode (Condition B). Schlagowski and colleagues found that musicians experienced similar levels of flow in both conditions while experiencing increased levels of co-presence while jamming in mixed reality (Condition B) [39]. To evaluate our approach for the conditions both separately and combined, we split the Dataset by Schlagowski and colleagues into a dataset that is solely comprised of physiological data from Condition A (data-split A from hereon), Condition B (data-split B from hereon) and a dataset that contains all available data (split AB).

Table 1 lists the 13 sessions, individual traits, and the data-splits (A, B, and AB) they are included in. As can be seen, some data for some sessions got lost or needed to be excluded due to technical problems during data collection.

From the GSR and PPG streams, which were recorded using the SSI framework (source), we extracted features using the *NeuroKit2* Toolkit by Makowski et al. [24]. Table 2 lists the features that we extracted along with their descriptions, source signals and domain (time, frequency or nonlinear).

**Table 1.** The *test sessions*, that each featured two musicians in two jam sessions, one while wearing VR headsets (A) and one while only hearing each other (B).

| No. | Age    | Gender | Instruments         | Expertise                  | A | B | AB |
|-----|--------|--------|---------------------|----------------------------|---|---|----|
| 01  | 67, 66 | m, m   | drums, guitar       | advanced, advanced         | X | X | X  |
| 02  | 27, 28 | m, m   | guitar, guitar      | advanced, advanced         | X | X | X  |
| 03  | 45, 41 | m, m   | guitar, guitar      | advanced, advanced         | X | X | X  |
| 04  | 63, 51 | m, m   | guitar, guitar      | professional, professional | X |   | X  |
| 05  | 30, 45 | m, m   | ukulele, ukulele    | advanced, advanced         | X | X | X  |
| 06  | 46, 42 | f, f   | bass, voice         | professional, professional | X | X | X  |
| 07  | 21, 20 | m, m   | bass, guitar        | beginner, beginner         | X | X | X  |
| 08  | 26, 25 | f, f   | violin, guitar      | advanced, advanced         | X | X | X  |
| 09  | 18, 18 | f, m   | guitar, bass        | advanced, advanced         |   | X | X  |
| 10  | 63, 65 | m, f   | guitar, ukulele     | advanced, advanced         |   |   |    |
| 11  | 69, 36 | m, m   | mandolin, accordeon | advanced, professional     | X | X | X  |
| 12  | 39, 30 | m, m   | voice, bass         | advanced, advanced         | X |   | X  |
| 13  | 30, 30 | m, m   | guitar, guitar      | advanced, advanced         | X |   | X  |

## 4 Results

We report our k-fold cross-validation results for three individual data splits:

A: The physiological data we gathered from participants who jammed with each other while only hearing each other (condition audio-only).

B: The physiological data we gathered from musicians jamming while seeing each other through mixed reality headsets (condition mixed reality).

AB: Data-splits A and B combined.

### 4.1 K-Fold Cross-Validation Results

Table 3 lists the aggregated results of the k-fold cross-validation. Figure 3 plots the total weighted average F1-score for data-splits A, B and AB. Additionally, Figs. 4, 5 and 6 show confusion matrices for the three splits.

### 4.2 Feature Ablation Results

Additionally, we evaluated individual features by training models using a slightly modified architecture (the input layer dimension reduced to) and singular features exclusively. Figures 7 and 8 display the weighted average F1 scores, both for conditions A and B separately. Tables 4 and 5 list these scores together with the confusion matrix values, macro average F1 and accuracy for conditions/modalities A and B.

