Emotional Sentience in AI-Created Music

LITERATURE REVIEW

NEIL KALE

Table of Contents

Introduction

What defines sentience? The word is derived from the Latin root *sentire* which means to sense or to feel. Indeed, sentience is to be "responsive to or conscious of sense impressions" (Merriam-Webster, n.d.). In other words, it is simply the ability to sense one's surroundings. However, a thermometer can sense the temperature around it, yet humans would not call it sentient. A light switch can sense whether it is on or off, hold that information, and even send it to a connected bulb. That said, it is still not sentient. Perhaps, what makes humans unique, and sentient, is not the ability to feel, but rather the ability to empathize (Keltner et. al, 2019). Regardless of the definition, it is taken for granted that humans are sentient. On the other hand, by the status quo, computers are not at all sentient (Allen, 2016). However, if Artificial Intelligence (AI) can read human creations that express emotion, and machine learning allows AI to create unique compositions modelled on whatever it reads, then artificial-intelligence creations will propagate emotions as well. This project focuses on whether algorithmic music compositions can induce an emotional response in humans. If so, an AI equipped with these algorithms is capable of propagating emotions, and humans are one step closer to asking the question - Can AI be sentient?

Humans express their emotions through a diverse set of mediums such as facial expression, speech, literature, music, and art (Naar, n.d.). Even if an AI can algorithmically reproduce these mediums, there remains the question of whether the emotions it evokes are truly its own. After all, the AI is merely selecting its emotion based on an algorithm. However, humans also derive their emotions from internal algorithms, albeit extremely complicated ones (Shapshak, 2018). This does not reduce the value of human emotions or sentience. In fact, it is all the more impressive that complex emotions like love and happiness can arise from an algorithm, rather than spontaneously materializing. If the human mind is a complex algorithm, then human's emotional expression is also algorithmic. Perhaps the great works of Mozart, Michelangelo, Shakespeare, and Van Gogh were not spontaneous feats of creativity; instead, it may be possible that they were algorithmically generated by the minds of their creators. The difference between AI and humans may simply be the complexity of their algorithms.

Ultimately, as humans progress through the 21st century, the line differentiating human and AI is fading. Advances in neuroscience are showing that humans may be more algorithmic than was previously believed (Shapshak, 2018). Furthermore, machine learning is allowing AI to step closer to humanity than ever before (Shabbir & Anwer, 2018). There is a need for research that enters the no-man's land (and no-AI's land) between humanity and AI, exploring what it truly means to be sentient as well as what it means to be human.

History of Algorithmic Music

While the first computer-based algorithms date only to the 1980s, algorithmic music has existed for almost a millennium. The first known algorithmic music is credited to an Italian monk named Guido d'Arezzo. In 1026 CE, d'Arezzo assigned a pitch to each vowel in the Latin alphabet, and thus converted Latin chants into music (Edwards, 2011). His creation falls into the category of a deterministic algorithm. This means that for any given input chant, the algorithm always outputs the same song as a result. By contrast, Mozart's *Musikalisches Würfelspiel*, or *Musical Dice Game*, combines fragments of music randomly according to the results of dice throws (Edwards, 2011). This is considered a stochastic algorithm, where the same set of rules can output a different composition each time they are applied.

Deterministic and stochastic aside, neither Mozart nor d'Arezzo ever faced backlash for their algorithms. However, when renowned composer Lejaren Hiller published the first computer-aided algorithmic composition (CAAC) in 1959, he faced bitter criticism from the media and the musical community. Hiller's *Illiac Suite*, so named after the Illiac Computer which created it, was dismissed as false music, and Hiller was deemed a false musician since he did not write the music himself. What his critics failed to realize was that algorithmic music had already existed for 1000 years. Hiller simply opened the door to new techniques which employed the power of computing (Edwards, 2011).

