Let's Jam: A Machine Learning Framework For Sample Pairing and Filtration Using Melodic and Tonal Movement Classification

Abhinav Damera and Hritik Patel

ABSTRACT

The search engines of music sampling services are often inadequate when addressing samples that tonally and sonically pair well with each other. Using search filters and other functions within these services, one is able to find samples that are within the same key as the desired sample, and with the same sonic qualities. Current technology is either hyperspecific with the sample selections, or focuses too much on the instrumentation of the sample alone. Using a machine learning model (MLM), we plan to create a sample-sorting system that categorizes samples, not on the instrumentation or sonic similarity, but on the tonal qualities, like the chord structure the sample would fall on when assigned chords based on common practice chord assignments (I, II, III, and so on). This would allow the model to adequately categorize the samples by their movement (by breaking up the given sample into chord progressions such as I-IV-V). This delivers further classification advantages (such as filtering samples by what chords they would fall under) and allows sample libraries to provide sample pairings that are not sonically identical, but still fall under the same movement. Samples with similar chord structures are far more likely to sound harmonious without sounding exactly the same. This would lead to more diversity and efficiency with sample pairing and selection, while also allowing the user to exercise more tonal specificity when searching, improving workflow.

BACKGROUND & PROPOSAL

The broad scope and subjectivity of music makes the categorization of samples in a digital sample library insubstantial and often confusing. Finding the “right” sample is an issue many producers face. A paper from the Los Angeles Film School details how sampling other copyrighted music depends on, “...how you manipulate the sounds and how creatively you transform them into something new.” While the samples in a library are royalty-free, it provides insight into how, “finding the right source material is always a challenge,” (Costa & Los Angeles Film School, 2025). Good samples maximize utility and align with the producer’s vision. Whilst many sampling services are able to master the use of instrumentation agreement (having features that specify instrument categories), and key agreement (having systems that use the tonal scale the sample falls under in its classification process), an area they have yet to address is the tonal specificities of each individual sample. The movement of the sample: the way the tones are structured (chord progressions) and the general harmonic structure is an important consideration that is overlooked by sampling services (Avise, 2013). Features like these are not dependent on the scale it is in, nor is it dependent on the instrumentation it has. Two samples that are in the same key are not guaranteed to sound good together, even with manipulation (Grey, 2025). Current sampling services fail to match samples through movement similarity, and instead rely on instrumentation.

Recently, several advancements have been made to the field of music regarding machine-interpretation of music. Advances in machine-learning models for genre classification using vocal, harmonic, and melodic features have improved in-depth music analysis (Bhattacharjee et. al, 2019). Sampling service Splice's use of a machine-learning model has released features such as “similar sounds” which offer a newer way to filter samples (Splice Records, 2022). However, the algorithm favors instrumental similarity over melody, resulting in mismatched or redundant samples. Artists want to find samples that are variations of the sounds they already have, while still maintaining a similar movement to promote chordal agreement (de Haas et al., 2011). If a sample is identical to what the producer already has, it does not help because it offers no new creative value. If a sample differs in movement, instrumentation similarity does little to help, as it will sound discordant and will have dissonance.

We propose a new algorithm to be used to categorize samples, in conjunction with the preexisting methodology already in use. This will hereon be referred to as Harmonic Progression Analysis of Sample Tonation (HPAST). HPAST is a model based around a Convolutional Neural Network(CNN) process in which a MLM will analyze a sample by analyzing the harmonic content of the sample itself to derive the most likely chord progression the sample falls under. These chord progressions would then be used when comparing samples in order to better judge how compatible they would be. This allows for a comparison of the melodic content of the sample, not just the instrumentation or “sound texture”, like sampling services do currently. This circumvents preexisting issues of sample libraries giving “too similar” or “too different” samples as it is using an entirely different algorithm with new methods of analysis to organize samples, while not compromising other algorithms.

