I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14258

Measuring National Life Satisfaction with Music

Emmanouil Benetos
Alessandro Ragano
Daniel Sgroi
Anthony Tuckwell

APRIL 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14258

Measuring National Life Satisfaction with Music

Emmanouil Benetos
Queen Mary University of London and The Alan Turing Institute

Alessandro Ragano
The Alan Turing Institute and University College Dublin

Daniel Sgroi
The Alan Turing Institute, University of Warwick and IZA

Anthony Tuckwell
The Alan Turing Institute and University of Warwick

APRIL 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Measuring National Life Satisfaction with Music

National life satisfaction is an important way to measure societal well-being and since 2011 has been used to judge the effectiveness of government policy across the world. However, there is a paucity of historical data making limiting long-run comparisons with other data. We construct a new measure based on the emotional content of music. We first trained a machine learning model using 191 different audio features embedded within music and use this model to construct a long-run Music Valence Index derived from chart-topping songs. This index correlates strongly and significantly with survey-based life satisfaction and outperforms an equivalent text-based measure. Our results have implications for the role of music in society, and validate a new use of music as a long-run measure of public sentiment.

JEL Classification: C8, N3, N4, O1, D6

Keywords: historical subjective wellbeing, life satisfaction, music, sound data, language, big data

Corresponding author:
Daniel Sgroi
Department of Economics
University of Warwick
Gibbet Hill Road
Coventry CV4 7AL
United Kingdom
E-mail: daniel.sgroi@warwick.ac.uk

Introduction

Aside from being a direct measure of the well-being, the average level of life satisfaction in a nation (or simply "national life satisfaction") 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 life satisfaction at the macro level is therefore an important area of research, with the most popular method in recent decades being surveys of subjective well-being. 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 (*1*). Like language, music can also encode emotional information: it has been described as a "language of the emotions" (*2*), with studies demonstrating that different people can recognise the same patterns of emotion in a song (*3*). Moreover, it is the emotional experience that music offers that primarily motivates individuals to listen to it (*4*). 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 life satisfaction 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 (*5*). Our methods are described in the Methods section. We find that the MVI displays a significant degree of similarity with the survey-based measure of life satisfaction, indicating that audio features embedded within the sound of popular music have the potential to describe national well-being. 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 regression analyses that feature a "horse race" between the MVI and the Text Valence Index (TVI) (*1*), the MVI emerges as a stronger predictor of life satisfaction.

Many papers have discussed the validity of self-reports of subjective well-being as a measure of national life satisfaction or national happiness, and have concluded that they are on the whole are fairly reliable (*6*). Going beyond survey-based measures and into the realm of natural language processing, mentioned already is the paper of (*1*), whose TVI measure (based on a text analysis of the valence of words in books and newspapers) is discussed in more detail and compared with the MVI below. (*7*) also conducts a text analysis and links this to well-being, but provides an individual-level analysis, measuring the well-being of three famous composers using the text of their personal letters. To the best of our knowledge, we are the first to use measured emotions in music derived from the audio features contained within sound to make any sort of inference about life satisfaction at the national level.

Our work is supported by 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 (*3*). 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 (*8*). 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 life satisfaction), but our results (and our success in developing a measure of national life satisfaction) 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 Online Appendix (Table S1)) which demonstrates that the chart-topping song is better able to capture national life satisfaction 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, (*1*) suggest a similar mechanism for the TVI in relation to publishing houses and books and argue that this is strongly suggestive of causation going from national mood to books/newspaper articles (via the selection of publishers/editors), rather than the reverse, which might also make sense in the music context.

Our paper also relates to the data science literature on music emotion recognition, a branch of music information retrieval (*9*). We provide a new application of these methods: correlating the emotions extracted with socio-economic variables.

Methods

Our methods involve first training a machine learning model to recognise high and low valence in a training set using 191 audio features. This model is then used to construct a Music Valence Index (MVI) based on the predicted valence of the most popular song of the year in the UK from 1973-2010, a time period that enables comparisons with the leading survey-based measure and text-based measure together with a set of controls as detailed below.

Data. 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 (which include downloads from 2004 onwards). 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 Online Appendix (Table S2), along with each song's predicted valence).

Training. 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 Online Appendix (Valence Prediction)). 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 (*10*).

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 (*5*) 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.