## 5 Discussion

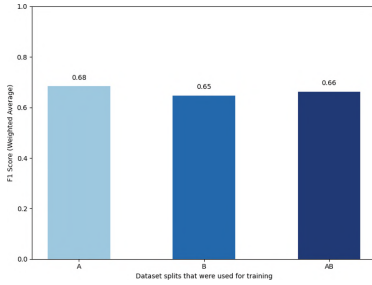
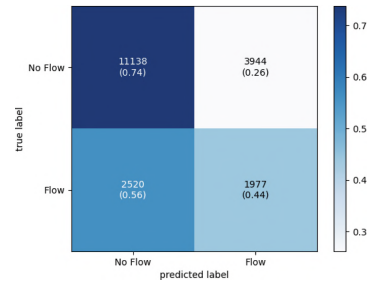
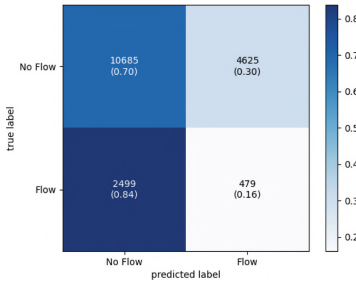
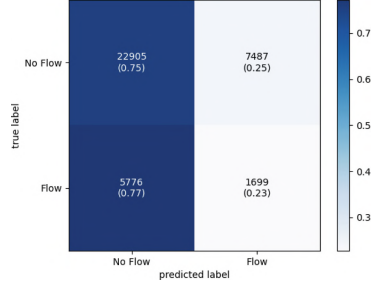
In previous work, we investigated correlations between individual physiology-based features and the self-annotated flow labels for conditions A and B, respectively. Doing

**Table 2.** Features we extracted using the *NeuroKit2* Toolkit.

| Abbreviation | Signal | Domain           | Description                                                                                    |
|--------------|--------|------------------|------------------------------------------------------------------------------------------------|
| ppg_mean     | PPG    | time domain      | mean heart rate                                                                                |
| meanNN       | PPG    | time domain      | mean of normal R to R interval values                                                          |
| sdnn         | PPG    | time domain      | mean of normal R to R interval values                                                          |
| sdsd         | PPG    | time domain      | standard deviation of successive RR interval differences                                       |
| cvNN         | PPG    | time domain      | coefficient of variation of R-R intervals                                                      |
| cvSD         | PPG    | time domain      | coefficient of variation of differences between adjacent normal RR intervals                   |
| medianNN     | PPG    | time domain      | median of normal R to R interval values                                                        |
| madNN        | PPG    | time domain      | median absolute deviation of the normal RR intervals                                           |
| pnn20        | PPG    | time domain      | The percentage of adjacent normal RR intervals that differ from each other by more than 20 ms  |
| pnn50        | PPG    | time domain      | The percentage of adjacent normal RR intervals that differ from each other by more than 50 ms  |
| pnn80        | PPG    | time domain      | The percentage of adjacent normal RR intervals that differ from each other by more than 80 ms  |
| HTI          | PPG    | time domain      | HRV triangular index                                                                           |
| TINN         | PPG    | time domain      | Triangular Interpolation of the normal RR Interval Histogram                                   |
| sd1          | PPG    | nonlinear        | The standard deviation of the each point from the $y = x$ axis in Poincare plot                |
| sd2          | PPG    | nonlinear        | The standard deviation of each point from the $y = x + \text{average R-R interval}$            |
| sd1_sd2      | PPG    | nonlinear        | The ratio between sd1 and sd2                                                                  |
| s            | PPG    | nonlinear        | The area of the ellipse which represents total HRV in Poincare Plot                            |
| lf           | PPG    | Frequency Domain | The total power spectral density in frequency spectrum between 0.05–0.15 Hz                    |
| hf           | PPG    | Frequency Domain | The total power spectral density in frequency spectrum between 0.15–0.5 Hz                     |
| lf_hf        | PPG    | Frequency Domain | The ratio between low frequency and high-frequency component                                   |
| lf_perc      | PPG    | Frequency Domain | The percentage of low-frequency component                                                      |
| hf_perc      | PPG    | Frequency Domain | The percentage of high frequency component                                                     |
| mean         | EDA    | time domain      | average value of given electrodermal signal window                                             |
| max          | EDA    | time domain      | maximum value of given electrodermal signal window                                             |
| min          | EDA    | time domain      | minimum value of given electrodermal signal window                                             |
| std          | EDA    | time domain      | standard deviation of given electrodermal signal window                                        |
| perc20       | EDA    | time domain      | 20th percentile of given EDA window                                                            |
| perc80       | EDA    | time domain      | 80th percentile of given EDA window                                                            |
| QuartDev     | EDA    | time domain      | the difference between the first quartile and the third quartile in the frequency distribution |

