Contextual Chord Embeddings for Stylistic Analysis

Matt Chiu, Eastman School of Music

Handout: https://www.shrunken.com/ChordEmbeddingsHandout
Code: https://www.shrunken.com/ChordEmbeddingsCode

---

[1] Hi all, and welcome to my talk on chord embeddings. The link to the handout is located on the bottom of the first few slides. You won't need the handout, but if you're interested in looking over the visuals, using my code for yourself, or reading some of the "mathy" details that I didn't include in the talk, you might consult it.

Acknowledgements
[2] First off, I'd like to thank Diane Litman for introducing me to the field of natural language processing, Eron Smith and Davy Temperley for feedback and revisions, and especially Noah Kahrs for going back and forth exchanging code during the projects' brainstoriming phase.

I'd also like to acknowledge that I'm recording from Rochester, NY—and from unceded land of the Seneca peoples. I ask you to join me in recognizing the lands founded upon exclusion and erasure of many indigenous peoples, and to acknowledge the Seneca community, past present and future generations.

[3] The ultimate aim of this paper is to study harmony through what I call *computational hermeneutics*—that is, chords contain contextual meaning—be it implications, references, contexts of occurrence, etc.—so how do we encode such deep levels of meanings into a computer? Before getting into the nitty gritty details of the theory and this paper, I want to take a step back to talk about complications in harmonic studies, the assumptions we make, and how we (meaning theorists) talk about them.

When we talk about chords, we have levels of representations (**Figure 1**). [4.0] For example, the "G major triad" is really a *class* for labeling objects [4.1] that fulfill certain criteria. [5.0] But that class also gains conditional meaning when discussed within different keys [5.1]: I in G major or bII in F# major have different *contextual* representations. [6] We might zoom out once again and see that most theorists talk about these meaning within another context: of *style*, and [7] where each style occupies its own set of contextual chord meanings.

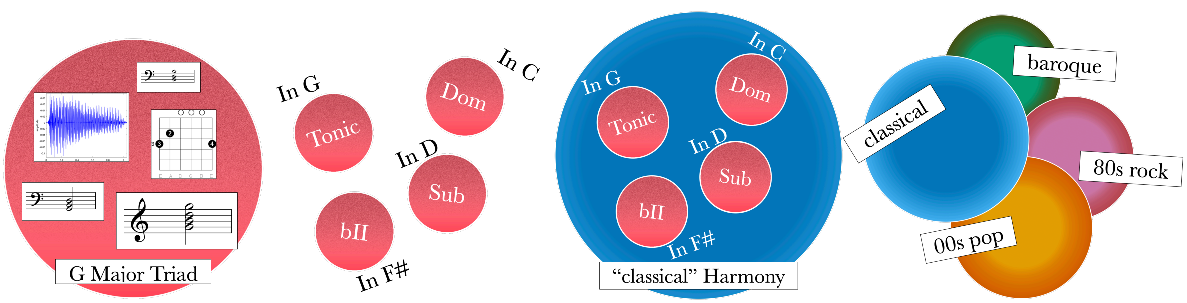

Figure 1. Different representational levels: (from left to right) class, contextual, stylistic, inter-stylistic

[8.0] Most music theory discussing harmony has concerned itself with common-practice western music. In this context, we often attribute chord meaning, or identity, with *function*—as shown in [8.1] Agmon (1995), Harrison (1994), Tymoczko (2011), and others. In fact, as discussed in Kopp (1995), Hugo Riemann's initial term *Funktion* concerns the [9] "meanings of chords." The term "function" has been used in so many different ways, but to oversimply the discussions [10]: chord class, and context are often grouped together under discussions of function.

[11] Prescriptive definitions of chord meaning are powerful and predictive, but we need to recognize their limited applicability: these chord meanings aren't universal and are specified to their stylistic milieu. [12] For example, this chord loop in Guns N' Roses "Sweet Child O' Mine" is totally normative in 80s rock, but aberrant in classical harmony [12.1. Let's listen. Reacting to theories incompatible with a certain musical style, [13] pop/rock music researchers have instead argued for separation of chord class and function, by syntactically-informed functions.

[14] Perhaps the most systematic study on chord context and function has been [14.1] White and Quinn (2018). They showed that there are differences between functional categories for different musical styles using Markov models and probability.

