WARWICK

ECONOMICS

Measuring national happiness with music

Emmanouil Benetos, Alessandro Ragano, Daniel Sgroi & Anthony Tuckwell

(This paper also appears as CAGE Discussion paper 537)

January 2021 No: 1326

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Measuring National Happiness with Music*

Emmanouil Benetos,[†] Alessandro Ragano,[‡] Daniel Sgroi,[§] Anthony Tuckwell[¶]

Abstract

We propose a new measure for national happiness based on the emotional content of a country's most popular songs. Using machine learning to detect the valence of the UK's chart-topping song of each year since the 1970s, we find that it reliably predicts the leading survey-based measure of life satisfaction. Moreover, we find that music valence is better able to predict life satisfaction than a recently-proposed measure of happiness based on the valence of words in books (Hills et al., 2019). Our results have implications for the role of music in society, and at the same time validate a new use of music as a measure of public sentiment. *JEL codes: N30, Z11, Z13*

Keywords: subjective wellbeing, life satisfaction, national happiness, music information retrieval, machine learning.

*We are grateful for helpful discussions with Mahendra Mahey, Filipe Bento, David Beavan, Mark Sandler and David De Roure. Funding was provided by the ESRC CAGE Centre (grant reference RES-626-28-0001). All authors are grateful for additional support from the Alan Turing Institute, including a scholarship provided to Ragano and Tuckwell as part of their PhD Enrichment Programme (EPSRC grant: EP/N510129/1). Note that authors are listed in alphabetical order. Corresponding author: Daniel Sgroi: daniel.sgroi@warwick.ac.uk; University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.

[†]Queen Mary University of London and The Alan Turing Institute.

[‡]University College Dublin.

[§]University of Warwick, ESRC CAGE Centre and IZA Bonn.

[¶]University of Warwick and ESRC CAGE Centre.

1 Introduction

One of the most fundamental human concerns, happiness, has also become a key focus of policymakers, who have recognised its positive effects for health and productivity as well as individual quality of life. Measuring happiness at the macro level is therefore an important area of research, with the most popular method in recent decades being surveys of subjective wellbeing. Recently, in response to historical gaps in such survey data, a new measure was developed which utilised the psychological valence of the words in books (Hills et al., 2019). Like language, music can also encode emotional information: it has been described as a "language of the emotions" (Cooke, 1959), with studies demonstrating that different people can recognise the same patterns of emotion in a song (Juslin, 2013). Moreover, it is the emotional experience that music offers that primarily motivates individuals to listen to it (Juslin and Laukka, 2004). This paper demonstrates that the valence of a country's most popular songs (extracted using techniques from music information retrieval) can also be used to measure national happiness and can be more robust than a text-based measure.

Our focus for this study is the UK, for which we constructed a Music Valence Index (MVI) using the valence of the most popular song of each year since the 1970s (according to the official music charts). This valence was predicted by a machine learning model (Support Vector Regression) that had been trained to learn audio features associated with high/low valence according to a separate set of songs that had been annotated by human subjects (Soleymani et al., 2013). We find that the MVI displays a significant degree of similarity with the survey-based measure of life satisfaction. First, the MVI appears to mirror key aspects in life satisfaction's variation over time. Second, the two have a significant pairwise correlation, which persists after controlling for GDP, the effect of time and a battery of other controls. Finally, in a horse race between the MVI and the Text Valence Index (TVI) of Hills et al. (2019), the MVI emerges as a stronger predictor of life satisfaction.

The rest of the paper is organised as follows. Section 2 reviews the relevant literature. Section 3 describes the data. Section 4 presents the results. Section 5 wraps up and offers some final thoughts.

2 Literature

First, our paper relates to the literature in economics that tries to measure happiness. Many papers have discussed the validity of self-reports of subjective wellbeing as a measure, which on the whole are fairly reliable (Diener et al., 2018). Mentioned already is the paper of Hills et al. (2019), whose TVI measure (based on the valence of words in books) is discussed in more detail and compared with the MVI below. To the best of our knowledge, we are the first paper to use measured emotions in music to make any sort of inference about national mood (including happiness).