**Table 3.** Aggregated k-fold cross-validation results.

| Data-split | F1 (weighted) | F1 (macro) | Accuracy | True Negatives (TN) | False Positives (TP) | False Negatives (FN) | True Positives (TP) |
|------------|---------------|------------|----------|---------------------|----------------------|----------------------|---------------------|
| A          | 0.684         | 0.577      | 0.670    | 11138               | 3944                 | 2520                 | 1977                |
| B          | 0.647         | 0.434      | 0.610    | 10685               | 4625                 | 2499                 | 479                 |
| AB         | 0.663         | 0.490      | 0.650    | 22905               | 7487                 | 5776                 | 1699                |

**Fig. 3.** Total F1 scores.**Fig. 4.** Confusion Matrix for split A.**Fig. 5.** Confusion Matrix for split B.**Fig. 6.** Confusion Matrix for split AB.

so, we found a Pearson correlation of 0.44 between BPM and flow in the mixed reality condition (data-split B), which was substantially stronger than in the audio-only condition A (Pearson's  $r$  of 0.12) [39]. Based on this observation, we proposed real-time flow classification for jamming musicians in mixed reality by using physiology-based features. To follow up on this idea, we evaluated a DNN-based classification approach on the data for conditions A and B individually and the data combined.

However, we did not observe a substantial increase in the weighted average F1 score for data-split B and saw similar results for data-splits A, B, and AB instead (see Fig. 3). Looking at the individual confusion matrices, we saw a substantial increase in false positives for split B. However, as the overall performance is similar to the performance of split A, we conclude that our models must have learned feature-agnostic no-flow classification, as our unbalanced dataset contained more no-flow data. Our feature ablation results further confirm this hypothesis. While the weighted average F1 scores for our extracted features vary for split A (See Fig. 7), we see an identical value for 17 features

**Table 4.** Results per feature for condition A.

| Feature  | F1(weighted) | F1(macro) | Accuracy | True Negatives (TN) | False Positives (FP)) | False Negatives (FN) | True Positives (TP) |
|----------|--------------|-----------|----------|---------------------|-----------------------|----------------------|---------------------|
| ppg_mean | 0.64         | 0.43      | 0.70     | 13477               | 132                   | 4365                 | 1605                |
| meanNN   | 0.63         | 0.42      | 0.68     | 13196               | 134                   | 4363                 | 1886                |
| sdnn     | 0.63         | 0.45      | 0.65     | 12155               | 492                   | 4005                 | 2927                |
| sdsd     | 0.61         | 0.42      | 0.62     | 11860               | 345                   | 4152                 | 3222                |
| cvNN     | 0.57         | 0.40      | 0.57     | 10816               | 375                   | 4122                 | 4266                |
| cvSD     | 0.64         | 0.45      | 0.66     | 12516               | 440                   | 4057                 | 2566                |
| medianNN | 0.62         | 0.43      | 0.64     | 12349               | 279                   | 4218                 | 2733                |
| madNN    | 0.65         | 0.51      | 0.65     | 11596               | 1109                  | 3388                 | 3486                |
| pnn20    | 0.65         | 0.44      | 0.70     | 13612               | 187                   | 4310                 | 1470                |
| pnn80    | 0.64         | 0.47      | 0.64     | 11814               | 760                   | 3737                 | 3268                |
| pnn50    | 0.62         | 0.45      | 0.63     | 11748               | 573                   | 3924                 | 3334                |
| sd1      | 0.61         | 0.45      | 0.62     | 11479               | 609                   | 3888                 | 3603                |
| sd2      | 0.63         | 0.45      | 0.65     | 12329               | 420                   | 4077                 | 2753                |
| s        | 0.60         | 0.44      | 0.59     | 10904               | 694                   | 3803                 | 4178                |
| sd1_sd2  | 0.67         | 0.46      | 0.73     | 14180               | 183                   | 4314                 | 902                 |
| HTI      | 0.64         | 0.45      | 0.67     | 12677               | 381                   | 4116                 | 2405                |
| TINN     | 0.60         | 0.40      | 0.63     | 12221               | 126                   | 4371                 | 2861                |
| lf       | 0.65         | 0.44      | 0.71     | 13699               | 182                   | 4315                 | 1383                |
| hf       | 0.64         | 0.43      | 0.69     | 13400               | 157                   | 4340                 | 1682                |
| lf_hf    | 0.62         | 0.41      | 0.67     | 13060               | 86                    | 4411                 | 2022                |
| lf_perc  | 0.65         | 0.45      | 0.71     | 13625               | 212                   | 4285                 | 1457                |
| hf_perc  | 0.64         | 0.43      | 0.71     | 13817               | 74                    | 4423                 | 1265                |