Methods of Algorithmic Composition

At its core, algorithmic composition is a mathematical endeavor. The power of the computer lies in its ability to quickly perform the immense number of calculations required to produce music. While many different computer algorithms exist for the task, almost all, if not all, of them use one or more of the following techniques: modelling non-algorithmic methods, selecting algorithms from another discipline, or modelling new procedures (Supper, 2001).

Lejaren Hiller's first forays into CAAC simply involved assigning probabilities to predetermined musical events and letting the computer select one, as a human composer would do. This approach is emblematic of modelling non-algorithmic methods of music composition (Supper, 2001). Other examples include d'Arezzo's chant conversion and Mozart's dice game which, as previously discussed, simply involve the application of certain rules on a basis of probability. That said, in the d'Arezzo system, there is a one-to-one mapping between notes and pitches, therefore probability is not a factor (Edwards, 2011). Even in Mozart's system, which involves rolling two 6-sided dice, there is 1 set of rules, with at most 11 possibilities per decision (Mozart, 1956). In contrast, Hiller's earliest compositions make use of 16 separate rules, each of which models a different aspect of music. With the power of computers, Hiller's compositions make a leap in complexity that even surpasses the creations of Mozart, one of the greatest composers of all time.

As mentioned earlier, composition algorithms are based in mathematics. However, modelling natural processes with math is a tricky process. It requires that the algorithm is both structured enough that the output sounds musical and random enough to produce diverse results (Smith et al., 2012). Algorithms of this nature are also required in the field of computational biology. For example, L-system is a type of algorithm developed by biologist Astrid Lindenmeyer to model plant growth. These algorithms possess a quality called self-similarity, which means that the minute aspects of a structure generated by an L-system share the characteristics of the structure at large. In the case of music, this means that a single measure of a piece will exhibit similar patterns to those exhibited over the course of the entire piece. In practice, running an L-system on small musical motifs consisting of pitch and duration produces self-similar music that technically balances structure and randomness well. However, the self-similarity can sometimes be too blatant, enough so that a listener can discern that the music was not created naturally (Supper, 2001).

While many early algorithmic compositions use one of the previously described methods, either copying algorithms designed for another field or attempting to simulate traditional composition, other methods attempt to do more than simply recombine motifs into compositions. One goal of these compositions is to utilize computational inspiration wherein non-musical cues can be converted into music. In some cases, the goal is merely to create music that synchronizes with an input audio file which could be the sound of a bird chirping or Barack Obama's 2004 DNC speech (Smith et al., 2012). However, computational inspiration can also be interpreted as capturing the emotional state of an input file. In this case, the input does not need to be audio. Theoretically, it could be any format that expresses emotion, even another music file (Smith et al., 2012).

Emotion Theories

Human emotions are notoriously difficult to quantify. Many starkly different classification schemes exist to that end; however, most of them are offshoots of two opposing points of view: the dimensional and categorical theories. The dimensional theory of emotion holds that there are infinitely many emotional states each of which is represented by a point on an N-dimensional space, where N is the number of axes that define an emotion (Lerdahl, 2001). For example, one version of this theory holds that emotions are determined by the following axes: Gaiety-Gloom, Tension-Relaxation and Solemnity-Triviality. By this variant of the dimensional theory, any emotional state can be classified by 3 values – one for each axis (Juslin, 2013). In contrast, the categorical theory holds that emotional states fall into distinctly labelled categories. One widely accepted form of this theory is Plutchik's wheel of emotions (Juslin, 2013). The wheel divides emotions into 4 pairs: anger-fear, disgust-trust, sadness-joy, and surprise-anticipation. Within each emotion category, there are different magnitudes. For example, anger contains the subcategories of annoyance, anger, and rage (Juslin, 2013). There are also mixed approaches to the dimensional-categorical dispute. One such theory holds that emotions do exist in an N-dimensional space, however, the space can be divided up into sectors such that each section corresponds to a different emotion category. Given the two examples above, this mixed theory may hold that a high level of gloom, tension, and solemnity is equivalent to sadness (Juslin, 2013).

