Machine Learning Prediction of Just Dance Exergame Enjoyment from Mobile Sensor Data

Joshua Audibert Elijah Gonzalez Ryan Orlando Nicholas Wong Emmanuel Agu Mark Claypool
Computer Science Department Worcester Polytechnic Institute (WPI), Worcester, MA 01609, USA
{jaudibert, egonzalez, rorlando, nwong,emmanuel,claypool}@wpi.edu

***Abstract*—Many young adults do not exercise enough, choosing instead to spend time on electronic media (e.g., smartphone, Internet). Exergames, which gamify physical activity, have been shown to be effective at increasing physical activity in an enjoyable way. For exergames to remain effective, sustained user engagement is key. However, sustaining long-term engagement in games (including exergames) is a challenging research problem – 95% of all new game players stop playing within 3 months, and 85% of new players stop after just one day. We posit that if detected early, waning player exergame enjoyment can be countered by recommending new, more enjoyable games before the player quits playing. In this paper, we investigate machine learning to predict user enjoyment of the *Just Dance* exergame by analyzing data gathered from the player's smartphone. Specifically, "ground truth" scores for the players' enjoyment obtained from the Immersive Experience Questionnaire (IEQ) E-scores are inferred from user behaviors such as increased excitement and gameplay frequency. These, in turn, are predicted by data gathered by the phone's sensors – accelerometer, gyroscope and game features. Analysis of data from a user study shows the Naive Bayes classification algorithm achieves the best results, achieving 75% accuracy for binary classification (enjoying vs. not enjoying the exergame) of enjoyment E-scores. The most predictive features were the energy in the 0.5 to 3 GHz range, windowed energy in the 0.5 to 3 Hz range and radio spectral peak using a Discrete Cosine Transform (DCT). Our results are preliminary but encouraging and we plan to improve on our results by collecting more data and utilizing state-of-the-art neural networks approaches.**

***Index Terms*—exergames, enjoyment, prediction, machine learning classification**

I. Introduction

Motivation: Physical inactivity increases the risk of many ailments including obesity and is the leading cause of death in the United States. Many young adults do not exercise enough, choosing instead to spend time on electronic media (e.g., smartphones and the Internet). Exergames, which gamify physical activity, have been shown to be effective at increasing physical activity in an enjoyable way. Several studies have confirmed that exergames can effectively mitigate sedentary habits especially in young people. For exergames to remain effective, sustained user engagement is key. However, sustaining long-term engagement in games (including exergames) is a challenging research problem – 95% of all new game players stop playing within 3 months, and 85% of new players stop after just one day. Our overall vision is to research and develop Cypress, a cyber-physical recommender system that actively monitors user enjoyment of exergames played, learns the types of games they like and recommends similar games when they get bored with their current game. This paper focuses on investigating whether users' enjoyment or boredom levels while playing the *Just Dance* exergame can be detected by analyzing the sensor data from their smartphone.

979-8-3503-1458-8/23/$31.00 ©2021 IEEE

Challenges: Proactively detecting users' enjoyment or boredom levels is a key task, which presents unique challenges including that smartphone sensor data can be noisy. Exergame enjoyment may also be confounded by other conditions, such as environment or player mood. Moreover, exergame enjoyment may manifest slightly differently in different users, which causes intra-class variability. Finally, enjoyment class boundaries such as between enjoying and not enjoying, may not be very distinct, causing fine-grained classification challenges.

Fig. 1. Screenshot from Just Dance Exergame

Approach: In this paper, we investigate whether machine learning analyses of data from mobile phone sensors on which the user is playing the game can be used to predict their enjoyment levels of the Just Dance exergame. To generate labeled data for supervised machine learning modeling, we conducted a study in which users played the Just Dance exergame and reported their enjoyment levels using the E-scores provided by the Immersive Experience Questionnaire (IEQ) [25]. Statistical features extracted from smartphone sensor data were classified using traditional machine learning classification algorithms. This paper explores binary classification of enjoyment scores into two bins (enjoying vs. not enjoying the exergame) as a first step that is actionable. Machine learning regression to estimate actual enjoyment scores will be explored in future work.

Key Findings: Out of 31 statistical features extracted, we found energy-based features extracted from the smartphone's accelerometer and gyroscope to be most predictive. The Naive Bayes classification algorithm achieved the best results, achieving 75% accuracy for binary classification of enjoyment E-scores. The most predictive features were the energy in the 0.5 to 3 GHz range, windowed energy in the 0.5 to 3 Hz range and radio spectral peak using a Discrete Cosine Transform (DCT).

