REVIEW ARTICLE

Music, Ultrasound, and Artificial Intelligence: Recent Advances and Future Challenges

Helena Dukić

Received on: 02 May 2022; *Accepted on:* 10 June 2022; *Published on:* 18 October 2022

Abstract

Our emotional connection to music is apparent in all living societies and cultures and has as such sparked intrigue regarding its purpose and origin. In recent years, there have been attempts to model and explain our emotional reactions to music. The issue has been approached in two different ways: (1) From a perspective of our prenatal development and early sound perception experiences and (2) From the perspective of artificial intelligence development, recreating the way humans would react to a certain piece of music. Both of these approaches are presented in this paper, along with the possible similarities between the two, as well as the assumed further developments.

Keywords: 4D ultrasound, Artificial intelligence, Fetal emotions, Fetal hearing, Music.

Donald School Journal of Ultrasound in Obstetrics and Gynecology (2022): 10.5005/jp-journals-10009-1933

Introduction

Why do we perceive music in an emotional way? Is there an adaptive value to appreciating music pieces? Music is a valued and widespread component of all known cultural groups, permeating everyday and spiritual life. Researchers[1] have looked for the origin of musicality in the human gestation period, specifically prenatal development,[2] to understand its pervasive and profound influence on individuals. They hypothesized that the complicated web of connections between patterns of sound, movement, and emotion that typify music could have prenatal origins. Humans may learn about these relationships through passive exposure before they are born; the intricate patterns of sound and movement to which they are subjected (the mother's voice, breath, heart-beat, digestion, and body movements) depend in consistent ways on her physical and emotional state.[3] Internal patterns of sound and movement to which the fetus is repeatedly exposed (the mother's voice, breath, heartbeat, digestion, and body movements) depend on her emotional and physical condition in consistent ways.[3] By classical conditioning, a fetus affiliates maternal patterns of sound and movement that are heavily influenced by the mother's emotional state with the mother's hormonal changes. Because behavioral correlates of maternal emotion change (sound and movement patterns) are perceived before physiological correlates (hormonal shifts), the former may predict the latter. The fetus may begin to adapt emotionally to modifications in sound and movement patterns that occur in a similar context after so many repetitions of these patterns in a particular context. To put it another way, it would learn to expect the corresponding hormonal changes that accompany the changes in a sound environment.[1] Postnatal music-like sequences can be created when the patterns of sound and movement formed in motherese mimic natural patterns of sound and movement heard before birth.[1] This is why early human development is such a promising area of study, both for the roots of emotional perception of music and for emotionally caused biochemical changes in pregnant women, which together influence postnatal human development.

Heartbeat, respiration, mother's voice, and footfalls are examples of music features based on sounds found in the uterus. They are analogous to musical structural elements.[1,4] Heart rate is compared to musical pulse, pulse amplitude contour to musical sound amplitude contour, combined heartbeat and footfalls could portray the source of musical meter, maternal voice spoken syllables can be matched to musical notes, and maternal speech prosody could be compared to melodic phrasing Lastly, the frequency range of an adult female voice is comparable to that of melodic instruments.[1]

Department of Musicology, Music Academy, University of Zagreb, Zagreb, Croatia

Corresponding Author: Helena Dukić, Department of Musicology, Music Academy, University of Zagreb, Zagreb, Croatia, Phone: 00385 91 552 4086, e-mail: helenadukic@yahoo.com

How to cite this article: Dukić H. Music, Ultrasound, and Artificial Intelligence: Recent Advances and Future Challenges. Donald School J Ultrasound Obstet Gynecol 2022;16(3):187–191.

Source of support: Nil

Conflict of interest: None

© The Author(s). 2022 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and non-commercial reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Stress causes physiological, psychological, and behavioral variability, changing the usual pulse rate, breathing, digestion, and voice patterns.[5] When subjects are exposed to a stressful stimulus, they typically exhibit a faster heart rate, faster breathing, more active digestion, and a shaky or uneven voice.[6] These changes in physiological traits alter the sounds the unborn child perceives in the uterus, and we expect the fetus to perceive and react to those changes on an evolutionary basis (the significance of bonding and attachment for child survival).

