University of California
Santa Barbara

Music Information Retrieval Driven Algorithmic Music

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Music

by

Henry Stewart Engart III

Committee in charge:

Professor João Pedro Oliveira, Chair
Professor Curtis Roads
Professor JoAnn Kuchera-Morin

March 2022

The Dissertation of Henry Stewart Engart III is approved.

Professor Curtis Roads

Professor JoAnn Kuchera-Morin

Professor João Pedro Oliveira, Committee Chair

March 2022

Music Information Retrieval Driven Algorithmic Music

by

Henry Stewart Engart III

Acknowledgments

This dissertation, along with the years in graduate school that preceded it, would not be possible without the support of others.

I would first like to thank my advisor, João Pedro Oliveira for his compositional guidance, expertise, and patience throughout our time working together. Your attention to detail and compositional thoroughness is a standard I plan to uphold throughout my career.

I would like to thank my other committee members, Curtis Roads and JoAnn Kuchera-Morin for their expertise, suggestions, and time throughout this process.

I would like to thank my qualifying exam committee members, Clarence Barlow, Curtis Roads, Benjamin Levy, and Andrew Tholl, for their guidance through the process of exploration and narrowing of topics.

I would like to thank Karl Yerkes and Marc Evanstein for their time and efforts in bringing my technical skills to a level that allowed me to conduct this research, compose any music that I conceived of, and for opening up worlds of possibilities that previously were unthinkable.

I would like to thank Thom Blum for his enthusiasm, experience, and music that ignited this research.

I would like to thank my classmates from the Departments of Music and Media Arts and Technology for their feedback and enthusiasm have been not only helpful, but inspirational.

This degree would not have been possible without the tireless work of the UCSB Music Department staff, most notably Carly Yartz whose student advocacy, institutional knowledge, and friendship made this work possible. From our first contact during the interview process, her professionalism and organizational ability have guided me through this degree.

I would like to thank Emily, Mini, and Marc Evanstein for their hospitality, housing, and friendship throughout my time at UCSB.

I would like to thank my friends and family for their understanding and patience as pursuing this degree required me to move across the country and miss out on countless important milestones.

I would like to specifically thank my parents who instilled in me at a young age the importance of music and who, throughout my life, have supported my pursuits.

Lastly, I would like to thank my partner Beth who, more than anyone else, has had to deal with a stressed partner throughout this process and has not hid but rather uprooted her life and moved across the country to stand by my side throughout this process.

Curriculum Vitæ

Henry Stewart Engart III

Education

2022	Ph.D. in Music (Expected), UC Santa Barbara.
2017	MM in Music Composition, University of Georgia.
2014	BM in Music and BA in Political Science, University of North Carolina.

Employment

2021	Organist, Summerland Presbyterian Church. Summerland, CA.
2019–21	Organist, Christ the King Lutheran Church, Newbury Park, CA.
2017–21	Audio Engineer, UC Santa Barbara Special Research Collections.
2018–21	Graduate Teaching Assistant, UC Santa Barbara.
2017–19	Pianist, St. Mark's University Parish. Isla Vista, CA.
2015–17	Graduate Research Assistant, Ideas for Creative Exploration (ICE), UGA.
2014–17	Pianist, Christus Victor Lutheran Church. Athens, GA.
2012–14	Audio Engineer, UNC-CH Music Department.

Awards

2018, 2019, 2021	Dorothy and Sherrill C. Corwin Award for Excellence in Music Composition, UCSB.
2017	Clementi Holder Opportunity Award, UGA.
2017	A2RU Travel Grant: Emerging Creatives Student Conference.
2014	Thelma Thompson Composition Award, UNC.
2010–2014	Andy and Cindi Griffith Scholarship, UNC.

Affiliations

International Society of Music Information Retrieval.

International Computer Music Association.

Society for Electro-Acoustic Music in the United States.

Society of Composers Inc.

Abstract

Composing in Latent Space: Music Information Retrieval Driven Algorithmic Composition

by

Henry Stewart Engart III

The modern composer spends a large amount of time collecting and then sorting though digital audio files. There have been many attempts to make use of the field of Music Information Retrieval to organize and make use of the large databases of sound. This paper will present methods to both deal with large amounts of data and to harness that data for music creation itself. This document also showcases the compositional growth during the my studies at UCSB and the technical and aesthetic challenges that I had along the way.

In this dissertation I will first layout the theory and history of work that has come before. I will then put my musical output in a greater musical context, explain my implementation of the theoretical concepts, presenting my own strategies, and specifically elaborating on the application of these techniques in two of my most recent compositions, Usynlig and A More Sound Output®.

Contents

Chapter 1

Motivation and Summary

Even before I started composing music I was recording audio and storing it for later use. Originally this was a result of my practice as a performer but also I began collecting sounds of interest from around my house and songs off of the radio. When the digital music revolution happened, my computer became filled with audio ripped from CDs, purchased from online stores, or elsewhere. Around this time I began composing, exclusively remixing pre-existing songs and other pre-existing pieces of music. When I began my work as an audio engineer, I was recording all sizes of ensembles and musical groups, but on the weekend I began a practice of field recording. As a result, of the past fifteen years, I have amassed a large personal library of audio.

When I began to compose electronic music I would collect new sounds on a piece-by-piece basis, rarely diving into the audio library I was also generating. This practice would often leave many audio files unused and added to the detritus on my harddrive, in folders accompanying the final composition. Once I began working at an audio archive that freely shared audio online for everyone to listen and create new music, I started to think about how to most effectively use and utilize my audio library without spending the time, potentially years, to create the metadata by hand, as the archive did.

At this point I tried available software and looked to available scholarship as I felt that I could not be the only composer with this struggle. Looking into the available material

I was unable to find software or even a paradigm that satisfied my creative needs. As a composer of algorithmic music I decided that there had to be a way that I could solve this problem for myself and offer a potential solution to others in a similar situation to myself.

What started as a project to compose with huge sound databases has lead me to a reconsideration of treating music as data, through the exploration of the field of Music Information Retrieval, and data, in turn, as music, through an application of my established practice as a composer of algorithmic music.

Chapter 2

Theory

Information retrieval can be broadly defined as the process of finding unstructured data from a large collection.[23] While this definition is quite broad, in practice the unstructured data is usually text, with interdisciplinary subfields working on challenges specific to their discipline. Music Information Retrieval is generally approached with two goals in mind: enforcing copyright and music recommendation.[35] These two goals are both approaches to creating automatic measures of similarity between audio files, with similar music being useful for recommendation algorithms and music with too many similarities being flagged for copyright violations.[5]

2.1 Segmentation

Audio must be segmented temporally into windowed frames for analysis by fast Fourier transform (FFT) to extract low-level features.[11] These frames, measured in samples are set at a fixed size and based on this analysis, higher-level segmentation can be established. These higher-level segments can be tied to onset or transient detection or could be the entire audio file. Utilizing each of these three levels of abstraction can be useful depending on the use-case. Additionally, when working with music specifically, one can segment by tempo or score.[27]

Working at the analysis frame level is advantageous when working on a textural section of a piece. The modulation or parameterization on the size of the analysis window can often lead to interesting results as the sounds are cut from their context mid attack and re-contextualized in a new cloud of sound. The drawbacks to working at this level is that the size of the analysis becomes a part of the piece as a quick tactus.

Working with segmentation by onset results in more percussive results. While the length of each audio segment is based on a set algorithm, usually a combination of amplitude thresholds, change in amplitude from the prior analysis frame, and transient detection, the length and characteristic of each audio segment is based on the original context that it was recorded. As this type of segmentation prioritizes the onsets of sounds, it is very helpful in the creation of micro-montage and creation of granular textures.

When one uses the whole audio file as an analysis frame, the resulting sound is less accurately represented by whatever number was returned through the algorithm. This is because when features are averaged across the duration of a file, there is a higher possibility for outliers. However, treating the audio file as a cloud that has a given center, it is possible to compose with less specificity. Unlike granular clouds, unless using specifically synthesized audio, the audio files will not be distributed in a logical way. This results in a less clear audible transmission of underlying logic but results in music that sounds more natural.

In my music, I have found it useful to use segmentation schemes from the analysis frame to the whole file at the same time as to not tire the listener's ear. In my piece *Usynlig*, you can hear throughout the piece where the segmentation logic changes, but the most drastic examples occur when two segmentation schemes are juxtaposed either temporally, as when the texture changes to a micro-montage, or when two schemes are used simultaneously.

2.2 Feature Extraction

Analysis done with the short periodic sampling necessary for FFT analysis constitute 17 low-level features, such as fundamental frequency, spectral centroid, and harmonicity.[22] Further analysis on these features can be abstracted to mid-level features that comprise of tonal and temporal features while high-level features describe semantic characteristics.[39] Every MIR analysis library prioritizes and offers different mid and high-level features while low-level features are generally offered uniformly. The most prominent mid-level feature being used for analysis is the use of Mel Frequency Cepstrum Coefficents (MFCCs) because of the separation of periodic components and timbral features.[24] This allows for the matching timbre independently of pitch. An example of a unique offering of a high-level feature is the Essentia library's "danceability" descriptor.[8]

In my audio library, I keep a database of audio indexed for as many features as possible that gets appended as files get added. This allows the flexibility to prioritize any feature needed in the compositional process, as I am not prioritizing for salience, without the need to run a batch process during the composition session.

In addition to automatic feature extraction, manual feature extraction can be done by tagging metadata to a file, as done in Trevor Wishart's Globalalia. In this case, low-level (pitch or pitch glide), mid-level (start consonant or cluster, vowel or vowel glide, and end consonant or cluster), and high-level features (original language, gender of speaker, and vocal quality) were assigned to each audio file manually.[37] In my own work, I have rarely used manual metadata tagging because of scale and available time. I do, however, keep manually tag each file that I use in a piece so that I could potentially use that as a parameter in the future.

2.3 Similarity

Similarity in MIR is measured in Euclidean distance in n-dimensional spaces where n is the number of features extracted from the audio file. Finding the distance in high dimensional spaces is often done through K-Nearest Neighbor (KNN).[13] Without weighting of features, either as a result of heuristics or through a measure of salience of features, undesired results can occur because KNN assumes that the closer two points, the more similar they are.[16] While it is relatively trivial to measure similarity in a high dimensional feature space, it is more complicated to measure dissimilarity effectively. While this does not have many commercial applications, when composing with MIR, there are many artistic applications for desiring the measurement of dissimilarity.

Additionally, when working in high dimensional space, it is important to avoid the "curse of dimensionality" that can occur when using high dimensional feature spaces for analysis. The result of this is less accurate analysis and increased computational time.[25] The ideal feature space represents the audio as compactly as possible while throwing away irrelevant and/or redundant noise.[5] It is difficult for the end user to always know which features are important when presented with a long list of available analysis.[7]

The process of weighting features is a productive practice when there is a specific outcome desired. While this can occur in the compositional process, often the process reveals a possibility that would be more favorable. This process does take extensive time as there is a long amount of time spent tinkering.

When it comes to composing similarity is not often the most important parameter of a sound. There are applications that would benefit from querying for similarity, most prominently bridging the gap between performer and tape in a mixed music setting, but unless the audience knows what the target sound is, they have no conception of how similar or not a given sound is to it. If the composer does give the audience both the target sound and the most similar sound, then the audience can more fully understand the process happening in the music. This is something that I used extensively in my piece *Usynlig* to create choruses of metal rakes and symphonic extensions on party horns.

2.4 Dimension Reduction

To solve the problems implicit in working in high dimensional space, either the analysis library can be restricted to features that the designers designate as perceptually significant or desired; usually these are at minimum; pitch, loudness, duration, and a way to measure timbre such as MFCCs; or through analysis, such as principal component analysis (PCA), can be run on the extracted features for a reduction of dimension.[22] The process of dimensional reduction also allows for the potential for visualization of the audio corpus in two and three dimensions that allows for a graphical interface for visual browsing and querying. In my piece *Usynlig*, I did this reduction both through analysis, PCA, and through compositional intuition. While there is little application of human intuition in most applications of MIR, I have found it useful when composing to maintain a perceivable inner-logic to my music.

Another strategy for dealing with the problems implicit in working in a high dimensional space is through the use of clustering. One of the most popular algorithms for clustering is K-means, which assigns each point to a cluster with the nearest centroid.[12] The result of this, in addition to dimension reduction, is the ability to work within sound categories instead of only relying on Euclidean distance between audio segments.[18] I have found this to be extremely fruitful and use it extensively, see Chapter 2 for more.

2.5 Concatenative Synthesis

Concatenative synthesis is a data-driven method for synthesis that deconstructs a corpus of audio to best match a desired target signal.[27] I will reference concatenative synthesis throughout this paper rather frequently as it is the main technique that musicians are applying MIR techniques to composition and interacting with large corpuses of sound.

Concatenative synthesis consists of four steps: segmentation, analysis, segment selection, and re-synthesis.[27] Concatenative synthesis traces its roots to concatenative text-to-speech, which can be thought of as being 10 years ahead of the field of concatenative synthesis.[28] It is important to differentiate concatenative synthesis from other synthesis techniques that might create audibly similar results, such as granular synthesis, which does not necessitate analysis. Additionally, the convention adopted by Sturm for concatenative synthesis, that for concatenative synthesis to be classified as synthesis requires the output to bear little resemblance to the corpus, distinguishes this technique from other methods of working with MIR for composition.[32]

Historically, concatenative synthesis's can also be traced back to the *musique concrète* of Pierre Schaeffer and *Groupe de Recherche Musicale* (GRM). While this process was completely manual, that is manual descriptors, selection, and concatenation, the progress in the field of concatenative synthesis is the move toward the automated and the real-time application of these techniques.

While I have experimented with concatenative synthesis extensively, I have only utilized it once in a piece of music, *Je me suis trouvé*. This piece was composed with a soprano, working remotely. The workflow would be for me to write a section and send it to the performer for feedback and a reading and we would very much communicate through voice memos. For this reason, in this piece I used the sound of the soprano speaking as a corpus with her singing as the target sound. This lead to interesting outcomes throughout the piece as the two sets of audio were quite different from each other. Other experiments of concatenative synthesis have left me feeling removed from the process of composition and at the will of the design of the program and algorithms.

