EarTune: Exploring the Physiology of Music Listening

Kayla-Jade Butkow
University of Cambridge
Cambridge, UK
Nokia Bell Labs
Cambridge, UK
kjb85@cam.ac.uk

Andrea Ferlini
Nokia Bell Labs
Cambridge, UK
andrea.ferlini@nokia-bell-labs.com

Fahim Kawsar
Nokia Bell Labs
Cambridge, UK
fahim.kawsar@nokia-bell-labs.com

Cecilia Mascolo
University of Cambridge
Cambridge, UK
cm542@cam.ac.uk

Alessandro Montanari
Nokia Bell Labs
Cambridge, UK
alessandro.montanari@nokia-bell-labs.com

ABSTRACT

Music is a universal component of human culture, which influences emotion, and mental state and also has a direct impact on the physiological functioning of the body. Heart rate, breathing rate and heart rate variability have been shown to be impacted by music listening, although the exact impact of combinations of musical features on these parameters is mostly unexplored. However, exploring how these musical features influence physiology enhances our understanding of the potential of music to be used as a tool to regulate and provide interventions. In this paper, we present EarTune, a system for predicting changes in physiological parameters and subjective categorisation of the 'feeling' of a song using only the vital signs that can be collected using earables. With an accuracy of 70% for predicting the change in physiology due to music listening and an accuracy of 92% in predicting the user's 'feeling' of the song, EarTune paves the way towards the development of a system that can tailor music suggestions based on an individual's current physiological state, contextual state and emotional needs.

CCS CONCEPTS

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

KEYWORDS

Music, Physiological Features, Earable, Vital Signs

ACM Reference Format:
Kayla-Jade Butkow, Andrea Ferlini, Fahim Kawsar, Cecilia Mascolo, and Alessandro Montanari. 2024. EarTune: Exploring the Physiology of Music Listening. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3675094.3680519

UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3680519

1 INTRODUCTION

Music is a universal component of human culture, which has long been used to impact human emotion and mental states [18]. In childhood, lullabies are used to stop crying; in workouts, pop music is used to motivate and to overcome tiredness, as well as to improve exercise efficiency [15]. However, the influence of music goes beyond emotion and extends into the physiological functioning of the body.

Body functions, such as heart rate, heart rate variability and breathing rate, are regulated by the autonomic nervous system (ANS), which consists of the sympathetic nervous system (SNS), responsible for the 'fight or flight' response, and the parasympathetic nervous system (PNS), which promotes relaxation and recovery. Music has been shown to stimulate either system, depending on the properties of the music itself such as its tempo, intensity, and rhythm. This stimulation of one system over another results in the increase or decrease of involuntary bodily functions. For example, fast-paced, rhythmically intense music has been shown to increase heart rate, and breathing rate and decrease heart rate variability (HRV), mimicking the body's natural response to physical activity or stress [17, 18]. On the other hand, slow, melodic songs often have a calming effect, reducing these rates and inducing a state of relaxation. There is also a growing body of literature indicating that music affects exercise - both in terms of exercise performance, and the individual's rate of perceived exertion (RPE) as well as enjoyment during exercise [15]. Additionally, there is evidence that using music with the wrong features can actually hinder exercise enjoyment and perception of exhaustion [19]. It is also evident from literature, that the physiological effects of music listening differ based on the familiarity of the music [3]. Suppose a person likes a particular song and feels an emotional connection to that piece of music. In that case, they might have a different emotional, and thus physiological response to listening to that song.

Exploring how musical features influence physiology enhances our understanding of the potential of music to be used as a tool to regulate and provide interventions. For example, music could be tailored for exercise based on a person's mood and physiology. Music could also be used for non-pharmacological interventions such as to reduce anxiety or panic attacks.