Second, our work also relates to a literature on the relationship between music and emotions. The fact that over a hundred studies report that different listeners can hear the same emotions in a song illustrates music's potential to express emotions (Juslin, 2013). It therefore stands to reason that listeners might choose songs based on their emotional content to help them work through their own emotions. Indeed, previous work shows how music is used to assist with the emotional processing of significant events, to heighten or strengthen the emotional significance of an activity or ritual, and to manage mood (Sloboda and Juslin, 2010). Our results add to this evidence base by showing that the emotions in the most popular songs reflect how people are actually feeling in the population. The psychology of music literature distinguishes between perceived and induced emotions, and it is important to emphasise that the MVI relates only to perceived emotions; however, this makes it consistent with the notion of music, like a language, being able to describe an emotion to the listener. Whether or not the music has an emotional impact on the listener is therefore not gauged by the MVI (and of course we make no claim that popular music is actually affecting national happiness), but our results (and our success in developing a measure of national valence) support the idea that the emotional content of popular music reflects the expressed emotions of listeners. We remain agnostic as to the cause, but one idea could be that people are more likely to buy a record if it is in tune with how they are feeling, which would imply that the most popular record is then the one that is best able to capture the public mood; this is at least consistent with additional evidence (presented in Appendix A) which demonstrates that the chart topping song is better able to capture national happiness than tracks further down the charts that are less popular. Note, such a process could be further facilitated by record labels, who would be motivated to promote tracks and artists that tap the public mood if such a strategy is favourable to selling records (indeed, Hills et al. (2019) suggest a similar mechanism for the TVI in relation to publishing houses and books).

Finally, our paper relates to the data science literature on music emotion recognition, a branch of music information retrieval (Kim et al., 2010). We provide a new application of these methods: correlating the emotions extracted with socio-economic variables.

3 Data

3.1 Music Valence

3.1.1 Popular Music

We identified the most popular song of the year in the UK using the official singles chart (`www.officialcharts.com`), which is based on record sales. Only weekly charts are available before 2005 so we applied the following transformation to determine annual scores. Let x_i be a track's chart position in a given week (1st, 2nd, etc.) and y be the lowest possible position on the weekly chart during the year (e.g. 50th, 100th); a track's popularity score for that year would be calculated as $\sum_{i=1}^{52}(y + 1 - x_i)$, with the highest-scoring then selected as the most popular. Note, it could be the case that people buy more music during certain weeks of the year (e.g. around Christmas time), so the track we identify as most popular might not have actually obtained the most record sales during the year; rather, the score picks up songs which had lasting popularity over the whole year. The most popular songs were then purchased from Amazon Music or the Apple iTunes Store depending upon availability (the song list is available in Appendix B, along with each song's predicted valence).

3.1.2 Valence Prediction

To predict the valence scores of each song we trained a machine learning model to learn audio features that best predicted valence using a separate set of tracks that had been annotated by human subjects. The annotated dataset comes from Soleymani et al. (2013) (http://cvml.unige.ch/databases/emoMusic/). It consists of 45-second clips of 744 songs from the Free Music Archive (https://freemusicarchive.org/) that span a variety of popular genres (blues, electronic, rock, classical, folk, jazz, country, pop). Each clip was annotated by a minimum of 10 participants on a 9-point valence scale, the average of which is our target measure. We computed our own audio features (191 in total) using the 45-second clips (details are provided in Appendix C). Because the valence target exists on an approximately continuous scale (after averaging across participants), we use a regression framework for prediction. Specifically, we use a Support Vector Regression (SVR) which has displayed relatively good performance for predicting valence in comparison to other regression methods (Yang et al., 2008).

To arrive at our predictive model, we first used a 5-fold cross validation procedure to optimise the SVR algorithm's parameters and the number of features (using R^2 to assess performance on the validation sets). We then trained a model using a fraction ($619 \approx 83\%$) of the annotated songs and tested its performance on the remaining 125 songs to see how well it might generalise; we were able to achieve a reasonably high R^2 on the test set in comparison to machine learning methods from other papers (0.33). Note that we used the same train-test split as in Soleymani et al. (2013) so we could benchmark the model's performance. Finally, we re-trained the model on the full sample of 744 annotated songs and used it to predict the valence scores of the UK's most popular songs (using 45-second clips extracted from the middle of each song as input data), which generates what we call the MVI.