2.6 Referential and Iconic Sounds

It would be remiss to discuss the use of sound corpuses with out addressing both the concept of musical borrowing and the referential nature of recorded sound. When working with MIR, it is assumed that there is a large corpus to analyze. While it is possible for the corpus not to contain any musical material, when dealing with sound as a material for composition, it is important to the compositional integrity of the composition to address the following questions, posed by Burkholder when discussing acoustic music quotation:

1. What is the relationship of the existing piece to the new work that borrows from it?
2. What element or elements of the existing piece are incorporated into or alluded to by the new work, in whole or part?
3. How does the borrowed material relate to the shape of the new work?
4. How is the borrowed material altered in the new work?
5. What is the function of the borrowed material within the new work, in musical terms?
6. What is the function or meaning of the borrowed material within the new work in associative or extramusical terms, if any?[10]

Self-reflecting, the answers to these questions vary extensively from piece to piece. In *Come On, America!*, the piece is based entirely on a pre-existing composition with the same name from Kenneth M. Murchison from 1918. The borrowed material was the chorus from the song, which was orchestrated throughout the ensemble based on combination tones so that the actual melody could not be heard unless the performers played together. The borrowed material shapes every section of the piece, which is constructed as a theme and variation where the variations are real-time exploration of the bitonic sorting algorithm. The original song, written as a pro-WWI propaganda song about the United States needing to do their part, in the original context it was about going to war. In my new piece I wanted to re-contextualize the piece's provenance through the various systems within the piece that relied on cooperation, to make a statement on the responsibility that the United States has on the globe.

Creating music from corpuses is often compared to, and in papers about the concept often cite, John Oswald's concept of plunderphonics as either an inspiration or an artistic lineage for reference. However, it seems that these comparisons miss the importance of source recognition, as Oswald put it, "The plundering has to be blatant..."[1] Additionally, the use of sounds from specific sources is so important to the conception of his music that they are listed in his CD liner notes. This is often the case because the use of the sampled music provides commentary both on the artist and the music industry, simultaneously satisfying the questions posed by Burkholder above and supplying inner-logic to his compositions.[21]

To use Sturm's convention for the definition of concatenative synthesis, within the process of concatenative synthesis, it is imperative that any referential or iconic nature of the sound should be avoided. This is often done in practice by restricting the length of segments to under 100ms, the threshold for recognition of instrumental timbre. This decision has the effect that the synthesis process works in the sample and micro sound timescales.[26] The creation of sound objects is not a concern of a concatenative synthesis where gesture is controlled by the target sound file. The process of concatenative synthesis then ultimately dissects a heterogeneous sound object, with all its time-varying properties, into several smaller homogeneous notes that represent the state of the sound object from the target file at a given frame of analysis.[27] The combination of the lack of reference and the imposition of homogeneity result in a cold, distant approach to composition.

2.7 Spectral Music

Composers have used timbre and spectral data as compositional materials in the past, both in acoustic and electronic works. Looking back to Henry Cowell's New Musical Resources published in 1919, the use of harmony, and really timbre, has been used compositional material for over 100 years. In the years following Cowell's publication, and not necessarily as a result of it, several composers drifted into what would now be considered proto-spectralist composition, see Olivier Messiaen's *Couleurs de la Cité Céleste*, Karlheinz Stockhausen's *Mantra*, and György Ligeti's *Atmospheres.*[4]

In the 1970s, two centers, Paris and Cologne, L'Itinéraire and Feedback Studios, were founded that continued work in producing spectrally informed music. During this time, the composers at these centers attempted different approaches at dealing with problems intrinsic to spectral music. On the role of melody in spectral music, for example, some L'Itinéraire composers chose to avoid it entirely, see Gérard Grisey's *Partiels* and Tristan Murail's *Gondwana*, while some Feedback composers put melody at the center of their compositional process, such as in Clarence Barlow's *cogluotobiisisletmesi* and Johannes Fritsch's String Quintet.

Several composers, notably, Murail and Grisey, actively embraced a new attitude toward composition, "spectral music", as designated by Hugues Dufourt in 1979. These composers that shared this broad attitude towards composition view "music [as] ultimately sound evolving in time."[15] Their composition practice relied heavily on the technological breakthroughs in software, historically this software developed at IRCAM, for generating the analysis that was required for their music. In the fifty years since the advent of spectral music, the software that used to be a barrier to entry in the field, has now largely been democratized and is mostly available online as free and open source software and as a result, there is a varying field of applications of spectral music techniques as well as an embrace of the techniques of the spectralists without a similar embrace of the label.

As far as technical aspects of spectralism, it is generally quite difficult to generalize as each piece operates on its own inner-logic. However, "from its beginnings, this music has been characterized by the hypnotic power of slowness and by a virtual obsession with continuity, thresholds, transience and dynamic forms."[17]

Because of the narrow description of spectralism from Grisey above, I had never considered my music to be remotely related to spectral music, mostly because of my compositional and aesthetic lack of enthusiasm toward slowness and continuity, all of my music is conceived as sound evolving in time with special attention given to the spectrum and timbre of the sounds. Working in the field of synthesis, specifically analog modular synthesis, I have felt the hypnosis of the slow evolving nature of the medium. However, the hypnosis ends when I step away from the synthesizer as a performer and instead listen critically. However, within all of my music is the influence of spectralism. In my entire approach to MIR driven algorithmic music, this influence is quit obvious, the main difference is the speed of the rate of change and purposeful lack of continuity. In my early works, the importance of spectral relationships and psychoacoustics can be found in every piece. In *vizual* the mapping of spectrum to the visual representation on the oscilloscope were tightly bound to the conception of the piece and guided every compositional decision made in the piece. Techniques such as formant synthesis and the use of combination tones in *Come On, America!* can be traced directly back to the influence of spectralism.

Chapter 3

Implementation

In 2015 I started to work in the field of interactive audio-video installation. This practice required some form of interface between the user and installation, I opted for the use of buttons and switches attached to an Arduino. After working a while within this field and being inspired by experiencing installations from interactive public art installations, I wanted to utilize microphones for user input. At that time the only method I knew to harness of sound was amplitude, which I used with an amplitude gate. I attempted to use this in concert pieces, specifically my piece Tears of Tesla 2 : Electric Boogaloo for accordion, electronics, and video. The piece moved through time with predefined parameters that were only affected by the amplitude of the accordion, where sounds were triggered and the amplitude was measured within Max/MSP and then visuals were produced using Processing with analysis data being sent over OSC.

I improved upon this later by adding an envelope follower, analysis of spectral centroid, and analyzing the amplitude of the seven octaves of the piano as frequency bands. Having more parameters to analyze meant that I could create more detailed visualizations. [1]

After attempting to use available concatenative synthesis software to create music, I

[1] It was during this time that I read the NIME paper for Adam Stark's Sound Analyzer and through that I dove deep into Stark's Music Information Retrieval library, Gist, that ran the analysis, as well as the field of Music Information Retrieval as a whole.[30]

Piece	Techniques
Tears of Tesla 2: Electric Boogaloo	analysis of amplitude for visualization
Come On, America!	analysis, formant synthesis, parameter modulation
Peace? Through the Night	time stretched use of pre-existing material
The Problem with Pandas	extraction of rhythmic contours from tape
vizual	analysis/re-synthesis, visualization (oscilloscope)
Je me suis trouvé	concatenative synthesis, analysis of sound grains

was unable to find a suitable software for my personal aesthetics and technical concerns. It was during this time I created my own implementation of concatenative synthesis, Catatonic. My attempt to improve upon the available offerings by reducing the iterative latency by implementing it in C++ and visualizing the Euclidean distance between samples based on audio features in a 3D space using Allolib and implementing clustering.[2] Ultimately, using both pre-existing and my bespoke software, I was unable to produce musically significant results. The issue was not with any given implementation of the algorithm but the algorithm itself falling short of my goals. But the way of working with sound had led to a breakthrough in my compositional practice by changing how I worked with sound material.

3.1 Precursors

Prior to my current compositional practice, in hindsight, I had attempted through several preceding pieces of mine to solve problems that would ultimately become part of compositional aesthetics and practice. See the below table for a quick description of compositional experiments or techniques explored in each of these earlier pieces.

These precursory pieces all were necessary in my compositional progression as each experiment led to a better understanding of the resulting material. The process of composing *Tears of Telsa 2: Electric Boogaloo* the motivation to continue down exploring possibilities of working with sound that did not exist in my prior work. Both *Peace?*

Through the Night and *Come On, America!* were written around the same time and share similar themes and process but more importantly to my compositional progression, the process of composing these piece led me to more possibilities of working with pre-existing written music as musical material. *The Problem with Pandas* was an attempt at bridging the gap between performer and tape rhythmically and that process led to a more complex conception of rhythm that was a place that I felt compositionally deficient. Both *vizual* and *Je me suis trouvé* were earlier experiments with feature extraction and analysis/re-synthesis.

3.2 Recent Work

3.2.1 Techniques for Composing with MIR

While concatenative synthesis has prioritized total automation and real-time synthesis, I have found both of these goals to be at odds with both the full potential of composing with MIR data and large databases of sounds as well as my personal aesthetic goals.

Often automation is thought of as relinquishing decisions of the user to the computer. This can be helpful when dealing with large amounts of data or choices that a user might act in a predictable way. However, it also results in the lack of control of the user and assumes that the user would like to interface repetitively in the same manner. In the world of digital instrument design, the lack of control as a result of automation creates a wall between the computer and the human which ultimately leads to a less convincing performance as the chain of cause and effect cannot be established. The role of the composer, specifically in fixed media pieces, is akin to that of a performer or interpreter and therefore should be given priority to a given algorithm or process. Therefore, the following strategies and techniques for working with sound that follow are all methods that allow for human arbitration at some point in the composition workflow.

3.2.1.1 Similarity Strategies

While concatenative synthesis uses a target file for similarity calculations, there seems to be no reason that a composer could not create a trajectory through latent space against which similarity calculations could be made.

The simplest example of this would be a scenario where a composer has a pure sine wave and white noise and in effort to bridge the extremes, wants to query a database for a file, based on some subset of extracted features, that would fall equal distance between the two poles. It is quite elementary to ramp up the complexity in this trajectory, for example, by adding more equally spaced intermediary files or changing the navigation through the database from a linear progression to one that is logarithmic or exponential.

Limiting of databases can be done in passes where either a specified number of entries, e.g. the closest 100 entries, or a specified percentage of the whole, e.g. the closest 10%, by a given feature may be used. When done in an initial pass, it is possible to limit queries with audio files that are varied in several salient independent features. This practice is useful in creating clouds of audio for a specified feature in a structure that the composer wishes to stay relatively static but maintain inner motion. The choice of whether to restrict through a given number of entries in a database or through a percentage of a whole is dependent on both the context established in the piece and the compositional limitations of the remaining files. This is often established through trial and error on the part of the composer.

I used this in *Usynlig* as one of the strategies to restrict the amount of material that I was working on to something more manageable. As this piece revolved around the use of iconic sounds and their relationship to each other, I used referential sounds as anchor units and restricted the nearest-neighbors to an amount that was manageable without wasting too much time on auditioning sounds but maintaining enough alternate sounds to layer in the background.

Another technique for working with similarity is the act of preserving order as a musical parameter.[31] Working in this way, the composer can create a control curve that could be mapped to order through time. Continuing with the example from the paragraph above, each query for each subsequent file in a gesture could be controlled by the similarity curve. For example, given a similarity curve of a sine wave and four queries of a database, the first query could be restricted to the closest 50 files, the second to the closest 100 files, the third to the closes 50 files, the fourth to the closest 25 files. I have found this useful when composing from the top-down and a compositional parameter is both the density and how tightly packed the files are to each other, but also useful when auditioning a section of a piece of music and finding a lack of diversity of material, controlling the span of files can provide the needed material for the section.

"Like other judgements, similarity depends on context and frame of reference" and it is important as a composer to understand that the breadth of possibilities of digital audio, when purposely attempting to convey the similarity between several files, it is important to prepare and frame the given gesture.[34] A technique for this could be to, at the beginning of the piece or section, have a gesture that spans the extremes of a given feature. This technique is similar to the practice by J.S. Bach to hide within a fugue subject key centers for the remainder of the section. However, with the conceit that the listener will come to a piece with fresh ears, a composer can limit the audio features used in the composition to further emphasize a planned similarity comparison. This initial compositional similarity comparison, i.e. gesture that goes from a dark to a bright sound, can aid in training the listener to the operating rules of a given composition. In my own music, I have found it quite difficult to give the frame of reference as an initial statement, instead I conceive of the frame of reference as something that is acquired through several listenings.

3.2.1.2 Interpolation of Anchor Units

The process of creating a trajectory through a sound database of MIR data requires the composer to create limits on their searching and querying. This can be done by selecting a referential or iconic sound, a sound that has desired sonic characteristics, or any other heuristic method. This creates a hierarchy where the audio at the beginning and ending of the trajectory are elevated and the navigation through the database is ultimately an interpolation between selected files. Stoll calls these selected files of importance, anchor units, in that the piece, gesture, or structure can be built anchored to the selected files. The interpolation between the anchor units can be linear, logarithmic, exponential, etc.[31]

A challenge with working in this way with digital audio is that the sonic possibilities are beyond human perception and therefore transversing any given parameter lacks importance unless the listener is presented with a frame of reference. The use of anchor units, when properly prepared, can provide the listener with a frame of reference which allows the underlying gestural material to shine through and for the arrival at an anchor unit to feel fulfilling.

3.2.1.3 Use of Analysis as Control

Taking the data straight from an analysis frame for gestural or structural shape is most akin to the process of data sonification. Because of the limitations to low-level analysis, the data will be error-laden and it is advised to filter the data through a low pass filter to remove the resulting noise.

What often is more musically rich than a direct mapping of sonified data to MIDI pitch, is the use of analysis curves to parameter curves. This way of working is quite similar to that of a modular synthesizer with the big difference being that the curves are not based on elementary wave shapes but rather from analysis of a given sound. This leads to more interesting curves, more artistic cohesion, and more nuanced approach to manipulating control signals as maintaining the integrity of the analysis intrinsically becomes important to the conception of the piece.

In 3.1 the process of linearly sampling a sine wave is visualized. This is a representation of how a sample and hold circuit attached to a low-frequency oscillator might function. I have not found this technique to be extremely useful in contexts where the extremes of the trajectory are not known to the listener.