**Table 5.** Results per feature for condition B.

| Feature  | F1(weighted) | F1(macro) | Accuracy | True Negatives (TN) | False Positives (FP)) | False Negatives (FN) | True Positives (TP) |
|----------|--------------|-----------|----------|---------------------|-----------------------|----------------------|---------------------|
| ppg_mean | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| meanNN   | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| sdnn     | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| sdsd     | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| cvNN     | 0.75         | 0.46      | 0.80     | 14479               | 77                    | 2901                 | 831                 |
| cvSD     | 0.75         | 0.45      | 0.81     | 14859               | 7                     | 2971                 | 451                 |
| medianNN | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| madNN    | 0.76         | 0.45      | 0.83     | 15266               | 0                     | 2978                 | 44                  |
| pnn20    | 0.71         | 0.43      | 0.72     | 13188               | 68                    | 2910                 | 2122                |
| pnn80    | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| pnn50    | 0.76         | 0.45      | 0.83     | 15141               | 5                     | 2973                 | 169                 |
| sd1      | 0.76         | 0.45      | 0.83     | 15157               | 0                     | 2978                 | 153                 |
| sd2      | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| s        | 0.76         | 0.46      | 0.84     | 15298               | 0                     | 2978                 | 12                  |
| sd1_sd2  | 0.75         | 0.48      | 0.78     | 14126               | 162                   | 2816                 | 1184                |
| HTI      | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| TINN     | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| lf       | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| hf       | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| lf_hf    | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| lf_perc  | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |
| hf_perc  | 0.76         | 0.46      | 0.84     | 15310               | 0                     | 2978                 | 0                   |