Earables, or sensor-equipped earbuds, are the optimal device for such a system. Through their fundamental functionality, earables provide music playback to the user and so can deliver the intervention. Modern earables, such as the OmniBuds[1] provide continuous vital signs from PPG sensors on the device. These vitals include heart rate, breathing rate and heart rate variability. Other works [2, 4, 8, 9, 12, 14, 16] also show the possibility of obtaining vital signs using PPG and microphones on earables.

In this paper, we present EarTune, a profiler for predicting changes in physiology and subjective categorisation of the type of song using only the vital signs that can be collected using earables. Specifically, we collected physiological data from users while listening to music of two categories (with the songs defined by the user): calming and energetic. We extract music features and physiological features and employ machine learning to identify the resulting change (increase or decrease) in physiology based on music listening. We also use the features to predict which category the user will place the song in. In doing so, we develop a system that can both understand the resultant 'feel' of a song for the user and also the change in their physiology due to listening. This forms the basis for a music system which can select specific music for a user based on their context (*i.e.,* during exercise, a user will want an energetic song), or mood (*i.e.,* when stressed, a user will want a calming song).

This paper makes the following contributions:

- We developed, for the first time, a profiler of physiological changes to music listening based on physiological features that can be extracted with earables. EarTune paves the way towards developing systems that can tailor music recommendations based on an individual's current physiological state, contextual state and emotional needs.
- We collected a novel dataset of physiological data from 15 participants while listening to music.
- We performed an analysis of the impact of music listening on physiology and found that we can predict the direction of change of physiological parameters with a 70% accuracy. We can also predict the user's categorisation of a song with an accuracy of 92%.

2 SYSTEM DESIGN

Our overall system pipeline is provided in Figure 1. EarTune uses features extracted from music to predict the direction of change (*i.e.,* increase or decrease) of physiological parameters due to music listening.

2.1 Study Design

In this work, we aim to study the link between musical features and changes in human physiology. Due to the strong individual preference for music, we center the study around users' preferred music, rather than music selected by the investigators. This ensures that strong physiological reactions are not due to dislike of the song. To achieve this, prior to the data collection, we asked participants to make two playlists, each consisting of 10 songs. The first playlist contained songs that make the user feel excited and energetic and the second, songs which make the user feel calm and relaxed.

From these two playlists, we curated a playlist per user containing two exciting songs from the playlist they provided and two calming songs from their playlist. We also played the user one song that was not present in any of the user's playlists, and which was common to all users *i.e.,* an unseen song. This song was included to analyse whether we could still make accurate predictions of physiological changes for a song without a predefined category from the user. Between each song of the same type, we included a 30-second silent baseline (BL) period. Between songs of different types, we included a three-minute silent baseline period. These baselines were included to allow the physiological parameters to return to a resting state. They also enabled us to compare the changed state during the song to the state of the most recent baseline. This playlist structure is indicated in Figure 2.

2.2 Data collection playlist generation

To select the songs for the data collection sessions, we extracted features from the music in the user's playlists (as discussed in Section 2.4). We then applied principle component analysis for dimensionality reduction to two dimensions and performed k-means clustering on the music features per user to generate two clusters: one of which should contain the energetic songs and one which should contain the calming songs. The intuition behind this is that although music is personal, the music that makes a person feel a specific way should have similar musical features and thus should cluster together. Based on the clustering, we selected two songs per cluster that lie close to one another, but far from the other cluster to try and achieve measurable changes in physiology due to vastly different music. The K-means clustering for one user is provided in Figure 3. From the figure, it is clear that the intuition mostly holds, since all the calm songs lie within one cluster. However, two of the energetic songs lie within the calm cluster. For this specific user, we selected songs E4 and E9, and C7 and C8.

2.3 Data collection

During the data collection sessions, participants wore a Zephyr BioHarness 3.0 chest strap. The Zephyr captures ECG signals and breathing signals using a pressure sensor. While we envision the final system running on a sensor-equipped earable, we collected data from the Zephyr to ensure system development using a high-fidelity data source. While wearing the Zephyr, participants sat and relaxed in a room while wearing their earphones. During the session, they listened to their custom playlist as detailed in Section 2.1. In total, we collected data from 15 participants, with each session lasting approximately 30 minutes.