II. Related work

Smartphone sensing of enjoyment: Smartphone sensing [1], [2] in which data from the smartphone's sensors (e.g., accelerometer, gyroscope) is used to infer user behaviors or emotions, is related to our goal of sensing of exergame enjoyment. Kuhn et al. [3] sensed whether party attendees were enjoying the music by classifying data from their smartphone's accelerometer to infer the energy and synchronization (from audio features) of user movements (dancing). Bao et al. [4] sensed users' facial expressions to infer how much they were enjoying YouTube videos on their smartphones. Other smartphone sensing examples include daily moods [5], human personality [6], happiness [7], emotion [8], stress [9], and sleep patterns [10]. Context-aware systems, behavior-aware computing [11] and activity recognition [12] are also related. Our work focuses on enjoyment of exergames.

Game analytics [13]: Analyzing game player data quantitatively in order to understand behavior is related to our work. Prior work includes using analytics to: drive the gamification of web applications [14], quantify user engagement in games [15], and detect player disengagement [16] to predict imminent player attrition (quitting). Predictions using these metrics were up to 90% accurate. Playnomics proposes an engagement score that combines attention (minutes played), loyalty (mean days played) and intensity (mean in-game actions per minute) [17]. Our work explores exergame session attributes (e.g., session duration and frequency) that predict imminent player attrition.

III. Background

A. Taxonomy of exergames

Figure 2 shows a taxonomy of exergames adapted from Agu et al. [18] that encompasses hardware (consoles, smartphones), locations of play (indoor, outdoor), number of players (solo, small team, large team) and genres (dance, sports/fitness, adventure, treasure hunt and pedometer gamification). Cypress targets smartphone exergames in all genres, locations and socialization levels. However, initially its focus is on smartphone exergames where sensor data is easily obtained and adequate processing power is available to deploy machine learning models.

B. Implicit indicators of enjoyment/interest:

Various indicators of exergame interest can be placed on an explicit/implicit continuum depending on how much action inference/interpretation is required. Explicit expressions of interest such as Likert-type ratings require minimal inference. Implicit indicators of interest [19] require some inference to deduce user interest. For instance, implicit user interest may be inferred from long, frequent gaming sessions or excited gameplay inferred by rapidity of user interactions measured on the user's smartphone sensors. We envision that Cypress will use implicit interest indicators since they impose minimal user burden beyond installing the client app on their smartphones. This paper's goal is to accurately predict user gameplay enjoyment by using machine learning to classify implicit interest indicators (smartphone sensor features).

C. Physiological indicators of enjoyment/interest:

Implicit interest indicators can be contrasted with affective ludology in which the exergamer is connected with probes to gather psychophysiological signals [20]–[22]. Affective gaming combines physiological signals with behavioral affect measures (e.g., facial expressions [24] and emotive body gestures) [23]. This paper does not consider physiological indicators of interest or enjoyment, but such approaches could be complimentary to our work.

D. Just Dance Gameplay

Players start by selecting the song they want to dance to. Then, while holding their smartphone, players follow the moves of the on-screen dancer and their choreographed routine. Players are judged by their animated score icons on a ranking scale for the accuracy of each of their moves in comparison to that of the on-screen dancer, and receive points. We decided upon using Just Dance as our primary exergame for collecting sensor data. Just Dance was selected because its three- to five-minute songs were an ideal length for experiments and there have been successful studies related to detecting enjoyment from people's dancing patterns [3].

E. Immersive Experience Questionnaire

To generate ground truth enjoyment labels (or E-Scores) for supervised machine learning, participants were administered the Immersive Experience Questionnaire (IEQ), developed by Jennett *et al.* [25]. The IEQ produces a scale used to subjectively measure immersion in games, initially developed with the purpose of investigating whether immersion could be defined quantitatively.