3.2 Other Happiness Measures

3.2.1 Life Satisfaction

To validate the MVI, we use Eurobarometer life satisfaction data (the average per year of all individuals surveyed). This is the longest-running measure of subjective wellbeing (available since 1973), and is also the one used to validate the TVI in Hills et al. (2019). The question asked is, "On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?", with responses given on a 4-point Likert scale.

3.2.2 Text Valence

The TVI measure from Hills et al. (2019) was constructed using the Google Books corpus (Lin et al., 2012). They derived annual valence scores for the UK using the average valence of words in books published in Great Britain during a particular year (weighted by their word frequencies). The valence norms used were for 14,000 English words (each an average of valence ratings by 20 participants on a 9-point scale (Warriner et al., 2013)).

3.3 Controls

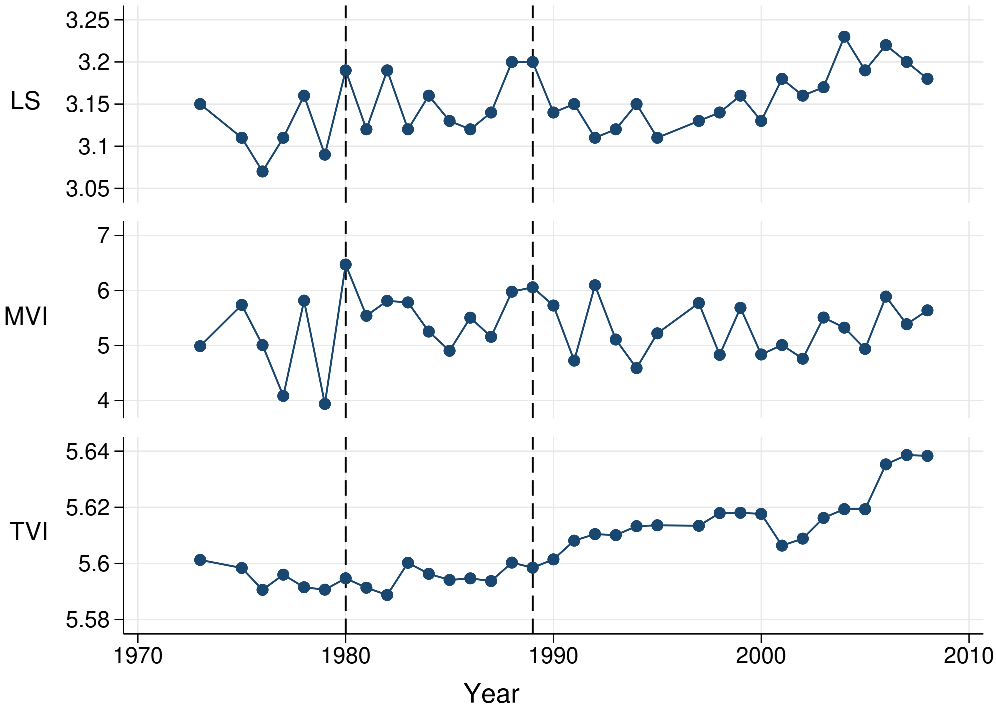
Incorporated in the analyses below are traditional controls used in the subjective wellbeing literature. Firstly, our measure of GDP is from the Penn dataset (in 2005 international dollars, adjusted for purchasing power parity). We also use a set of measures from the OECD: life expectancy at birth (as a measure of health); education inequality (measured as a GINI index); total gross central government debt as a percentage of GDP (as a measure of public expenditure); and inflation.

4 Results

4.1 Time Series of Life Satisfaction, MVI and TVI

As seen in Figure 1, the MVI displays a high degree of similarity with life satisfaction over time, mirroring key elements in its variation. For example, local peaks in life satisfaction in 1980 and 1989 are picked up by the MVI, which also appears to match well the frequency of the life satisfaction data. The TVI on the other hand does less well at picking up such peaks, with its frequency resembling that of a smoothed series. These "eyeballing" observations are confirmed by formal statistical analysis, to which we will now turn.