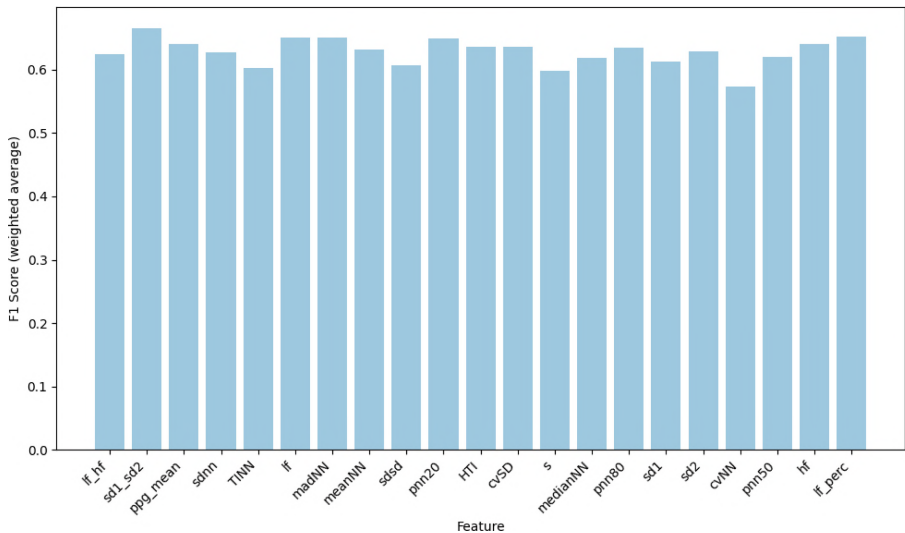


**Fig. 7.** F1 scores (weighted average) for the feature ablation in condition A.

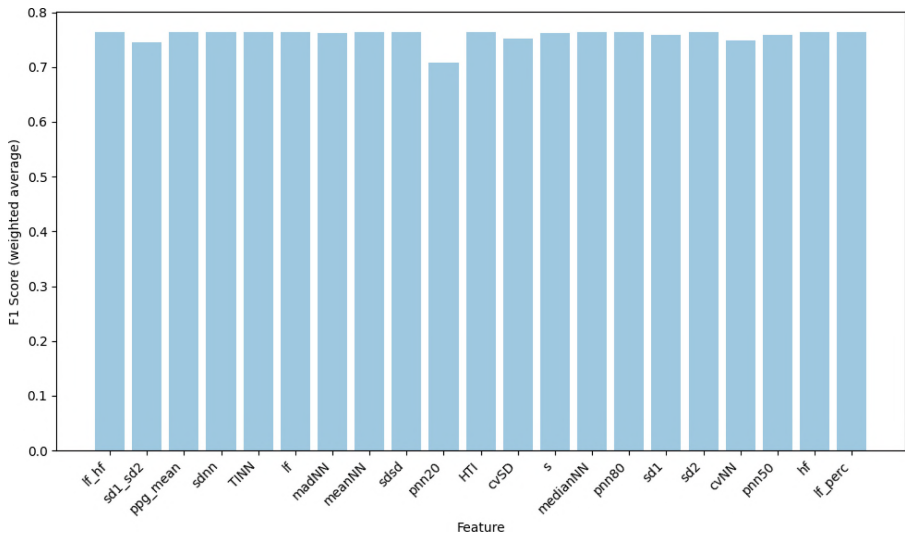


**Fig. 8.** F1 scores (weighted average) for the feature ablation in condition B.

in split B (see Fig. 8). In these 17 instances, our model learned to predict no-flow exclusively due to the biased dataset in split B, which is also reflected in the zero-values for false and true positives in Table 5.

One reason for this prediction strategy change for split B is the drop-outs from the individual data-splits. For instance, missing data from test session 4 in split B (see Table 1) featured two professional guitarists playing on a high musical level and flow-

ing continuously, thus shifting the overall label balance heavily towards flow for split A when compared to split B. This illustrates a key challenge that would need to be overcome in order to train reliable flow classifiers: The retrospective annotation of the gathered data by study participants varied greatly, not only in regards to annotation strategy, but also annotation frequency, which varied with the skill of musicians that took part in the study. We thus aim to extend our dataset by gathering data from more musicians that are also more similar in regards to proficiency with their instruments.

Overall, our approach performed best and substantially above chance in split A, where the occurrence of flow-/no-flow labels was more balanced compared to split B. As such, balanced datasets seem to be a prerequisite for achieving physiology-based flow prediction using DNNs. However, gathering balanced data remains quite a challenge, as participants reported long intervals of no-flow, independent of whether they were wearing HMDs or not. As we did not observe an increase in model performance in the mixed reality condition (split B), we cannot confirm our previous hypotheses, which expected a better prediction performance for classifiers using features from condition B. Our feature ablation results show relatively evenly distributed performances for individual features, which contradicts the previously observed Pearson correlation of 0.44 between BPM and annotated flow in condition B [39].

## 6 Conclusion

In this paper, we reported on our experiments that explored a deep neural network-based classification approach for the flow state of jamming musicians wearing PPG and GSR sensors while making music. While we found challenges imposed by our dataset, we can confirm that our approach performed substantially above chance while not employing trivial strategies such as feature-agnostic prediction if the dataset is not too heavily biased for the no-flow class. Our feature ablation results did not reveal any physiology-based feature that stood out regarding model performance, which could not confirm hypotheses from previous work that identified heart rate as a predominant feature for flow classification. Furthermore, we did not observe a substantial increase in model performance for data stemming from musicians jamming while wearing Mixed Reality (or AR/VR) headsets compared to musicians who were solely jamming while only hearing each other. Being the first study to investigate physiology- and DNN-based flow classification, we can confirm that despite our dataset problems, our model could still classify flow to some extent and perform above chance. Overall, we encourage further research in this field by evaluating more model architectures and working with a larger dataset that we aim to gather in collaboration with more musicians.

**Acknowledgments.** This paper was partially funded by the DFG through the Leibniz award of Elisabeth André (AN 559/10-1).

