



Kent Academic Repository

Jordanous, Anna (2026) *Emovectors: assessing emotional content in jazz improvisations for creativity evaluation*. In: 2025 IEEE International Conference on Big Data. . ISBN 979-8-3315-9448-0. E-ISBN 979-8-3315-9447-3.

Downloaded from

<https://kar.kent.ac.uk/112076/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1109/BigData66926.2025.11402279>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Emovectors: assessing emotional content in jazz improvisations for creativity evaluation

Anna Jordanous
School of Computing
University of Kent
Canterbury, UK
a.k.jordanous@kent.ac.uk

Abstract—Music improvisation is fascinating to study, being essentially a live demonstration of a creative process. In jazz, musicians often improvise across predefined chord progressions (leadsheets). How do we assess the creativity of jazz improvisations? And can we capture this in automated metrics for creativity for current LLM-based generative systems? Demonstration of emotional involvement is closely linked with creativity in improvisation. Analysing musical audio, can we detect emotional involvement? This study hypothesises that if an improvisation contains more evidence of emotion-laden content, it is more likely to be recognised as creative. An embeddings-based method is proposed for capturing the emotional content in musical improvisations, using a psychologically-grounded classification of musical characteristics associated with emotions. Resulting ‘emovectors’ are analysed to test the above hypothesis, comparing across multiple improvisations. Capturing emotional content in this quantifiable way can contribute towards new metrics for creativity evaluation that can be applied at scale.

Keywords: music, emotion, emovectors, creativity evaluation, improvisation

Index Terms—H.5.5 Sound & Music Computing, I.2 Artificial Intelligence, O.2 Modeling human emotion

I. INTRODUCTION

Music improvisation is fascinating to study for musical creativity, as it is a domain which is inherently the live demonstration of a creative process as it happens. A common scenario in jazz improvisation is where a group of musicians play together through multiple repeats of a set of predefined chords (a leadsheet) for a song or melody, typically with one musician soloing (playing an improvisation) at a time, and the other musicians playing accompanying backings to that solo.

How do we detect or evaluate more creative jazz improvisations, particularly through computational metrics? Automated metrics for creativity, that can scale up to analysis of Big Data, would be particularly valuable in assessing and guiding output of current LLM-based music AI systems, if they can be grounded in understanding of what it means to be creative as an improviser. One area with strong links towards enhanced perceived creativity is the demonstration of emotional involvement by the improvising musician [1]. It is recognised that music content can express emotions [2], [3]. In scenarios where the only data we have available for analysis is music audio data such as music recordings, we can detect indicators of emotional content within that audio data that can be treated as proxy evidence for emotional involvement.

This study investigates the research hypothesis that: if an improvisation contains more evidence of emotion-laden content, it is more likely to be recognised as creative.

An embeddings-based method is proposed to capture the emotional content in music improvisations, using a psychologically-grounded classification of musical characteristics associated with emotions to generate emotion embeddings from audio features [4]. The resulting embeddings, or ‘emovectors’, are analysed to test the above hypothesis across multiple improvisations. Results so far suggest this approach could provide a method to capture and measure emotional content, towards metrics for creativity evaluation at scale.

II. BACKGROUND

This work is motivated by the need for automated tests for music improvisation creativity, which can be applicable for large data. Tools for generating music and audio have been an area of research for decades e.g. [5], [6]. Older tools typically could generate output at a reasonable scale; however the advent of recent LLM-based music generative AI tools such as MusicGen [7] has transformed the scale at which credible outputs can be generated. With greater scale comes greater difficulties for any evaluative processes which require human input or more lengthy, involved tests which cannot be applied practically at scale. For example, evaluating the creativity of the output of such music generators is a complex and unsolved problem [8], typically relying on tests which require manual intervention, interactivity (causing a fitness bottleneck when applied at scale), and/or application of automated tests that do not necessarily correlate with what is important for musical creativity itself. This is driven by pragmatism; creativity is challenging to define and test for.