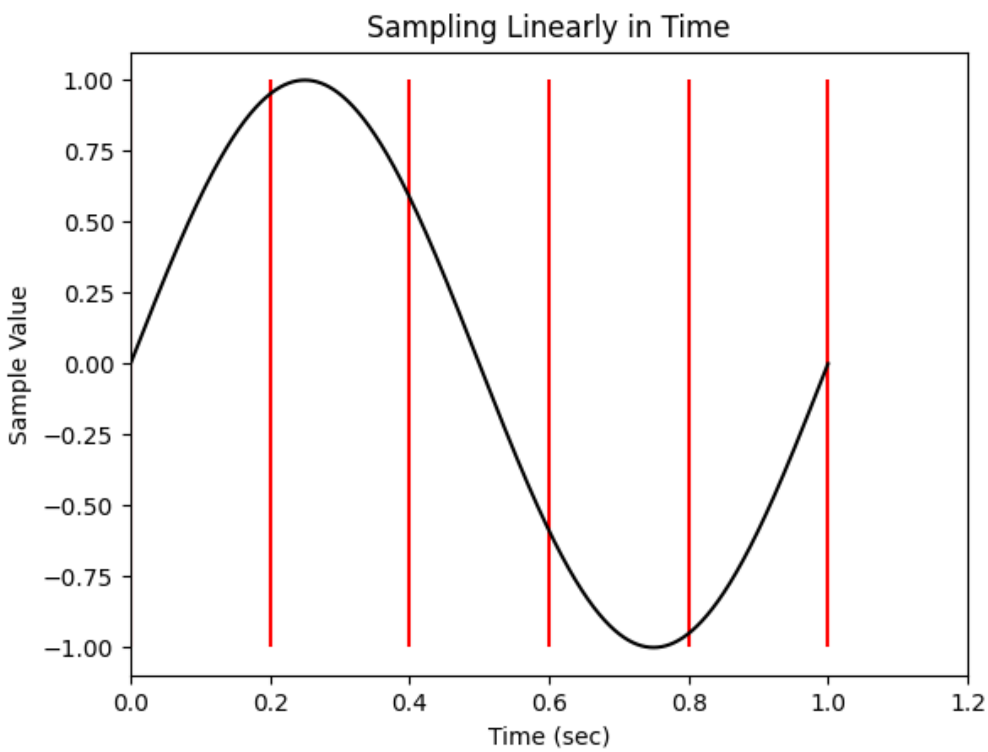


Figure 3.1: Use of sine wave as a stand in for data from analysis, sampled linearly.

3.2 has the visualization of a non-linear sample of a wave. I have found this to be useful in adding complexity to trajectories through the feature space as it allows for a more organic sounding result, as opposed to something sampled at a regular interval. This is a technique I first used in my piece *Come On, America!* where I used the data from a low-frequency sine wave, sampled at the beginning of a section of music, to determine the speed of which the section would be played. While this was written for acoustic instruments, to further showcase the technique, I also transposed the music in each section as if it were a record being played back at a different speed.

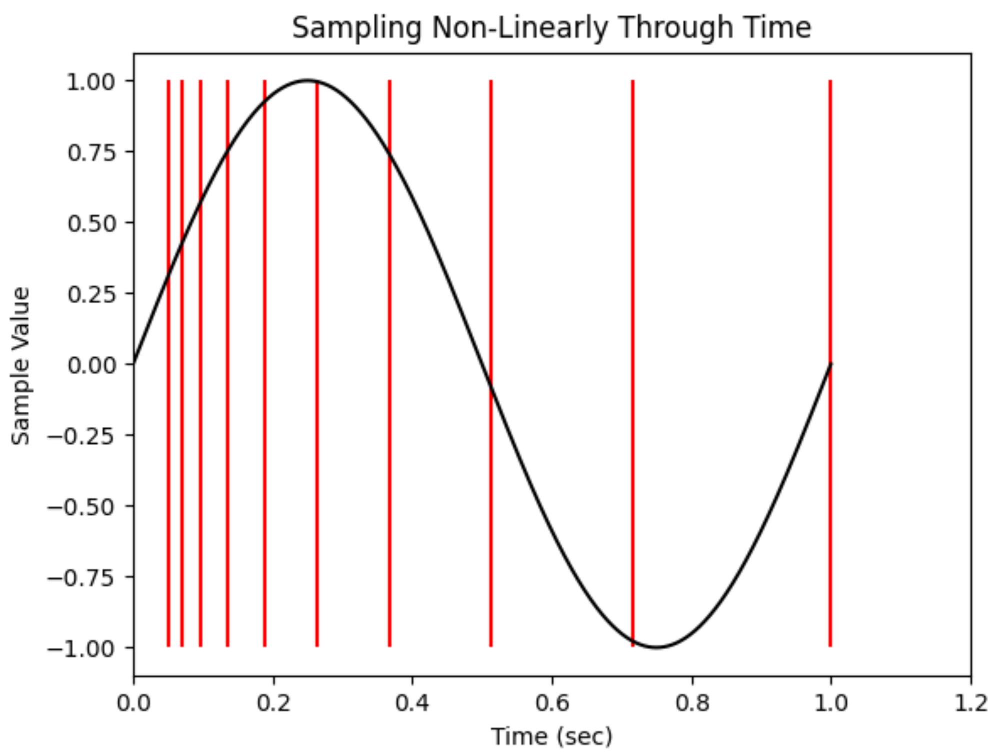


Figure 3.2: Use of sine wave as a stand in for data from analysis, sampled non-linearly.

3.3 is a visualization of a sampling of a sine wave with a Gaussian distribution. I have found this, and similar techniques, useful for non-teleological sections of music.

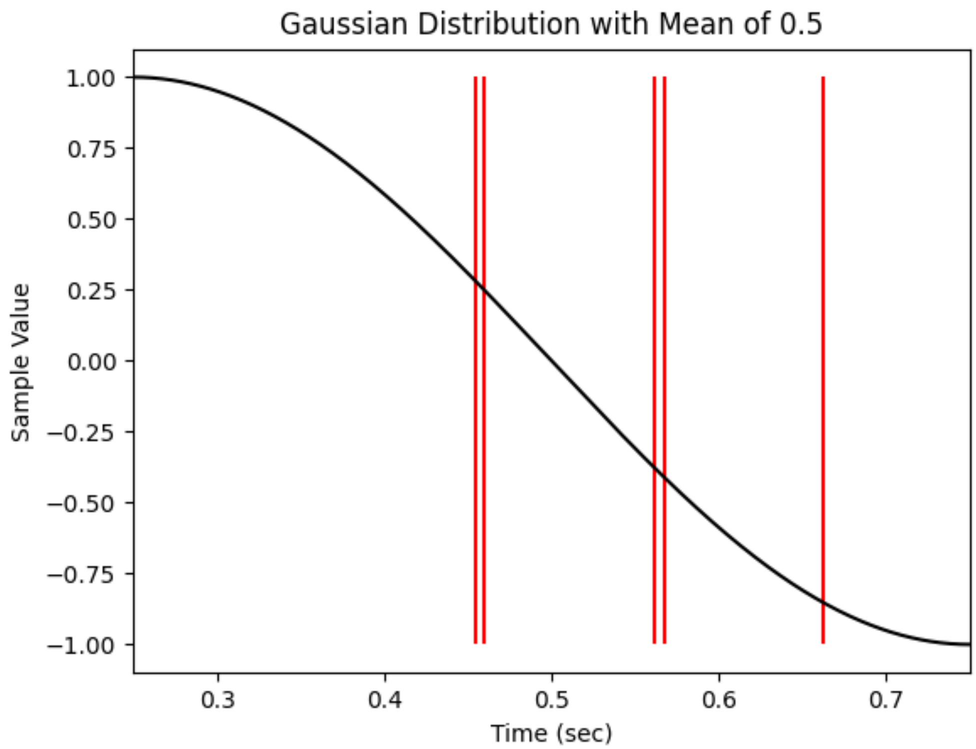


Figure 3.3: Use of sine wave as a stand in for data from analysis, sampled through the use of Gaussian distribution.

3.2.1.4 Querying Strategies

One problem with working with large amounts of data, music or otherwise, is a lack of a strategy, in addition to the unlimited amount of decisions and possibilities, in dealing with the data in a meaningful way.[19] When considering this and constructing my own strategies and techniques for working with large amounts of audio data, I found it important to consult Anderson and Knees's survey of expert users of MIR, who found that:

1. Surprise and serendipity in recommendation and retrieval are important to support creative work
2. Users have personal mental images of sound
3. There is a need for semantic representations of sounds for retrieval, which are not just tags and words but rather reflect those mental images (which can be visual or haptic)
4. Instead of "more of the same" recommendations, desired features are surprise, opposition, individuality, and control over the recommendation process [3]

With these considerations in mind, the shortcomings of the available approaches were blatant. It is of course not the fault of any of the predecessors in the field, but is a result of the four findings not being useful or profitable for development and research. However, for someone preparing to establish querying strategies for navigating a large sound database, their findings served both as motivation and validation.

Working within a database does not often breed "surprise and serendipity" but that is because usually that is not the reason for using a database. However, with this new goal in mind, I was able to come up with querying strategies that would do just that.

To prototype this technique, I started with a dictionary of words with one feature, their index in the dictionary. I started with a target phrase from Kurt Vonegut, "We are what we pretend to be so we must be careful about what we pretend to be." Then I selected 18 random words from the dictionary and interpolated through their index to the target phrase for a given number of intermediary sentences. For any given word from the random phrase to the target phrase was not super interesting, apart from any archaic words that might arise, but the most effective use of this technique was when I would read each intermediary phase on the way to the target phrase. This of course required a frame of reference to fully appreciate but led to surprise and perhaps even serendipity.

The work on translating a mental image of sound to something that is able to be queried requires an intimate knowledge of the features that can be extracted so that searching for the imagined sound can be done with relative ease. However, I also found that trial and error and iteration could greatly improve the translation of imagined sound into realized.

The need for semantic representations that reflect the mental image of a sound can be accomplished through clustering, see below, or through measurement of similarity of the most salient features, or more obscure features however the relationships between the sounds would be obscured. Again in this scenario, an intimate understanding is required to interpret the relationships between sounds, but this can be easily automated.

Finally, the desire for a more individual recommendation process or a process where the user has more control over the process as a desire of expert users of MIR is surprising as it goes against the design philosophy of many technologies that use MIR. However, maintaining flexibility and even allowing for the modulation of a parameter to preserve order can allow for more fluency in this domain. While working in this field, I have been consciously attempting to prevent from boxing myself, and therefore my artistic output, into a set sound or technique. I have found that implementing the analysis of this data in the Python programming language to easily provide me with the flexibility to bring in many different types of analysis, data transformations, and machine learning that could help craft a new piece of music.

The findings were paramount in designing how I wanted to interact with MIR data and are helpful to refer back to reconsider my relationship to a given process. While occasionally my process includes randomness that might induce serendipity, see Gaussian distribution of sampling a feature, my approach is not one that is extremely strict nor often preoccupied on reproducing the same or similar music. The modularity to the design of the tools that I use allow for rerouting of parameters, this is also influenced by my work with modular synthesizers, that allow for discoveries and layers of complexity that often go beyond what I was planning on or conceiving of at the beginning of the process. This modularity also gives the possibility for individuality and control over the music, I often find myself changing how I conceive of a gesture or structure in the middle of composing it and by allowing myself the freedom to give in to a whim, there is often more interesting material produced as a result.

3.2.1.5 Clustering

Clustering of sonic material can be a helpful method in navigating and working with large databases of sound. The application of clustering of sound can be traced back to Stockhausen's KONTAKTE where the material of the sound sources were parameters for serialization, flesh, wood, etc.

Clustering be done by supervised and unsupervised learning both of which have musical applications. Supervising a clustering algorithm means that you can designate number of groups and their classification, for example, sounds that have a certain similarity in MFCC dimensions can be classified as being wood. However, it is perhaps more interesting to cluster sounds through unsupervised learning, as with the K-means algorithm. This will create n-number of clusters based on distance from centroid. Navigating these, can lead to both surprise and revelation.

The most direct application of clustering from a compositional standpoint is in the creation of textural motion and a maintenance of stasis. In my piece *Usynlig* I used clustering both for moments of stasis and for limiting the type of sounds used. I have found through my composition that it is often more fruitful to cluster by unsupervised learning instead of predetermining centroids as the clusters can change drastically based on small changes to parameters. In 3.4 I visualized the clustering of large audio files clustered into three groups. While this clustering is not super effective because of the scale of the audio database it was applied to, it could be a first step in limiting sounds for a gesture. In 3.5 I took the same audio files and features but clustered them into twelve clusters. This gives perhaps a more compositionally interesting result because there is specificity to the clusters. I have found that clustering in passes and excluding all files outside of a given cluster, to be a very effective way to finding audio that is uniquely similar and useful for composition. In 3.6 I took the original large audio database and clustered it multiple times to get to a more manageable amount of files. This visualization is very similar to AudioStellar and it excels at exploration of clusters.

3.2.1.6 Segmentation Strategies

While analysis relies on analysis frame level segmentation, this often does not have musical significance because the size and phase of the analysis frame is independent of any audio content. Concatenative synthesis techniques offer both this fixed frame level segmentation and segmentation by onset detection. This algorithm usually is based on a combination of threshold and change in amplitude over a certain number of analysis frames.

