Proceedings Paper:
Irrgang, Melanie, Steffens, Jochen and Egermann, Hauke Wolfgang orcid.org/0000-0001-7014-7989 (2018) Smartphone-Assessed Movement Predicts Music Properties : Towards Integrating Embodied Music Cognition into Music Recommender Services via Accelerometer. In: Proceedings of the 5th International Conference on Movement and Computing, MOCO 2018. ACM Proceedings . ACM Digital Library .

https://doi.org/10.1145/3212721.3212852

Smartphone-Assessed Movement Predicts Music Properties

Towards Integrating Embodied Music Cognition into Music Recommender Services via Accelerometer Motion Data

Melanie Irrgang
Technische Universität Berlin
Berlin, Germany
melanie.irrgang@tu-berlin.de

Jochen Steffens
Technische Universität Berlin
Berlin, Germany
jochen.steffens@tu-berlin.de

Hauke Egermann
University of York
York, United Kingdom
hauke.egermann@york.ac.uk

ABSTRACT

Numerous studies have shown a close relationship between movement and music [7], [17], [11], [14], [16], [3], [8]. That is why Leman calls for new mediation technologies to query music in a corporeal way [9]. Thus, the goal of the presented study was to explore how movement captured by smartphone accelerometer data can be related to musical properties. Participants (N = 23, mean age = 34.6 yrs, SD = 13.7 yrs, 13 females, 10 males) moved a smartphone to 15 musical stimuli of 20s length presented in random order. Motion features related to tempo, smoothness, size, regularity, and direction were extracted from accelerometer data to predict the musical qualities “rhythmicity", “pitch level + range" and "complexity“ assessed by three music experts. Motion features selected by a 20-fold lasso predicted the musical properties to the following degrees “rhythmicity" (R^2 : .47), pitch level and range (R^2 : .03) and complexity (R^2 : .10). As a consequence, we conclude that music properties can be predicted from the movement it evoked, and that an embodied approach to Music Information Retrieval is feasible.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**; **Gestural input**; • **Information systems → Music retrieval**; • **Computing methodologies →** *Cognitive science*;

KEYWORDS

Embodied cognition and movement, Music Information Retrieval, Accelerometer, Smartphone

ACM Reference Format:
Melanie Irrgang, Jochen Steffens, and Hauke Egermann. 2018. Smartphone-Assessed Movement Predicts Music Properties: Towards Integrating Embodied Music Cognition into Music Recommender Services via Accelerometer Motion Data. In *MOCO: 5th International Conference on Movement and Computing, June 28–30, 2018, Genoa, Italy*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3212721.3212852

ACM ISBN 978-1-4503-6504-8/18/06…$15.00
https://doi.org/10.1145/3212721.3212852

1 INTRODUCTION

Listeners of music are assumed to internally mimick body movements when perceiving music and its emotional or intended qualities [7], [17]. As a consequence, the body and movements are an active contributor in meaning formation [11].

Movement in music has been shown to be synchronized to rhythmic (e.g., meter and tempo; [15], [3]) as well as tonal or expressive qualities of music (e.g., melody, timbre, sound intensity; [2], [5],[10], [16], [12]). Sievers, Polansky, Casey, and Wheatley asked participants from the U.S. and Cambodia to adjust the features rate, jitter, consonance/smoothness, step size and direction for five different emotions[14]. For one group, the adjustment of features led to different movements and appearances of a bouncing ball. For the second group, adjusting features changed the melodic features and expression of a piano piece. The settings used for different emotions were highly similar for motion and music in both cultures. Accordingly, the authors conclude that emotion expression in music and movement seem to be based on the same universal features. Amelynck, Grachten, van Noorden, and Leman [1] investigated whether motion features can be used to predict experienced emotion of listeners. They asked participants to perform arm gestures while holding a Wii remote controller in order to describe their music listening experience. Afterwards, the emotional qualities of musical excerpts presented were rated on the dimensions of valence and arousal. Using motion features recorded with the Wii controller generated fairly good predictions for the arousal dimension, but performed less accurate predictions for the valence dimension. The authors argue that this might be due to people rating sad music as pleasant [4], and conclude that the Circumplex Model [13] might be unsuitable to be used with musical emotions.