IV. Approach

A. Pilot Data Gathering Study

To establish the start-to-finish procedure of generating a prediction model for exergame enjoyment, we conducted a pilot study to see if the model could work in extreme cases where we attempted to exaggerate movements while enjoying and not enjoying gameplay, which biased the data, making Machine Learning analyses and enjoyment score prediction

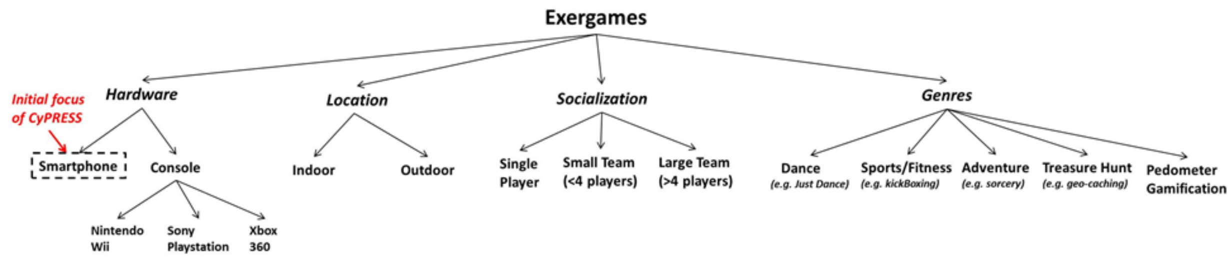

Fig. 2. A taxonomy of exergames (adapted from Agu et al. [18])

easier. In the pilot study, we participated in two Just Dance sessions using AndroSensor to gather sensor data from the smartphone. In one session we acted "excited" by exaggerating our movements. In the other session we acted "bored" by minimizing our movements. After we gathered the data, we entered it into MATLAB and performed pre-processing steps, which included segmenting the sensor data into 1-second segments. Statistical features were the extracted per-segment including features such as skewness, kurtosis, min-max difference, standard deviation, and root mean squared. The data for each session was then entered into Weka, where the first enjoyment score (E-score) prediction model was then created. We investigated whether the machine learning model could accurately predict participants' E-score from smartphone sensor data gathered during gameplay.

B. Just Dance Experiments

We gathered fifty two individuals to participate in a controlled Just Dance Now user study. Participant ages ranged from eighteen to twenty-three and included an equal number of males and females. Experiments were performed one participant at a time. The experiment procedure was as follows:

1) Users signed a consent form.[1]
2) We described the experiment procedure to participants.
3) We administered a pre-experiment survey, where participants provided demographic information such as age, gender, and weekly amount of exercise.
4) We activated the Androsensor application on the phone that the participant used. Androsensor is a smartphone app that captures sensor data including the gyroscope and accelerometer, from a user's smartphone and saves it to a `.csv` file. We set the application to capture user motion data every 10 milliseconds.
5) We had the participants play the song "Taste the Feeling". The participant stood seven to ten feet from the screen, and all proctors provided some privacy by not watching the participant play.
6) Once the song was completed, the participant answered the IEQ survey to gauge how much they enjoyed the game using strongly disagree to strongly agree questions.
7) The participant began their second Just Dance game session, in which they were instructed to pick of choice and dance.
8) The participant filled out the IEQ survey again to gauge how much they enjoyed the second song.
9) The experiment ended with a simple post-survey which explicitly asked whether the participant enjoyed the Just Dance sessions.

[1]The study was approved by our Institute Review Board (IRB).

C. Overview of our Machine Learning Approach

Figure 3 presents and overview of our proposed approach. Raw data from the smartphone of the user is segmented, features are extracted and then classified using traditional machine learning classifiers.

D. Data Pre-Processing

Various pre-processing approaches such as different data bin/window sizes, were explored as summarized in Table 4.

E. Feature Extraction

A total of 31 time, frequency, wavelet, statistical and information theory features were extracted from mobile sensor data during our experiments. Of the 31 features, 27 were calculated based on the accelerometer (x, y, and z) and 4 were calculated based on the gyroscope (x, y, and z). Depending on the song data used to create the prediction model, features correlated differently to the E-score. Here, we only expound on the most predictive features utilized in our E-Score classification work.

1) Ratio of Spectral Peak using DCT and FFT: This feature is defined as the ratio of the energies of low and high frequency bands, defined in Equation 1. Various Discrete Fourier Transform (DFT) methods were explored including the default Welch transform, Fast Fourier Transform (FFT), and using Discrete Cosine Transform (DCT).

$$\text{Ratio Spectral Peak} = \frac{\max(\text{power}_{freq})}{\text{mean}(\text{power}_{freq})} \tag{1}$$

2) Energy in Band 0.5 to 3 Hz (energy in 0.5 to 3): This feature is defined as the energy in a frequency band and describes parts of distinct frequencies in the signal. The frequency range is recommended as 0.5 Hz to 3 Hz. Typical frequency bands for specific movements can be defined. It is considered here because of its promising performance in