Firstly, the HPAST would separate the sample into its various components: Vocals, Drums, Instruments, and Bass (with the drums, which do not provide harmonic content). The HPAST would use frequency analysis to be able to separate the samples into its various parts. Then, these stems will be analyzed for frequencies that "stick out" and are the most prominent. From there, the most commonly occurring notes would have a chord assigned to it through a vast database of possible chords ranging from multiple scales (major, minor, harmonic, melodic, blues, etc). From there, a chord progression can be extracted from the sample, which can be used for better sample pairing.

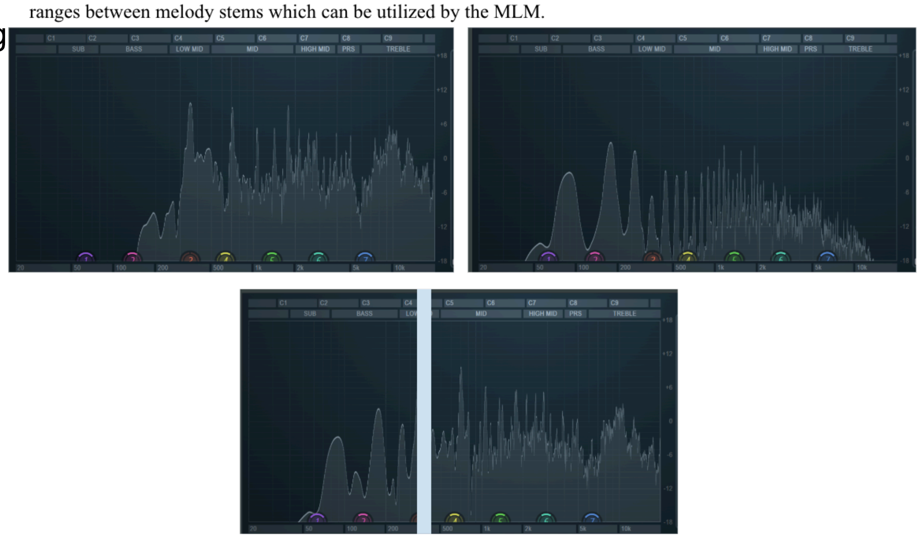
This innovation has potential to improve the lives of music producers. For example, SageJournals states, "81% rated their financial stress as high or overwhelming" and "Over 45.4% endorsed significant occupational stress" (Berg et. al, 2022). Stress of producers may be caused by their inability to develop high quality pieces efficiently enough. One cause of this stress is the time-intensive process that is required to combine harmonically compatible samples. HPAST helps to reduce the time needed to find the "right" sample. Aden Russell, a lead at sampling service EDMProd states, "If it takes you 2 minutes instead of 4 to find a sample that fits during the production process, and you use an average of 15 samples per song, that's 30 minutes that you've saved" (Russell, 2023). Increased efficiency can accelerate project completion and improve releasing or licensing work, which can aid in reducing financial stress.

STRUCTURE & ORGANIZATION

The model would categorize and pair musical samples based on harmonic movement and functional chord structure. The core idea is to extract harmonic information from audio samples, map that information to functional chord labels (I, ii, IV, V, etc.), and use an MLM to group samples with similar tonal movement. Our algorithm would contain four main stages: First, audio preprocessing and feature extraction; Next, chord and key estimation; Third, functional chord encoding; Finally, machine-learning-based classification and pairing.