2.4 Music Feature Extraction

To extract features from the songs themselves, we leverage both high and low level features. We extract the high-level features from the Spotify API[2], and low-level features using Essentia[3]. The selected features are depicted in Figure 4. These features were chosen as they relate both to the high-level feel of the song (*e.g.,* energy is a measure of the intensity and activity level of a song; valence is a measure of how positive sounding the song is), and

[1]https://www.omnibuds.tech/

[2]https://developer.spotify.com/documentation/web-api/reference/get-audio-features

[3]https://essentia.upf.edu/index.html

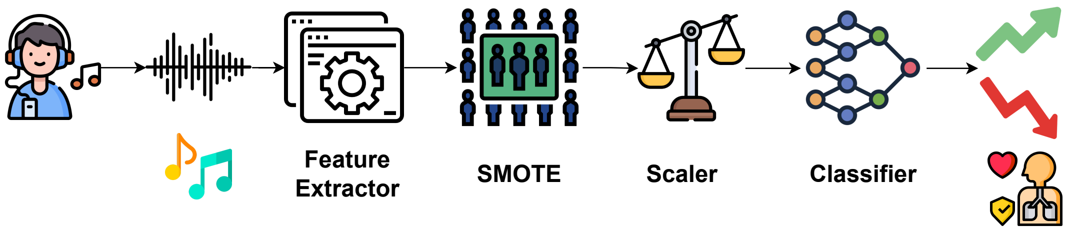


Figure 1: EarTune System Overview.

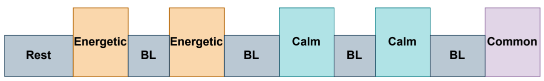


Figure 2: Data collection playlist structure.

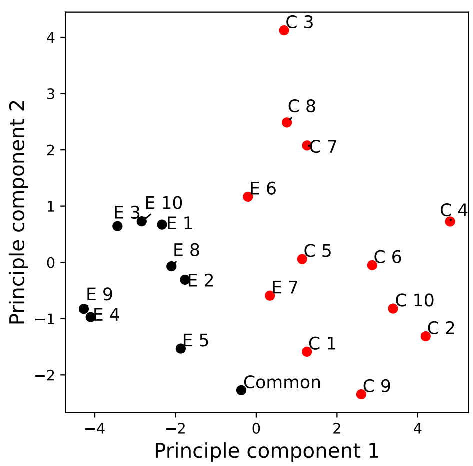


Figure 3: K-means clustering of musical features, where E refers to the songs from the energetic playlist, and C to the songs from the calm playlist.

the low-level features of the signal (*e.g.,* MFCCs which relate to the frequency content of the song). This results in 88 features being extracted per song. These features are listed in Appendix A.

2.5 Physiological Feature Extraction

To assess the impact of music on physiology, we extract a large number of physiological features. We extract physiological features for 30 seconds before each song (during the silent period before the song *i.e.,* the baseline), and then for the last 30 seconds of each song and subtract them to calculate the change in the physiological parameter. According to literature, heart rate, breathing rate and HRV are the physiological parameters most influenced by music [17, 18]. They are also parameters that can easily be extracted from an Earable, either from a PPG sensor or from a microphone [4, 8, 9].

We computed heart rate using the ECG signal and used the respiratory frequency (Rf) provided by the Zephyr chest strap. We also compute the HRV by computing the inter-beat interval from the

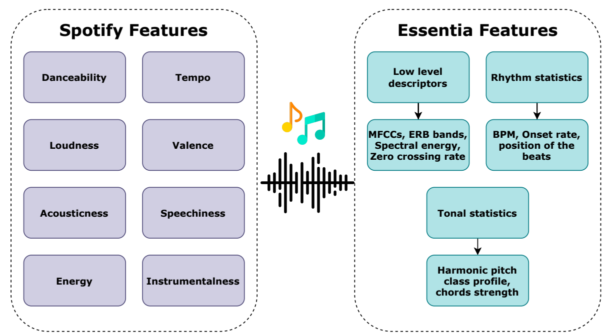


Figure 4: The features extracted from the music.