Fetal Correlates of Emotions

On the assumption that the fetus cannot reflect and in that sense cannot "experience" anything, investigation of "fetal emotions" is restricted to physiological correlates of emotions on the assumption that these do not change fundamentally during development and can therefore be extrapolated from adults or children to the fetus. Developments in the field of obstetrics and gynecology have introduced a 4D sonography that enables us to observe fetal movements over time, as well as fetal facial expressions.[7,8] As the unborn baby develops, its facial expressions become more complex; Expressions associated with crying and laughing can already be visible in the third trimester.[7,8] Facial expressions are important in postnatal fetal and maternal interaction and bonding, as well as in the regulation of parental care.[9] Thus, fetal facial movements may reflect internally generated physiological reflex patterns.[10] Even with total forebrain damage or transection, brainstem stimulation can induce smiling, screaming, and crying.[11] In addition to promoting different fetal body orientations, data show that fetal movements and facial expressions convey fetal emotional responses and representations of intentions.[12]

In general, the amygdala mediates emotional memory, attention, arousal, and the experience of love, fear, pleasure, and joy. It has facial recognition neurons, which allow us to acknowledge the emotional importance of the various facial expressions. Furthermore, the amygdala is responsible for the evaluation of faces in social processing. Because the amygdala develops from early embryonic life to an advanced maturation stage during the first postnatal year,[11,13] facial emotion-like expressions or emotion-like behaviors may represent a reliable behavior or some emotional state of the fetus *in utero*.[14] For example, fetuses may be capable of feeling pain and changing their facial expression, such as "frowning" when stressed and "smiling" when happy. Trevarthen and Delafield-Butt[15] proposed that fetuses' facial expressions, in addition to physical movements, provide verifiable data of emotions such as discomfort, curiosity, or pleasure, adapted for communication of preferences and feelings. Reissland et al.[9] discovered that in the third trimester, fetal facial expressions displayed by 4D ultrasound develop into structures that denote a "cry-face-gestalt" or a "laughter-gestalt," expressing emotions that can be conveyed to parents or caregivers right after birth. According to the research, fetal facial expressions are fairly reliable indicators of fetal emotion. However, the specifics of the connection between fetal facial expressions and fetal emotion require further research.

Fetal Response to Induced Maternal Emotions

Araki et al.[16] investigated the link between fetal movement and acute maternal emotional changes during pregnancy. They used two empirically supported feature film clips to elicit happy and sad emotions. Two real-time ultrasound machines (4D ultrasound) were used to monitor separate fetal arm, leg, and trunk movements, while a film clip presentation was used to manipulate maternal emotions. They studied the emotional responses of 22 pregnant women between the ages of 28 and 36 weeks. To avoid any interactions between the effect of emotions elicited by the happy film and the effect of emotions elicited by the sad film, participants were randomly assigned to one of two independent groups (happy film group: $n = 11$ or sad film group: $n = 11$). During the tests, the pregnant women lay semi-supine in a 30 lux-lit room. Film clips were shown on a 21-inch monitor placed in front of the women on a table. The sound of the film was presented to the women *via* headphones. Films were shown in the order: neutral and happy for the happy film group, or neutral and sad for the sad film group. Following each film, the women were asked to rate their emotional state on the two emotional dimensions of happiness and sadness using the visual analog scale (VAS [30]). The number of fetal arm movements increased, but not the duration, when pregnant women were shown a happy film, according to the findings. With the sad film presentation, both the number and duration of fetal arm movements decreased. The presentation of happiness or sadness had no effect on fetal leg or trunk movements. The researchers came to the conclusion that induced emotions in pregnant women mainly affect fetal arm movements, and that positive and negative emotions have contrary effects on fetal movement.