Previous work conducted in our lab followed the model of Amelynck et al., but applied a different emotion model (Geneva Emotion Music Scales, [18]) and used smartphones and their inherent motion sensors to capture movements [8]. However, while a subset of the GEMS-9 was predicted to a meaningful degree by the motion features, participants did not provide enough variance in the data for some other dimensions like *transcendence*. Due to the problems associated with self-report measures of perceived emotion, our idea was to test the direct connection between musical properties and movement. Thus, the goal of the presented study was to explore how gestures captured by smartphone-assessed accelerometer data can predict musical qualities presented to listeners. Perspectively, these findings will contribute to develop corporeal querying of music databases in the field of Music Information Retrieval as called for by Leman [9].

2 METHOD

2.1 Stimulus selection

In the course of a pre-study, a set of 31 music stimuli was rated by three experts using 10 characteristics also used in Gomez and Danuser [6, p. 379] plus three new characteristics added by the authors:

- rhythm (1 = vague, 10 = outstanding)
- tempo (1 = slow, 10 = fast)
- accentuation (1 = light, 10 = marcato)
- articulation (1 = staccato, 10 = legato)
- melodic direction (1 = descending, 10 = ascending)
- pitch level (1 = low, 10 = high)
- pitch range (1 = narrow, 10 = wide)
- mode (1 = minor, 10 = major)
- complexity (1 = simple, 10 = complex)
- consonance (1 = dissonant, 10 = consonant) item backbeat (1 = vague, 10 = outstanding)
- downbeat (1 = vague, 10 = outstanding)
- syncopation (1 = accent on beat, 10 = accent on off-beat)
- beat position of bass/snare (1 = laid back, 10 = up front)

Stimuli consisted of 15 music excerpts à 20 seconds and were selected from the 31 original pieces so that their musical characteristics would not change (much) over time. In order to select 15 stimuli representing all musical characteristics in the larger set of 31 pieces, we computed k-means clustering for 15 clusters from the expert ratings and selected one stimulus from each cluster. Table 1 depicts the final list of samples which were presented in random order in the main study.

2.2 Main Study

Twenty-three persons with a mean age of 34.6 yrs (SD = 13.7 yrs, 13 females, 10 males) participated in the main study. In the beginning, they were asked for their written consent to participate. Their motions were tracked by an Optitrack[1] motion capture system in order to have a reference measure for the accelerometer data (not shown here). Participants wore a motion capture suit equipped with 37 markers and holding the smartphone equipped with another three marker points as rigid body. Markers were tracked by eight cameras. A video camera recorded the participants' movements in order to disambiguate occluded marker points and to record short interviews about the music after each excerpt. Music was presented to participants via loudspeakers. An Android App was developed to capture motion and to define a random order for the music stimuli. The App was controlled by the investigator via remote access with AirDroid[2]. The whole study was conducted on a Samsung Galaxy S6. A sling around the wrist served as a safety measure for the phone not to be slipped. Except for one person, participants held the phone in the right hand since they were right handed. Note that the phone was kept ergonomically (front surface facing to the torso's side). Figure 1 and 2 show the smartphone's position relative to the body and the experimental setting of the Motion Capture. In the beginning, participants chose one of their own songs to warm up and get familiar with the study's procedure. They were