ECG signal. In addition, we developed a feature related to the rate of change of the breathing (Rc) by computing the difference between each consecutive pair of values divided by the time difference. For each extracted feature, we compute the mean, standard deviation, minimum, maximum and median. Our overall feature set contains 17 features, each of which is representative of a net change in the statistical measure of the feature.

2.6 Physiological change prediction

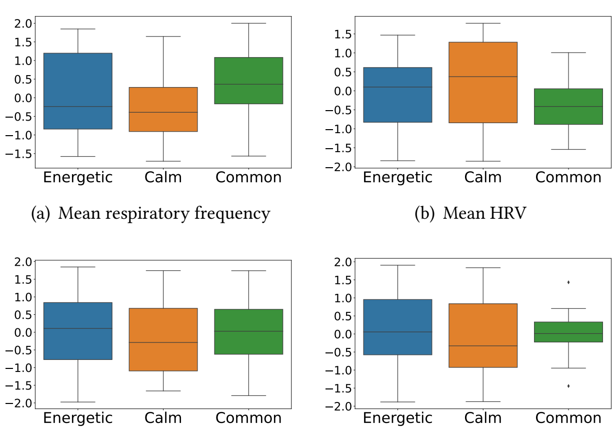
To formalise the relationship between musical features and changes in physiology, we predict whether a physiological parameter will increase or decrease upon music listening relative to the baseline. To achieve this, we train several classifiers and assess their ability to predict whether the parameter increases or decreases based on the musical features. The pipeline for this system is presented in Figure 1. We use the features extracted from the music (Section 2.4) and apply the Synthetic Minority Oversampling Technique (SMOTE) [5] as a data augmentation technique and to reduce the class imbalance in the dataset for predictions of physiological changes. We then apply Z-score standardisation and use this as input to the classifier. We compare SVM with various kernels, Decision Tree, AdaBoost, and K-nearest neighbours classifiers. Due to the small number of participants, we train and test using stratified k-fold cross-validation with 8 splits.

3 ANALYSIS

Here now follows an analysis of the performance of EarTune. We first focus on empirical data analysis. We then aim to predict physiological changes, and finally, we attempt to categorise different songs. We assess the pipeline's performance and its applications in these areas.

3.1 Empirical analysis

To analyse the impact of music on physiology, we first present boxplots showing the changes in various parameters over all participants when listening to music of the three categories: energetic, calming, and the common song. These boxplots are provided in Figure 5. From Figure 5(a), we see that for the mean breathing rate, the median value is lower for the calm condition than the energetic condition. We also see a much larger range of breathing rates and durations in the energetic condition. Interestingly, both the energetic and calm have negative median breathing rates, implying that the breathing rate decreases while listening to a preferred song compared to a resting period. However, the median breathing rate of the common song is much higher than the baseline implying that the breathing rate increased compared to the baseline. In Figure 5(b), we see that the mean HRV increases more for the calm scenario than the energetic but both increase compared to the baseline. The mean and standard deviation (std) in heart rate (Figures 5(c) and 5(d)) have consistent patterns where the energetic category shows positive median values and a narrower spread than the calm. Under calm, the median values for both are negative showing a decrease compared to the baseline. Interestingly, the common song shows a very narrow range in the standard deviation of heart rate and a much narrower range in the mean heart rate, showing that the non-preferred song does not impact the heart rate as much as the preferred song compared to the baseline.



(a) Mean respiratory frequency

(b) Mean HRV

(c) Mean HR

(d) Standard deviation in HR

Figure 5: Boxplots showing the changes in physiological parameters over all participants with various music types.