Another study, conducted by DiPietro et al.,[17] looked at the effect of induced maternal stress on the fetus. Stress was induced in mothers at 24 and 36 weeks gestation using the Stroop color-word task. The responses of the mother (heart rate and skin conductance) and the fetus (heart rate, heart rate variability, and motor activity) were assessed. The findings suggested that fetal responses to maternal sympathetic activation induced by a cognitive stressor were routinely disrupted by maternal environmental intrusions: fetuses responded to maternal stress with increased variability in heart rate and decreased fetal movement, contrary to expectations. The magnitude of the fetal response increased during pregnancy.

Finally, during the 32nd week of pregnancy, DiPietro et al.[18] investigated the fetal response to induced maternal relaxation. The mother's heartbeat, skin conductance, respiration period, and respiratory sinus arrhythmia changed significantly after an 18-minute guided imagery relaxation session. Major changes in fetal neurobehavior were also observed, including decreased fetal heart rate (FHR), increased FHR variability (which showed a linear increase from prebaseline to recovery, indicating a prolonged effect of maternal rest), and suppression of fetal motor activity (FM), and increased FM-FHR coupling.

Future Challenges

These examples demonstrate that the mother's induced emotions do have an influence on fetal behavior and reactions. However, to our knowledge, there has been no research done evaluating the fetus's facial expressions to induce maternal emotions. Due to the most recent developments in sonography, it is now possible to evaluate the facial expressions that emerge, so a step further in the exploration of the mother-fetus emotional relationship would be to evaluate the influence of induced motherly emotions on the emotional state of the fetus. This could be done by inducing a particular emotion in the mother (positive valence and high arousal vs negative valence and low arousal) using music and simultaneously monitoring fetal facial expressions and movements to see if there are any correlations between fetal emotional expressions and in mother's emotional reactions. If there is indeed a correlation between fetal emotional expressions and the mother's emotional states, the next step would be to evaluate if the intrauterine sound is the transmitter of the emotional content from the mother to the fetus. Apart from offering an insight into the possible prenatal developments of our emotional relationship to music, this future research could also offer an explanation of the way in which fetuses learn to adapt and prepare for possible events in the future.

Music and Artificial Intelligence (AI)

Music has been described in various ways, but one of the most common in musicological research is the idea that music is an intellectual process; the ability to identify patterns and imagine them altered by actions while having emotional responses that emerge from this intellectual activity.[20] This ability is not limited to music; it is an essential part of the human mind. Researchers attempt to decode the inner workings of both music and intelligence by studying music models. AI, on the other hand, has a rather similar task: to gain a deeper understanding of human intelligence by studying "intelligent agents"; systems that comprehend their surroundings and take action to increase their likelihood of success. Various subfields of AI research are centered on concrete objectives and the use of specific tools. Reasoning, knowledge representation, planning, learning, natural language processing, and perception are some of the most common goals of AI research. One of the field's long-term goals is to develop a model of general intelligence (the ability to solve any problem). AI researchers have adapted and incorporated a number of different techniques and tools to resolve these issues, including search and mathematical optimization, formal logic, artificial neural networks, statistical methods, probability, and economics. AI also incorporates knowledge from psychology, linguistics, philosophy, and a variety of other disciplines. The field was founded on the premise that "human intelligence can be so precisely described that a machine can be built to simulate it."[21] Music, which involves both cognitive tasks and emotional reactions, is thus among the most intriguing areas of AI research. The ultimate objective of music and AI research is to make computers behave like professional musicians, performing highly specialized tasks such as composition, analysis, improvisation, instrument playing, and so on, as well as less specialized tasks such as being able to have an opinion on a piece of music, engage in conversation with fellow musicians, and have emotional reactions to music.[20] To be able to do just that, music AI machine would need to understand basic human emotions of sadness and joy.

How does AI Learn Music?