[1]www.optitrack.com
[2]www.airdroid.com

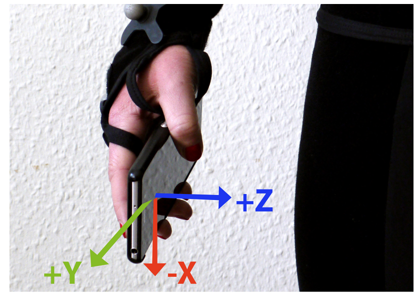

Figure 1: Position of Smartphone in Hand

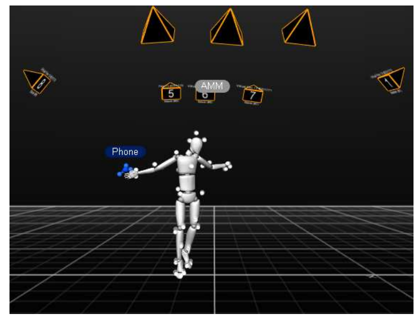

Figure 2: Skeleton View of Participant

instructed as follows: "Move the smartphone to the music. You can move the rest of the body intuitively along with it but keep in mind that the characteristic motion must be captured by the phone. Please stay in the delineated area of 2x2 metres". For every music excerpt, the procedure was as follows:

(1) multiple test listening to the music excerpt in order to develop movement strategy
(2) accelerometer recording of the movement during stimulus presentation
(3) short interview about the music excerpt
   (a) Which musical quality did influence your movement most?
   (b) Did you have to decide between different suitable movement patterns or not? If so, can you perform one alternative pattern for the camera?
   (c) How much did you like moving/dancing to the song?
   (d) How much did you like the song?
   (e) How well did you know the song before?

Subsequent to the study's main part, participants were asked to fill out a questionnaire on experience in music and movement/dance etc.

2.3 Data Analysis

As ratings of the single music properties were highly correlated, we calculated a principal component analysis for those properties for which the inter-rater reliability was acceptable (inter-rater correlation r > .50). A scree-plot indicated a 3-factor solution. These factors were labeled *rhythmicity*, *pitch level + range* and *complexity* (see Table 2). Using Matlab, spectral and temporal motion features related to tempo, size, regularity and smoothness were extracted from

Table 1: List of Music Stimuli.

Artist	Title	Time Slot
Alicia Keys feat. Nicki Minaj	Girl on Fire (Inferno Version)	00:00-00:20
Marcus Miller	Detroit	00:00-00:21
Sia	Chandelier	00:30-00:55
Alicia Keys	You don't know my name	00:33-00:57
Fever Ray	Dry and Dusty	01:20-01:42
David Bowie	Aladdin Sane	02:00-02:30
Stevie Wonder	Another Star	02:21-03:00
Andy Allo	People Pleaser	00:18-00:40
Michael Jackson	Bad	00:00-00:20
Röyksopp	Monument (The Inevitable End Version)	00:40-01:02
Igor Stravinsky	Le Sacre du Printemps/Part1	00:20-00:40
Chris Garneau	The Leaving Song	00:00-00:22
Jherek Bischoff & Amanda Palmer feat. Neil Gaiman	Space Oddity	01:36-02:00
David Bowie feat. Tina Turner	Tonight	00:18-00:48
Florence + the machine	Cosmic Love	00:20-00:43

smartphone-generated accelerometer[3] data and averaged over time as in [8]. Spectral features (e.g. *max frequency in Hz* or *magnitude of the most dominant frequency relative to the median magnitude of all frequency peaks*) were computed by applying a FFT to the first principle component of the PCA transformed accelerometer data. Temporal features (e.g. *median peak amplitude*, *midcrossings*, *rise and fall time*) were extracted from non-transformed acceleration. Subsequently, a within subject-centralization of motion features was applied to account for inter-individual differences in feature use. We selected features by the least absolute shrinkage and selection operator (LASSO) with nfolds = 20 for all three factors, i.e. *rhythmicity*, *pitch level + range* and *complexity*.[4]

3 RESULTS

Rhythmicity was predicted best by motion features, followed by *complexity* and *pitch level + range* (see Table 3). In particular, *rhythmicity* was predicted by *irregular* movement in the direction of the first principle component (max_freq_mag_rel) and regular movement in x direction (std_dist_midcrossings_x), *large* movement in x (median_peak_x) and y (median_peak_y) direction, and sharp movement in z direction (median_rise_z). *Pitch level + range* was characterized by regular movement in x direction (std_dist-_midcrossings_x), and small movement in y (median_peak_y) direction. Equal to *rhythmicity*, *complex* music was predicted by irregular movement along the first principle component. Furthermore, tempo (fast) and irregularly large movement in x (std_peak_x) direction were significant features for the prediction of the *complex* music property.