Objectives
[15] So, our goal is to create a generalized methodology for studying harmony systematically, and to do that, we need to create chord representations for computers. From what I just showed, [16] we want the method to group harmonies into chord *classes*, derive chord meaning from contextual placement, and recognize separations of style. Using the following methodology, I show that [17] *chord embeddings* can do just this.

Outline
[18] In this presentation, I present the [19] theory side of things first. I'll explain what [19.1] embeddings are broadly and then talk about my methodology [19.2] and how embeddings might be used in music research. [20] I'll then give some implementations of the theory, demonstrating applicability on [20.1] common-practice music and then on [20.2] rock music.

Embeddings
[21] So I've been using this word "embedding"—where does it come from and what does it mean? [22] Embeddings originated in the field of [22.1] *natural language processing*, or NLP for short. NLP is an intersection of linguistics, computer science, and artificial intelligence. In NLP research, [22.2] embeddings are numerical vector representations for words—kind of like a unique label for each word. [22.3] Vectors are lists of numbers where there are as many dimensions as the vector is long (**Figure 2**). An interval vector is technically a 6-dimensional embedding for pitch-class sets! Embeddings are usually in too many dimensions for us to imagine, so rather than just study the vectors themselves, it's mostly useful to *relate* them to one another. So you can think of embeddings as vectors, and in this talk I'll often use the terms interchangeably. Okay, but why do we need to represent words as vectors?

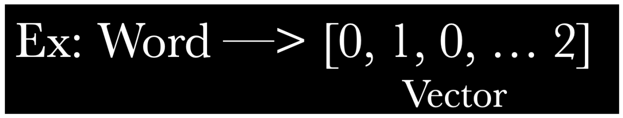

Figure 2. Word embedding

[23] Well, text might be easy to code into a computer, but representing *meanings* and relationships between words is [23.1], to say the least, tough. [24] NLP found, however, that computers are much better at understanding words if they're all encoded as vectors [24.1]. Enter, the word embedding! [25] You make use of word embeddings every day on your phones, for autocomplete on Facebook, for speech recognition, and countless other processes. Note, though, that there are are tons of ways to actually make these word embeddings. In this paper, I'll be using a common method from machine learning [26] called word2vec, pioneered by Mikolov et al. (2013).

[27] In music, word2vec has been used in studies on music information retrieval or cognitive experiments. For example, it's been used to [27.1] create embeddings that are then compared to subject responses of harmonic tension (Nikrang, Sears, and Widmer 2017), or [27.2] create the encodings to train a Bach choral generator (Liang, Gotham, Johnson, and Shotton 2017) .

[28] The thought process behind word2vec is that things gain meaning from the context they appear. This is echoed through a lot of other scholarship on language, such as [29] J.R. Firth when he said "You shall know a word by the company it keeps," and similarly Ludwig Wittgenstein's quote "The meaning of a word is its use in the language."

Take this sentence, from a Philip K. Dick short story [30]: "They had been playing in the shade of an immense **bengelo**, whose ancient branches drooped and swayed…" To my knowledge, the word "bengelo" doesn't exist outside of this story. But from context, we can infer it conveys some "tree-ness" quality. [31] So because [31.1] **bengelo** and **tree** both occur near words and phrases like [31.2] "branches," "sway," and "in the shade," [31.3] we're able to infer they have [32] similar meanings or occur in similar contexts (**Figure 3**). [33] Using word2vec in music suggests that chord meaning is also developed through contexts.

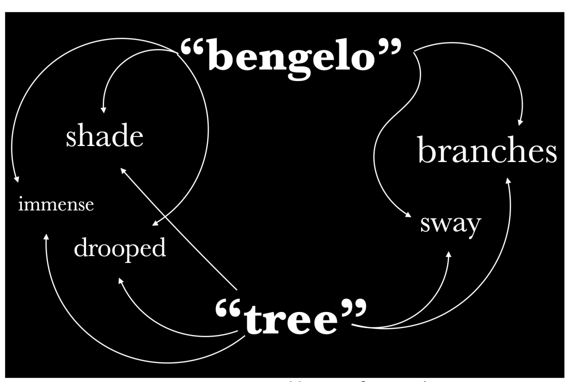

Figure 3. Contextually inferred meaning

Methodology

[34] That's the underlying theory to word2vec and chord embeddings—next we need the actual methodology. As I said earlier, rather than going into the mathematical

details of word2vec here, I'll focus on interpretive takeaways. For those interested in the math, I've put a few more details in the handout.