In order to replicate this particular type of human behavior (having emotional reactions to musical material), researchers need to understand two points: the building blocks of music along with the rules that bind them (musical grammar), and the type of emotional reactions particular "grammatical" constructions elicits. The notion of formal grammar sprung from AI research and spread to the musical domains, giving rise to new musical theories.[22] They first appeared in the 1950s, when linguist Naom Chomsky released a book called Syntactic Structures,[23] which proposed that individuals can only speak and understand language because they have mastered its grammar. In order to fully describe how this grammar works, it must be based on mathematical rules. Chomsky also argues that it is possible to define a universal grammar that could be applied to all languages. The connection between music and language has always been studied, so it is not unusual that linguistics, particularly formal grammar, has had a strong influence on music and AI research. Many researchers believe Chomsky's theories also apply to music.

Parncutt & Chuckrow's theory of prenatal sound patterns has some parallels with the AI notions of "musical grammar": in both cases, there is a learning curve in which the subject (AI or a fetus) associates sound patterns (not yet thought of as music) with certain emotions. By means of repetition and continuous exposure to a huge amount of similar stimuli (data), the subjects acquire "knowledge" that enables them to react "instinctively" to sound patterns (music). Since we have already looked at the ways in which prenatal sound patterns relate to music, it would be time to examine how machines come to learn the musical grammar that they later use to compose, play or understand music. In order to define a grammar for music, one should consider the constitutes of the rules: notes, the direction of the consecutive notes, intervals, duration, harmony, etc. There are two types of rules: generative rules (generates notes from scratch) and transformational (changes the already generated notes).[20] For example, a generative rule might indicate the melody in the following way: The notes of the sequence are calculated from a given reference note in this order: the interval is a perfect fourth and the direction of the interval is upward. A transformational rule of the above passage could for example create a new sequence by joining two sequences into a single event. In this way, a computer could be programmed to produce entire musical compositions by successive activation of generative and transformational rules.[20] This system of acquiring information is similar to what a fetus might be exposed to prenatally and how it might perceive it: noise patterns consisting of sounds that have different durations and frequencies, connected to each other by certain grammatic rules. The fetus learns these rules by exposure and cultivates them subconsciously into, what we call, an innate musical ability.

AI Music Emotion Recognition

What lacks in the explanation of the AI acquisition of musical skills, however, is the acquisition of emotional reactions to the generated musical material. As we have seen, a fetus presumably acquires emotional reactions by a process of conditioning: a hormonal surge that causes the emotional reaction is preceded by a specific sound pattern and after many repetitions, the connection forms. The fetal emotional reactions are presumably very simple at this stage, consisting of what we might describe as happy or relaxed states and unhappy or tense states.[7] Stern[24] explained that early experiences of interaction between a mother and a baby also reflect this emotional exchange: a line of dramatic tension based on the difference between waiting for food (tension) and being fed (relaxation).[25] He called those short interactions "tiny narratives" or "proto-narrative envelopes" and proposed that they represent a basic unit or a building block for constructing the representational world. Narmour's[26] implication-realization model of music used the idea of a musical plot comprising alternating tension and relaxation to develop a musical grammar that created expectations and then resolutions, mirroring the proto-narrative exchange Stern proposed. Marvin Minsky, an AI researcher, proposed in a similar manner that "Our listening machine would have to understand music well enough to recognize from each moment to the next which problems in music have been solved (relaxation) and which remain open (tension, expectation)."[27]

Much like with the fetuses, in order to teach machines which music features create the basic emotional responses of tensions and relaxations, we need to provide them with a sort of "parental guidance." This is done in the following way: To represent the acoustic property of a music piece, the building blocks of music such as timbre, rhythm, and harmony are typically extracted from pieces. Psychologists have mapped out the musical parameters which affect the perceived emotion of music. A subjective test is typically used to gather the true data required for training the computational model of emotion prediction. The subjects are told to report their emotional reactions to the music they are listening to. Because emotion perception is subjective, each piece is annotated by multiple subjects, and the ground truth is determined by popular vote. After gathering emotional data and extracting music features, several classification algorithms are used to teach machines the relationship between music attributes and emotion labels. Neural networks, Gaussian mixture models, support vector machines, and other machine learning methods are among these algorithms.[28] Machines, like human children and fetuses, learn better when they are exposed to more data, examples, and situations that are reiterated to them. Following training, the automatic model is used to classify the emotion of a music piece.