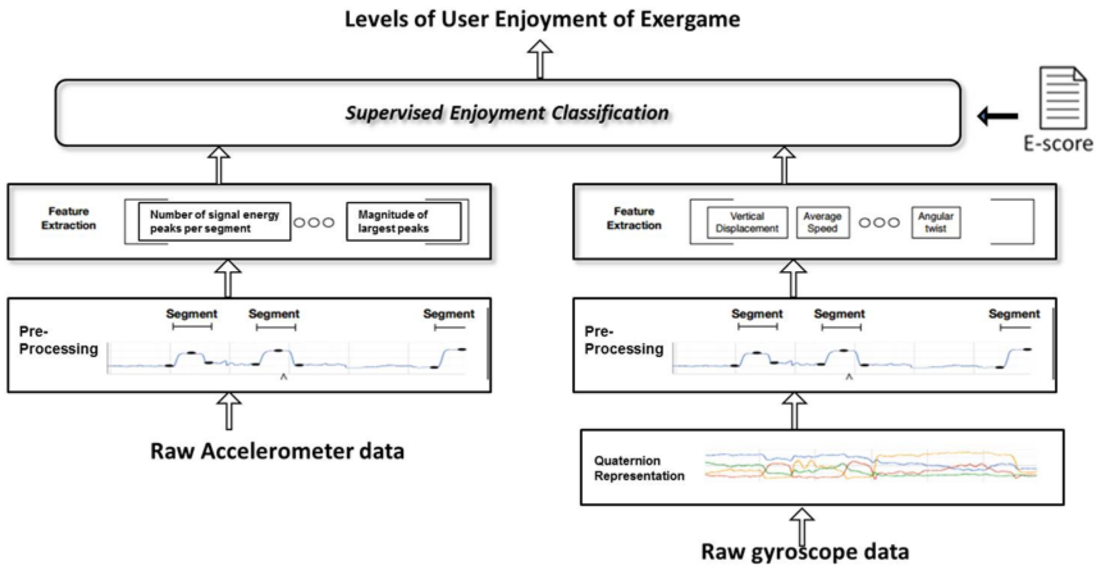

Fig. 3. Overview of our Machine Learning Pipeline for Predicting User Enjoyment of the Just Dance Exergame

Data Processing Method	Possible Values	Description
Number of Bins	2, 3, 6, 10	The number of bins (buckets) for categorized the E-scores into.
Bin range	Bins split at the median of our observed E-scores (i.e. 0-59, 60-80) , bins split at the median possible E-score (i.e. 0-40, 41-80), bins with same number of possible E-scores (i.e. 0-26, 27-53, 54-80)	The range or size of each categorized E-score bin.
Sample of song sessions	Both songs (100), first songs (49), sampling of both songs (24)	The song sessions included in the dataset.
Features	All features, statistically-significant features, Weka-chosen feature selection, combinations of other statistically significant features	The features used to create the prediction model

Fig. 4. Summary of Pre-processing Approaches Explored for Predicting User Enjoyment of the Just Dance Exergame

prior work. It was applied to gyroscope data and not the acceleromter data.

$$\text{energy in 0.5 to 3} = \int_{0.5}^{3} \text{psd}_f df \quad (2)$$

where psd_f refers to the power spectral density of frequency, and the frequency range is from 0.5 Hz to 3 Hz. In discrete signal processing as in the accelerometer data analyzed in this paper, the integral is converted into a sum.

3) Windowed Energy in Band 0.5 to 3 Hz (windowed energy in 0.5 to 3): This feature is defined as the energy in a frequency band of 5 second windows with an overlap of 2.5 seconds, where windows from complete signal sequences are averaged. This feature is considered because of its promising performance in prior work, it was applied to gyroscope data and not to accelerometer data.

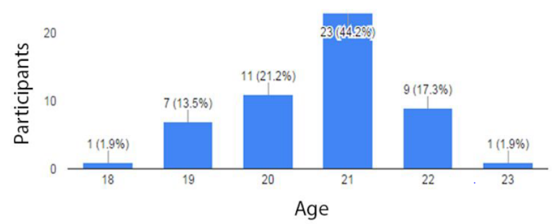

Fig. 5. Age Distribution of Participants in Just Dance Study

$$\text{windowed energy in 0.5 to 3} = \int_{0.5}^{3} \text{windowed psd}_f df \quad (3)$$

where windowed psd_f refers to the windowed power spectral density of frequency, and the frequency range is from 0.5 Hz to 3 Hz. In discrete signal processing the integral is converted into sum.