Though these three are very different approaches, all have merit. The best theory is sometimes arbitrary. For example, expression marks in music are closely linked to categorical representations of emotion, such as "dolce" to tenderness and "furioso" to anger, but no such connection is established between expression mark and various dimensions of emotion. A theory can be selected based on the type of data to be collected (Mauss and Robinson, 2009). This is discussed in greater detail in Section 5: "Measuring Emotional Response".

Emotion in Music

Considering the many theories about emotions themselves, it is not at all surprising that there are also several different views on how music expresses emotions. One major area of study is determining which emotions are most strongly evoked by music. This research assumes a categorical theory of emotion; hence, results are confined to a finite group of labels rather than an infinite number of points. Various studies have produced different results (See Fig. 1); however, happiness and sadness often rank first and second, respectively (Juslin, 2013).

As described earlier, each emotional category contains different intensities. To restate the previous example, anger contains the subcategories of annoyance, anger, and rage (Juslin, 2013). Music tends to elicit less intense emotions such as annoyance rather than anger, serenity instead of joy, and pensiveness instead of sadness. Slight variations in emotion are harder to detect; the consequences of this are discussed in Section 6 (Juslin, 2013).

Emotional coding in music can be separated into 3 categories: iconic, intrinsic, and associative. Iconic coding refers to similarity between music and another signal such as voice or movement. For example, a slide whistle may be reminiscent of jumping, which is why it connotes surprise (Dowling and Harwood, 1986). Iconic coding is understood in a very similar manner by people of many different backgrounds. It is for this reason that music is sometimes termed "the universal language" (Juslin, 2013). Intrinsic coding is a term for emotions evoked by the qualities of the music itself. One example is the relationship between the notes that are played. A minor chord, for instance, may evoke sadness, whereas a major chord may evoke joy. However, this differs between cultures; the Tsimane people of Bolivia, who have little to no experience with Western culture, did not express this view (Undurraga et al, n.d.). Associative coding depends on personal connections to the music. For example, the Star Spangled banner may evoke patriotism (Dowling and Harwood, 1986). The sound of a church organ may represent solemnity or spirituality, simply because of the semantic or episodic knowledge that it ties back to (Juslin, 2013). However, associative coding depends on the listener having certain knowledge. A South African boy who has never heard the Star Spangled banner is unlikely to feel much patriotism due to associative coding when he listens to it.

Table 1 | Ratings of the extent to which specific emotions can be expressed in music.

	Kreutz (2000)	Lindström et al. (2003)	Juslin and Laukka (2004)
Subjects	**50 students**	**135 expert musicians**	**141 volunteers**
No. of emotions Subjects	**32**	**38**	**38**
RANK ORDERING			
1.	**Happiness**	**Joy**	**Joy**
2.	**Sadness**	**Sadness**	**Sadness**
3.	Desire	**Anxiety**	**Love**
4.	Pain	**Love**	Calm
5.	Unrest	Calm	**Anger**
6.	**Anger**	Tension	**Tenderness**
7.	**Love**	Humour	Longing
8.	Loneliness	Pain	Solemnity
9.	**Fear**	**Tenderness**	**Anxiety**
10.	Despair	**Anger**	Hate

Only the ten most highly rated emotions in each study have been included in the Table. Those emotion categories that correspond to the basic emotions are set in bold text. (Anxiety belongs to the "fear family," and tenderness to the "love" family, see, e.g., Shaver et al., 1987.) The original lists of emotion terms contained both "basic" and "complex" emotions, as well as some terms commonly emphasized in musical contexts (e.g., solemnity).