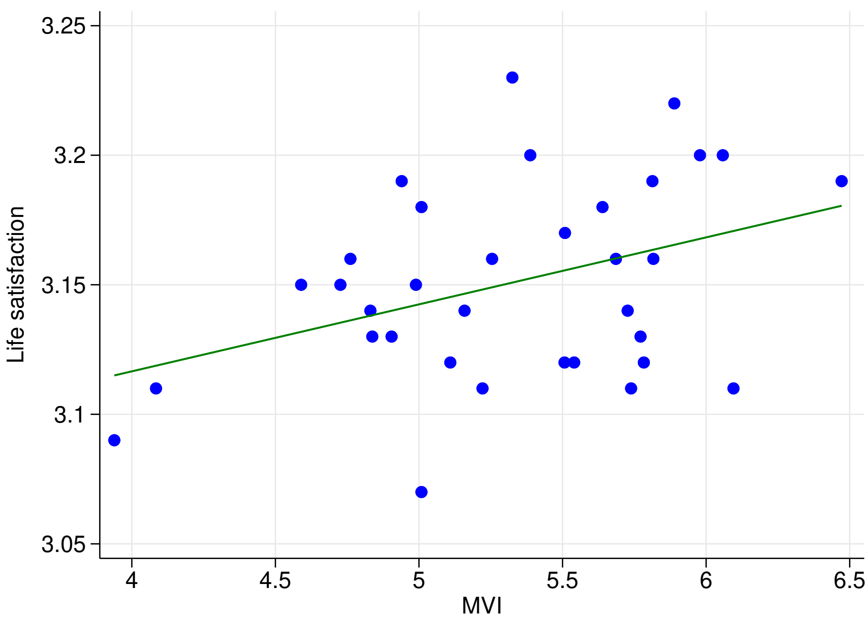
Figure 1: Time Series of Life Satisfaction (LS), MVI and TVI



4.2 Correlation of Life Satisfaction and MVI

Figure 2 shows a scatter plot of life satisfaction and the MVI. As can be seen, they display a significant positive correlation ($r = 0.39$; $p = 0.02$).

Figure 2: Scatter Plot of Life Satisfaction and MVI



The analysis in Table 1 then shows that this positive relationship between MVI and life satisfaction is robust to the introduction of GDP, a time trend and various other controls ($p = 0.003$ without the additional controls; $p = 0.008$ with them). In all regression analyses we report (White) standard errors that are robust to heteroskedasticity, but there are no substantive differences in the results with regular standard errors.

Table 1: The MVI Predicts Life Satisfaction

Marginal effects	Life satisfaction	
	(1)	(2)
MVI	0.392^{***}	0.388^{***}
	(0.122)	(0.135)
GDP	6.645^{*}	6.840
	(3.828)	(4.700)
Trend	Yes	Yes
Other controls	No	Yes
Observations	34	34

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Marginal effects with robust (White) standard errors in parentheses. Life satisfaction and MVI are standardised; GDP is the logarithm of gross domestic product per capita. Other controls include life expectancy, education inequality, public debt and inflation.

4.3 Comparing the MVI and TVI

As shown in Table 2, when included in the same regression, the MVI emerges as a stronger predictor of life satisfaction than the TVI for the UK, with only its coefficient remaining significant. This holds true whether the full set of controls (life expectancy, education inequality, public debt and inflation) are included or not ($p = 0.004$ without the additional controls; $p = 0.007$ with them).

Table 2: MVI a Stronger Predictor of Life Satisfaction than the TVI

Marginal effects	Life satisfaction	
	(1)	(2)
MVI	0.394*** (0.125)	0.405*** (0.139)
TVI	-0.099 (0.236)	-0.276 (0.347)
GDP	6.677* (3.861)	6.666 (4.642)
Trend	Yes	Yes
Other controls	No	Yes
Observations	34	34

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Marginal effects with robust (White) standard errors in parentheses. Life satisfaction, MVI and TVI are standardised; GDP is the logarithm of gross domestic product per capita. Other controls include life expectancy, education inequality, public debt and inflation.

5 Discussion

In this paper we have provided evidence that the valence of a country's most popular songs can provide a reliable indication of average happiness in the population. Moreover, for the UK at least, it appears that the valence of popular music provides a more accurate depiction of its happiness than the valence of books, which supports the idea of music as a specialised "language of the emotions" (Cooke, 1959). A nice feature of the measure is that it only requires collecting information on one song each year (the most popular), which makes it relatively cheap and easy to implement. We support this further in Appendix A where we show that using the valences of tracks that are less popular (including an average of the top 10 songs) does not work as well as focusing only on chart-topping songs.

