Similarities Between Language and Music May Inform Text-Generating AI

Language and music do not strike the casual observer as similar; it would be hard to confuse a Mozart symphony with an Obama speech. But structural parallels between linguistic and music theory and recent findings in neuroimaging show that the way we learn, understand, and produce music and language are not all that different. Capitalizing on their likeness may have far-reaching implications for text-generating AI systems.

As they relate to the human mind, language and music both (1) rely on syntactic structure and (2) involve the same brain regions for processing. Musical and linguistic syntax are hierarchical structures that organize inputs into predictable, meaningful patterns. In language, speech sounds are combined to form words, words are arranged to form sentences, meaning is derived from sentences, and contexts are interpreted from meaning. It is the linguistic syntax -- the set of rules that arranges words and phrases -- that connects a language's structure with its meaning (Jackendoff, 2009). Similarly, in music, multiple levels of organization give rise to meaningful sequences. Tones are combined to form scales, specific notes of scales are harmonized to create chords, chords that share scalar properties comprise a key, a key's tones and chords are organized into song-form[1], and from song-form the listener derives a sensuous experience. The musical syntax -- song-form, the ordering of tones and chords -- links music's structure with its meaning in the same way that linguistic syntax does (Patel, 2010).

Mapping language and music processing in the human brain has revealed surprising overlap and breadth. Research during the 1960s found that patients with damage to certain areas

[1] Song form is the structure of a song that outlines its repetitions, harmonies, melodies, and rhythm.

of the left hemisphere developed aphasia, while those with damage to certain areas of the right hemisphere suffered from amusia (Kleist, 1962). These results led to the conclusion that the left hemisphere was strictly responsible for language and the right for music. Specifically, it was thought that language was restricted to Broca's area, which governs the meaning of words, and Wernike's area, which dictates syntactic structure. But with the advent of fMRI, scientists discovered that Broca's and Wernike's areas were also activated when processing music both aurally and visually (Steinbeis et al., 2008). Further fMRI studies have shown that regions of the right hemisphere operate during language and music processing as well. It has since been suggested that the left and right hemispheres have broad functions that language and music both recruit. The left hemisphere specializes in hierarchical sequencing like syntax and meaning while the right processes contour-based patterns and emotional affect. Put simply, the left hemisphere identifies denotation and the right interprets connotation (Harvard-Smithsonian, 2010).

The similarities between language and music reveal symbolic and physical structures that the human brain uses to create and communicate: rigid syntax organizes structural components, the left hemisphere processes its literal meaning, and the right hemisphere extrapolates this meaning.

Understanding and modeling these building blocks has allowed music-producing AIs to get off the ground. IBM Watson Beat, a neural network that composes original music, is a foremost example. The system learns by consuming audio files broken down into their core syntactic elements, which are linked with information on emotions and musical genres. These structural reference points allow IBM Watson Beat to interpret and classify music; they also resemble the building blocks of language and music used in the human brain. The project started

in 2016, and by 2018 IBM Watson Beat was reliably producing coherent, original music (IBM, 2018). Today, some music artists like Taryn Southern use music-producing AI as inspiration (Deahl, 2018) while others are using it to create entire songs or albums, such as the first AI-produced album "Hello World" (Hello, n.d.). There is still room to grow and perfect music-producing AI, but recent progress is reflective of an accurate understanding of how the human brain processes music.

IBM Watson Beat's demonstrated comprehension of music theory is remarkable, but its ability to form connections between musical elements and emotions are what allowed it to pioneer artificial music production. Previous attempts to create a music-producing AI overlooked the right hemisphere's role in music contextualization and as a result could not create cohesive pieces. The connection between musical data points and emotion is so integral because the human brain does not create music out of thin air. Rather, the musician draws on sensuous experience as the muse for a piece. When music is created, human experience is translated to musical syntax using emotion. For instance, "broken up with" is represented by minor keys and "sexual pleasure" is represented by major keys. Conversely, when music is listened to, musical syntax is translated to an affective human experience by interpreting emotion in the right hemisphere of the brain. The right hemisphere uses emotion as the mediator between human experience and musical syntax, which gives a musical piece coherence (Madell, 2002).

Text-generating AIs, on the other hand, have not seen as much success. GPT-2, produced by OpenAI, is the most advanced text-generating AI to date. Given a prompt, the neural network's objective is to predict the most likely next word. GPT-2 produces impressive text, often many paragraphs long, that can convince the naive reader. This step forward in narrative

capacity has put GPT-2 ahead of other text-generating AI like Word2vec and Google Translate, which struggle to create any narrative at all. But there is still an element of creativity and coherence missing. GPT-2 needs a thorough prompt to generate text from, and even with that crutch, its logicality eventually falls off the rails (Radford, 2019). GPT-2 moves as quickly as possible between words to decide which is most likely to follow in sequence, but strict mathematical probability omits any consideration of the meaning that mediates language (Hofstadter, 2020).

I propose that the shortcoming of GPT-2 and other text-generating AIs may be that they overlook the right hemisphere's role in understanding and creating language. Since the right hemisphere is responsible for extrapolating connotations, a lack of coherence among outputs of text-generating AIs could be attributed to a lack of context that the right hemisphere usually provides. Like music, words and sentences are not spontaneously created. They are guided by emotions, experience, and context, which the right hemisphere uses in translation with linguistic syntax. When music-producing AIs adopted mechanisms to imitate the right hemisphere, the coherence of their outputs dramatically improved. While music is more mathematical than language, and thus music-producing AIs are easier to program than text-generating AIs, language's lack of mathematical structure only means that understanding it requires relying even more so on the right hemisphere's contextual processing. Given the similarities between language and music processing in the human brain, and the need for a meaning-mediator in text-generating AI, perhaps the same solution that worked for music-producing AI can be applied to text-generating AI. Further work will need to be done to identify the specific roles of the right hemisphere in language processing and how it mediates language.

Works Cited

Deahl, D. (2018, August 31). How AI-generated music is changing the way hits are made. https://www.theverge.com/2018/8/31/17777008/artificial-intelligence-taryn-southern-amper-music

Harvard-Smithsonian. (2010). *The Neuroscience of Teaching and Learning*. Annenberg Media. Un. 1 Sec. 3.

Hello World Album (n.d.). https://www.helloworldalbum.net/

Hofstadter, D. (2020, January 23). The Shallowness of Google Translate. Retrieved from https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/

IBM Watson Beat. (2018). https://www.ibm.com/case-studies/ibm-watson-beat

Jackendoff, Ray. (2009). Parallels and Nonparallels between Language and Music. Music Perception - MUSIC PERCEPT. 26. 195-204. 10.1525/mp.2009.26.3.195.

Kleist, K. (1962). Sensory aphasia and amusia: The myeloarchitectonic basis. Pergamon.

Madell, Geoffrey (2002). Philosophy, music and emotion. Edinburgh University Press.

Patel, A. D. (2010). *Music, language, and the brain*. Oxford: Oxford University Press. Ch. 5.1 - 5.3.

Radford, A. (2019, December 13). Better Language Models and Their Implications. https://openai.com/blog/better-language-models/

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and FMRI provides evidence for similar and distinct neural representations. *PloS one*, *3*(5), e2226. https://doi.org/10.1371/journal.pone.0002226