[35] The end goal of the methodology is to [35.1] develop unique chord embeddings for each chord, and, from that, study the resulting chord space. [36] Since chord embeddings are multidimensional vectors, [36.1] we can go into the model and navigate how the chords are related [36.2] within this really high-dimensional space. The hope is that the chord space should tell us something about chord proximity and expectation.

[37] So to develop each set of chord embeddings, we need corpora to use word2vec on, and a way to encode the chords into the machine to identify it.

[38 In this paper, I'll use the same encoding methodology on two different corpora: The [38.1] Yale Classical Archives Corpus and the [38.2] Rolling Stone corpus to train two sets of chord embeddings.

[39] The encoding methodology used here follows other vertical-segmentation procedures like [39.1] Cuthbert and Ariza (2010) or White and Quinn (2018). Given a musical corpus made up of pieces [39.2], chords are encoded as pitch classes of vertical slices [39.3]. The process chunks the pieces so anytime a new pitch class is introduced, it records all pitch classes present at that current time as a chord slice. [39.4] My encodings also includes a "key index," which records the current key as well (**Figure 4**). So, playing on the terminology "salami slicing," [39.5] I call each of these verticalities "key-lime" slices.

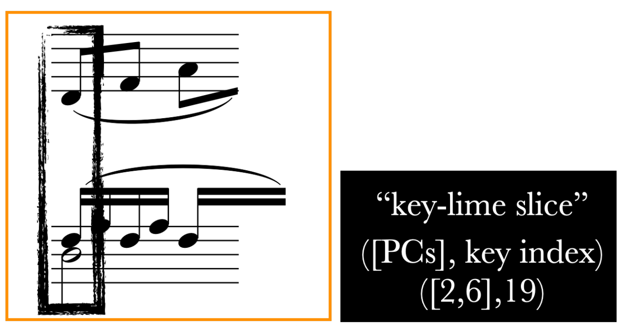

Figure 4. Key-lime slice example

[40] To create chord embeddings, first, each possible chord [40.1] in the corpus is initially assigned an embedding with a randomized 30-dimensional vector [40.2]. The randomized vectors are just a placeholder for the embeddings we want eventually. Word2vec is going to scan over each piece in the corpus to update and "learn" embeddings—this is where the machine learning tools will come in handy (**Figure 5**). [41] To learn embeddings, we go through [41.1] every chord in the corpus calling each one a "target chord" and using the surrounding chords as "context chords." [42] Using just the target chord's chord embedding, the model tries to predict [43] the surrounding "context chords." These predictions are based on the "dot product" between vectors—which is just a mathematical way of measuring vector similarity. The actual learning happens next. [44] Word2vec uses those predictions and compares the predictions to what chords *actually* occur in the music. [45] The embeddings updated so that chord embeddings that occur near each other are now closer together in their vector space.

[46] We repeat this process for every chord in each piece, and, on top of that, [46.1] iterate through the corpus a few times. Afterwards, we have a set of chord embeddings that capture each chord's contextual placement in the style of music, where the "meaning" of a chord based on where we would expect to hear it. [47] With the methodology covered, we can move into the application.