Therefore, it is evident that changes in physiology occur due to listening to music of different subjective categorisations. It is also clear that listening to a new song, or a non-preferred song, alters physiology differently to preferred music. However, further analysis is required to identify whether this is the effect of non-preferred music listening, or whether the musical features of this song are significantly different from the other clusters, making this song fall into a third, undefined, emotional category. We also leave clustering this song with the user's other music for future work.

3.2 Predicting changes in physiology

Overall, the best classifier is the decision tree classifier which predicts changes in physiology with an overall balanced accuracy of 0.7. However, upon further analysis of the results, we see that different classifiers achieve different performances for the four physiological parameter sets. This is shown in Table 1, where it is evident that decision tree has the best performance for HRV, Rc and respiratory frequency. However, the best results for heart rate are achieved using the k-nearest neighbours (KNN) classifier. It must be noted that using the decision tree classifier, balanced accuracy for heart rate prediction is 0.53, which is not a significant performance decline to that obtained with the KNN.

From Table 1, it is evident that not all physiological changes can be accurately predicted based on music features. Respiratory frequency (its mean, median, minimum, maximum and standard deviation) are all directly influenced by musical features and changes can be accurately predicted. However, for heart rate and HRV, overall, we do not achieve results that are significantly better than a random guess.

Table 1: Best classification results per physiological feature set.

Parameter set	Classifier	**Balanced Accuracy**
Heart rate	KNN	0.58
HRV	Decision Tree	0.54
Rate of change of breathing	Decision Tree	0.77
Respiratory frequency	Decision Tree	1

We further break this down in Figure 6, where we provide the accuracy of predicting each physiological parameter based on the best-performing classifier for that parameter set from Table 1. Here we see that, again, we achieve good performance in predicting the breathing-related features, but worse performance for the heart-related features. We see that no HRV-related feature achieved a classification accuracy above 56%, showing that inter-beat interval was not directly affected by the extracted musical features. According to Figure 5(b), there is a change in the distribution of HRV values with different music listening conditions. However, we see that both calming and energetic music results in an increased median HRV from the baseline, and there is little change in the minimum HRV with music listening. As such, we can conclude that listening to preferred music increases the inter-beat interval, irrespective of whether the listener finds it calming or not and irrespective of the music features.

We achieve better performance on the heart rate-related metrics, with the worst performance being achieved on the standard deviation of the heart rate. From Figure 5(d), we see that again the median of the standard deviation in heart rate changes sign based on the song condition, however, the minimum and maximum remain fairly consistent. Thus, it is expected that these outliers in the dataset lead to ambiguities in the dataset, resulting in poor classification results.

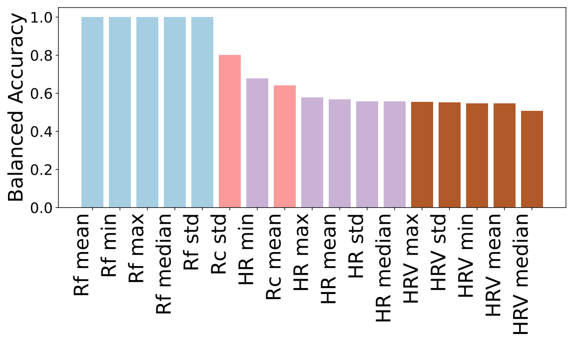


Figure 6: Classification performance per parameter using the best classifier for each parameter group.

3.3 Predicting song categorisation

Finally, we try to assess whether it is possible to predict the user's categorisation of the song based on the physiology, musical features and a combination of both. We present the results of this classification using a decision tree classifier in Figure 7. When using only the changes in physiology, we achieve a 50% accuracy in predicting whether the user finds the song to be calming or energetic. Using only the musical features, this increases to 70%. Thus, it is evident that while the musical features are more related to the user's categorisation than their changes in physiology, neither achieves excellent results in isolation. However, when using a combination of the physiological changes and the musical features, we can predict the categorisation with an excellent accuracy of 92%. This indicates that musical features alone are not enough to understand how a piece of music makes a person feel, but that when viewed in combination with how their physiology changes while listening, we are able to predict the emotional response to the song.