A ‘standard definition of creativity’, the combination of producing novel output which is valuable [9], was proposed by Runco and Jaeger to facilitate the process of creativity evaluation with a workable definition. This definition captures two fundamentals of creativity, which can be treated as necessary - but not sufficient [10]. Value is context dependent, as is novelty, and what is important for creativity in one type of area is not necessarily important for creativity in another area. As Plucker and Beghetto highlight, although creativity is essentially domain-general (fundamentally the same across all

types of creativity), in practical terms we need to add domain-specific context to how creativity is assessed:

‘Creativity is the interplay between ability and process by which an individual or group produces an outcome or product that is both novel and useful as defined within some social context.’ [11, p. 156]

What social context should we take into account when evaluating creativity in music improvisation? Improvisation is the generation of music in real-time, and can be considered as the live demonstration of musical creativity: a creative process where the outputs are closely tied to the live process. In fact, improvisation and creativity are often conflated by authors rather than being distinct behaviours [12]–[15].

While there have been several useful contributions on creativity in musical improvisation (e.g. [1], [13]–[15]), no overall consensus emerges for how that creativity is manifested. However one theme which emerges is the importance of emotional involvement and emotion-driven process [10], [15], [16]. Lewis coins improvisation as a process such that while improvising, ‘one hears something of oneself’ [16].

How can we model human emotions in a computational way, so computers can detect emotional content in musical content? While much work exists in music mood/emotion recognition [17], it can be argued that work in this area does not always align with psychologically-informed models of emotion. Two leading psychological models of emotion are Russell’s circumplex model [18] and Ekman’s model of basic emotions [19]. Russell’s two-dimensional model, treating valence (the positive/negative sentiment) and arousal (the amount of active energy perceived in a sound) as two axes by which all emotions can be plotted, has been shown useful for music emotion categorisation e.g. [20], [21]. However it does not capture the magnitude of emotions present, only the presence (or absence) of individual emotions detected in data.

The Ekman model has been less explored in practical terms in music emotion research, probably due to the relative difficulty of detecting Ekman’s basic emotions compared to the Russell model, which requires tests for two audio features (valence and arousal) that have been implemented computationally e.g. as evaluated in [22]. Perhaps somewhat overlooked, the work of Juslin and Laukka in identifying and categorising acoustic cues for different types of emotion [4] collects together different audio-based cues for emotional content in audio (music performance and speech), through a review of 41 studies of musical performance and 104 studies of vocal expression. Juslin and Laukka’s review identifies how acoustic cues from the reviewed studies fall into categories (anger, fear, happiness, sadness, tenderness) that align reasonably well with Ekman’s model of basic emotions (anger, fear, happy, sad, disgust, and the later addition of surprise as a sixth basic emotion). The final category by Juslin and Laukka, tenderness, could be argued to be the antithesis of the disgust emotion named in Ekman’s model.

In contrast to the binary nature of emotion detection in Russell’s model, the acoustic cues identified in Juslin and Laukka’s review can be tested for in a way which allows us

to measure the strength or magnitude of the presence of that acoustic cue. This allows us to go beyond categorisation of emotional content, to measure magnitudes at which different emotional indicators are present. It could also be argued that the Juslin and Laukka model is more grounded in psychological plausibility. Unlike many computational implementations of valence and arousal, which typically test for audio features that can be used as a proxy for the two features, the Juslin and Laukka model is the result of collating and combining multiple studies on acoustic cues for the presence of emotion in audio context. Given such psychological validation, this model seems useful for investigation in this work’s goal: for detecting and comparing the presence and quantity of emotion-laden content. If the cues in Juslin and Laukka’s model can be successfully implemented as tests which can be applied to audio at scale, then this gives us a psychologically validated set of metrics that can be applied to detect musical cues for emotion, at the scale needed for evaluation of current music generative AI tools. This provides the step forward for this current study: evaluating a contextually important factor of musical improvisation creativity using automated tests which can be applied at scale to large datasets.