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Barrass, S., Barrass, T.: Musical creativity in collaborative virtual environments. *Virtual Reality* **10**(2), 149–157 (2006)
2. Bian, Y., et al.: A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. *Pers. Ubiquit. Comput.* **20**(5), 821–832 (2016)
3. Csikszentmihalyi, M.: *Beyond Boredom and Anxiety*. The Jossey-Bass Behavioral Science Series (1975)
4. Csikszentmihalyi, M.: *Beyond Boredom and Anxiety*. Jossey-Bass (2000)
5. Davis, C.A., et al.: Dopamine for “wanting” and opioids for “liking”: a comparison of obese adults with and without binge eating. *Obesity* **17**(6), 1220–1225 (2009)
6. De Manzano, Ö., Theorell, T., Harmat, L., Ullén, F.: The psychophysiology of flow during piano playing. *Emotion* **10**(3), 301 (2010)
7. Diener, E., Emmons, R.A.: The independence of positive and negative affect. *J. Pers. Soc. Psychol.* **47**(5) (1984)
8. Dietrich, A.: Neurocognitive mechanisms underlying the experience of flow. *Conscious. Cogn.* **13**(4), 746–761 (2004)
9. Dunbar, R.I., Kaskatis, K., MacDonald, I., Barra, V.: Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* **10**(4), 147470491201000420 (2012)
10. Elvers, P., Fischinger, T., Steffens, J.: Music listening as self-enhancement: effects of empowering music on momentary explicit and implicit self-esteem. *Psychol. Music* **46**(3), 307–325 (2018)
11. Forbes, M.: Giving voice to jazz singers’ experiences of flow in improvisation. *Psychol. Music* **49**(4), 789–803 (2021)
12. Ford, J.L., Vosloo, J., Arvinen-Barrow, M.: ‘pouring everything that you are’: musicians’ experiences of optimal performances. *Br. J. Music Educ.* **37**(2), 141–153 (2020)
13. Fredenburg, H.A., Silverman, M.J.: Effects of music therapy on positive and negative affect and pain with hospitalized patients recovering from a blood and marrow transplant: a randomized effectiveness study. *Arts Psychother.* **41**(2), 174–180 (2014)
14. Fritz, B.S., Avsec, A.: The experience of flow and subjective well-being of music students. *Horiz. Psychol.* **16**(2), 5–17 (2007)
15. Groarke, J.M., Hogan, M.J.: Listening to self-chosen music regulates induced negative affect for both younger and older adults. *PLoS ONE* **14**(6), e0218017 (2019)
16. Hamilton, R., Platz, C.: Gesture-based collaborative virtual reality performance in carillon. In: *Proceedings of the 2016 International Computer Music Conference*, pp. 337–340 (2016)
17. Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., Ullén, F.: Physiological correlates of the flow experience during computer game playing. *Int. J. Psychophysiol.* **97**(1), 1–7 (2015)
18. Jackson, S.A., Csikszentmihalyi, M.: *Flow in sports. The keys to optimal experiences and performances: human kinetics* (1999)
19. Keeler, J.R., et al.: The neurochemistry and social flow of singing: bonding and oxytocin. *Front. Hum. Neurosci.* 518 (2015)
20. Kivikangas, J.M., et al.: Psychophysiology of flow experience: an explorative study (2006)
21. Kusunoki, F., Sugimoto, M., Hashizume, H.: *Symphony-q: a support system for learning music in collaborative learning*. In: *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 6–pp. IEEE (2002)
22. Lindlbauer, D., Wilson, A.D.: Remixed reality: manipulating space and time in augmented reality. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (2018)