The above approaches are rather limiting and once the composer is free from working exclusively on such small time scales, both of which can have musical applications, the composer can analyze a whole audio file, of indeterminate time, and extract feature average and feature change from the audio file. The former is a simple way to get higher level features like spectral flux and the later replicates a stochastic approach to

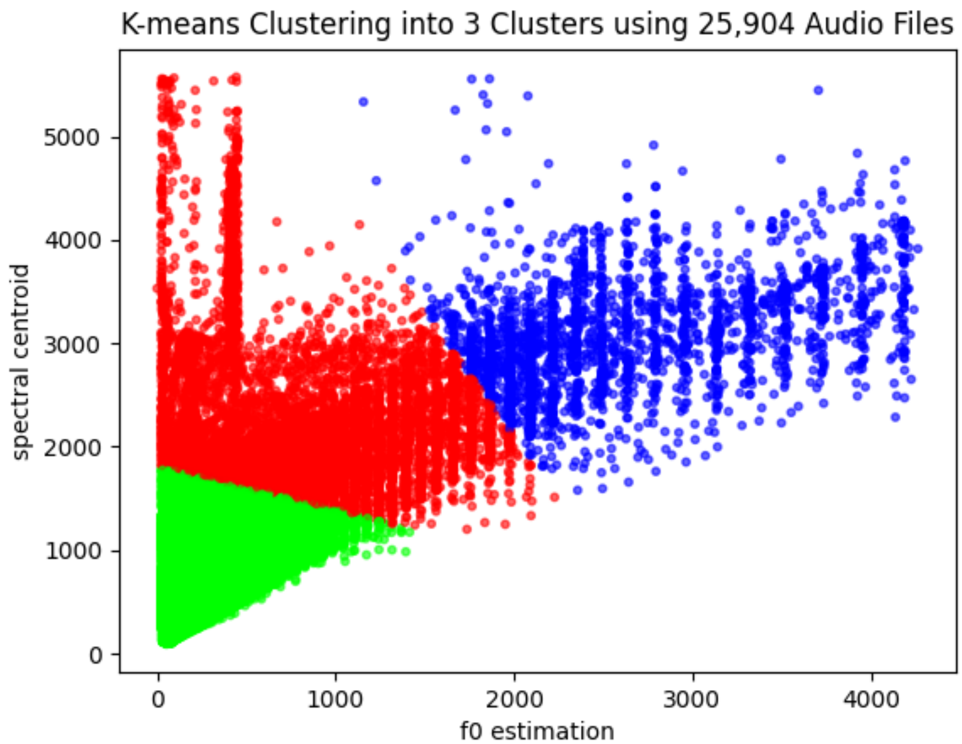


Figure 3.4: Clustering of audio files by f0 and spectral centroid. Unsupervised learning into three clusters.

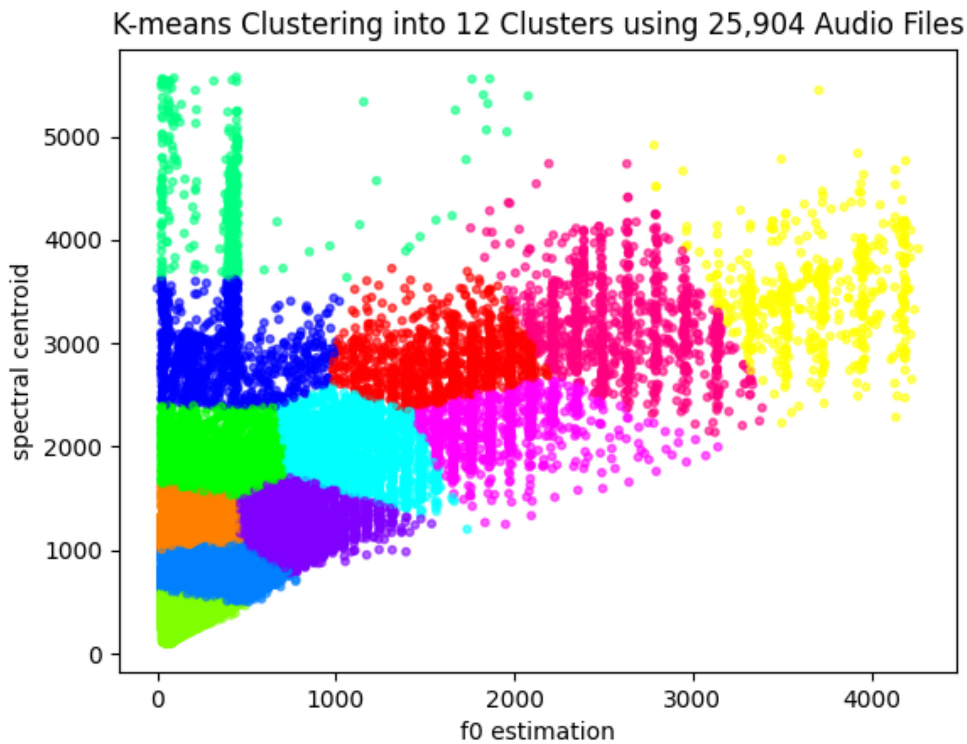


Figure 3.5: Clustering of audio files by f0 and spectral centroid. Unsupervised learning into twelve clusters.

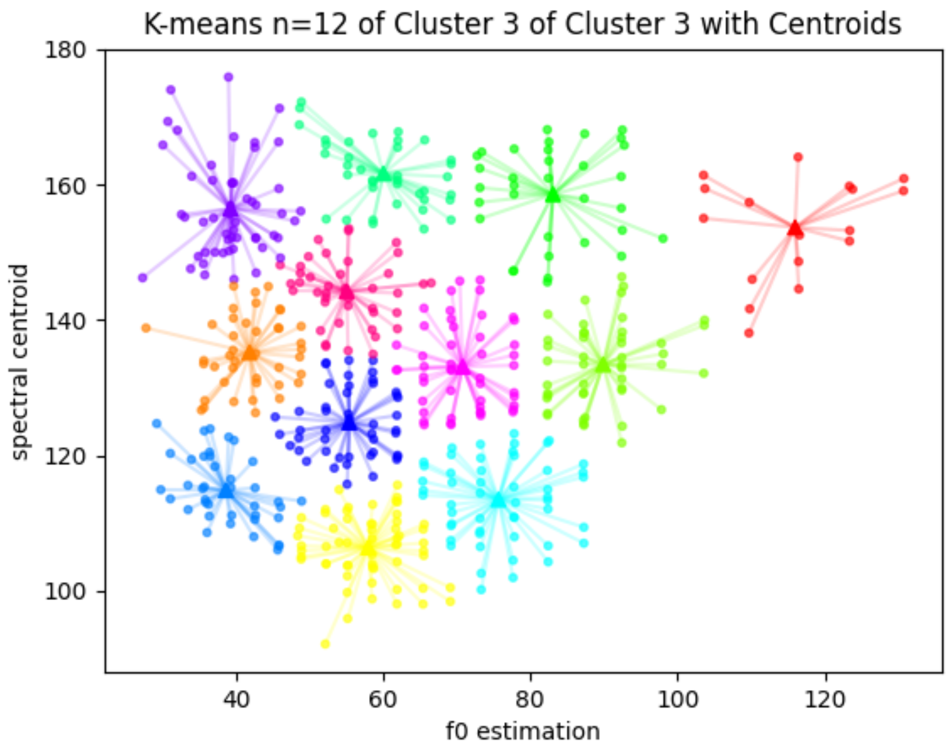


Figure 3.6: Clustering of audio files by f0 and spectral centroid. Unsupervised learning into twelve clusters with centroids.

composition where the sound mass has a given spectral centroid and a spectral spread, but local events are generalized.

3.2.1.7 Conception to Execution

By utilizing both in a bottom up, gestural creation, and in a top down, large forms and structures, manner, the composer has more flexibility to harness MIR data to span larger structures that give the composition both a high level of detail and movement through time in the piece. Working in multiple time scales simultaneously, again adds to the given complexity of an individual gesture but contextualizes it within a larger structure. Pieces of music composed exclusively with concatenative synthesis are exclusively bottom up compositions that are working on a segment by segment basis without any requirement of memory from the previous match. All temporal motion is derived from the target file, and while this would seem to give the composer control over the outcome, that controls comes at the cost of lack of serendipity in the compositional process.

Pieces composed exclusively in a spectralist manner, often do so in a way that relinquishes control and focuses instead on exploitation or representation of a psycho-acoustical property. Spectralism “is in essence outside the world of linear time” and as such, is usually not teleological in nature nor detail-focused.[20] However, by combining approaches, techniques, and attitudes from composers working in the fields of concatenative synthesis and spectralism with those from a background in micro-montage, the composers is able to realize the musical potential while avoiding, as best as possible, downsides inherited from their influences.

3.2.1.8 Establishing a Personal Corpus

Another potential of maintaining a searchable database of audio for use with MIR, is the ease of which tagging can be added to the workflow to further refine a composer’s practice. First and foremost, it makes it easy to add tags to sounds based off of their characteristics, this is time-consuming but can be automated through unsupervised and supervised clustering. Additionally, it allows for the composer to keep track of audio used in past compositions. Compositionally, this is efficient and removes the redundant process of collecting sounds on a piece-by-piece basis but also allows for previously used audio to be excluded from searching. It also allows for exclusion of everything except from pre-existing material. This could be useful for the composer making a retrospective, remix, or just a composer interested in using their past sound choice as a parameter for future musical composition.

3.2.1.9 Use of Salient Features

As the process of composing with MIR does not necessitate a recognition by the listener of any similarity, composition does not need to be driven merely by the most salient features for similarity but instead these features can be restricted. When a database is filtered based off of an anchor unit and given range of the most salient features, this is either by limiting audio files by distance to the anchor unit or by returning a desired number of most similar files, navigating the remaining files by less salient features producing an aural result where the focus of gesture is superseded by a concern of reference or source by the listener. The applications of these audio files produce an overall feeling of stasis and result in the listener returning to source identification as a major mode of listening. This can be a useful technique to employ in the barrier between two different methods of working with MIR data as a refresher to the ear and as a closing of a section.

Researchers working in the MIR field prioritize the most salient audio features because the tasks that these researchers usually work on rely on accurately modeling how humans perceive sound, see playlist recommendation tasks or copyright enforcement. Additionally, researcher-composers that work in the field of sonification rely on psychoacoustics in mapping their data because the task that they are charged with is conveying information in the audio domain.[9] As a composer of music that does not necessarily require the audio transmission of data, I am not bound to rely on such features, however, the hierarchy of audio features is important to know by the composer to prevent unintended results.

3.2.1.10 Gestural Construction

I have found it particularly fruitful to apply Dennis Smalley's Spectromorphologies, specifically the motion and growth processes, to the creation of gestures when working with MIR. While these labels were originally meant to serve a descriptive role rather than prescriptive, composers, such as Manuella Blackburn, have found these labels to be effective for creating music and limiting decisions in the compositional process.[6]

Additionally, I have found that creating these gestures through the use of MIR data has been significantly more straight forward and maintains the referential aspect of the sounds used.

The most straight forward spectromorphology are the group classified as unidirectional and I have found applying the kNN algorithm to an ideal trajectory through the feature space to be the most straightforward method for creating these gestural shapes, see 3.7.

Reciprocal gestures are slightly more involved. Oscillation being a navigation through the feature space with one cycle of a sine wave, eliminating any duplicates from the audio selection algorithm while oscillating. Undulation gestures can be crafted through using a wave-folded sine wave to navigate the feature space, again removing any duplicate files. Parabolic gestures start at a given place in the feature space, descend and then return to where the trajectory started, removing any duplicates that might be present.

Cyclic and centric gestures circle a given point in the feature space, removing any

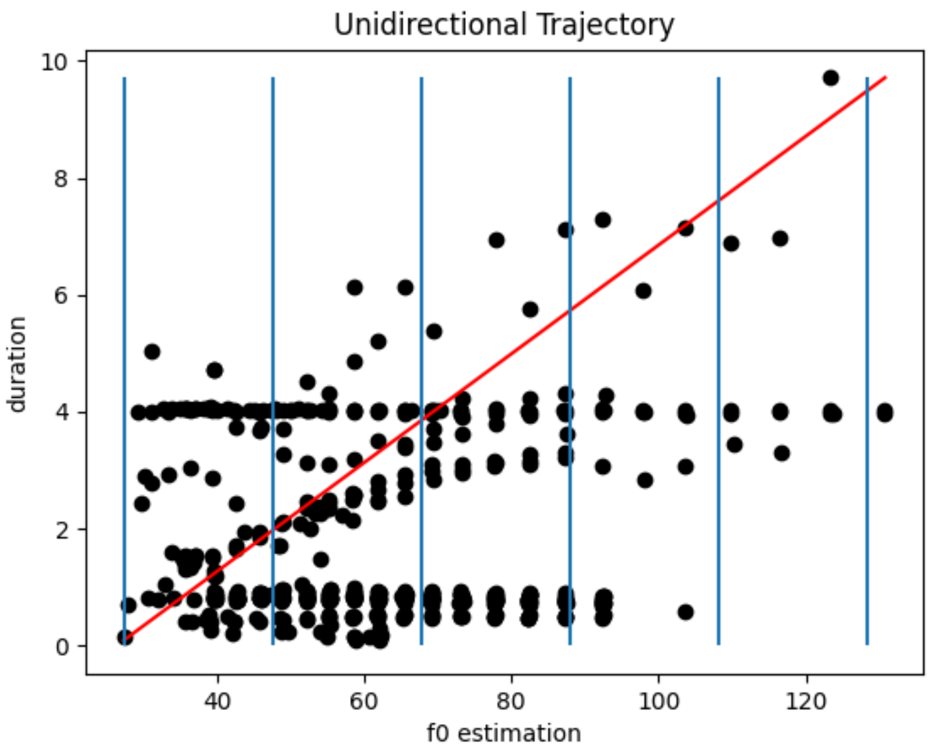


Figure 3.7: Example of a unidirectional trajectory through f0 and duration. An approach to Smalley's ascent.

duplicates from the file list. Bi/Multidirectional gestures like dilation and contractive gestures can be accomplished through taking the top n most similar for a given feature or features and then sort by effective duration (total duration – silence).[29]

The rationale for elaborating here on the application of the motion and growth processes through MIR data and not, for example, the structural functions described in the same paper, is because of the convenience that MIR affords to this process and the creativity, and potential serendipity afforded by the wide variety of sounds used for these techniques.

3.2.2 *Usynlig*

Usynlig was commissioned with the prompt of “In the now”. While I had quite a difficult time of figuring out what that meant to me, much less how I would apply that to a piece of music, I eventually realized that it could be a great application of composing algorithmic music with MIR data. In this piece I challenged myself to leverage my entire audio collection to create a piece of music, 65,000 audio files. By not limiting myself to a subset of audio files for the entirety of the piece I was opening myself up to the potential for a confusing musical experience, given the vast origins of the sounds included in my corpus. My response to the prompt, and my response to the wide variety of audio queried in my database, was to construct structures and gestures that were constructed of an inner logic. In this way the piece asks the listener to stay “In the now” and not get distracted by the referential nature of the sounds and instead focus on the trajectories through the feature space and the relationship of disparate audio files to each other.