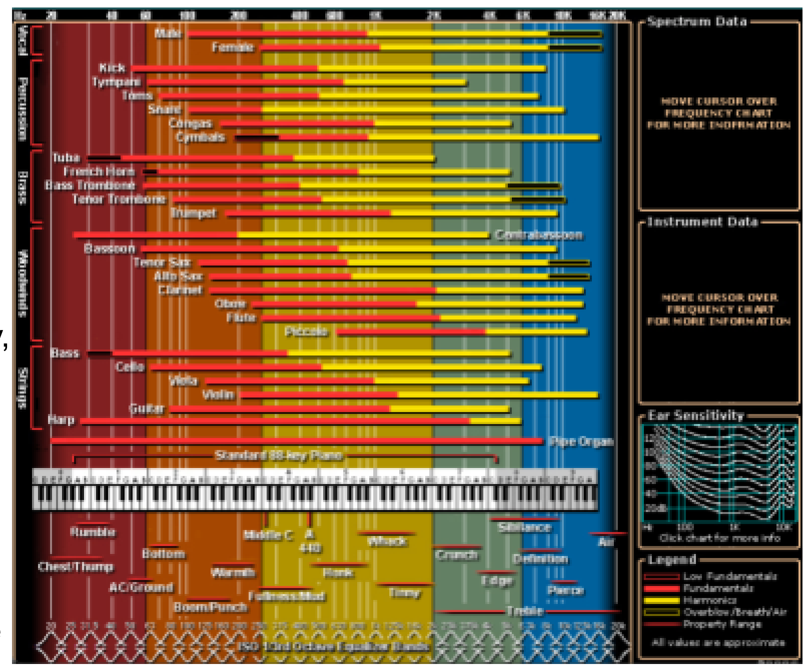
We would first convert each audio sample into a standardized format (mono, fixed sampling rate). We would do this in order to compare samples based on just harmonic content which allows for cross-sampling pairing. HPAST will use audio(spectrogram) patterns from being fed vocals and specific instruments to isolate frequencies in the sound that are occupied by a sound that appears to be an instrument, drums, or vocal (Hilsdorf, 2023). By mapping out frequencies onto a spectrogram, the model will be able to visualize these inputs graphically. This can be achieved through supervised inputs of the parameters and the

Note F3 on a female vocal(left), and on a bass guitar(right), and played together(bottom) on an equalizer(advanced spectrogram) in FL Studio(Dambrin & Schaack, n.d.) and reverse-engineered decision boundary separating the frequency ranges shown. This demonstrates the inherent differences in frequency ranges between melody stems which can be utilized by the MLM.



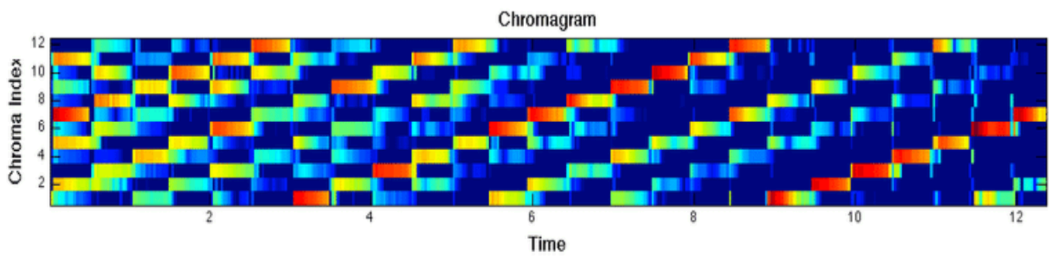
assignment of malleable decision boundaries to sound inputs for the model to separate these inputs. For example, vocal voicing often has varied frequency ranges(75-400 Hertz-Bass/Tenor, 165-1100 Hz-Alto/Soprano)(Huss, 1983). Additionally, instruments have general frequency classifications that can be used to feed a predictive model what frequency range(s) the model should separate the audio into.

HPAST will then utilize methods such as phase cancellation (combining frequencies in the waveform with sound that is the same frequency, but an opposite polarity, i.e., the shape of its waveform would be flipped, it would silence the sound) to isolate and prepare the various "stems" (Gagneja, 2020). The system would then extract chroma-based harmonic features, which represent the energy of the twelve pitch classes independent of octave (Ellis & LabROSA, 2007). Using frequencies from the spectrogram, the MLM would use a Mel-spectrogram-like classification system to "pick out" the most prominent frequencies during each time interval, which would then be isolated to form a "guess" as to what the notes of the audio input most likely are. We would do this by computing the short-time Fourier transform (STFT) of the audio signal.