III. METHODOLOGY

Audio data of various improvisations was collected for analysis. The goal was to have a dataset of more creative improvisations matched with a dataset of less creative improvisations. Ideally it would be helpful to have data on rankings of highly creative improvisations (or conversely, improvisations which are deemed to be of low creativity). To the best of my knowledge, such a dataset does not exist. For the purposes of this study, a working assumption is made that if we compare a set of notable improvisations by famous jazz musicians against a set of improvisations generated by an AI tool, or co-created by a human musician working with an AI tool, we can treat the notable improvisations as more creative than the co-created improvisations.¹ The source of the co-created improvisations in this study was a set of tracks generated using the Impro-Visor music improvisation software [6].

For a fair comparison between computer-generated and human-performed improvisations, it is important to detect the emotional content directly in the music data, as distinct from emotional content added during individual performance. Hence for this study, MIDI tracks generated by Impro-Visor were used for all audio data, being converted to a signal-based audio representation for more straightforward analysis. This removed any additional emotional indicators added by a human performer, allowing analysis to focus purely on the musical data. Data were collected from <https://www.cs.hmc.edu/~keller/jazz/improvisor/Solos/>, comparing transcriptions of improvisations by Charlie Parker, Clifford Brown and Dave Liebman with a set of improvisations generated either

¹While this working assumption gives us an approach that allows us to proceed with data, it is perhaps somewhat unfair on the process involving the less notable musician and AI generator. It should be noted that the human musician involved, Bob Keller, was regarded as a competent musician.

by Impro-Visor or by a human musician (Bob Keller) working with the Impro-Visor software.

Juslin & Laukka features were implemented to analyse audio data, as reported in the next section. As part of this implementation, a benchmark identification phase was necessary; each of the acoustic cues used by Juslin and Laukka rely on descriptive terms to distinguish between how musical features are used in each emotional category. For example, the anger, fear and happiness emotions are categorised in [4] as typically associated with ‘fast’ tempos, with the sadness and tenderness emotions typically associated with ‘slow’ tempos. Hence a set of emovectors was calculated for a benchmark dataset (the jazz subset of the GTZAN dataset [23]), to provide percentiles as thresholds for each acoustic cue as per Table II.

With the calculated thresholds, the collected improvisations data were analysed to judge differences in the acoustic cues being measured. The output of the analysis process are *emovectors*: 5-dimensional embeddings or vectors, one for each song in analysis data. The five dimensions represent the five emotions categorised by Juslin and Laukka: anger, fear, happiness, sadness and tenderness.

Emovector results across sets of improvisations were compared across the two sets of improvisations collated for comparison, to investigate the hypothesis that more creative improvisations contain more emotion-laden content.

IV. OPERATIONALISING THE ACOUSTIC CUES

The acoustic cues identified by Juslin and Laukka were compiled into Table I, representing how each acoustic cue differed per emotion. The Table I were then converted into a set of indicative measurements per acoustic cue for each emotion, representing concepts such as ‘high’, ‘little’ etc in a numeric form, as represented in Table II.

Evaluation of each of the acoustic cues was operationalised as follows, using the Librosa [24] Python library (see <https://github.com/annajordanous/emovectors> for code and data).

- tempo: calculated using Librosa’s `beat_track()` function.
- microstructural irregularity: treated as beat irregularity, using Librosa’s `frames_to_time()` function to convert beats to a time-based representation, then calculating the extent to which each beat deviates from expected tempo.
- sound level: calculating the root-mean-square (RMS) value for each frame (sample segment) of the audio using Librosa’s `rms()` function, to represent the amount of energy present in the audio signal at each frame.
- sound level variability: using the RMS calculations for sound level per frame and calculating the extent to which frames differed from each other over the entire audio.
- high-frequency energy: calculating the mean spectral bandwidth for each audio; the wider (larger) the bandwidth on average, the more high-frequency energy is present, as the bandwidth represents the span from highest to lowest frequencies present in a sound signal.
- pitch level: calculated using Librosa’s `yin()` function, which calculates the fundamental frequency of the sound signal (the lowest frequency present) for each phrase.

- pitch variability: calculating the extent to which the pitch level differs across all frames of the audio.
- tone attacks: calculating attack times (the initial part of a sound ‘envelope’, which represents how sound changes over time for individual sounds such as a note being played) by using the Librosa `rms()` function to identify frames with levels of energy indicating an attack (implemented as above a threshold of 10% of the maximum energy detected in the audio) and calculating the number of frames for which an attack lasts.