Here we have only shown that music can predict happiness within a country. Future research might wish to consider the potential of music to explain between-country differences in happiness. Music has the potential to be a good between-country predictor since it is not only an emotional language, but a "universal" one (Longfellow, 1835) and is found in every society with a stable set of functions (Mehr et al., 2019). In general, we hope to encourage a closer look at the emotions in music as potentially representative of underlying social and cultural patterns.

References

Bogdanov, Dmitry, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, O Mayor, Gerard Roma, Justin Salamon, JR Zapata, and Xavier Serra (2013), "Essentia: an audio analysis library for music information retrieval." In *Proceedings of the 14th Conference of the International Society for Music Information Retrieval Conference (ISMIR)*.

Choi, Keunwoo, György Fazekas, Mark Sandler, and Kyunghyun Cho (2017), "Transfer learning for music classification and regression tasks." In *Proceedings of the 18th Conference of the International Society for Music Information Retrieval (ISMIR)*.

Cooke, Deryck (1959), *The language of music*. Oxford University Press.

Diener, Ed, Shigehiro Oishi, and Louis Tay (2018), "Advances in subjective well-being research." *Nature Human Behaviour*, 2, 253.

Hills, Thomas T, Eugenio Proto, Daniel Sgroi, and Chanuki Illushka Seresinhe (2019), "Historical analysis of national subjective wellbeing using millions of digitized books." *Nature Human Behaviour*, 15 October.

Juslin, Patrik N (2013), "What does music express? basic emotions and beyond." *Frontiers in Psychology*, 4, 596.

Juslin, Patrik N and Petri Laukka (2004), "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening." *Journal of New Music Research*, 33, 217–238.

Kim, Youngmoo E, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull (2010), "Music emotion recognition: A state of the art review." In *Proceedings of the 11th Conference of the International Society for Music Information Retrieval (ISMIR)*.

Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, William Brockman, and Slav Petrov (2012), "Syntactic annotations for the google books ngram corpus." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*.

Longfellow, Henry Wadsworth (1835), *Outre-mer: a pilgrammage beyond the sea*. Harper.

McFee, Brian, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto (2015), "Librosa: Audio and music signal analysis in python." In *Proceedings of the 14th Python in Science Conference*, volume 8, 18–25.

Mehr, Samuel A, Manvir Singh, Dean Knox, Daniel M Ketter, Daniel Pickens-Jones, Stephanie Atwood, Christopher Lucas, Nori Jacoby, Alena A Egner, Erin J Hopkins, et al. (2019), "Universality and diversity in human song." *Science*, 366.

Sloboda, John A and Patrik N Juslin (2010), "At the interface between the inner and outer world." *Handbook of Music and Emotion*, 73–97.

Soleymani, Mohammad, Anna Aljanaki, Yi-Hsuan Yang, Michael N Caro, Florian Eyben, Konstantin Markov, Björn W Schuller, Remco Veltkamp, Felix Weninger, and Frans Wiering (2014), "Emotional analysis of music: A comparison of methods." In *Proceedings of the 22nd ACM International Conference on Multimedia*, 1161–1164.

Soleymani, Mohammad, Micheal N Caro, Erik M Schmidt, Cheng-Ya Sha, and Yi-Hsuan Yang (2013), "1000 songs for emotional analysis of music." In *Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia*, 1–6.

Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert (2013), "Norms of valence, arousal, and dominance for 13,915 english lemmas." *Behavior Research Methods*, 45, 1191–1207.

Yang, Yi-Hsuan, Yu-Ching Lin, Ya-Fan Su, and Homer H Chen (2008), "A regression approach to music emotion recognition." *IEEE Transactions on Audio, Speech, and Language Processing*, 16, 448–457.

Table: The Most Popular Song is the Best Measure of Life Satisfaction

Correlations *(p)*	Life Satisfaction
Valence of #1 Song (MVI)	0.386^{**} (0.024)
Valence of #2 Song	0.128 (0.471)
Valence of #3 Song	0.235 (0.180)
Valence of #4 Song	0.344^{*} (0.054)
Valence of #5 Song	-0.161 (0.364)
Valence of #6 Song	0.022 (0.902)
Valence of #7 Song	0.017 (0.924)
Valence of #8 Song	-0.157 (0.375)
Valence of #9 Song	0.308^{*} (0.077)
Valence of #10 Song	0.017 (0.924)
Average Valence of #1-#10 Songs	0.307^{*} (0.077)

Pairwise correlations with p-values in parentheses. Statistically significant measures presented in bold: $^{**}p < 0.05$; $^{*}p < 0.1$.