23. Loepthien, T., Leipold, B.: Flow in music performance and music-listening: differences in intensity, predictors, and the relationship between flow and subjective well-being. *Psychol. Music* **50**(1), 111–126 (2022)
24. Makowski, D., et al.: Neurokit2: a Python toolbox for neurophysiological signal processing. *Behav. Res. Methods* 1–8 (2021)
25. Men, L., Bryan-Kinns, N.: Lemo: supporting collaborative music making in virtual reality. In: 2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE), pp. 1–6. IEEE (2018)
26. Merry, M., Silverman, M.J.: Effects of patient-preferred live music on positive and negative affect and pain with adults on a post-surgical oncology unit: a randomized study. *Arts Psychother.* **72**, 101739 (2021)
27. Morgan, E., Gunes, H., Bryan-Kinns, N.: Instrumenting the interaction: affective and psychophysiological features of live collaborative musical improvisation (2014)
28. Morgan, E., Gunes, H., Bryan-Kinns, N.: Using affective and behavioural sensors to explore aspects of collaborative music making. *Int. J. Hum. Comput. Stud.* **82**, 31–47 (2015)
29. Park, S.: ARLooper: a mobile AR application for collaborative sound recording and performance. In: Kronland-Martinet, R., Ystad, S., Aramaki, M. (eds.) *CMMR 2019. LNCS*, vol. 12631, pp. 554–562. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-70210-6\\_35](https://doi.org/10.1007/978-3-030-70210-6_35)
30. Peifer, C., Schulz, A., Schächinger, H., Baumann, N., Antoni, C.H.: The relation of flow-experience and physiological arousal under stress-can u shape it? *J. Exp. Soc. Psychol.* **53**, 62–69 (2014)
31. Pejsa, T., Kantor, J., Benko, H., Ofek, E., Wilson, A.: Room2room: enabling life-size telepresence in a projected augmented reality environment. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 1716–1725 (2016)
32. Pfaff, S., Lervik, O.: Games in concert. In: 2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), pp. 1–4 (2015)
33. Poupyrev, I., et al.: Augmented groove: collaborative jamming in augmented reality. In: *ACM SIGGRAPH 2000 Conference Abstracts and Applications*, vol. 17, p. 77 (2000)
34. Regenbrecht, H., Ott, C., Park, N., Duncan, S., Collins, J.: Voxelvideos for entertainment, education, and training. *IEEE Access* **9**, 68185–68196 (2021)
35. Regenbrecht, H., Park, J.W., Ott, C., Mills, S., Cook, M., Langlotz, T.: Preaching voxels: an alternative approach to mixed reality. *Front. ICT* **6**, 7 (2019)
36. Schäfer, K., Saarikallio, S., Eerola, T.: Music may reduce loneliness and act as social surrogate for a friend: evidence from an experimental listening study. *Music Sci.* **3**, 2059204320935709 (2020)
37. Schellenberg, E.G.: Music and cognitive abilities. *Curr. Dir. Psychol. Sci.* **14**(6), 317–320 (2005)
38. Schlagowski, R., Gupta, K., Mertes, S., Billinghamurst, M., Metzner, S., André, E.: Jamming in MR: towards real-time music collaboration in mixed reality. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 854–855. IEEE (2022)
39. Schlagowski, R., et al.: Wish you were here: mental and physiological effects of remote music collaboration in mixed reality. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16 (2023)
40. Schlagowski, R., Wildgrube, F., Mertes, S., George, C., André, E.: Flow with the beat! human-centered design of virtual environments for musical creativity support in VR. In: *Proceedings of the 14th Conference on Creativity and Cognition*, pp. 428–442 (2022)
41. Tan, L., Sin, H.X.: Flow research in music contexts: a systematic literature review. *Music Sci.* **25**(4), 399–428 (2021)

42. Tozman, T., Magdas, E.S., MacDougall, H.G., Vollmeyer, R.: Understanding the psychophysiology of flow: a driving simulator experiment to investigate the relationship between flow and heart rate variability. *Comput. Hum. Behav.* **52**, 408–418 (2015)
43. Turchet, L., Hamilton, R., Çamci, A.: Music in extended realities. *IEEE Access* **9**, 15810–15832 (2021)
44. Welch, G.F., Biasutti, M., MacRitchie, J., McPherson, G.E., Himonides, E.: The impact of music on human development and well-being. *Front. Psychol.* **11**, 1246 (2020)
45. Wrigley, W.J., Emmerson, S.B.: The experience of the flow state in live music performance. *Psychol. Music* **41**(3), 292–305 (2013)
46. Yoon, B., Kim, H., Lee, G.A., Billinghamurst, M., Woo, W.: The effect of avatar appearance on social presence in an augmented reality remote collaboration. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 547–556. IEEE (2019)