4 DISCUSSION AND CONCLUSIONS

The presented study aimed at predicting properties of music by the movements it evoked. The better prediction results for *rhythmicity* suggest that qualities related to *rhythmic entrainment* are easier to model. Rhythmic entrainment works as a rather spontaneous and subconscious synchronization to the beat of the music. Therefore, rhythm-related properties of music might evoke more similar movement among participants. The musical properties *complexity* and *pitch level + range* are stronger related to tonal and expressive qualities of music, and hence describe more complex, conscious processes of music perception that also include empathy. Thus, these properties of music might evoke less universal but more individual movement patterns among participants according to their music preferences, and hence are more difficult to predict. Furthermore, we assume that rhythmicity is more related to *arousal* whereas expressive properties are more related to *valence* which has been shown to me more difficult to predict by movement (cf. Amelynck et al. [2012] or Irrgang and Egermann [2016]). Another restriction to predict complex and tonal properties stems from the fact that we did not model all changes over time in accelerometer motion data and music features since both were averaged over time. In particular, qualities related to non-periodic "gestalt", as are typical for expressive properties, might have to be assessed via time series analyses. The presented study showed that music properties can be predicted from the movement it evoked. Thus, the results confirm the feasibility of an embodied approach to Music Information Retrieval.

ACKNOWLEDGMENTS

The authors would like to thank Balthasar Effmert and Dmitry Grigoriev for their assistance in conducting the study.

REFERENCES

[1] D. Amelynck, M. Grachten, L. van Noorden, and M. Leman. 2012. Toward E-Motion-Based Music Retrieval a Study of Affective Gesture Recognition. *IEEE Transactions on Affective Computing* 3, 2 (April 2012), 250–259. https://doi.org/{10.1109/T-AFFC.2011.39}

[2] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toiviainen. 2013. Relationships between Perceived Emotions in Music and Music-Induced Movement.

[3] corrected linear acceleration as described for Sensor.TYPE_LINEAR_ACCELERATION in: https://developer.android.com/guide/topics/sensors/sensors_motion

[4] For transparency data and code will be provided in the following repository: https://github.com/mirrgang/motion2music

Table 2: Factor loadings according to scree-test.

original musical property	rhythmicity	pitch level + range	complexity
rhythm	**.912**	-.096	.226
backbeat	**.718**	.133	-.605
tempo	**.457**	.168	**.631**
accentuation	**.918**	.110	.175
articulation	**-.888**	.219	-.127
pitch level	-.138	**.914**	.088
pitch range	.044	**.895**	.252
complexity	.169	.380	**.829**

Table 3: Results for the fitted linear regression models.

	name	category	β	b (Sig.)
rhythmicity ($R^2 = .47$)***				
magnitude of most dominant frequency relative to median magnitude of all frequency peaks	max_freq_mag_rel	regularity	-1.7	-4.5***
std of midcrossing's distribution in x	std_dist_midcrossings_x	regularity	-0.9	-2.4*
median peak amplitude in x	median_peak_x	size	1.7	3.0*
median peak amplitude in y	median_peak_y	size	2.0	3.5**
median rise time in z	median_rise_z	smoothness	-1.1	-3.2*
pitch level and range ($R^2 = .03$)*				
std of midcrossing's distribution in x	std_dist_midcrossings_x	regularity	-0.7	-2.9*
median peak amplitude in y	median_peak_y	size	-0.5	-2.1*
complexity ($R^2 = .10$)***				
magnitude of most dominant frequency relative to median magnitude of all frequency peaks	max_freq_mag_rel	regularity	-0.7	-2.5*
median of distances between midcrosses in x	median_dist_midcrossings_x	tempo	-0.6	-2.3*
std of amplitude distribution x	std_peak_x	regularity	0.7	2.7*