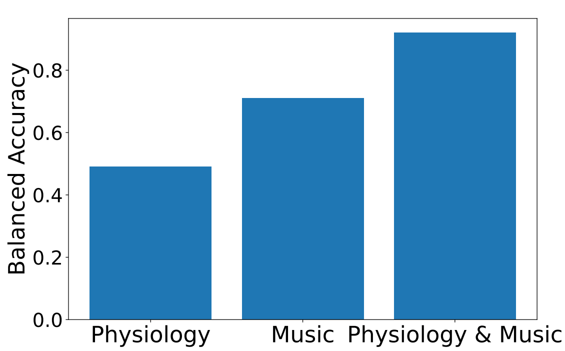


Figure 7: Results of predicting the song categorisation.

4 RELATED WORK

Limited work exists on profiling music with wearables based on physiological changes. Most of the body of research is on music recommendation systems, some of which use physiological signals. [1] recommends music based on emotion inferred using physiological sensors on wearables, and [11] recommend music based on user personality and physiological signals. However, these both require multiple sensors which are not present in commodity wearables, such as GSR sensors. [10] combine user context (*i.e.,* activity, weather) with heart rate to predict mood and thus whether the user will enjoy the song or not. However, the authors did not consider user's preexisting music preferences, which is very important in a person's physiological response to music. The authors also used only heart rate without incorporating other physiological parameters, and found that heart rate did not significantly improve performance.

Other works recommend music based on limited physiological data, such as [7] who select music with a tempo that matches walking pace which was determined using an accelerometer. [6, 13] used heart rate sensors on wearables to detect heart rate, and then recommend music which will adjust the user's heart rate back to a target. However, these only focus on heart rate and tempo, and do not consider any other features of music. They also do not take into account the user's subjective feeling about the piece of music which is essential in delivering music that is satisfying to the user.

Therefore, no works have looked at the direct link between changes in physiology and musical features and tried to link these without considering emotion or mood. Additionally, no works have looked at predicting the feeling a song evokes in a user based on physiological changes.

5 DISCUSSION AND CONCLUSION

In this work, we presented EarTune, an analysis of the impact of music listening on changes in physiology and a user's subjective categorisation of a piece of music. We found that musical features have a significant impact on breathing rate and its rate of change. Musical features are less strongly linked to heart rate and HRV. However, these findings should be further validated on a large dataset of more participants to ensure generalisability of the findings. An aspect left unexplored is the impact of music listening on longer-term changes in heart rate and HRV, since we only analysed the last 30 seconds of the song. We also leave exploring whether there is a time lag between music listening and changes in heart rate for future work. Through our analysis, we have also found that by using musical features and changes in physiological features, it is possible to accurately predict whether a song makes a user feel calm or energised. This valuable finding forms the first step towards achieving our ultimate goal: a system which can select specific music for a user based on the user's context (*i.e.,* during exercise, a user will want an energetic song), or based on a user's mood (*i.e.,* when stressed, a user will want a calming song). We envision this system being developed onto sensor -equipped earbuds which will serve a dual purpose of playing the music to the user, and also of profiling their changes in physiology.

ACKNOWLEDGMENTS

This work was partially supported by European Research Council (ERC) project 833296 (EAR).

REFERENCES

[1] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak. 2018. Emotion Based Music Recommendation System Using Wearable Physiological Sensors. *IEEE Transactions on Consumer Electronics* 64, 2 (May 2018).

[2] Ananta Narayanan Balaji, Andrea Ferlini, Fahim Kawsar, and Alessandro Montanari. 2023. Stereo-bp: Non-invasive blood pressure sensing with earables. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*.

[3] Christopher G. Ballmann. 2021. The Influence of Music Preference on Exercise Responses and Performance: A Review. *Journal of Functional Morphology and Kinesiology* 6, 2 (April 2021), 33.