Validation. 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 well-being (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.

The TVI measure from (*1*) was constructed using the Google Books corpus (*11*). 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 (*12*)).

Incorporated in the analyses in the Results section are traditional controls used in the subjective well-being 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.

Results

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.

Figure 2 shows a scatter plot of life satisfaction and the MVI. As can be seen, they display a significant positive correlation ($r = 0.39$; $p = 0.02$). Moreover, as shown in Figure 3, when we consider the annual change in the MVI as compared with the annual change in life satisfaction, we also see a clear positive correlation ($r = 0.46$; $p < 0.01$). These visual observations are confirmed by formal statistical analysis, to which we will now turn.

Regression analyses in Table 1 (specifications (1) and (2)) 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 heteroskedasticity-consistent (or Eicker–Huber–White) standard errors, but there are no substantive differences in the results with classical standard errors.

Next we consider the relative strength of our MVI to a text-based measure when the two are pitted against each other. To do so we perform a regression analysis with both of our candidate predictors, the MVI and TVI, situated on the right-hand side of the regression, which is commonly referred to in the literature as a “horse race”. Rather than attempting to suggest that either variable has a causal effect on life satisfaction (the more common use of a regression), this technique instead seeks to evaluate which is a stronger predictor, or alternatively which has a stronger correlation, measurable using p-value. As shown in specifications (3) and (4) of Table 1, 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).

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 life satisfaction in the population. This might be considered surprising: not everyone listens to music and indeed listening to “chart-topping” music might even be considered largely a teenage pass-time. However, it is clear from our results that the audio features embedded within the sound of chart-topping music do correlate well with national well-being. This could be because the most popular chart hit in any given year goes beyond the traditional pop music demographic and is more representative of national mood, it could be because those who buy popular music do in fact provide a reasonable sample of the population, or it might provide a reasonable proxy for some other reason. What is clear is that for whatever reason the correlation between the MVI and national well-being as measured through more traditional survey-based measures is strong and highly significant.

Moreover, for the UK at least, it appears that the valence of popular music provides a more accurate depiction of national life satisfaction than the valence enshrined within books, which provides even greater support for the idea of music as a specialised “language of the emotions” (*2*). A nice feature of our 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 Online Appendix (Table S1) 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. It might also be interesting to note that the pairwise correlation between the MVI and life satisfaction falls to only 0.15 (and becomes insignificant) when we consider life satisfaction lagged by one year. This is in stark contrast to the TVI which improves when we lag life satisfaction. This suggests that music is also a more immediate measure of national mood.

Here we have shown that music can predict life satisfaction within a country. Future research might wish to consider the potential for music to explain between-country differences in life satisfaction. Music has the potential to be a good between-country predictor since it is not only an emotional language but a “universal” one (*13*) and is found in every society with a stable set of functions (*14*). Data availability is improving over time: for the UK downloads were incorporated in music chart data in 2004, streaming was partially added from 2008 and fully incorporated from 2014 onwards. With downloads and streaming becoming increasingly prevalent it will be easier to measure listening behaviour accurately. There is also scope for examining both the role of different genres of music (as they compete for an audience) and the changes in valence within genres (which might link to the mood of specific groups who are more likely to listen to these genres). In general, we hope to encourage a closer look at the emotions contained within music as potentially representative of underlying social and cultural patterns.