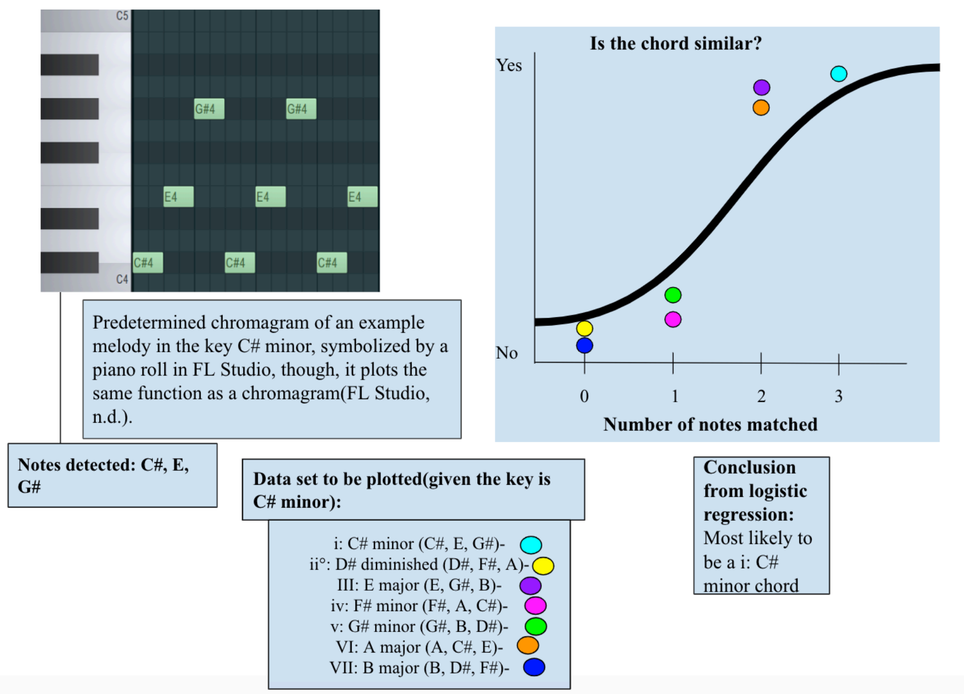
Equalizer frequency table showcasing general frequency ranges for instruments and percussion(Martin, 2025)

In doing so, the signal will be isolated into its component frequencies across many independent time-intervals, which the model will then isolate frequencies that appear above a certain relative threshold. The frequencies are then inputted into folding frequency bins, using a 12-dimensional chroma vector: $c(t) = [c_1(t), c_2(t), \ldots, c_{12}(t)]$ (Lenssen et. al, 2014). These chroma vectors will be mapped onto a chromagram over a function of time, where the Y axis is the note values of an octave in ascending order (each chroma vector corresponds to a note along that octave).



Chromagram of an ascending scale. As shown, each note is able to be translated to a chroma vector(indicated by a chroma index value) over a course of time, creating a "timeline" of notes, visualizing the melody(Lenssen et. al, 2014).

We would then perform chord detection by chroma vectors to predefined chord templates using cosine similarity of the various melody regions: $sim(c, t) = \frac{c^*t}{||c||||t||}$. The larger the difference calculated, the more likely there is to be a separation in the melody (Oudre et al., 2011). Using chroma features, the system estimates a local chord sequence and a global key for each sample. Chords and their corresponding notes will be fed into the algorithm(for example, if the sample was predetermined to be in G minor, the selection it would choose from would have "i-Chord=G, Bb, D; iidim-Chord=A, C, Eb" and so on). The chords will then be assigned one of these chord labels using the Logistic Regression, where the probability of the section belonging to each chord will be based on the occurrence of those respective notes in that section (Moazeni, 2024). This aspect can be trained by using “example melodies(labeled data)” that are predetermined to correspond to an existing chord. The detected chords are then converted into their respective functional chord labels relative to the key (I, ii, iii, IV, V, vi, vii°, etc). For instance, a G major chord in the key of C is encoded as V. Then, each sample will be converted to represent a sequence of functional chords in chord progression format: $S = [f_1 - f_2 - ... - f_n]$ (Music Theory Academy, 2025). This allows the sample to be compatible with machine learning because the chord sequences would be turned into computable values. Next, adjacent identical labels are merged together(e.g.