Looking back at the compositional process, I can recognize that this piece consisted of four distinct steps: querying sounds for a chosen trajectory through feature space, application of spatialization and other effects, additionally layering of gestures from the large corpus, and arbitrary compositional decisions. These four steps were always in flux and happened in different orders and frequencies given a specific musical instance.

Additionally, looking back at the composition, I recognize toward the end of the piece I purposely began restricting my sound selections to audio that was more iconic than others to further explore the prompt and to give the listener something more to relate to their experiences.

Aesthetically I challenged myself by attempting to, instead of applying a spectral transformation or apply a complicated sound design technique, that I would use MIR to find a missing sound or to fill a gap temporally or timbrally. This decision led to a composition that, consciously or subconsciously, sounds as if it is made of recognizable sounds and a composition that grows in interest on repeat listens as more processes can be heard. For an example of the balance between MIR and sound design, see 4.1 and 4.2.

3.2.3 A More Sound Outlook®

I challenged myself in this piece to replicate the interaction between sound and composition through MIR that I had established in the fixed electronic medium in a piece for acoustic instruments and performers. The first step was to collect sounds, as I would with an electronic piece. For this piece artistically I wanted to document the sound world that was confined to my apartment during the COVID-19 pandemic, as I yearned for a return to normalcy and outside of my apartment. I took survey of the twelve most common sounds that made up the soundscape of my apartment, recorded them, and assembled them in a list to my partner with the request for her to order the sounds based on the amount of stress that each sound caused. My original intention was to analyze each sound and map it to the equal tempered scale and extract gestural and formal structures from the MIR data. However, after analysis it became clear that I had too many streams of data to drive a relatively short piece of music, so I decided to only use the most stressful sound, as ordered by my partner, in attempt to re-contextualize and remove negative associations with the sound. The sound was the sound of her work email, Microsoft Outlook®.

As the piece was intended to be referential, I purposely restricted my analysis to features usually considered salient; FFT, f0 estimation, harmonic spread, noisiness, and power. The first step in analysis was to find the peaks in the FFT to extract harmonic data, I then quantized the analysis to the equal tempered scale as I was writing the piece

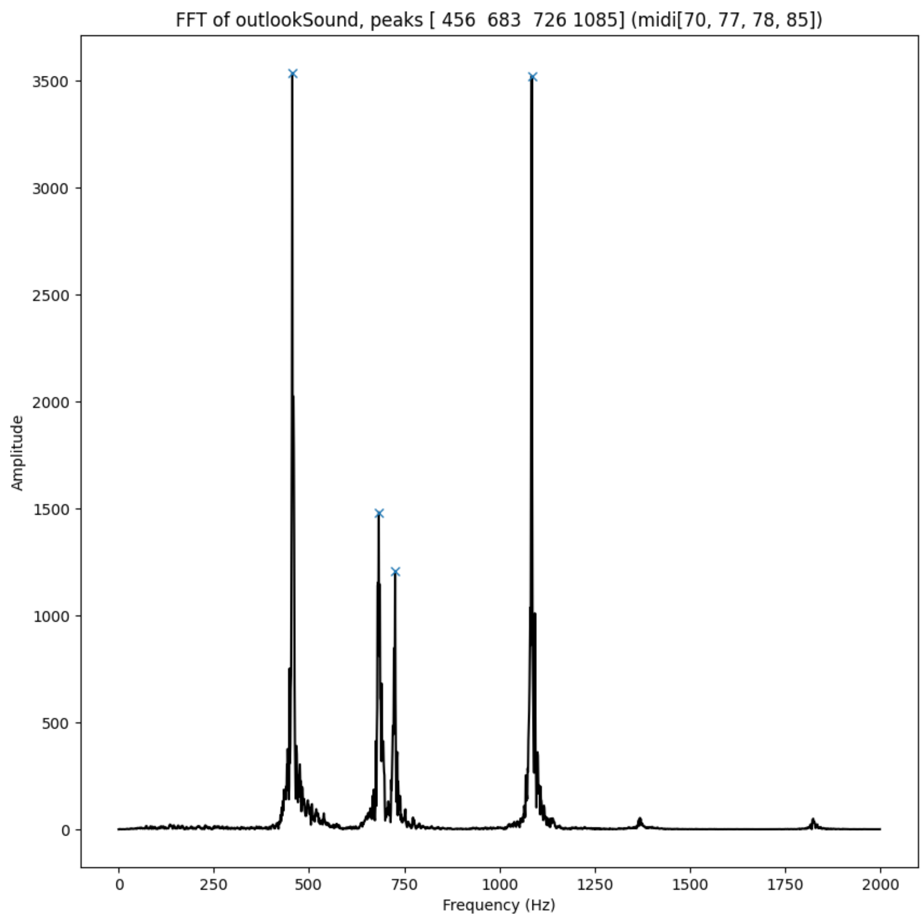


Figure 3.8: FFT analysis of the Microsoft Outlook email notification sound with peaks.

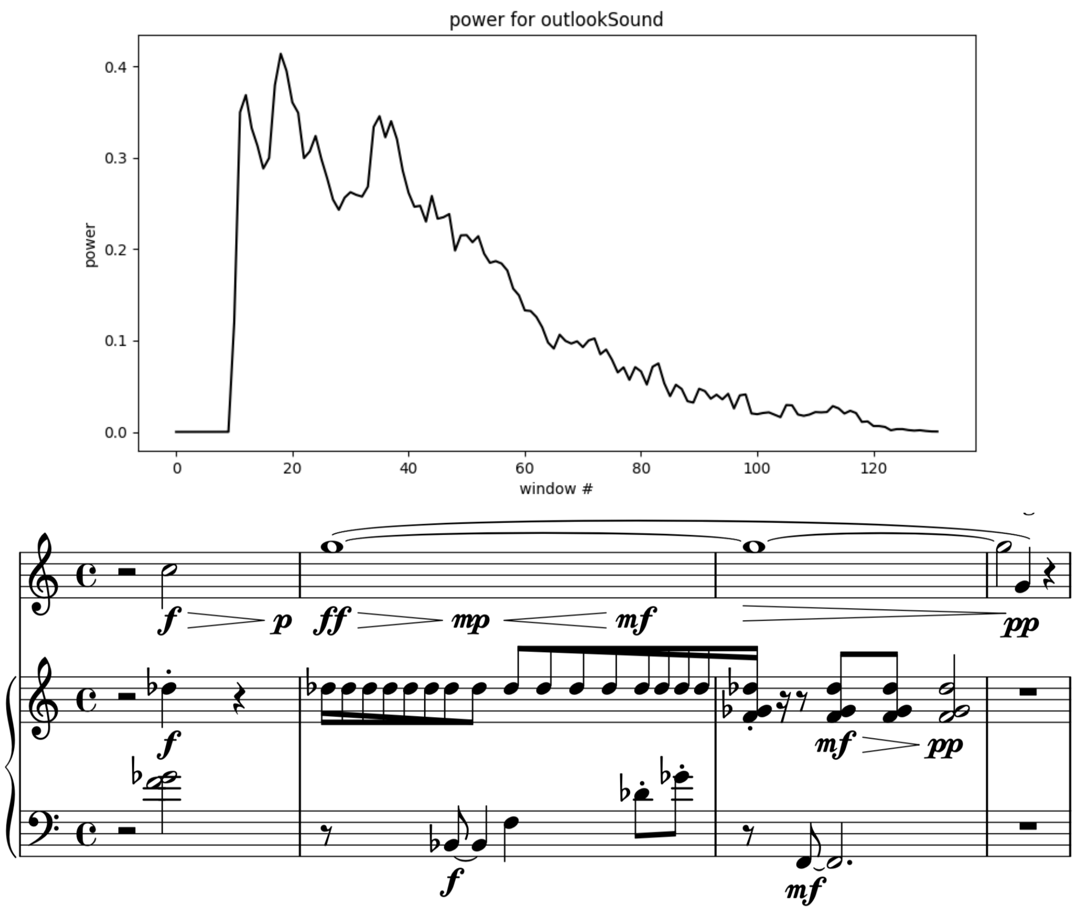


Figure 3.9: Power of the Microsoft Outlook email notification sound and artistic interpretation of data.

for bass clarinet and marimba and did not want to account for microtonal adjustments.

Following this initial step, I took the set produced by peak finding (0158) and applied pitch multiplication to the set to flesh out the harmony and went through the process of using the MIR data as control signals. Through this process, the returning theme in the A section of the piece was derived from the power data mapped on to the harmony extracted through the FFT.

The generation of the main theme, required a back and forth between composer and computer, using intuition and compositional training to artistically reinterpret and orchestrate the control signal into a recognizable motive in the composition. The rest of the piece was constructed in this manner, from the bottom-up.

This replicated the technique used when working with fixed electronic music and MIR where solutions are presented through algorithms but the choice on straying from the recommendations and the task of orchestrating and creating gesture falls on the shoulders of the composer.

Chapter 4

Future Work

4.1 Mixed Music

The prospect of applying the strategies presented in this paper to mixed music, that is music combining performer and electronics, is full of potential. Methods presented by Hans Tutschku using the FluCoMa library to extract data from a performance for control data, is one of the most innovative use of working in the medium and bridging the two sound worlds through a new level of interaction.[33]

Working with mixed media with fixed electronics also has a wealth of potential, providing solutions to deal with the some of the mediums most difficult challenges, bridging the two sound worlds and providing

4.2 Pedagogy

There is incredible potential for a rich composition, analysis, and sound design pedagogy grounded in MIR techniques. These fields, as far as they relate to digital audio, are constructed with the main forms of analysis taking place through inspection of the waveform and spectrogram. By demystifying what is happening beyond these analysis tools, the novice could interact directly with aspects of audio that require years of active listening and experience to fully grasp.

MIR has been used as a pedagogical tool, for an introduction to coding, however, not as means of elucidation of core, yet complex, skills for the modern composer.[38]

Through my experience of teaching composition at the undergraduate level, I have seen an apprehension to composing with sounds and sound objects directly and instead I have found that the novice is more inclined to work with this sounds through a layer of mediation, such as samplers that quantize the audio to both equal temperament and tempo. The use of samplers here, while they provide a quick way to trigger and transpose via MIDI, provide the novice with more information about the components of the sound, namely the fundamental pitch of a sound.

Fundamental pitch estimation is also what is used in sample packs available on the internet in places such as Splice. This additional information is helpful to the composer to limit the number of sounds auditioned, especially in the creation of contemporary pop music where there are a limited number of sound classifications. For example, a composer, or usually referred to as producer in this genre, will search a personal sample library or cloud based one for a kick drum sound in C. This can return hundreds of kick samples and at this point the producer does not have other means of refining a search or classifying sounds. What generally happens is an audition of sounds until a suitable one is found, discarding the remaining files. This approach priorities convenience over completeness and relies on past experiences with sound and expectations of a genre that can be difficult to teach when rationale for picking one sound over another is based on feeling, fatigue, and is subject to choice paralysis.

Contrasting this approach, it is relatively easy to teach both the vocabulary and theory to describe a desired sound and lower the barrier of entry into a field that, in general, relies on years of training and study.

Once a novice is equipped with the vocabulary, skills, and tools to extract musical features from audio, the task of analysis is a lot more feasible and aids in equipping the novice with the skills to go from liking a sound found on the radio to both understand how one might replicate the sound through sound design and why, within the sound's original context, the sound was appealing. In both my own and in my student's compositional paths, I have found that this task is one of the early and recurring road blocks to the creative process. By approaching sound in this way, it refocuses the priority from harmony and form as is usual in a university setting, and equips the novice with other means of creating music from the sound itself, not bound to ancient musical forms. However, it is quite easy to get taken up with manipulating the data and lose sight of the musical result.

Looking forward to continuing lowering the barrier of entry and creating a more equitable field, my future work in this field will, at least in part, be to create tools and beginner friendly material to use and apply MIR for composition as it holds a lot of promise in bridging the gap from metaphorical descriptions of sound and concrete audio features.

4.3 Analysis

Another application of MIR for musical composition is to analyze decisions made by the composer during and after the compositional process.

During the compositional process there are often moments when a structure is lacking for something and the composer then goes through several modes of listening to dissect and diagnose. Having more information can help guide a composer to make decisions and this diagnosis could be automated, if desired. The strength of this approach is that it is possible to analyze both temporal and discrete dimensions of sound and can reveal to the composer nested structures of various sizes that might have gone unnoticed.

In 4.1 I visualized the data for the sound files chosen from my software opening gesture from my piece *Usynlig*. My intent in doing this originally was to see if there was an obvious trend in my selection process that would become apparent. However, as the first gesture is quite varied and I meant for it to be a statement at the beginning of the piece this was not extremely useful. The comparison to 4.2 is quite useful, as it shows the data for the audio files in their final state in the piece given spectral and temporal transformations. It also showcases the conflict between a conceived sound in my head, an algorithm to generate the sound, and the process of applying my ear to shape the sound to what was desired. This type of analysis seems helpful after a piece has been composed but could lead to quite a change in a compositional process if used in the middle of composing. I have not attempted to compose while visualizing the features of the audio files I am manipulating in real-time but I do believe that it would be an interesting experience.

After the compositional process is completed, analyses can be made both on the audio selected from the larger corpus and how, through sound transformations, the audio selected was changed to work in the composition. Both of these analyses could, given enough data, be used as models for selecting from a corpus audio files that would fit a large compositional aesthetic. This could be useful in creating a sonic brand or identity and also extending a piece of music that would maintain a sense of sonic cohesion even if the structural organization would significantly change. This could also end horribly because of the lack of specificity and complexity of the features and models that are available now.