*** p-value < .0001, ** p-value < .001, * p-value < 0.05

Music Perception. *Frontiers in Human Neuroscience* 30 (2013), 517–533. Issue 5. https://doi.org/10.1525/MP.2013.30.5.517

[3] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toiviainen. 2014. Hunting for the beat in the body: on period and phase locking in music-induced movement. *Frontiers in Human Neuroscience* 8 (2014), 250–259. https://doi.org/10.3389/fnhum.2014.00903

[4] S. Garrido and E. Schubert. 2011. Negative Emotion in Music: What is the Attraction? A Qualitative Study. *Empirical Musicology Review* 6, 4 (2011), 214–230.

[5] L. B. Giordano, H. Egermann, and R. Bresin. 2014. The Production and Perception of Emotionally Expressive Walking Sounds: Similarities Between Musical Performance and Everyday Motor Activity. *PLoS ONE* 9 (2014).

[6] P. Gomez and B. Danuser. 2007. Relationships Between Musical Structure and Psychophysiological Measures of Emotion. *Emotion* 7, 2 (2007), 377–387.

[7] J. A. Grahn and J. B. Rowe. 2009. Feeling the beat: premotor and striatal interactions in musicians and non-musicians during bear perception. *Journal of Neuroscience* 29 (2009), 7540–7548. Issue 23. https://doi.org/10.1523/JNEUROSCI.2018-08

[8] M. Irrgang and H. Egermann. 2016. From Motion to Emotion: Accelerometer Data Predict Subjective Experience of Music. *PLoS ONE* 11(7) (2016). Issue 7. https://doi.org/10.1371/journal.pone.0154360

[9] M. Leman. 2007. *Embodied Music Cognition and Mediation Technology.* MIT Press, London.

[10] M. Leman and P.-J. Maes. 2014. The Role of Embodiment in the Perception of Music. *Empirical Musicology Review* 9, 3-4 (2014), 236–246.

[11] P.-J. Maes, M. Leman, C. Palmer, and M. Wanderley. 2014. Action-Based effects on Music Perception. *frontiers in Psychology* 4 (2014).

[12] Pieter-Jan Maes, E. Van Dyck, M. Lesaffre, M. Leman, and M. P. Kroonenberg. 2014. The Coupling of Action and Perception in Musical Meaning Formation. Vol. 32. University of California Press, 67–84.

[13] A. J. Russell. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178. http://psycnet.apa.org/psycinfo/1981-25062-001

[14] B. Sievers, L. Polansky, M. Casey, and T. Wheatley. 2013. Music and Movement Share a Dynamic Structure that Supports Universal Expression of Emotion. *PNAS* 110, 1 (2013), 70–75.

[15] F. Styns, L. Van Noorden, D. Moelants, and M. Leman. 2007. Walking on Music. *Human Movement Science* 26 (2007), 297–305. Issue 5.

[16] E. Van Dyck, D. Moelants, M. Demey, P. Coussement, A. Deweppe, and M. Leman. 2014. The impact of the Bass Drum on Body Movement in Spontaneous Dance. In *Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC11).*

[17] R. J. Zatorre, J. L. Chen, and V.B. Penhune. 2007. When the brain plays music: audiotory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8 (2007), 547–558. https://doi.org/10.1038/nrn2152

[18] M. Zentner, D. Grandjean, and K. R. Scherer. 2008. Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion (Washington, D.C.)* 8, 4 (2008), 494–521. https://doi.org/0.1037/1528-3542.8.4.494