I-I-IV becomes I-IV). This would allow us to produce a simplified chord progression. The global key is estimated by analyzing pitch-class and chord frequency over the full sample and selecting the key that maximizes tonal consistency. This step enables later abstraction from absolute pitch to functional harmony.

Now we would use these functional representations as data for an unsupervised learning model (hierarchical clustering). Unsupervised learning would be used over supervised learning for this phase because it allows the system to make natural groupings of the samples based on their tonal structure (Stryker, 2024). This is superior to supervised learning because labelling data like that would be expensive and subjective. We look for relative similarity so producers can combine samples that sound good, rather than enforcing strict classification. Using unsupervised learning (clustering), we would be making fewer assumptions and there would be less risk of bias from weak labels leading to precise harmonic pairings (Noble, 2024). The distance between two samples is computed using standard vector distance metrics: $d(S_a, S_b) = ||v_a - v_b||$. Samples that fall within the same cluster/below a distance threshold are considered harmonically compatible and are recommended first in a search output. This mechanism prioritizes tonal coherence while encouraging diversity in sound selection. By utilizing this sample filtration system, search engines in sample libraries can be upgraded to better pair samples internally in their library, an approach no sampling service has done in the past.

Our list of goals is as follows:

1. Acquire all resources needed, and generate baseline models for our algorithm
2. Train the first model(Functional chord encoding) map isolated stems onto a chromagram, 150 audio clips of precomposed melodies will be fed into the system, and spectrogram analysis will be provided as well as chromograms for what they would translate into. >80% success rate(chromagrams have an 80%

similarity to the inputted audio and its corresponding graphs) necessary to progress.

3. Train the second model to assign chord monikers to each section, 150 prerendered chromagram melodies will be fed into the system with preidentified chord assignments, >80% success rate indicates moving onto the next step.
4. Train the third model(Machine-learning-based classification and pairing) to pair premade chord progressions in proper notation by order of similarity, 25 instances of 4 different chord progression inputs(such as I-IV-I and II-IV-I) will be fed to the ai with a pregiven chord progression render. A predetermined similarity order will be used to compare. >80% success rate by human listening trials by an expert needed to progress.
5. Import all previous algorithms into a master model that organizes the previous functions into a sequence, following an organization structure seen in CNNs

There are several risks associated with developing this machine learning model. One potential risk is inaccuracies with chord detection. Collating multiple chord predictions, removing low-confidence detections and initially training the model with a wide range of samples can help reduce this. Another risk is the subjectivity of harmonic similarity. The system offers multiple outputs and prioritizes expert-validated samples because pairings are context dependent. Additionally, inaccurately identifying chord changes is a risk. Initially, it is addressed with cosine similarity, HPAST can also factor in time signature information to enable consistent chord separation(e.g., four-bar progressions in 4/4 time).

Discussion

Many of the components of the model are advanced versions of algorithms that exist in indirect forms. For example, algorithms like the STFT and methodology such as linear and logistic regression have been present in previous models (Bruin, 2021). However, the novelty of the algorithm is present in its combination of advancements made to the present algorithms, as well as its use of visualization as its main analysis. The use of categorization by order of importance(in this situation, similarity) is useful in search engines for sample libraries, as the organization structure for search engine results can be ordered based on said similarity. The translation of melodic analysis to search engine data is unconventional; however, using advanced algorithms working in a multi-step process can advance classification systems in a way that can integrate with existing search engine filtration.