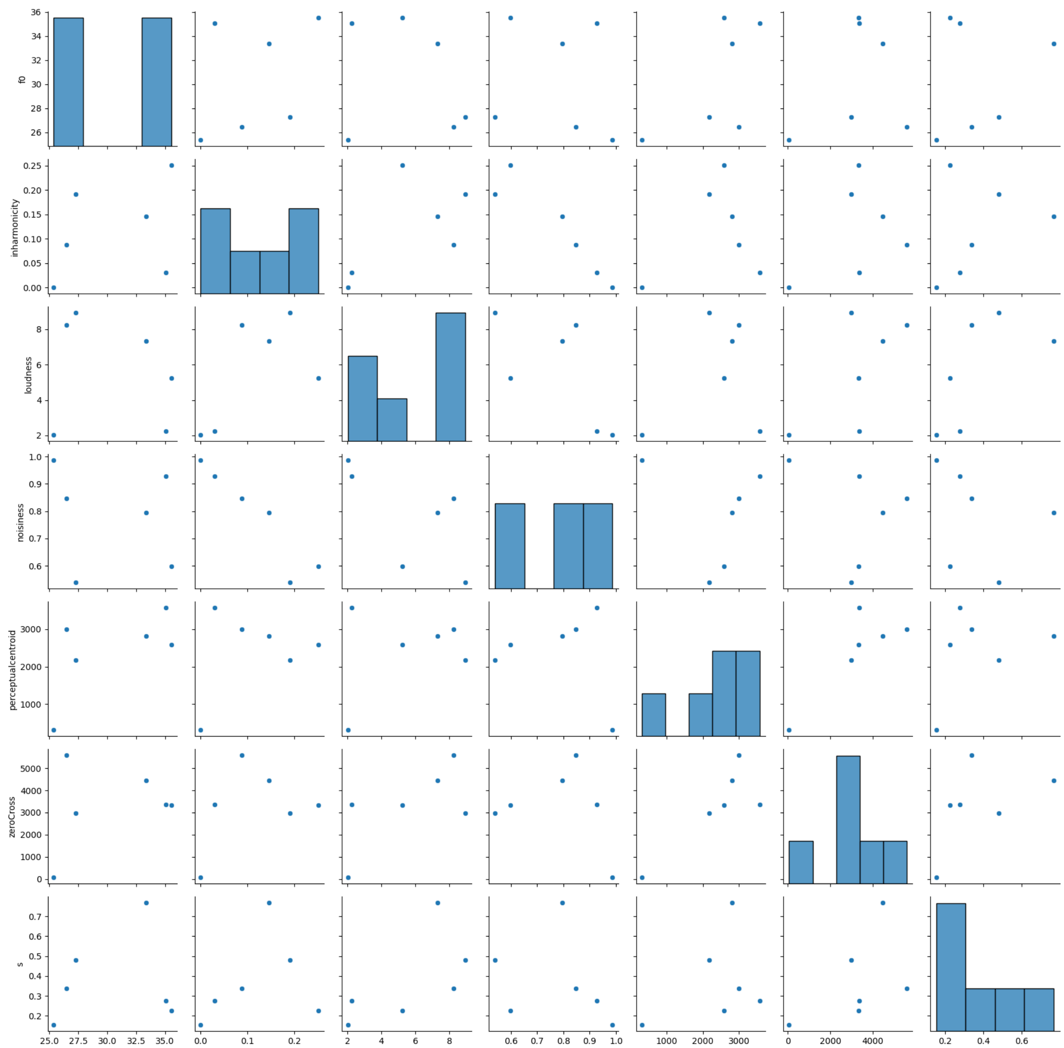


Figure 4.1: Analysis and comparison of salient audio features from the audio files chosen for the opening gesture of *Usynlig*.

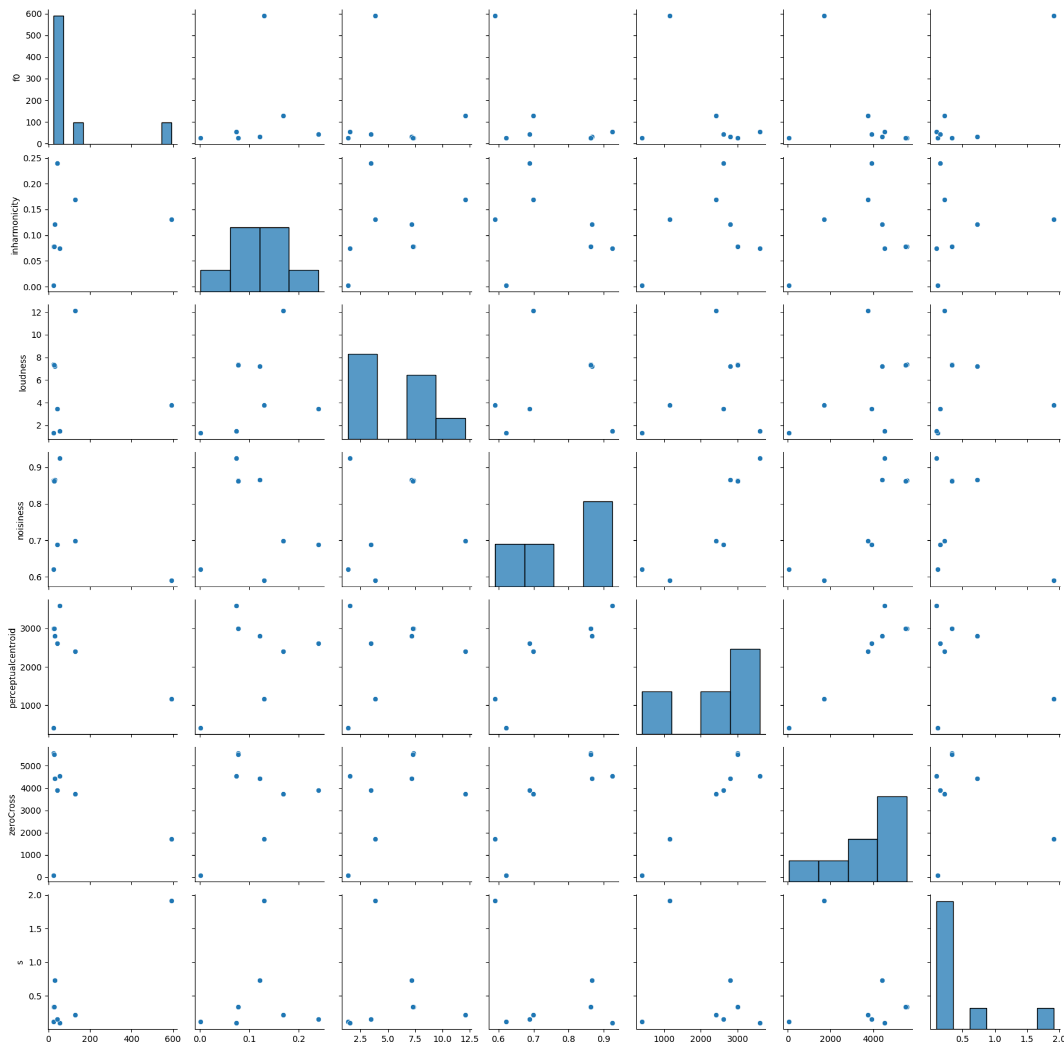


Figure 4.2: Analysis and comparison of salient audio features from the opening gesture of *Usynlig* as the sound files are presented in the piece.

4.4 Streamlining of Interface

My current workflow contains quite a bit of latency in the process of querying and sorting sound and actually manipulating that sound in the DAW. This slows down iteration and therefore slows down the compositional process.[14] There has been interesting work done within the DAW REAPER and its scripting possibilities that cut down on latency between database and audio. Specifically, nvk CREATE, which allows the composer to query a database based off of descriptive strings but also allows for quick auditioning and limited transformation of sound.[36] While this specific script was designed for game and sound designers to streamline their process, the capabilities that it showcases and the fact that the DAW is open source and scriptable is promising.

Future work will be made in creating a workflow, probably REAPER specific because of its scriptability, that minimizes the latency in feedback and allows for faster iteration of compositional material.

Appendix A

Portfolio of Compositions

A.1 A More Sound Outlook®

This piece was composed during the COVID pandemic. Unlike other pieces that I composed during this period, pieces that all served as a means of escape, I wanted to use this piece as a way of documenting what my life had been for the last year.

I started writing this piece by creating a list of the most common sounds heard around my apartment, I then asked my partner to order my list along a spectrum of anxiety.

I then analyzed each sound through standard MIR techniques; f0 estimation, harmonic spread, noisiness, and power; originally intending to use all of the sound samples as musical material.

Through the process of composing this piece, I realized that I had enough material in a single sound, the sound that causes my partner the most stress, the sound of her work email, Microsoft Outlook. This piece is both composed by data extracted from this stressful sound and serves as a way to re-contextualize and remove the anxiety associated with the sound. The title is a nod to the source material and a hopeful look to the future as the vaccine rollout continues.

This piece was composed for Transient Canvas.