Table: Most Popular Songs of the Year and their Predicted Valences (which form the MVI)

Year	**Title**	**Artist**	**Valence** *(1-9)*
1973	Tie a Yellow Ribbon Round the Ole Oak Tree	Dawn featuring Tony Orlando	4.99
1974	The Wombling Song	The Wombles	5.40
1975	Bye Bye Baby	Bay City Rollers	5.74
1976	Mississippi	Pussycat	5.01
1977	Evergreen	Barbra Streisand	4.08
1978	Rivers of Babylon	Boney M.	5.82
1979	Bright Eyes	Art Garfunkel	3.94
1980	Feels Like I’m in Love	Kelly Marie	6.47
1981	Birdie Song	The Tweets	5.54
1982	Come On Eileen	Dexy’s Midnight Runners	5.81
1983	Blue Monday	New Order	5.78
1984	Relax	Frankie Goes To Hollywood	5.25
1985	The Power of Love	Jennifer Rush	4.90
1986	So Macho	Sinitta	5.51
1987	Never Gonna Give You Up	Rick Astley	5.16
1988	Push It	Salt-N-Pepa	5.98
1989	Ride on Time	Black Box	6.06
1990	Killer	Adamski	5.73
1991	(Everything I Do) I Do It for You	Bryan Adams	4.73
1992	Rhythm Is a Dancer	Snap!	6.10
1993	No Limit	2 Unlimited	5.11
1994	Love Is All Around	Wet Wet Wet	4.59
1995	Think Twice	Celine Dion	5.22
1996	Return of the Mack	Mark Morrison	5.98
1997	I’ll Be Missing You	Puff Daddy & Faith Evans	5.77
1998	How Do I Live	LeAnn Rimes	4.83
1999	Heartbeat	Steps	5.69
2000	Amazed	Lonestar	4.84
2001	Whole Again	Atomic Kitten	5.01
2002	How You Remind Me	Nickelback	4.76
2003	In Da Club	50 Cent	5.51
2004	Left Outside Alone	Anastacia	5.33
2005	You’re Beautiful	James Blunt	4.94
2006	Hips Don’t Lie	Shakira featuring Wyclef Jean	5.89
2007	How to Save a Life	The Fray	5.39
2008	Rockstar	Nickelback	5.64
2009	Poker Face	Lady Gaga	6.01
2010	Empire State of Mind	Alicia Keys	4.45

Valence Prediction

We extracted commonly used acoustic features for music emotion recognition (Kim et al., 2010) using the music processing libraries Librosa (McFee et al., 2015) and Essentia (Bogdanov et al., 2013):

- Spectral Centroid
- Spectral Rolloff
- Spectral Contrast - 7 bands
- Mel-Frequency Cepstrum Coefficients (MFCC) - 24 coefficients
- Zero Crossing Rate
- Chroma Energy Normalized Statistics (CENS) - 12 chroma
- Beat Per Minute (BPM)
- Root Mean Square (RMS)
- Spectral Flux
- Onset Rate
- High Frequency Content (HFC)

All features were extracted at the frame level except for BPM, RMS, spectral flux, onset rate and HFC. For frame-level features, we used Hann windows of 46 ms, and computed the mean and variance of the frame values and first-order differences. In total there were 191 features.

We then trained a Support Vector Regression (SVR) on the annotated Free Music Archive dataset using radial basis functions as kernels. Features were preprocessed with z-score normalisation (removing the mean and scaling to unit variance) so features with large magnitude would not dominate the objective function. A 5-fold cross-validation procedure selected the optimal parameters of the SVR algorithm and number of features (100). Feature selection was carried out using the F-test which tests the individual effect of each feature by converting the correlation between each feature and the valence to an F score. Using the same train-test split as in Soleymani et al. (2013), our achieved R^2 on the test set compares favourably with other machine learning models:

Method	Valence R^2
This Paper	0.33
Baseline[a]	0.12
MFCC[b]	0.20
TUM[c]	0.42
UAizu[c]	0.35
UU[c]	0.31

[a] Soleymani et al. (2013). [b] Choi et al. (2017). [c] Soleymani et al. (2014).