References

1. T. T. Hills, E. Proto, D. Sgroi, C. I. Seresinhe, *Nature Human Behaviour* **3**, 1271– (2019).
2. D. Cooke, *The language of music* (Oxford University Press, 1959).
3. P. N. Juslin, *Frontiers in Psychology* **4**, 596 (2013).
4. P. N. Juslin, P. Laukka, *Journal of New Music Research* **33**, 217 (2004).
5. M. Soleymani, M. N. Caro, E. M. Schmidt, C.-Y. Sha, Y.-H. Yang, *Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia* (2013), pp. 1–6.
6. E. Diener, S. Oishi, L. Tay, *Nature Human Behaviour* **2**, 253 (2018).
7. K. J. Borowiecki, *Review of Economics and Statistics* **99**, 591 (2019).
8. J. A. Sloboda, P. N. Juslin, *Handbook of Music and Emotion* pp. 73–97 (2010).
9. Y. E. Kim, *et al.*, *Proceedings of the 11th Conference of the International Society for Music Information Retrieval (ISMIR)* (2010).
10. Y.-H. Yang, Y.-C. Lin, Y.-F. Su, H. H. Chen, *IEEE Transactions on Audio, Speech, and Language Processing* **16**, 448 (2008).
11. Y. Lin, *et al.*, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (2012).
12. A. B. Warriner, V. Kuperman, M. Brysbaert, *Behavior Research Methods* **45**, 1191 (2013).
13. H. W. Longfellow, *Outre-mer: a pilgrammage beyond the sea* (Harper, 1835).
14. S. A. Mehr, *et al.*, *Science* **366** (2019).
15. B. McFee, *et al.*, *Proceedings of the 14th Python in Science Conference* (2015), vol. 8, pp. 18–25.
16. D. Bogdanov, *et al.*, *Proceedings of the 14th Conference of the International Society for Music Information Retrieval Conference (ISMIR)* (2013).
17. K. Choi, G. Fazekas, M. Sandler, K. Cho, *Proceedings of the 18th Conference of the International Society for Music Information Retrieval (ISMIR)* (2017).
18. M. Soleymani, *et al.*, *Proceedings of the 22nd ACM International Conference on Multimedia* (2014), pp. 1161–1164.

Acknowledgments

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). All authors contributed equally to the production of this paper and are listed alphabetically. The authors declare no conflicts of interest.