Figure 1. A side-by-side comparison of 3 studies which asked participants to rank emotions based on how much they are expressed by music. Taken from: Juslin, P. N. (2013). What does music express? basic emotions and beyond. Frontiers in Psychology, 4, 596. doi:10.3389/fpsyg.2013.00596

Affective Computing

In the modern day, there is an increasing demand for AI that can communicate emotion. This has led to the rise of affective computing: the field of designing computer systems that can understand an input of human emotions and possibly output emotion as well. Affective computing improves AI's ability to convey and understand emotion. This in turn makes AI better able to connect with its users and gain their trust. In many cases, affective computing may improve AI's ability to complete a task. For example, the military currently uses AI to create training simulations. With affective computing, these simulations may be able to incite empathy or even fear in trainees. In other cases, affective computing will make humans more ready to allow an AI to conduct a task. For example, oftentimes psychological evaluations can be biased because people are not comfortable speaking negatively about themselves in front of a doctor. If an AI were to take on that task, it would be able to collect data privately. However, for patients to trust the AI, it must be able to understand their emotional states and empathize with them just like a human doctor would (McDuff and Czerwinsky, 2018).

The first affective agents, meaning AIs capable of understanding emotions, were dialogue systems. However, up until recent times, dialogue systems were solely text-based meaning the range of emotions that they could express was strictly limited to words. However, new systems such as XiaoIced employ text-to-speech to express a greater range of emotion through voice prosody, and others such as Leonardo can even produce facial expressions, albeit on a screen (McDuff and Czerwinsky, 2018). Affective computing has also led to a new way of looking at emotions. In contrast to dimensional or categorical models, cognitive-appraisal models focus on how emotions affect decision-making. However, faced with the same stimulus, different individuals may have the same emotion, but react differently depending on their personal experiences and the context. This suggests that the ideal emotional model for affective agents is a mixed dimensional-appraisal system. This type of system will be structured enough to measure emotions reliably, while remaining flexible to personalization (McDuff and Czerwinski, 2018).

Measuring Emotional Response

The simplest way of measuring a person's emotional response to a stimulus is to ask them. Formally known as self-report, this technique refers to any form of surveying or questioning where a participant is asked to describe their own emotions. Self-report is found to be more accurate when asking about a participant's current emotional state than when asking about overall emotional personality traits. (Mauss and Robinson, 2009) Furthermore, a disorder known as alexithymia, which afflicts approximately 10% of the population, reduces individuals' awareness of their own emotional state. This can affect their ability to self-report accurately. (Lumley et al., 2001) The Discrete Emotions Questionnaire (DEQ) is a self-report based measure of emotional response. As the name suggests, the survey is based on a categorical model of emotions. It is designed to ask participants about their current emotional state, rather than general personality traits. Generally, self-report is thought to generate better results if used with a dimensional model (Mauss and Robinson, 2009). However, the DEQ is specifically designed to resolve flaws in the Positive and Negative Affect Schedule (PANAS), a dimension-based survey which, until recently, was the industry standard (Harmon-Jones et al, 2016).

Another method of measuring emotion is through autonomic measures. The autonomic nervous system (ANS) regulates bodily functions largely subconsciously. As such, autonomic measures can pick up on slight variations in body state that self-report may not capture. ANS responses such as heart rate, blood pressure, and skin conductance level are good indicators of sympathetic and parasympathetic nervous system activity. As a result, they can measure the arousal level of an emotional response, meaning its magnitude, but they are not accurate measures of emotional category. Furthermore, ANS responses can easily be swayed by non-emotional factors such as external temperature and movement (Mauss and Robinson, 2009).

Closely tied to ANS responses are brain state measures such as EEG and fMRI. EEG studies have linked frontal lobe asymmetry to approach-avoidance tendencies. Greater activation of the left frontal lobe is linked to emotions associated with approach such as anger and worry. Anger often leads to confrontation, and worry leads to problem-solving which is an approach tendency. On the other side, greater activation of the right frontal lobe is linked to avoidance tendencies such as fear (Mauss and Robinson, 2009). fMRI measures blood flow in various parts of the brain. As such, it often has finer resolution than an EEG. fMRI studies focus in on whether activation of particular brain circuits can be linked to emotion. While certain results have emerged such as a link between negative emotions such as disgust and activation in the insular cortex, overall, further research is required to cement a link between brain circuits and specific emotions (Mauss and Robinson, 2009).