F. Machine Learning Classification

Machine Learning classifiers we explored for classifying the smartphone sensor data into target E-scores included Random Forest, J48, SMO and Naive Bayes. 10-fold Cross-validation approach was utilized with a 90-10 training-test set split. Model training was done in the Weka Machine Learning library.

V. Results

A. Participant Demographics

The majority of participants (30) were male, while 21 participants were female, and 1 participant preferred not to declare their gender.

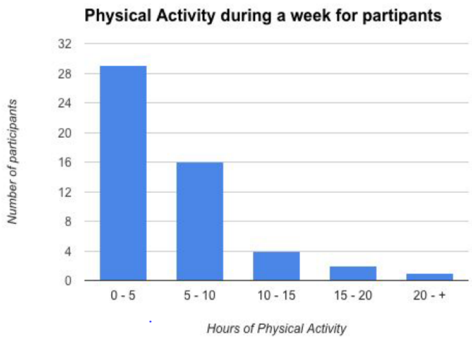

Fig. 6. Average Number of Hours of Physical Activity Per Week for Participants

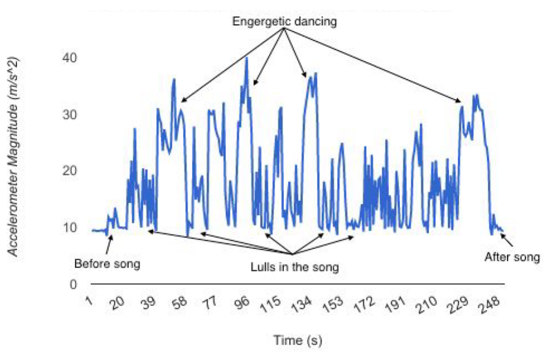

Fig. 7. Sample Accelerometer Sensor Data Magnitude for Just Dance over Time

Figure 5 shows the age distribution of participants in the Just Dance Study. Most participants were students at our university, hence, ages fell between the ages of 18 and 23 years, with 21 years being the most common age.

Figure 6 shows the number of hours of physical activity for participants per week. Most of the participants exercised between zero to five hours of physical activity per week.

B. Sample Sensor Data During GamePlay

Figure 7 shows sample sensor data collected during gameplay for Just Dance showing typical values for energetic dancing, and values before and after the song as well as lulls in the song.

C. Boxplots of E-Scores

Figure 8 shows gender-specific box plots of E-scores for Just Dance. There was no statistically significant difference between males and females at a 0.05 significance (p value 0.075).

Figure 9 shows boxplots of whether subjects enjoyed game vs. E-score. Enjoyment was determined using a question on the survey explicitly asking them whether they enjoyed the game or not. As shown by the E-scores from the second song, which was chosen by the participants, there is a sharp contrast between the two distributions. This suggests that the E-score calculation was a good gauge of whether a person enjoyed the game or not (p value < 0.0001) with a correlation of 0.68.