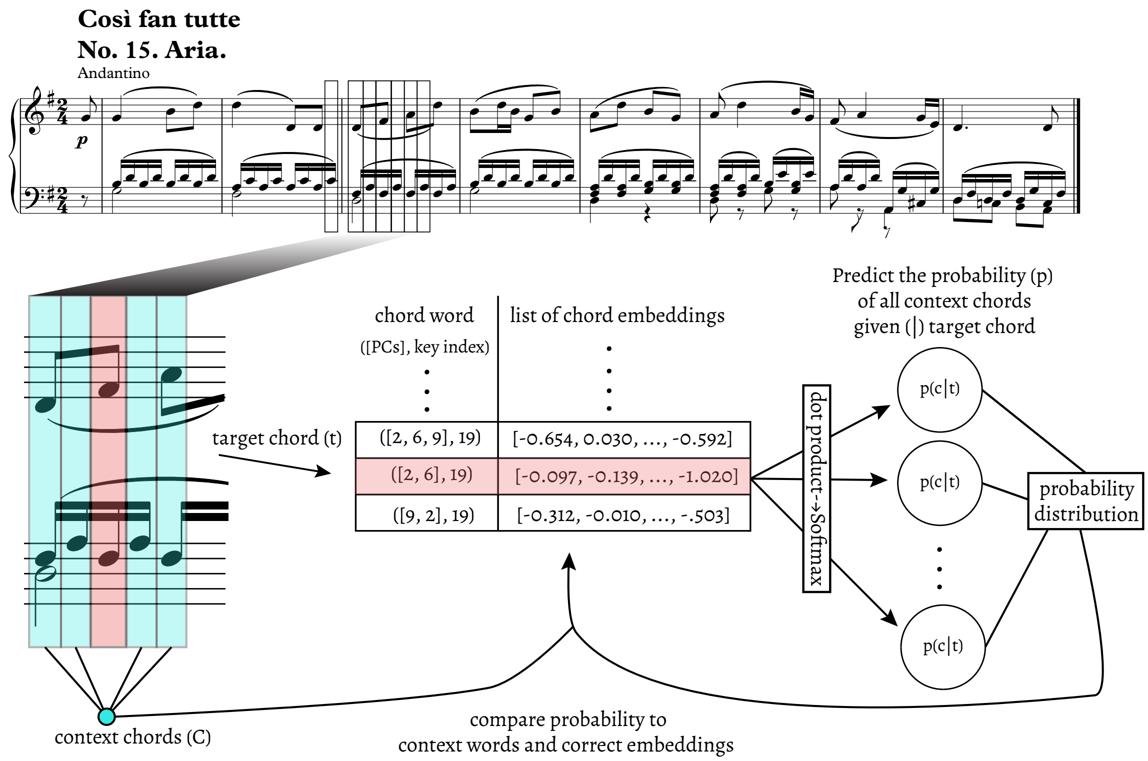

Figure 5. Chord embedding methodology

Mozart model

[48] The first set of chord embeddings was trained on a subset of the Yale-Classical Archives corpus, using only pieces by Mozart. [49] Let's examine some of the theoretical aspects of these chord embeddings, starting with [50] keys and key relationships. Since the embeddings are in a vector space, we can use mathematical operations to compare them. The three primary measures I'll be using here are called Euclidean distance, cosine distance, and cosine similarity—they're operations for calculating similarities between vectors. The distance measures show how different vectors are from one another, and cosine similarity shows how similar. [50.1] So I took the tonic triads from the twelve major keys, and [50.2] clustered them by distance (**Figure 6a**). The idea is that embeddings will be clustered together into groups of nearby keys. [51] Shown in this dendrogram, we see C major is closest to F major and G major—basically a confirmation of traditional music theory's fifth-based key primacy. In fact, [52] on closer inspection, key distances with embeddings mirrors the circle of fifths (**Figure 6b**).

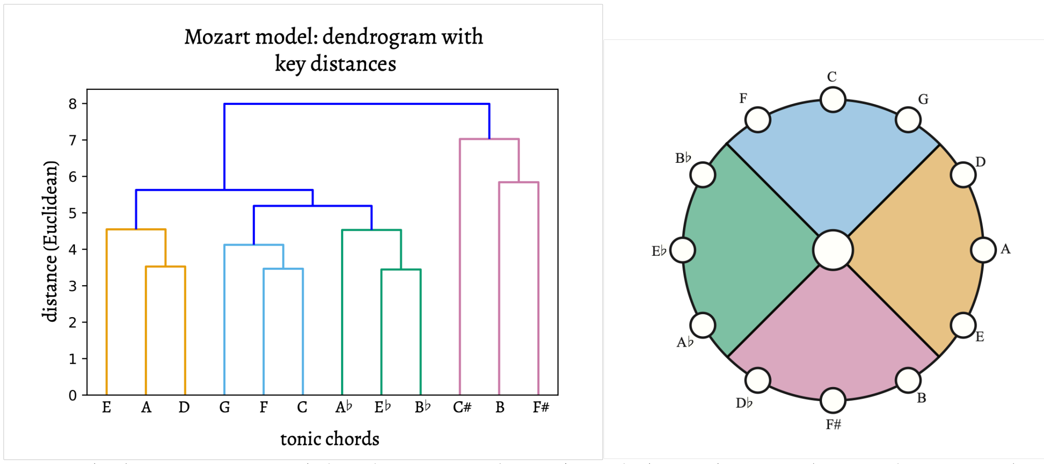

Figure 6a and 6b. Mozart model: Clustering keys (Euclidean distance), resulting in the circle of fifths