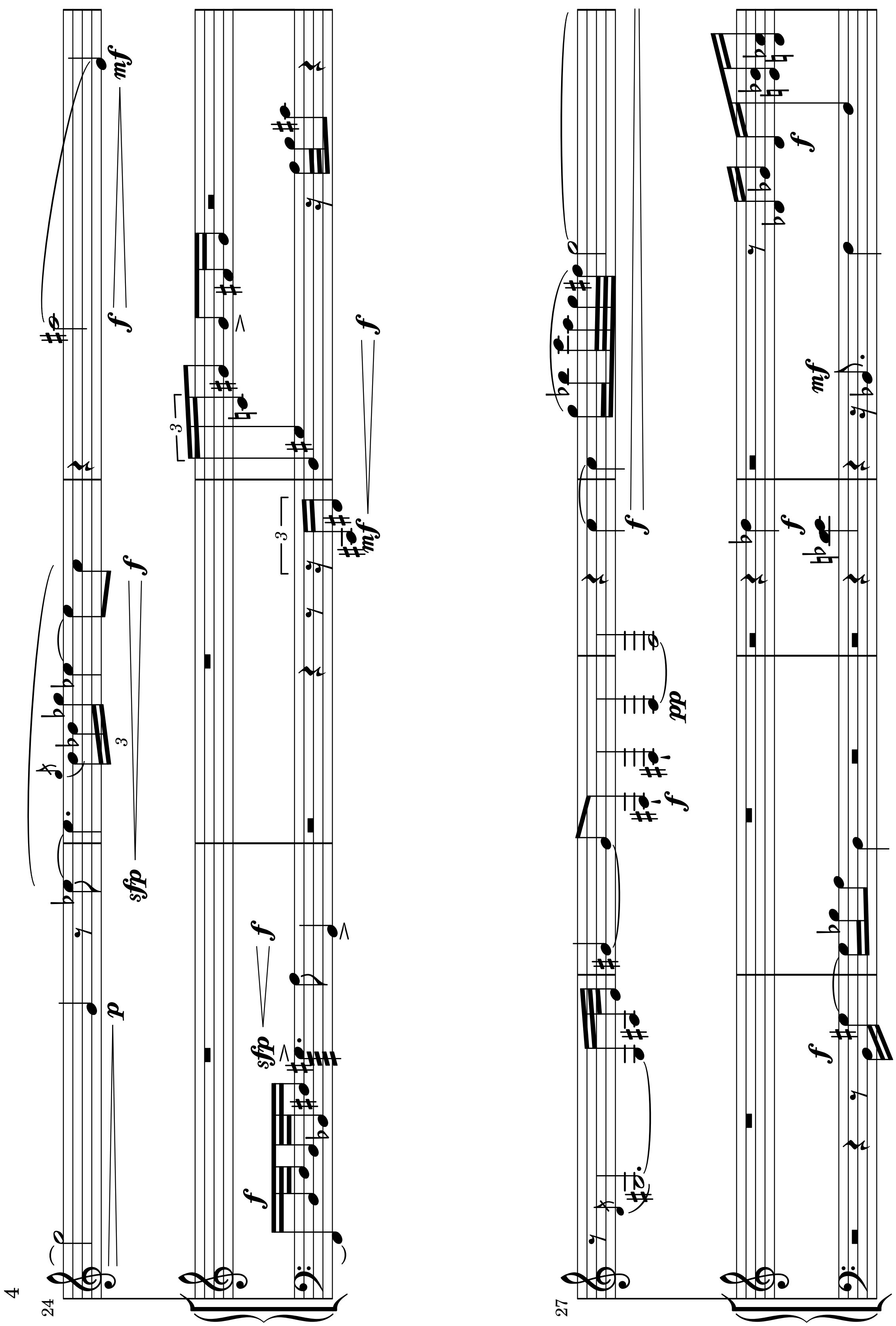
A MORE SOUND OUTLOOK ®

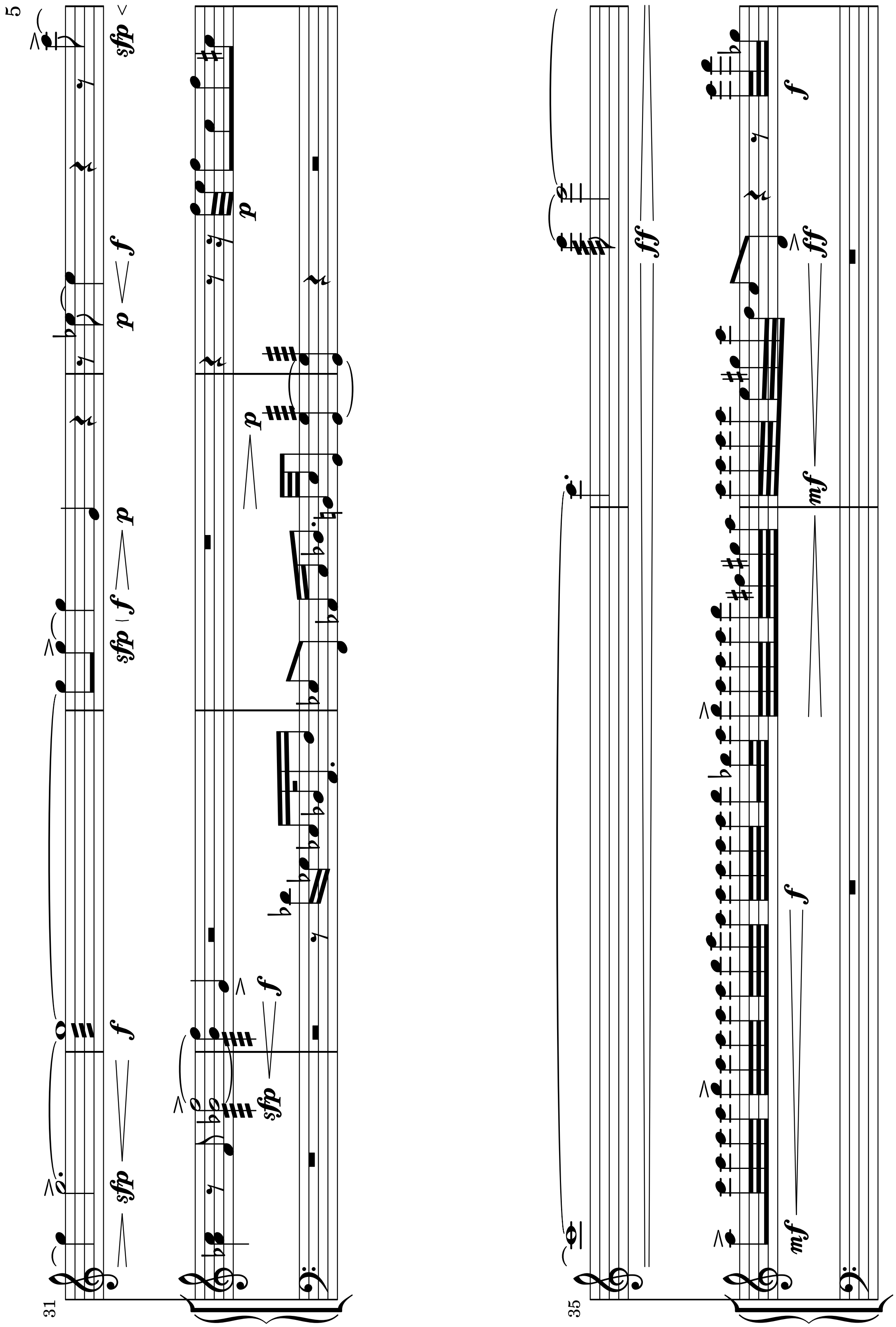
for Transient Canvas

Transposed Score

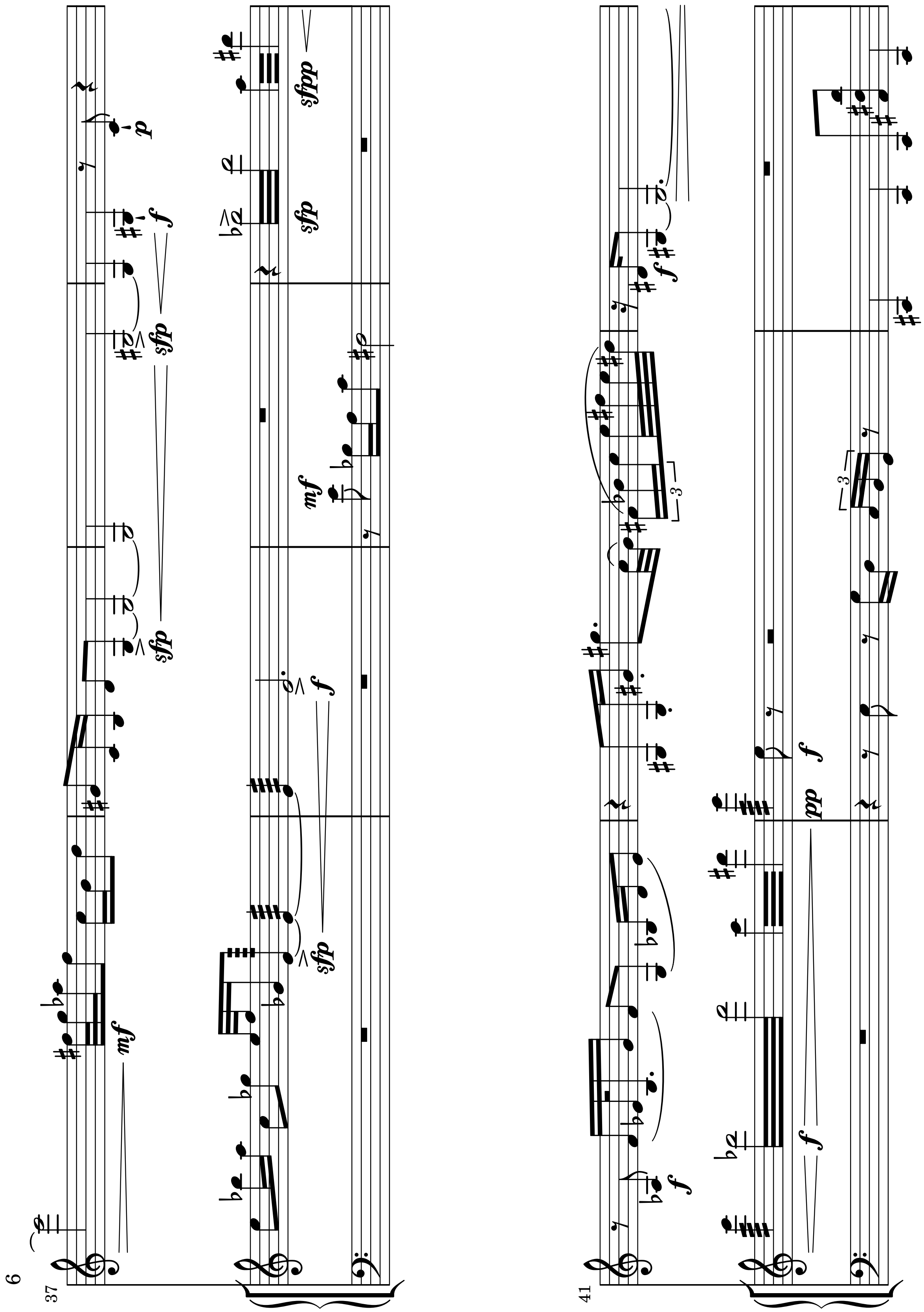
Stewart Engart

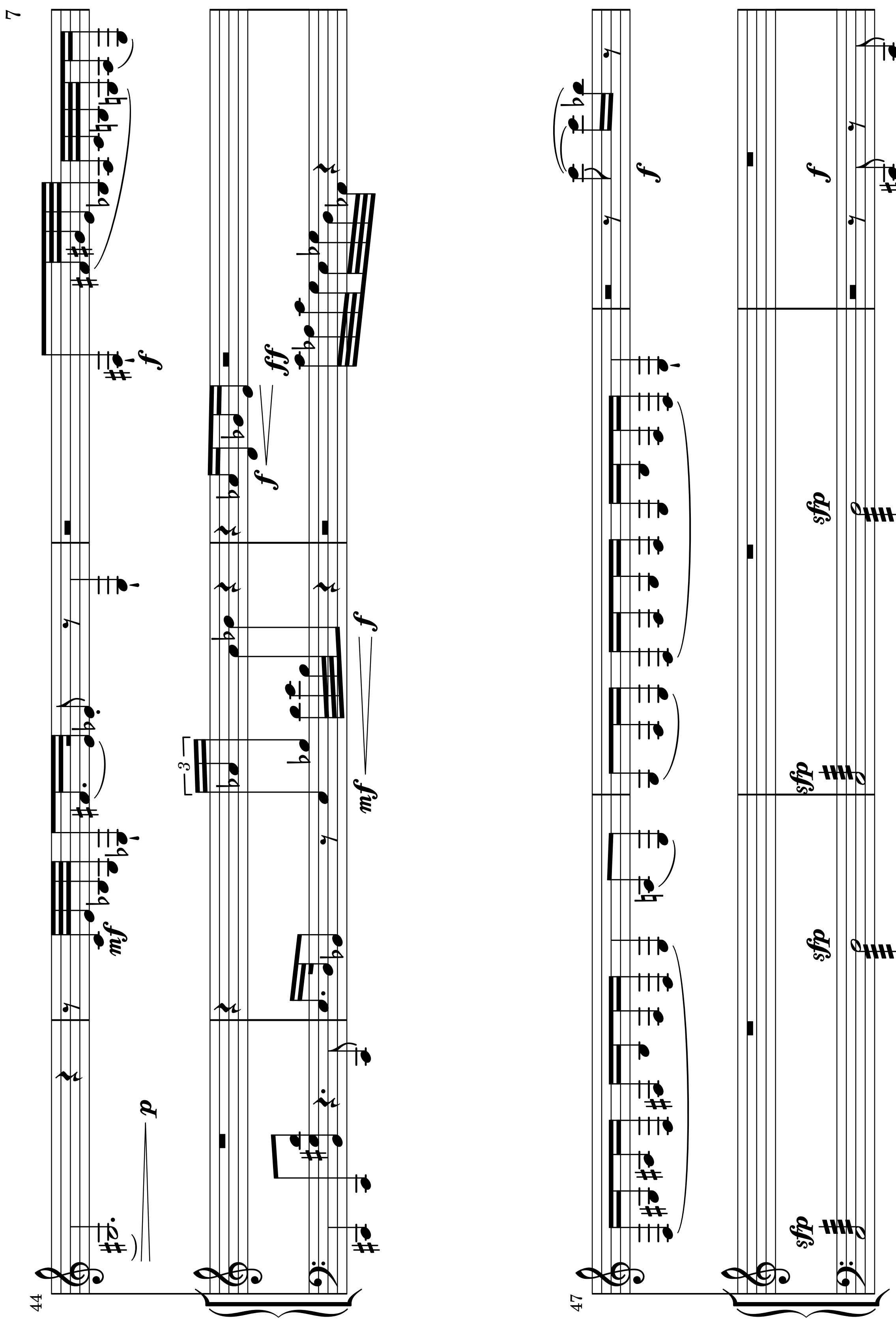

17
21