Future Challenges

Due to the subjective nature of the particular topic (music emotion recognition), developing a reliable and generalizable computer model for emotional recognition remains a significant challenge. Currently, research efforts are being made in order to understand what a particular algorithm is learning from the data presented to it and how the emotional predictions for a particular input are being made. For an uninformed person, this might seem like a trivial task, but it is no simpler than asking a human "How does your brain come to form sentences and words in a way it does?." Since none of us can answer this question easily or at all for that matter, we cannot expect a program to display a similar level of self-awareness. Recently, there have been significant developments in this field; Chowdhury et al. have found a way to debug a biased emotion prediction model by forming an intuitive connection chain between the input audio and the out emotion predictions, allowing them to get an explanation of why a program recognized certain emotions from the musical data presented.[29] This problem is not unique to musical AI research but persists in any algorithm that is trying to reach an independent decision of a presented data. To a certain extent, it mirrors the situation presented in the fetal emotional sound perception; We can see that certain emotional correlates to emotions are being presented (facial expressions), but we do not know how they came to be. Hence, efforts should be made in both fields to understand the "unconscious" workings of humans and machines, and to untangle the "intuitive" decision-making process.

Conclusion

In the end, we can see that the process of acquiring musical skills in humans and in machines has certain parallels. The evolution of AI music processing and emotional recognition can benefit the field of music research by offering new ideas and insights into what the learning process might entail. Thus, it is crucial that the seemingly separate fields of AI, music, and fetal research work together in order to feed on each other's insights. Individuals store knowledge in a highly complex manner: our brains are many-layered systems that cannot be neatly disassembled. Our intelligence is made up of both conscious and unconscious components that are distributed at various levels in our psyche. Although the conscious mind can be impartially accessed and manipulated, the unconscious elements are more difficult to access, despite the fact that they do the majority of our thinking.

The research concerning both fetal emotional perception and AI music emotion recognition concerns the study of processes that result in what we think of as intuitive or unconscious reactions. Since unconscious processes, as mentioned above, govern most of our behavior, it is crucial for us to understand how they come to be, and the fetal and AI data presented in this paper can pave the way to that understanding.