The oldest method of measuring emotion, first pioneered by Charles Darwin in the 1800s, is through observation. Darwin proposed that since emotions are a higher form of communication, they should be expressed through body language and speech in some form. Behavioral measures include vocal characteristics, facial expressions, and body language. Traditionally, quantifying these measures depends on observer ratings which may be biased (Mauss and Robinson, 2009). However, recently developed affective agents like those discussed in the last section can reliably measure behavioral response to emotion without observer bias.

Conclusion

To conclude, recent advances in technology have shown the power of affective agents, AI which can express and understand emotion. Through algorithm created music has existed since the 1960s, and has only grown more advanced over time, more research is warranted into whether this music can affect emotions. As stated in the introduction, it is possible that if AI can affect emotions, then it steps towards sentience. This is clear in the alternate name for affective agents: emotionally sentient agents (Edwards, 2011). Through the DEQ, human response to algorithmic music can be accurately logged, and it can finally be determined whether AI not only has the power to communicate with humans by copying their style of speech and language, but also through its own unique musical creations.

References

Allen, A. D. (2016). The forbidden sentient computer: Recent progress in the electronic monitoring of consciousness. IEEE Access, 4, 5649-5658. doi:10.1109/ACCESS.2016.2607722

Dowling, W. J., and Harwood, D. L. (1986). Music Cognition. New York, NY: Academic Press

Edwards, M. (2011). Edwards, M. algorithmic composition: Computational thinking in music. Retrieved from https://www.openaire.eu/search/publication?articleId=od______3094::bff6ccd36e04a8ad002e2ce6e6bce6cb

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. PloS One, 11(8), e0159915. doi:10.1371/journal.pone.0159915

Juslin, P. N. (2013). What does music express? basic emotions and beyond. Frontiers in Psychology, 4, 596. doi:10.3389/fpsyg.2013.00596

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. Journal of Nonverbal Behavior, 43(2), 133-160. doi:10.1007/s10919-019-00293-3

Lerdahl, F. (2001). Tonal pitch space. doi:10.2307/40285402

Lumley, M. A., Neely, L. C., & Burger, A. J. (2007). The assessment of alexithymia in medical settings: Implications for understanding and treating health problems. Journal of Personality Assessment, 89(3), 230-246. doi:10.1080/00223890701629698

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. Cognition & Emotion, 23(2), 209-237. doi:10.1080/02699930802204677

McDuff, D., & Czerwinski, M. (2018, Nov 20,). Designing emotionally sentient agents. Communications of the ACM, 61, 74-83. doi:10.1145/3186591 Retrieved from http://dl.acm.org/citation.cfm?id=3186591

Merriam-Webster. (2019). Definition of sentience. Retrieved from https://www.merriam-webster.com/dictionary/sentience

Mozart, W. A., & Taubert, K. H. (1956). Musikalisches würfelspiel. Mainz [u.a.]: Schott.

Naar, H. (n.d.). Art and Emotion. Retrieved December 5, 2019, from https://www.iep.utm.edu/art-emot/.

Shabbir, J., & Anwer, T. (2018). Artificial intelligence and its role in near future. Retrieved from https://arxiv.org/abs/1804.01396

Shapshak, P. (2018). Artificial intelligence and brain. Bioinformation, 14(1), 38-41. doi:10.6026/97320630014038

Smith, R., Dennis, A., & Ventura, D. (2012). Automatic composition from non-musical inspiration sources.

Supper, M. (2001). A few remarks on algorithmic composition. Computer Music Journal, 25(1), 48-53. Retrieved from https://search.proquest.com/docview/1255386

Undurraga, E. A., Emlem, N. Q., Gueze, M., Eisenberg, D., Huanca, T., Reyes-García, V., & Godoy, R. Musical chord preference: Cultural or universal? data from a native Amazonian society.