[4] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, Yang Liu, and Cecilia Mascolo. 2024. An Evaluation of Heart Rate Monitoring with In-Ear Microphones under Motion. *Pervasive and Mobile Computing* (2024), 101913.

[5] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *J. Artif. Int. Res.* 16, 1 (jun 2002), 321–357.

[6] Ming-Chuan Chiu and Li-Wei Ko. 2017. Develop a personalized intelligent music selection system based on heart rate variability and machine learning. *Multimedia Tools Appl.* 76, 14 (jul 2017), 15607–15639.

[7] Greg T. Elliott and Bill Tomlinson. 2006. PersonalSoundtrack: context-aware playlists that adapt to user pace. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI EA '06)*. Association for Computing Machinery.

[8] Xiaoran Fan, David Pearl, Richard Howard, Longfei Shangguan, and Trausti Thormundsson. 2023. APG: Audioplethysmography for Cardiac Monitoring in Hearables. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. ACM.

[9] Andrea Ferlini, Alessandro Montanari, Chulhong Min, Hongwei Li, Ugo Sassi, and Fahim Kawsar. 2022. In-Ear PPG for Vital Signs. *IEEE Pervasive Computing* 21, 1 (2022), 65–74.

[10] Jiayu Li, Zhiyu He, Yumeng Cui, Chenyang Wang, Chong Chen, Chun Yu, Min Zhang, Yiqun Liu, and Shaoping Ma. 2022. Towards Ubiquitous Personalized Music Recommendation with Smart Bracelets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (Sept. 2022), 1–34.

[11] Ruilun Liu and Xiao Hu. 2020. A Multimodal Music Recommendation System with Listeners' Personality and Physiological Signals. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. ACM, Virtual Event China, 357–360.

[12] Yang Liu, Kayla-Jade Butkow, Jake Stuchbury-Wass, Adam Pullin, Dong Ma, and Cecilia Mascolo. 2024. RespEar: Earable-Based Robust Respiratory Rate Monitoring. arXiv:2407.06901

[13] Shahriar Nirjon, Feng Zhao, Robert F. Dickerson, Qiang Li, Philip Asare, John A. Stankovic, Dezhi Hong, Ben Zhang, Xiaofan Jiang, and Guobin Shen. 2012. MusicalHeart: A *Hearty* Way of Listening to Music. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems - SenSys '12*. ACM Press, Toronto, Ontario, Canada, 43.

[14] Julia Romero, Andrea Ferlini, Dimitris Spathis, Ting Dang, Katayoun Farrahi, Fahim Kawsar, and Alessandro Montanari. 2024. OptiBreathe: An Earable-based PPG System for Continuous Respiration Rate, Breathing Phase, and Tidal Volume Monitoring. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*.

[15] Peter C. Terry and Costas I. Karageorghis. 2006. Psychophysical Effects of Music in Sport and Exercise: An Update on Theory, Research and Application. In *Proceedings of the 2006 Joint Conference of the Australian Psychological Society and New Zealand Psychological Society*, Mary Katsikitis (Ed.). Melbourne, Australia.

[16] Hoang Truong, Alessandro Montanari, and Fahim Kawsar. 2022. Non-invasive blood pressure monitoring with multi-modal in-ear sensing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.

[17] Ken Watanabe, Yuuki Ooishi, and Makio Kashino. 2017. Heart Rate Responses Induced by Acoustic Tempo and Its Interaction with Basal Heart Rate. *Scientific Reports* 7, 1 (April 2017), 43856.

[18] Shannon E. Wright, Valentin Bégel, and Caroline Palmer. 2022. *Physiological Influences of Music in Perception and Action* (first ed.). Cambridge University Press.

[19] Jianfeng Wu, Lingyan Zhang, Hongchun Yang, Chunfu Lu, Lu Jiang, and Yuyun Chen. 2022. The Effect of Music Tempo on Fatigue Perception at Different Exercise Intensities. *International Journal of Environmental Research and Public Health* 19, 7 (March 2022), 3869.