Much of the neural network development as well as organization of algorithms in the CNN are modeled after biological processes the human brain undergoes when leading a melody “from ear”. In a Harvard study, it was found that the brain utilizes filtration methods of sound in order to classify them as “familiar” and “unfamiliar”, when listening to the surrounding environment (Eck, 2024). The temporal lobe is responsible for this filtration, which can often lead to neural visualization following a process known as a “what pathway”(Freeman, 2011). This is why sound can often be described as having “textures”. Similarly, a filtration system was used first in HPAST, not only to separate stems, but to also account for variety in sound texture, similar to how the human brain utilizes the same process.

Additionally, acknowledgement of atonality was considered, similar to how the brain reacts to it. Atonality is perceived as a more unpredictable stimulus in the brain, and as a result, “clashing” chords tend to sound atonal, as they lack a central key and often sound astringent. Hence why two sounds in the same key are not guaranteed to sound cohesive, and why chord classification is necessary. A study in Frontiers utilized an MIR toolbox to model matching pitch-class profiles(similar to our quantification of chromograms) and pulse clarity(how easily a key or general tone can be recognized) as they evaluated varying chords in order of atonality. This modeling system closely followed the process the brain uses when discerning how synonymous two sounds sound together. Basic music theory, which is what our chord classification system is based off of, follows tones that surpass a certain value of pulse clarity, and can be accurately fit into the 12 standardized pitch-class profiles. Moreover, the use of concepts like anticipation, call-and-response, and reward systems in the brain also allow for music theory principles to aid in our classification process. This is why we chose to classify each sample with a chord progression following basic music theory, as well as classify them in order of similarity to the entered prompt.

In summation, the HPAST is highly achievable, although it requires resources to further advance the algorithms that exist that fulfill some of the tasks proposed. Additionally, the sequencing of these steps, as well as the intake of additional data provided in most sampling services, like key, tempo, and genre, are also important considerations that need to work in tandem with our model. However, when realized, HPAST will provide a significant impact to workflow, and innovate the way sampling services classify their samples. It will also provide producers with a newer understanding of the tonal structure of the samples they are using, which can allow them to maximize their efficiency and creativity to an extent and with a level of accessibility seldom seen previously.

References

Avise, J. C. (2013). *From perception to pleasure: Music and its neural substrates*. National Library of Medicine. Retrieved December 27, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC3690607/

Bruin, J. (2021, August 22). *Logistic Regression Analysis | Stata Annotated Output*. OARC Stats. Retrieved February 10, 2026, from https://stats.oarc.ucla.edu/stata/output/logistic-regression-analysis/

Berg, L., King, B., Koenig, J., & McRoberts, R. L. (2022). Musician occupational and financial stress and mental health burden. Psychology of Music, 50(6), 1801-1815. https://doi.org/10.1177/03057356211064642 (Original work published 2022)

Bhattacharjee, Utpal & Mannala, Jyoti. (2019). An Experimental Analysis of Speech Features for Tone Speech Recognition. International Journal of Innovative Technology and Exploring Engineering. 2. 4355-4360. 10.35940/ijitee.B7748.129219.

Costa, L., & Los Angeles Film School. (2025). *Music Sampling Explained: Creative Techniques and Copyright Laws Every Producer Should Know*. The Los Angeles Film School. Retrieved December 27, 2025, from https://www.lafilm.edu/blog/music-sampling-explained/

Dambrin, D., & Schaack, D. (n.d.). *Fruity Parametric EQ 2 - Effect Plugin*. FL Studio. Retrieved December 28, 2025, from https://www.image-line.com/fl-studio-learning/fl-studio-online-manual/html/plugins/Fruity%20Parametric%20EQ%202.htm

de Haas, W. B., Robine, M., & Hanna, P. (2011). *Comparing Approaches to the Similarity of Musical Chord Sequences*. Hal Open Science. Retrieved December 28, 2025, from https://hal.science/hal-01006469v1/document