[53] I also calculated the similarity of the tonic in C major to each of its diatonic chords *within* C major (**Figure 7a**). And, again, we see that this mirrors music theory's traditional fifth-related triads with F major and G major identified as the closest chords to the tonic. [54] To compare how these chord distances relate to other metrics, I correlated them to Lerdahl's *Tonal Pitch Space* chord distances (2001). His model relies on common-tones and fifth-relationships, whereas these chord embeddings are strictly based on likelihood of occurrence. [55] Turns out, they have a high correlation (**Figure 7b**). So despite the differences in methodology, both come to similar conclusions.

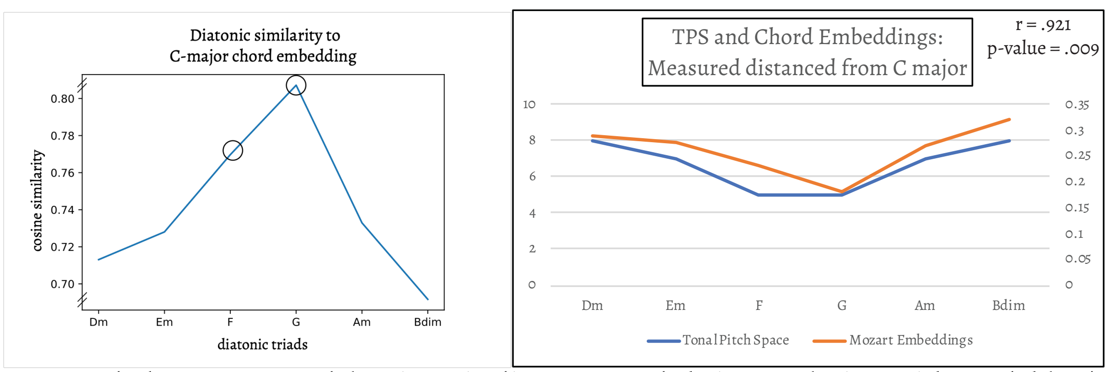

Figure 7a and 7b. Mozart model: Diatonic distances and their correlation with Lerdahl's (2001) chord distance

Mozart Analysis

[56] But how might these chord embeddings be used for analysis? [57] I'll demonstrate one method on the opening 8 bars of *Non siate ritrosi* from Mozart's opera *Cosi fan tutte*—[58] let's give it a listen. [58.1]

[59] For analysis, we can take each chord and its respective chord embedding and measure the distance [59.1] from one to the next to show a kind of "chord expectacty" (**Figure 8**).

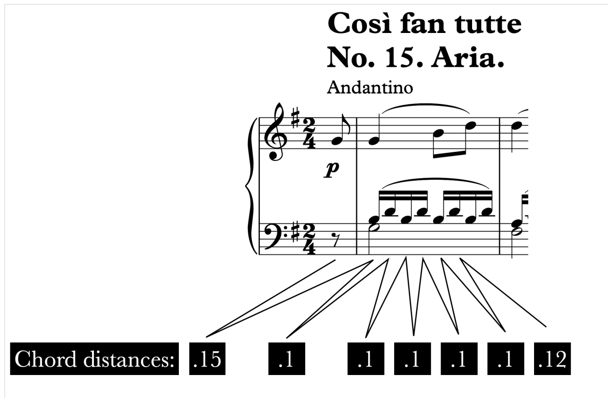

Figure 8. Chord distances between adjacent key-lime slices

[60] The result is this graph, where distance is the y-axis and time is on the x-axis (**Figure 9**). [60.1] Note how each peak corresponds to chord changes—this demonstrates that chord classes are being grouped together and that it's picking harmonic changes. [61] The increased frequency of peaks captures the harmonic acceleration. [62] The highest peak is on the applied chord, signaling that a chord "outside the key" is quantitatively further according to chord-embeddings. In this mode of analysis, the embeddings pick up the harmonic rhythm of the phrase, and show how the phrase builds up to the half cadence.