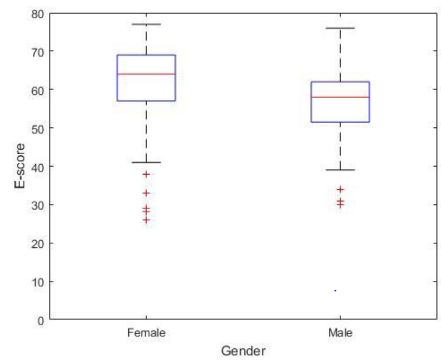

Fig. 8. Gender-specific Box Plots of E-scores

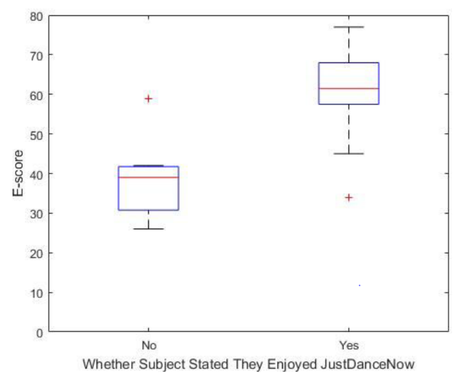

Fig. 9. Box Plots of Participant Enjoyment vs. E-score

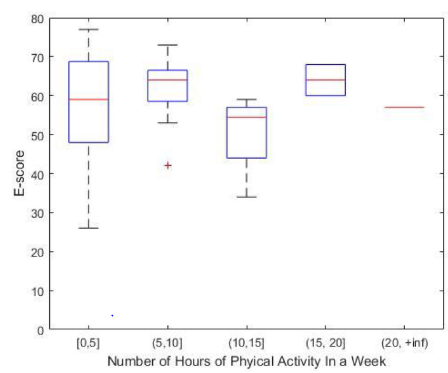

Fig. 10. Box Plots of Number of Hours of Physical Activity per Week vs. E-scores

Figure 10 shows boxplots of the number of hours of physical activity per week vs. E-scores. The goal here was to investigate whether players with different physical activity levels per week were more or less likely to enjoy exergames. Participants' E-scores were compared with their average number of hours of physical exercise during a week broken into 5-hour buckets. There were a few participants with significant amounts of per-week exercise, but the majority of participants exercised between zero and ten hours a week. There is no apparent relationship between the amount that the participant exercised, and the amount they they enjoyed the game.

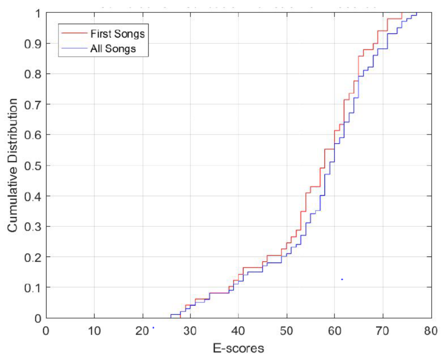

Fig. 11. Cumulative Distribution Function (CDF) of E-scores

Figure 11 shows the Cumulative Distribution Function (CDF) of E-scores. The median E-score overall was 59 for both song data, while the median E-score for the first song was 58. The average for both song data was approximately 57.3 while the average for the first song was approximately 55.9. There is little difference between these distributions, although the E-scores for the both songs are consistently slightly higher than for the first songs, indicating that the second songs (the ones participants chose) were enjoyed more.

D. Correlation between Feature Values and E-Score

Figure 12 shows the correlation between Features and E-Score (2 songs) in the Just Dance exergame. At the top are the features that are statistically significant (i.e., p values less than 0.05), while the rest are sorted by decreasing absolute correlation coefficient. The three correlated features are all based on maximum spectral density (power over frequency). These features could be most correlated (albeit weakly correlated) to E-score because they represent the amount of energy used in the phone's movements – i.e., a person expending more energy while playing could be enjoying it more.

E. Machine Learning E-Score Classification Results

Figure 13 shows machine learning classification results for the Just Dance Exergame. The J48, SMO, and Naive Bayes classifiers achieved the best prediction performance. Random Forest serves as a consistent baseline to compare against, and the overall best prediction model with 75% accuracy was generated from running Naive Bayes on the dataset consisting of both songs, 2 bins split between 0 to 40 and 41 to 80 of equal size (12 songs per bin) and the energy in 0.5 to 3, windowed energy in 0.5 to 3, and radioSpectralPeak_DCT features.

The confusion matrix for the best performing model is shown in Figure 14. Most values lie on the leading diagonal which indicates that there was not much confusion between the model's prediction of target classes.

VI. Conclusion and Future Work

This study explored the prediction of the Just Dance exergame enjoyment from accelerometer and gyroscope sensor data gathered from the user's smartphone as they played the game. Machine learning modeling utilized standard steps including data pre-processing, feature extraction and selection, and machine learning classification. The Naive Bayes classification algorithm achieved the best results, achieving 75% accuracy for binary classification of enjoyment E-scores. The most predictive features were the energy in the 0.5 to 3 Gz range, windowed energy in the 0.5 to 3 Hz range and radio spectral peak using a Discrete Cosine Transform (DCT). Our results are preliminary but encouraging.

While encouraging, our work has limitations that we plan to address future work. In our experiments, we restricted how players select songs, which may have affected our results. In the future, we will investigate other song selection rules. We also plan to collect more data, which would facilitate the use of deep learning approaches that are generally more accurate when adequate data is available. We would also like to deploy our machine learning enjoyment predictors and compare them to human raters. Finally, we would like to predict enjoyment of other exergame types.

References

[1] N.D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A.T. Campbell, "A Survey of Mobile Phone Sensing", IEEE Communications Magazine, Vol. 48, No. 9, pages 140-150, 2010.

[2] W. Khan, Y. Xiang, M. Aalsalem, Q. Arshad, "Mobile Phone Sensing Systems: A Survey", IEEE Communications Surveys & Tutorials, Vol. PP, No. 99, pages 1-26, 2013.