Eck, A. (2024). *How Music Resonates in the Brain | Harvard Medicine Magazine*. Harvard Medicine Magazine. Retrieved February 10, 2026, from https://magazine.hms.harvard.edu/articles/how-music-resonates-brain

Ellis, D. P.W., & LabROSA. (2007). *CLASSIFYING MUSIC AUDIO WITH TIMBRAL AND CHROMA FEATURES*. Columbia University. Retrieved December 28, 2025, from https://www.ee.columbia.edu/~dpwe/pubs/Ellis07-timbrechroma.pdf

FL Studio. (n.d.). *Piano roll Basics*. FL Studio. Retrieved December 28, 2025, from https://www.image-line.com/fl-studio-learning/fl-studio-online-manual/html/pianoroll.htm

Freeman J, Simoncelli EP. Metamers of the ventral stream. Nat Neurosci. 2011 Aug 14;14(9):1195-201. doi: 10.1038/nn.2889. PMID: 21841776; PMCID: PMC3164938.

Gagneja, N. (2020, October 8). *Understanding phase and polarity in sound recording*. Ux Collective. Retrieved December 17, 2025, from https://uxdesign.cc/understanding-phase-and-polarity-in-audio-101aa8cac2eb

Grey. (2025). *Major Scale Chord Function: How Chords Behave*. Hub Guitar. Retrieved December 27, 2025, from https://hubguitar.com/music-theory/chord-function

Hilsdorf, M. (2023, September 20). *AI Music Source Separation: How it Works and Why It Is So Hard*. Medium. Retrieved December 17, 2025, from https://medium.com/data-science/ai-music-source-separation-how-it-works-and-why-it-is-so-hard-187852e54752

Huss, P. J. (1983, 12). *Vocal Pitch Range and Habitual Pitch Level: The Study of Normal College Age Speakers*. Western Michigan University. Retrieved February 9, 2026, from https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=2616&context=masters_theses

Lenssen, Nathan & Needell, Deanna. (2014). An Introduction to Fourier Analysis with Applications to Music. Journal of Humanistic Mathematics. 4. 72-91. 10.5642/jhummath.201401.05.

Martin, T. (2025, June 5). *Detailed Frequency Ranges of Instruments and Vocals*. The Absolute Sound. Retrieved February 9, 2026, from https://www.theabsolutesound.com/freqchart/main_display.htm

Mencke, I., Omigie, D., & Wald-Fuhrmann, M. (2019, January 7). *Atonal Music: Can Uncertainty Lead to Pleasure?* Frontiers. Retrieved February 10, 2026, from https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2018.00979/full

Moazeni, S. (2024, October 17). *Logistic Regression: A Method for Classification | by S. Moazeni, PhD*. Medium. Retrieved December 28, 2025, from https://medium.com/@AILearningHub/logistic-regression-a-method-for-classification-649ca3a68608

Music Theory Academy. (2025). *5 basic rules of Chord Progressions*. Music Theory Academy. Retrieved December 28, 2025, from https://www.musictheoryacademy.com/understanding-music/chord-progressions/

Noble, J. (2024). *What is Hierarchical Clustering?* IBM. Retrieved December 28, 2025, from https://www.ibm.com/think/topics/hierarchical-clustering

Oudre, L., Grenier, Y., & IEEE. (2011). *Chord Recognition by Fitting Rescaled Chroma Vectors to Chord Templates*. Institut de Recherche en Informatique de Toulouse. Retrieved December 26, 2025, from https://www.irit.fr/~Cedric.Fevotte/publications/journals/ieee_asl_deterchord.pdf

Splice Records. (2022, December 1). *Use Similar Sounds to find samples on Splice - Blog*. Splice. Retrieved December 27, 2025, from https://splice.com/blog/introducing-similar-sounds/

Stryker, C. (2024). *What Is Unsupervised Learning?* IBM. Retrieved December 28, 2025, from https://www.ibm.com/think/topics/unsupervised-learning