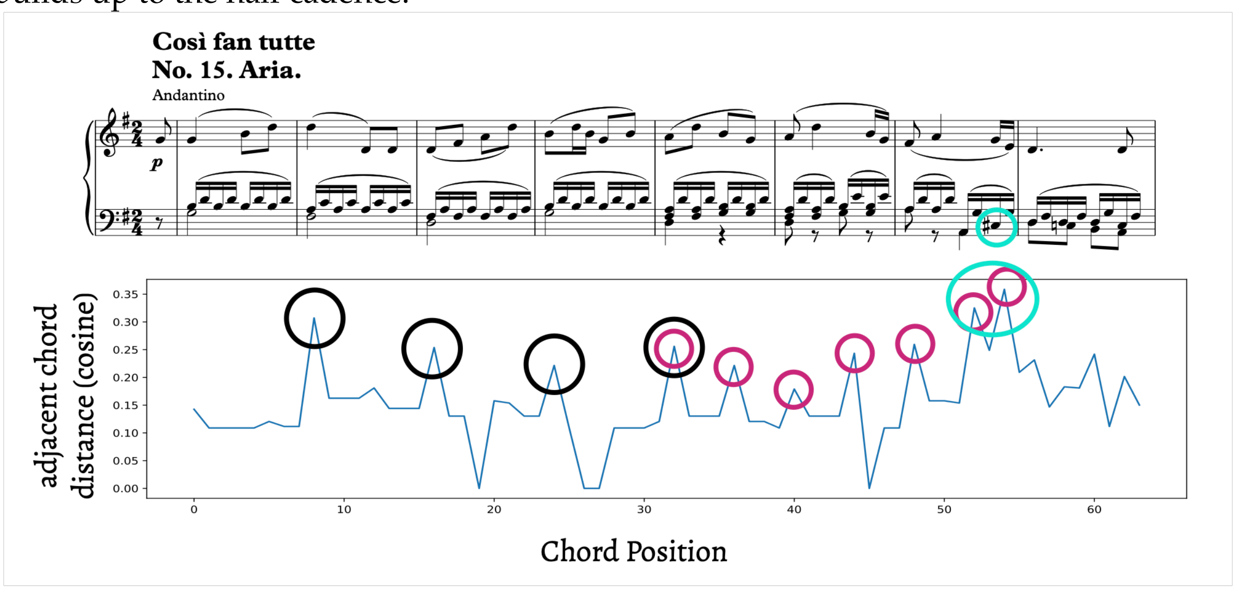

Figure 9. Analysis of Mozart's Non siate ritrosi with chord embeddings

[63] Using the same procedure for rock music shows its unique qualities, and points us to where harmonic languages and chord meanings are different.

Rock model
[64] I trained the second model on de Clercq and Temperley's (2015) Rolling Stone Corpus. [65] And unfortunately, there weren't enough modulations in the corpus to consistently judge key-relationship. But another takeaway is that keys are much more contained in rock and that, stylistically, songs tend to stay within one key.

[66] Within a key, the closest triads to the tonic are the submediant and the subdominant (**Figure 10**). [66.1] As Temperley (2018, 47) notes (in a table), I most frequently moves to IV, V, VI, and bVII, explaining why they're close in the vector space. [67] Unique to the rock embeddings, though flat VI and flat VII outside of the key, [67.1] they're relatively close to I. This shows that the model is picking up on progressions like the [67.2] "aeolian cadence" (Biamonte 2010) or "rogue dominants" (Doll 2007).

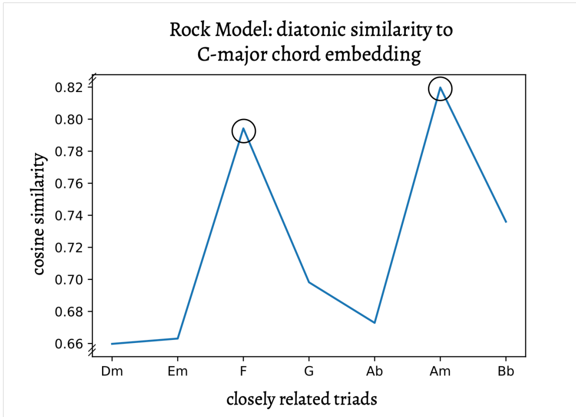

Figure 10. Rock model: diatonic distances from the tonic

Rock Analysis

[68] For the final analytical example, I want to go back to the chord progression from [69] Guns N' Roses. The chord loop V-bVII-I happens in the chorus right before the solo, where the music shifts from Db major to Eb minor. Though its cued with a snare build up, the key shift is really sudden compared to the rest of the chord loop. Let's listen—you'll hear the loop [69.1] twice before the shift. [69.2]

For a fun theoretical experiment, let's analyze this aria with the Mozart chord embeddings (**Figure 11**). [70] The initial chord loop, as noted in the introduction, is marked as aberrant; bVII-I doesn't follow harmonic norms per the galant conventions, and the Mozart model doesn't expect the progression. This also means that the key shift to the Eb minor is less pronounced in relation to the chord loop. [71] If we were to smooth over the chord distances, we would see a relatively flat line—this translates to saying: because *everything* is far, the modulation isn't pronounced.