Tables and Figures

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

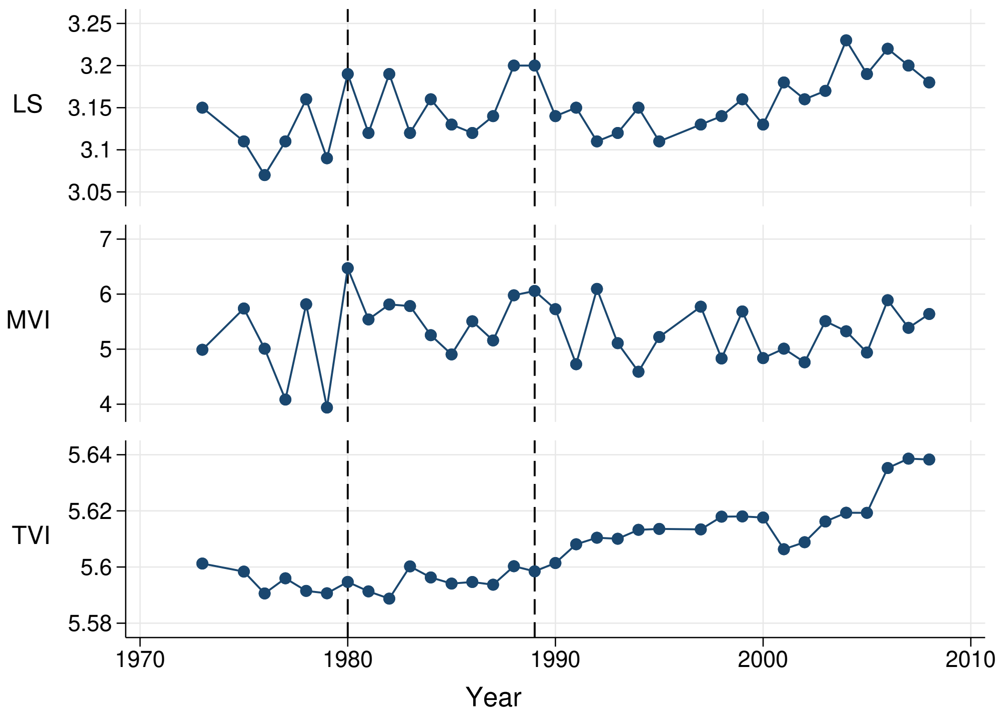

Figure 2: Scatter Plot of Life Satisfaction and MVI

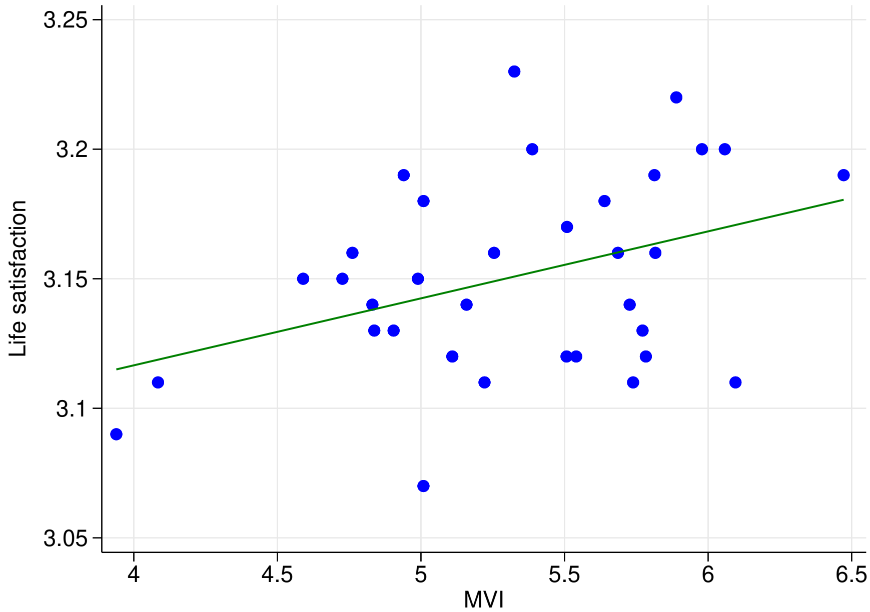

Figure 3: Scatter Plot of Annual Change in Life Satisfaction and Annual Change in MVI

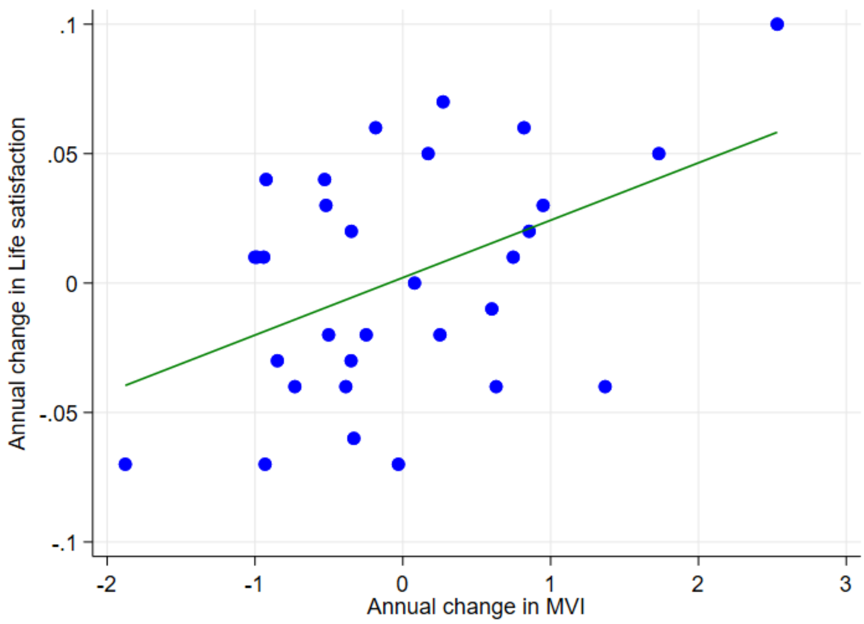

Table 1: The MVI Predicts Life Satisfaction

Marginal effects	Life satisfaction			
	(1)	(2)	(3)	(4)
MVI	0.392*** (0.122)	0.388*** (0.135)	0.394*** (0.125)	0.405*** (0.139)
TVI			-0.099 (0.236)	-0.276 (0.347)
GDP	6.645* (3.828)	6.840 (4.700)	6.677* (3.861)	6.666 (4.642)
Trend	Yes	Yes	Yes	Yes
Other controls	No	Yes	No	Yes
Observations	34	34	34	34

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Marginal effects with (heteroskedasticity-consistent) 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.

Online Appendix

Contents

Valence Prediction

We extracted commonly used acoustic features for music emotion recognition (*9*) using the music processing libraries Librosa (*15*) and Essentia (*16*):

- 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)

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. For spectral flux and HFC we computed only the mean and the variance of frame values. 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 (*5*), our achieved R^2 on the test set compares favourably with other machine learning models as indicated in the following table:

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] (*5*), [b] (*17*), [c] (*18*)

Table S1: 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 S2: 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