One acoustic cue has not yet been implemented, which is the pitch contour cue. This is complex to calculate for polyphonic audio which contains several instruments. In future work, this could be implemented if working with monophonic data such as a solo played by an instrument such as a saxophone or trumpet, which typically can only sound one note at a time.

V. RESULTS AND DISCUSSION

Results, as presented in Table III, showed that for each of the five emotions, on average (mean), each emotion was present to a greater extent than in the Impro-visor/Keller solos. While the datasets are too small for analysis of significance of results, this represents promising initial results for our hypothesis that if an improvisation contains more evidence of emotion-laden content, it is more likely to be recognised as creative.

Operationalising these tests give us an approach that can be run at scale, with no human interaction required. This is particularly relevant when wishing to analyse the (often vast) output of an AI generative music tool.

VI. FUTURE WORK

As mentioned above, one acoustic cue will be implemented in future work (pitch contour) with different data. Once larger datasets can be sourced for analysis, the evaluations can be rerun at larger scale, and this will allow for statistical analysis across songs matched to ranking to verify hypothesis. What we lack is any dataset where improvisations are somehow ranked for creativity; this represents a significant area for future data collection activity. Such a dataset would be a valuable resource for future work on music improvisation creativity.

Once tests can be implemented at scale and validated against other rankings of creativity, the hypothesis of this work can be fully evaluated. If successful, this method enables investigations such as analysis of how chord sequences influence creativity potential for improvisation, or analysis of creativity of an individual artist across different recordings, time periods or musical groups they play in.

VII. CONCLUSIONS

Demonstration of emotional involvement is closely linked with creativity in music improvisation. This study hypothesises that if an improvisation contains more evidence of emotion-laden content, it is more likely to be recognised as creative. This is difficult to test for in current LLM-based generative systems or for large music data sets, as automated metrics for emotion detection typically either require human interaction,

	Tempo	sound level	sound level variability	high-frequency energy	pitch level	pitch variability	pitch contour	tone attacks	microstructural irregularity
Anger	Fast	high	much	much	high	much	rising	fast	Irregularity
Fear	Fast	low	much	little	high	little	rising	n/a	a lot of irregularity
Happiness	Fast	medium-high	n/a	medium	high	much	rising	fast	very little regularity
Sadness	Slow	low	little	little	low	little	falling	slow	irregularity
Tenderness	Slow	low	little	little	low	little	falling	slow	regularity

TABLE I
MAPPING OF ACOUSTIC CUES TO EMOTION.

	Tempo	sound level	sound level variability	high-frequency energy	pitch level	pitch variability	pitch contour	tone attacks	microstructural irregularity
Anger	1	1	1	1	1	1	1	1	0.5
Fear	1	0	1	0	1	0	1	n/a	1
Happiness	1	0.75	n/a	0.5	1	1	1	1	0.25
Sadness	0	0	0	0	0	0	0	0	0.5
Tenderness	0	0	0	0	0	0	0	0	0

TABLE II

REPRESENTATION OF ACOUSTIC CUE AS A SET OF INDICATIVE MEASUREMENTS PER ACOUSTIC CUE FOR EACH EMOTION, IN THE INTERVAL [0, 1].

	Transcribed famous solos (n=6)	Impro-visor/Keller solos (n=11)
Anger	mean = 2.000 s.d = 0.894	mean = 1.545 s.d = 1.128
Fear	mean = 2.000 s.d = 1.095	mean = 1.364 s.d = 0.674
Happiness	mean = 1.333 s.d = 1.506	mean = 1.091 s.d = 0.831
Sadness	mean = 2.000 s.d = 1.897	mean = 1.727 s.d = 1.104
Tenderness	mean = 2.000 s.d = 1.549	mean = 1.546 s.d = 1.036

TABLE III

COMPARATIVE RESULTS BETWEEN EACH DATASET

which does not scale up for large datasets, or may be somewhat distinct from psychologically validated models of emotion.

The results of this study thus far are based on small datasets and with a number of limitations based on the lack of validated data ranking creativity of improvisations. Initial results do show promising results for the theory that higher emotional content in music can be used to contribute towards new metrics for creativity evaluation that can be applied at scale.