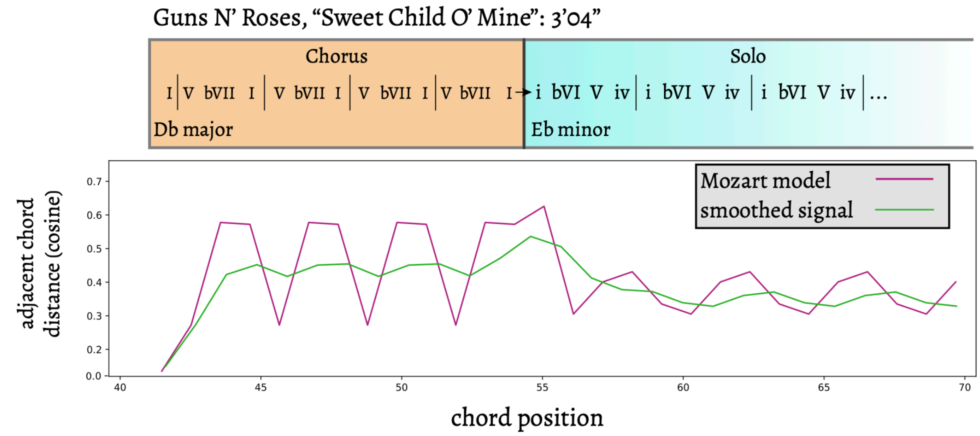

Figure 11. Mozart model: Analysis of "Sweet Child O' Mine" – 3'04

[72] Turning to an analysis using the rock model, we see that it recognizes the normativity of these chord loops (**Figure 12**). This helps promote the relative distance for the Eb minor key shift (which is still not normative in rock). [73] Smoothing over the rock model's distances shows flat lines and a peak at the key shift. So smoothing captures the expectation of the recurring chord loop and the change in formal position. Future work might use this methodology to segment pieces formally with these stylistically sensitive harmonic representations.

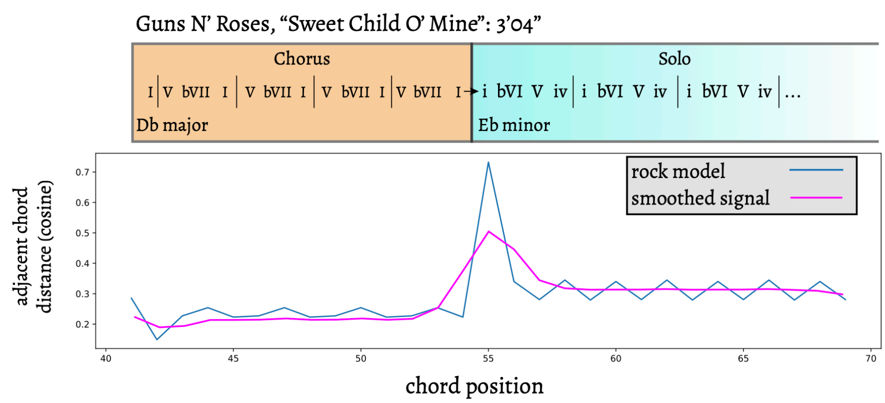

Figure 12. Rock model: Analysis of "Sweet Child O' Mine" – 3'04

Conclusion

[74] I've demonstrated that this chord embedding methodology is capable of reflecting our previous theories on music and captures insightful phenomena both theoretical and analytical. I like to remind myself that this is a computer with no pre-installed knowledge of what a chord is, so to extract takeaways equal to that of professional music theorists is pretty incredible. This means we might use the methodology to conduct systematic analyses reliably at large-scale levels.