[3] M. Kuhn, R. Wattenhofer, M. Wirz, M. Fluckiger, G. Troster, "Sensing Dance Engagement for Collaborative Music Control," in Processings of the 15th Annual International symposium on Wearable Computers (ISWC), pages 51-54, June 2011.

[4] X. Bao, S. Fan, A. Varshavsky, K. Li and R. Choudhury, "Your Reactions Suggest You Liked the Movie: Automatic Content Rating via Reaction Sensing", in Proceedings of Ubicomp, pages 197-206, 2013.

[5] R. LiKamWa, Y. Liu, N.D. Lane, and L. Zhong. "MoodScope: Building a Mood Sensor from Smartphone Usage Patterns", In Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys). ACM, New York, NY, USA, pages 389-402, 2013.

[6] G. Chittaranjan, J. Blom and D. Gatica-Perez, "Who's Who with Big Five: Analyzing and Classifying Personality Traits with Smartphones", in Proceedings of the IEEE International Symposium on Wearable Computing (ISWC), pages 29-36, 2011.

[7] A. Bogomolov, B. Lepri, and F. Pianesi, "Happiness Recognition from Mobile Phone Data", In Proceedings of the International Conference on Social Computing (SocialCom), 2013

[8] K.K. Rachuri, M. Musolesi, C. Mascolo, P.J. Rentfrow, C. Longworth, and A. Aucinas, "EmotionSense: a Mobile Phones-based Adaptive Platform for Experimental Social Psychology Research", In Proceedings of the 12th ACM International Conference on Ubiquitous computing (UbiComp), ACM, New York, NY, USA, pages 281-290, 2010.

[9] H. Lu, D. Frauendorfer, M. Rabbi, M.S. Mast, G.T. Chittaranjan, A.T. Campbell, D. Gatica-Perez, and T. Choudhury, "StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones", In Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp), ACM, New York, NY, USA, pages 351-360, 2012.

[10] Z. Chen, M. Lin, F. Chen, N. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, A. Campbell, "Unobtrusive Sleep Monitoring Using Smartphones," in Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pages 145-152, May 2013.

[11] J. Favela, "Behavior-Aware Computing: Applications and Challenges", in IEEE Pervasive Computing, pages 14-17, July-September 2013.

[12] J. Kwapisz, G. Weiss, and S. Moore, "Activity Recognition Using Cell Phone Accelerometers", ACM SIGKDD Explorations Newsletter, Vol. 12, No. 2, pages 74-82, Mar. 2011.

Feature Names	Features Coef	Features Coef(abs)	P-value	Predictable (p<0.05)
radioSpectralPeak_DCT	0.3095	0.3095	0.0017	1
radioSpectralPeak	0.2957	0.2957	0.0028	1
radioSpectralPeak_FFT	0.2622	0.2622	0.0084	1
spectralCentroid	-0.1694	0.1694	0.0921	0
averageStepTime	-0.1420	0.1420	0.1589	0
peakFreq	0.1394	0.1394	0.1667	0
coef of var of stepTime	0.1271	0.1271	0.2076	0
numSteps	0.1187	0.1187	0.2395	0
averageCadence	0.0981	0.0981	0.3316	0
averageStepLength	0.0981	0.0981	0.3316	0
std	0.0980	0.0980	0.3319	0
rms	0.0980	0.0980	0.3319	0
YZEllipse	0.0919	0.0919	0.3631	0
kurtosis	-0.0884	0.0884	0.3820	0
harmonic ratio	0.0868	0.0868	0.3906	0
energy in _5 to 3	0.0826	0.0826	0.4141	0
wavelet entropy	0.0802	0.0802	0.4275	0
wavelet band	-0.0800	0.0800	0.4289	0
averagePower	0.0776	0.0776	0.4427	0
windowed energy in _5 to 3	0.0769	0.0769	0.4468	0
entropy rate	0.0752	0.0752	0.4572	0
gaitVelocity	0.0738	0.0738	0.4656	0
cross correlation	-0.0736	0.0736	0.4670	0
snr	0.0725	0.0725	0.4732	0
thd	0.0699	0.0699	0.4898	0
XZEllipse	0.0561	0.0561	0.5795	0
calculatedVolume	-0.0503	0.0503	0.6194	0
skewness	-0.0494	0.0494	0.6255	0
minMaxDiff	0.0398	0.0398	0.6944	0
bandwidth	-0.0357	0.0357	0.7246	0
XYEllipse	0.0184	0.0184	0.8556	0

Fig. 12. Correlation between Features and E-Score (2 songs) in Just Dance Exergame

[13] Editors: M. El-Nasr and A. Drachen, Game Analytics: Maximizing the Value of Player Data, Springer publishers, 2013.