VIII. ETHICS STATEMENT

No experiments were conducted with human participants for this study. Training data were from licensed open data sources.

IX. ACKNOWLEDGEMENTS

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

REFERENCES

- [1] A. Jordanous and B. Keller, "What makes musical improvisation creative?," *Journal of Interdisciplinary Music Studies*, vol. 6, no. 2, pp. 151–175, 2012.
- [2] J. A. Sloboda and S. A. O'Neill, "Emotions in everyday listening to music," *Music and emotion: Theory and research*, vol. 8, pp. 415–429, 2001.
- [3] P. N. Juslin, "What does music express? basic emotions and beyond," *Frontiers in psychology*, vol. 4, p. 596, 2013.
- [4] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: different channels, same code?," *Psychological Bulletin*, vol. 129, no. 5, 2003.
- [5] M. Allan and C. Williams, "Harmonising chorales by probabilistic inference," in *Advances in Neural Information Processing Systems* (L. Saul, Y. Weiss, and L. Bottou, eds.), vol. 17, MIT Press, 2004.
- [6] R. M. Keller, "Automated Jazz Improvisation," in *Proceedings of the International Conference on Computational Creativity*, 2010.
- [7] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, and A. Défossez, "Simple and controllable music generation," 2024.
- [8] K. Agres, J. Forth, and G. A. Wiggins, "Evaluation of musical creativity and musical metacreation systems," *Comput. Entertain.*, vol. 14, no. 3, 2016.
- [9] M. A. Runco and G. J. Jaeger, "The standard definition of creativity," *Creativity research journal*, vol. 24, no. 1, pp. 92–96, 2012.
- [10] A. Jordanous and B. Keller, "Modelling creativity: Identifying key components through a corpus-based approach," *PLoS one*, vol. 11, no. 10, p. e0162959, 2016.
- [11] J. A. Plucker and R. A. Beghetto, "Why Creativity is Domain General, Why it Looks Domain Specific, and why the Distinction Doesn't Matter," in *Creativity: From Potential to Realization* (R. J. Sternberg, E. L. Grigorenko, and J. L. Singer, eds.), ch. 9, pp. 153–167, Washington, DC: American Psychological Association, 2004.
- [12] R. K. Sawyer, "Improvisation," *Journal of Linguistic Anthropology*, vol. 9, no. 1/2, pp. 121–123, 1999.
- [13] B. Thom, "Interactive improvisational music companionship: A user-modeling approach," *User Modeling and User-Adapted Interaction*, vol. 13, no. 1, pp. 133–177, 2003.
- [14] P. N. Johnson-Laird, "How jazz musicians improvise," *Music perception*, vol. 19, no. 3, pp. 415–442, 2002.
- [15] M. Biasutti and L. Frezza, "Dimensions of music improvisation," *Creativity Research Journal*, vol. 21, no. 2-3, pp. 232–242, 2009.
- [16] G. E. Lewis, "Improvising with creative machines: Reflections on human-machine interaction (keynote talk)," in *Proceedings of the 2nd International Conference on Computational Creativity*, 2011.
- [17] D. Han, Y. Kong, J. Han, and G. Wang, "A survey of music emotion recognition," *Frontiers of Computer Science*, vol. 16, no. 6, pp. 166–335, 2022.
- [18] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
- [19] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.
- [20] M. Soleymani, M. N. Caro, E. M. Schmidt, C.-Y. Sha, and Y.-H. Yang, "1000 songs for emotional analysis of music," in *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, 2013.
- [21] G. Langroudi, A. Jordanous, and L. Li, "Music emotion capture: sonifying emotions in EEG data," in *AISB Symposium on Emotion Modelling and Detection in Social Media and Online Interaction*, 2018.
- [22] D. Vidas, L. Nitschinsk, M. S. Osborne, and N. S. Rickard, "Validating Spotify's 'Valence', 'Energy', and 'Danceability' audio features for music psychology research," 2025.
- [23] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.
- [24] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," *SciPy*, vol. 2015, 2015.