Over the past 8 years, while embeddings have become a topic for serious research in machine learning, computer science, linguistics, and elsewhere, they have only been left in the peripherals of music theory. My work here suggests that chord embeddings can be used successfully to study harmonic progressions, phrases, expectation, and formal divisions through stylistically sensitive norms. Future work will pursue such questions at a larger scale, change embedding methodologies, and, ultimately, interrogate how such chord spaces reflect, imitate, or oppose conventional and perceptual assumptions.

[75] Thank you.

(References are on the next page.)

References

Agmon, Eytan. 1995. "Functional Harmony Revisited: A Prototype-Theoretic Approach." *Music Theory Spectrum* 17, no. 2: 196–214.
Bengio, Yoshua, Re'jean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. "A Neural Probabilistic Language Model." *Journal of Machine Learning Research* 3: 1137–1155.
Biamonte, Nicole. 2010. "Triadic Modal and Pentatonic Patterns in Rock Music." *Music Theory Spectrum* 32, no. 2: 95–110.
de Clercq, Trevor, and David Temperley. 2011. "A Corpus Analysis of rock Harmony." *Popular Music* 30, no. 1: 47–70.
Cuthbert, Michael Scott, and Christopher Ariza. 2010. "music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data." 11th International Society for Music Information Retrieval Conference (ISMIR).
Dick, Philip K. *The Philip K. Dick Reader*. Citadel Press, 1997.
Doll, Christopher. 2007. "Listening to Rock Harmony." PhD diss., Columbia University.
———. *Hearing Harmony: Toward a Tonal theory for the Rock Era (Tracking Pop)*. University of Michigan Press.
Firth, John Rupert. 1968 [1957]. "A Synopsis of Linguistic Theory, 1930–55." In F.R. Palmer (eds) *Selected Papers of J.R. Firth (1952–59)*: 168–205. Indiana University Press.
Foucault, Michel. 1972 [1969]. *The Archaelogy of Knowledge and The Discourse on Language*. Trans A.M. Sheridan Smith. Vintage Books, NY.
Harrison, Daniel. 1994. *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents.* University of Chicago Press.
Kopp, David. 1995. "On the function of function." *Music Theory Online* 1, no. 3: 1–15.
Krumhansl, Carol L., and Lola L. Cuddy. 2010. "A Theory of Tonal Hierarchies in Music." In Jones, Riess M., Richard R. Fay, Arthur N. Popper (eds.) *Music Perception* 51–87. Springer handbook on Auditory Research, vol 36. Springer, New York, NY.
Lerdahl, Fred. 2001. *Tonal Pitch Space*. Oxford University Press.
Liang, Feynman T., Mark Gotham, Matthew Johnson, and Jamie Shotton. 2017. "Automatic Stylistic Composition of Bach Chorales with Deep LSTM." In *ISMIR*, pp. 449–456.
Mikolov, Thomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." arXiv preprint: 1301.3781.
Nikrang, Ali, David R.W. Sears, and Gerhard Widmer. 2018. "Automatic Estimation of Harmonic Tension by Distributed Representation of Chords." In Aramaki, Mitsuko, Matthew E.P. Davies, Richard Kronland-Martinet, Sølvi Ystad (eds.) *Music Technology with Swing*. Lecture Notes in Computer Science 11265. Springer, Cham.
Nobile, Drew. 2016. "Harmonic Function in Rock Music: A Syntactical Approach." *Journal of Music Theory* 60, no. 2: 149–180.
Patel, Aniruddh. 2007. *Music, Language, and the Brain*. Oxford University Press
Riemann, Hugo. 2000 [1893]. *Harmony Simplified: The Theory of The Tonal Functions of Chords*. Augenert LTD.
Temperley, David. 2007. *Music and Probability*. MIT Press.
———. 2018. *The Musical Language of Rock*. Oxford University Press.
Temperley, David, and Trevor de Clercq. 2013. "Statistical analysis of harmony and melody in rock music." *Journal of New Music Research* 42, no. 3: 187–204.
Tymoczko, Dmitri. 2010. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press.
White, Christopher Wm., and Ian Quinn. 2016. "The Yale-Classical Archives Corpus." *Empirical Musicology Review* 11, no. 1: 50–58.
———. 2018. "Chord Context and Harmonic Function in Tonal Music." *Music Theory Spectrum* 40, no. 2: 314–335.
Wittgenstein, Ludwig. 1953. "Philosophical investigations." Trans. by G.E.M. Anscombe. Blackwell.