[14] Y. Xu¸ "Literature Review on Web Application Gamication and Analytics", CSDL Technical Report 11-05, Department of Information and Computer Sciences, University of Hawaii, Honolulu, HI, August 2011.

[15] L. Petersen, "Gabe Zichermann on Gamification, Fun and Metrics", E-Consultancy, LLC, 2011. http://econsultancy.com/us/blog/7283-q-a-with-gabe-zichermann-on-gamication-fun-and-metrics

[16] B. Harrison, and D.L. Roberts, "Analytics-driven Dynamic Game Adaption for Player Retention in Scrabble", Computational Intelligence in Games, 2013.

[17] "Player Engagement Study", Q3 July-September, 2012.

[18] E. Agu, A. Chokchaisiripakdee, N. Sujumnong, L. Krachonkitkosol, B. Tulu, "Making Exergames Appealing: An Assessment of Commercial Exergames", in Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice, IGI Global Publishers.

[19] W.P., J.H. Lin, and J. Crouse, "Is Playing Exergames Really Exercising? A Meta-Analysis of Energy Expenditure in Active Video Games", Cyberpsychology, Behavior, and Social Networking, Vol. 14, No. 11, 2011.

[20] L. Nacke and C. Lindley, "Affective Ludology, Flow and Immersion in a First Person Shooter: Measurement of Player Experience", The Journal of the Canadian Game Studies Association, Vol. 3, No. 5, 2009.

[21] L. Nacke and C. Lindley, "Flow and Immersion in First-Person Shooters: Measuring the Player's Gameplay Experience", in Proceedings of FuturePlay, 2008.

[22] H. Martinez, M. Garbarino and G. Yannakakis, "Generic Physiological Features as Predictors of Player Experience", Affective Computing and Intelligent Interaction, Vol. 6974, pages 267-276, 2011.

[23] T. Christy and L. Kuncheva, "Technological Advancements in Affective Gaming: a History Survey", International Journal on Computing (JoC), Vol. 3, No. 4, Apr. 2014.

[24] S. Larsson, "Facial Expression Recognition for Distinctive Game Events and Personality Profiles", Master Thesis - Information Studies: Game Studies, University of Amsterdam, 2014.

[25] Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T. and Walton, A. "Measuring and Defining the Experience of Immersion in Games," International Journal of Human-Computer Studies, 66(9), pp.641-661, 2008.

Both songs, 2 bins (0-59, 60-80), even distribution (50 songs per bin), 3 features (radioSpectralPeak, radioSpectralPeak_FFT, radioSpectralPeak_DCT)				
	Random Forest	**J48**	**SMO**	Naïve Bayes
Correctly Classified Instances	52.0%	**65.0%**	**65.0%**	62.0%
Kappa Statistic	0.040	**0.293**	**0.292**	0.232
ROC Area Weighted Average	0.546	**0.613**	**0.644**	0.599
First Songs, 2 bins, even distribution (24-25 songs per bins), all features				
	Random Forest	J48	SMO	**Naïve Bayes**
Correctly Classified Instances	59.2%	53.1%	65.3%	**69.4%**
Kappa Statistic	0.183	0.059	0.303	**0.389**
ROC Area Weighted Average	0.648	0.497	0.651	**0.677**
Both songs, 2 Bins (0-40, 41-80), even distribution (12 songs per bin), 3 features (energy in 0.5 to 3, windowed energy in 0.5 to 3, and radioSpectralPeak_DCT). *Note that we added radioSpectralPeak_DCT because it had a P-value of 0.0502, which is close enough to having a P-value below 0.05.				
	Random Forest	J48	SMO	**Naïve Bayes**
Correctly Classified Instances	66.7%	70.8%	66.7%	**75.0%**
Kappa Statistic	0.333	0.417	0.333	**0.500**

Fig. 13. Machine Learning Classification Results for Just Dance Exergame

		Predicted class	
		a	**b**
Actual class	**a = 40**	11	1
	b = 80	5	7

Fig. 14. Confusion Matrix for the Best Performing Model using Naive Bayes Classifier