References

1. Parncutt R, Chuckrow R. Chuckrow's theory of the prenatal origin of music. Musicae Scientiae 2019;23(4):403–425. DOI: 10.1177/1029864917738130
2. Parncutt R. Prenatal and infant conditioning, the mother schema, and the origins of music and religion. Musicae Scientiae 2009;13(2_suppl):119–150. DOI: 10.1177/1029864909013002071
3. Mastropieri D, Turkewitz G. Prenatal experience and neonatal responsiveness to vocal expressions of emotion. Dev Psychobiol 1999;35(3):204–214. DOI: 10.1002/(sici)1098-2302(199911)35:3<204::aid-dev5>3.0.co;2-v
4. Teie D. A comparative analysis of the universal elements of music and the fetal environment. Frontiers in psychology 2016;7:1158. DOI: 10.3389/fpsyg.2016.01158
5. Kim H, Lee MH, Chang HK, et al. Influence of prenatal noise and music on the spatial memory and neurogenesis in the hippocampus of developing rats. Brain Dev 2006;28(2):109–114. DOI: 10.1016/j.braindev.2005.05.008
6. Van Puyvelde M, Neyt X, McGlone F, et al. Voice stress analysis: a new framework for voice and effort in human performance. Front Psychol 2018;9:1994. DOI: 10.3389/fpsyg.2018.01994
7. Kadić AS, Kurjak A. Cognitive functions of the fetus. Ultraschall Med 2018;39(2):181–189. DOI: 10.1055/s-0043-123469
8. Kurjak A, Antsaklis P, Stanojevic M, et al. Multicentric studies of the fetal neurobehavior by KANET test. J Perinat Med 2017;45(6):717–727. DOI: 10.1515/jpm-2016-0409
9. Reissland N, Francis B, Mason J, et al. Do facial expressions develop before birth? PLoS One 2011;6(8):e24081. DOI: 10.1371/journal.pone.0024081
10. Merz E, Pashaj S. Fetal facial expressions: demonstration of the smiling, the sad and the scowling fetus with 4D-Ultrasound. Ultraschall Med 2015;36(1):1–2. DOI: 10.1055/s-0034-1369546
11. Joseph R. Fetal brain behavior and cognitive development. Developmental review 2000;20(1):81–98. DOI: 10.1006/drev.1999.0486
12. Delafield-Butt JT, Trevarthen C. 11 theories of the development of human communication. Theories and models of communication. 2013 Jan:199.
13. Humphrey T. The development of the human amygdala during early embryonic life. J Comp Neurol 1968;132(1):135–165. DOI: 10.1002/cne.901320108
14. Hata T, Hanaoka U, Tenkumo C, et al. Three- and four-dimensional HDlive rendering images of normal and abnormal fetuses: pictorial essay. Arch Gynecol Obstet 2012;286(6):1431–1435. DOI: 10.1007/s00404-012-2505-1
15. Trevarthen C, Delafield-Butt JT. Autism as a developmental disorder in intentional movement and affective engagement. Front Integrative neuroscience 2013;7:49. DOI: 10.3389/fnint.2013.00049
16. Araki M, Nishitani S, Ushimaru K, et al. Fetal response to induced maternal emotions. J Physiol Sci 2010;60(3):213–220. DOI: 10.1007/s12576-010-0087-x
17. DiPietro JA, Costigan KA, Gurewitsch ED. Fetal response to induced maternal stress. Early Hum Dev 2003;74(2):125–138. DOI: 10.1016/j.earlhumdev.2003.07.001
18. DiPietro JA, Costigan KA, Nelson P, et al. Fetal responses to induced maternal relaxation during pregnancy. Biol Psychol 2008;77(1):11–19. DOI: 10.1016/j.biopsycho.2007.08.008
19. book. Luger GF. Artificial intelligence: structures and strategies for complex problem solving. Pearson education; 2005/book
20. Miranda ER. (Ed.). Readings in music and artificial intelligence. Routledge; 2013.
21. McCarthy J, Minsky ML, Rochester N, et al. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine 2006;27(4):12. DOI: 10.1609/aimag.v27i4.1904
22. Lerdahl F, Jackendoff R. A generative theory of tonal music. Cambridge, MA; MIT Press; 1983.
23. Chomsky N. Syntactic Structures, The Hague/Paris: Mouton, 1957, ISBN 978-3-11-021832-9.
24. Stern DN. The 'pre-narrative envelope': an alternative view of 'unconscious phantasy'in infancy. Bull Anna Freud Centre 1992;15(4):291–318.
25. Stern DN. The first relationship. Harvard University Press; 2002.
26. Narmour E. The analysis and cognition of melodic complexity: The implication-realization model. University of Chicago Press; 1992./book
27. Balaban M, Ebcioğlu K, Laske O, editors. Understanding music with AI: perspectives on music cognition. MIT Press; 1992 Aug 17.
28. Yang YH, Chen HH. Music emotion recognition. CRC Press; 2011 Feb 22.ž
29. Chowdhury S, Praher V, Widmer G. Tracing back music emotion predictions to sound sources and intuitive perceptual qualities. arXiv preprint arXiv:2106.07787. 2021 Jun 14. DOI: 10.48550/arXiv.2106.07787