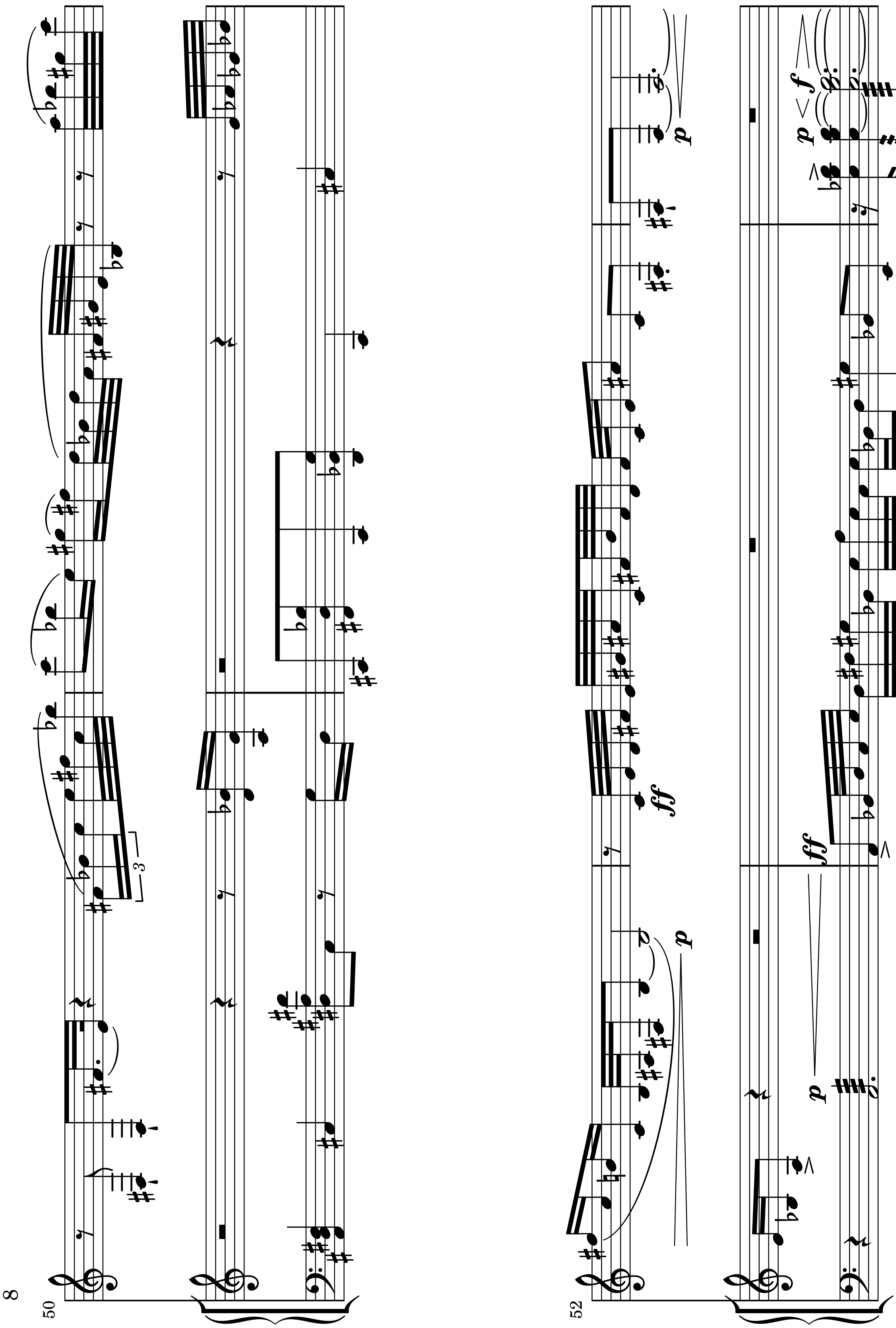


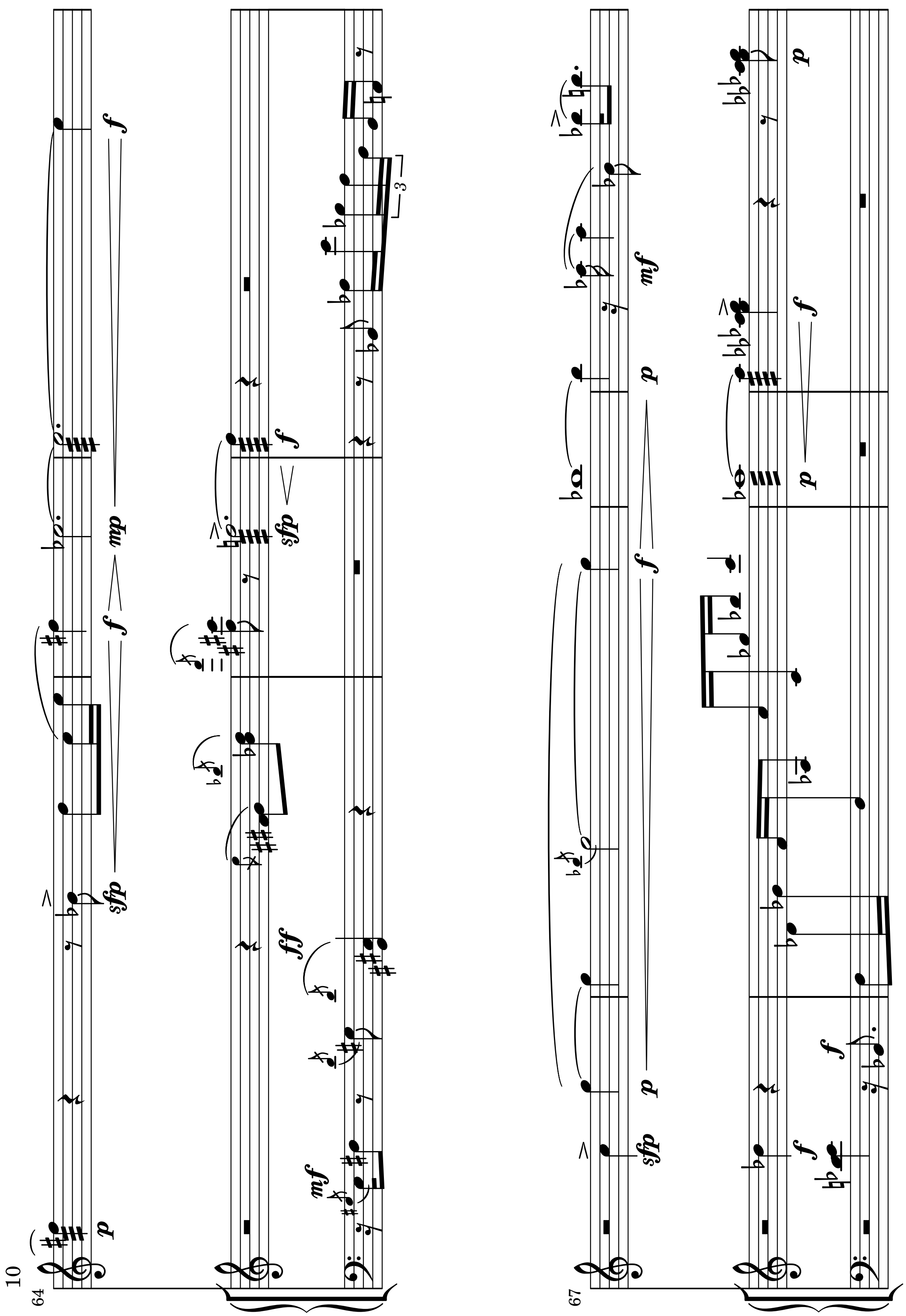


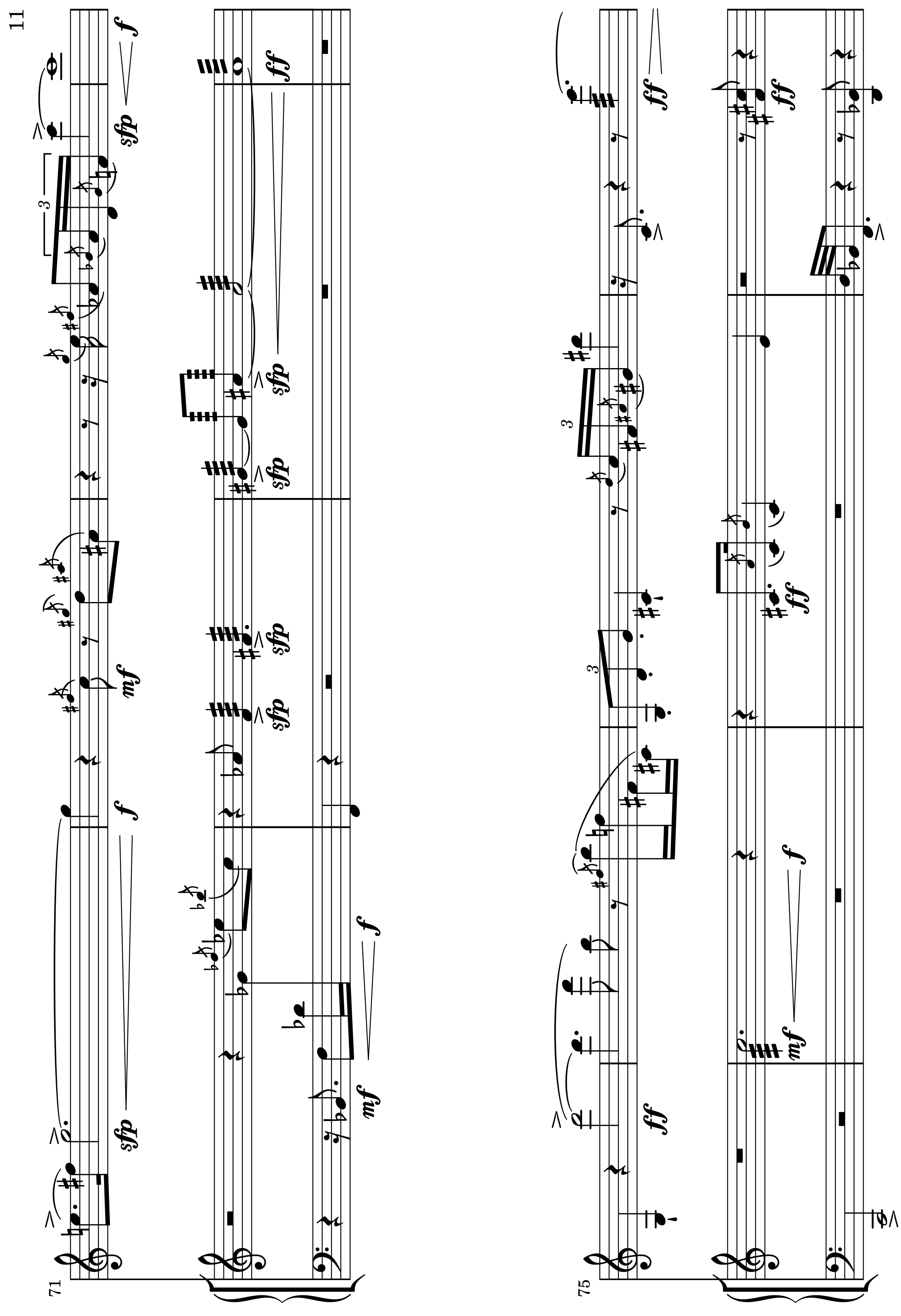
31
sfp
f
sfp
f
p
p
f
sfp
sfp
f
p
p
5
35
ff
mf
f
mf
ff
f

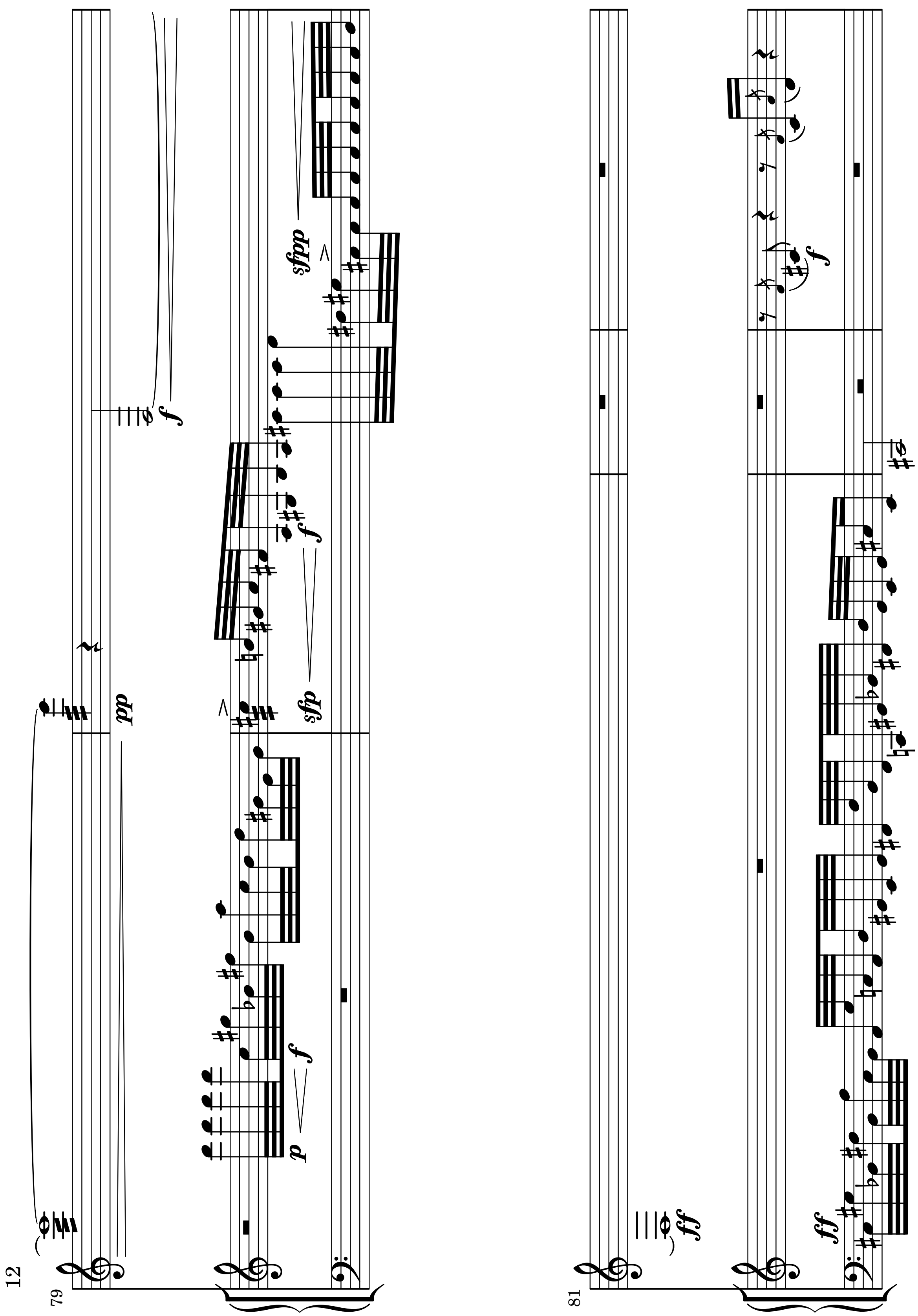


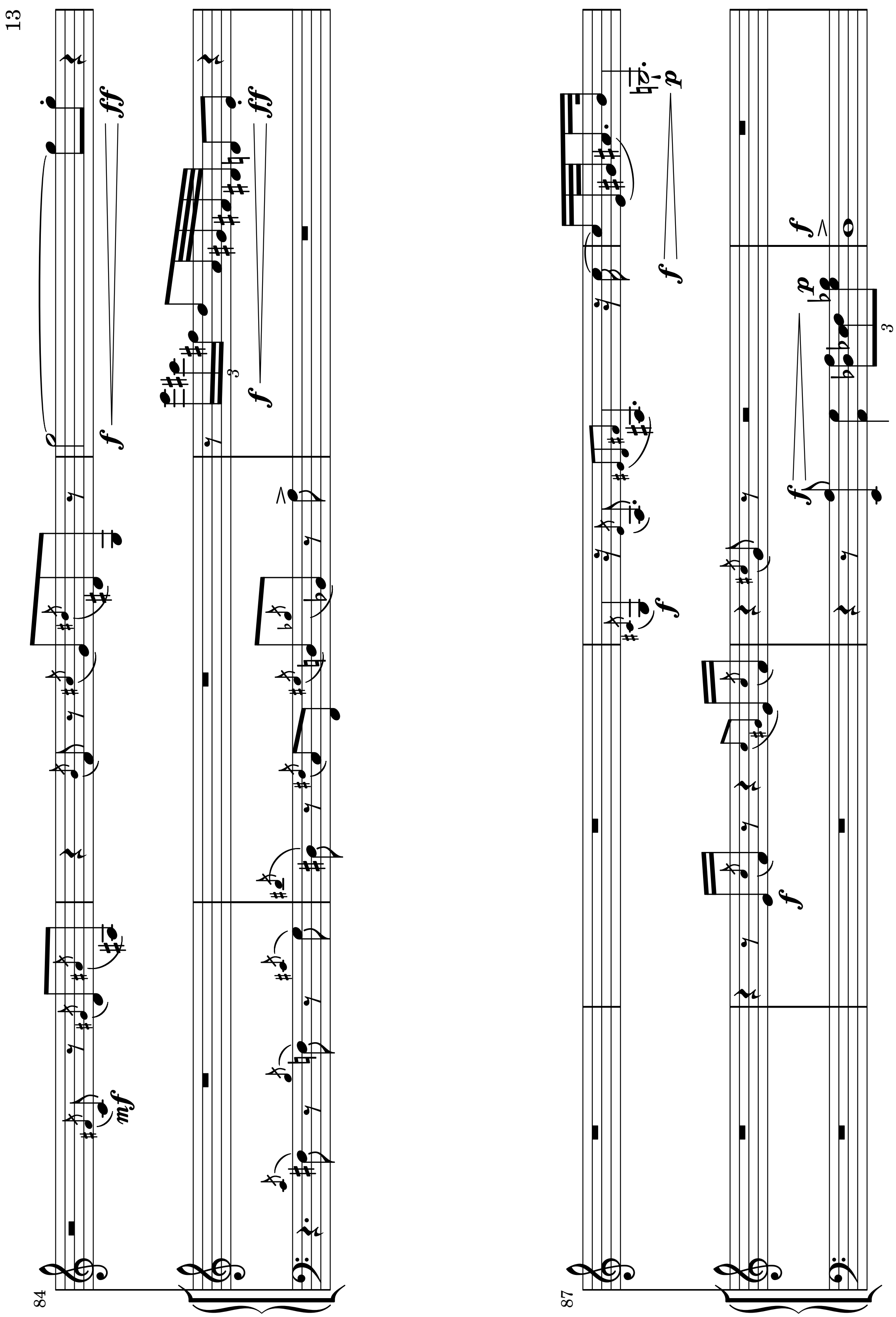


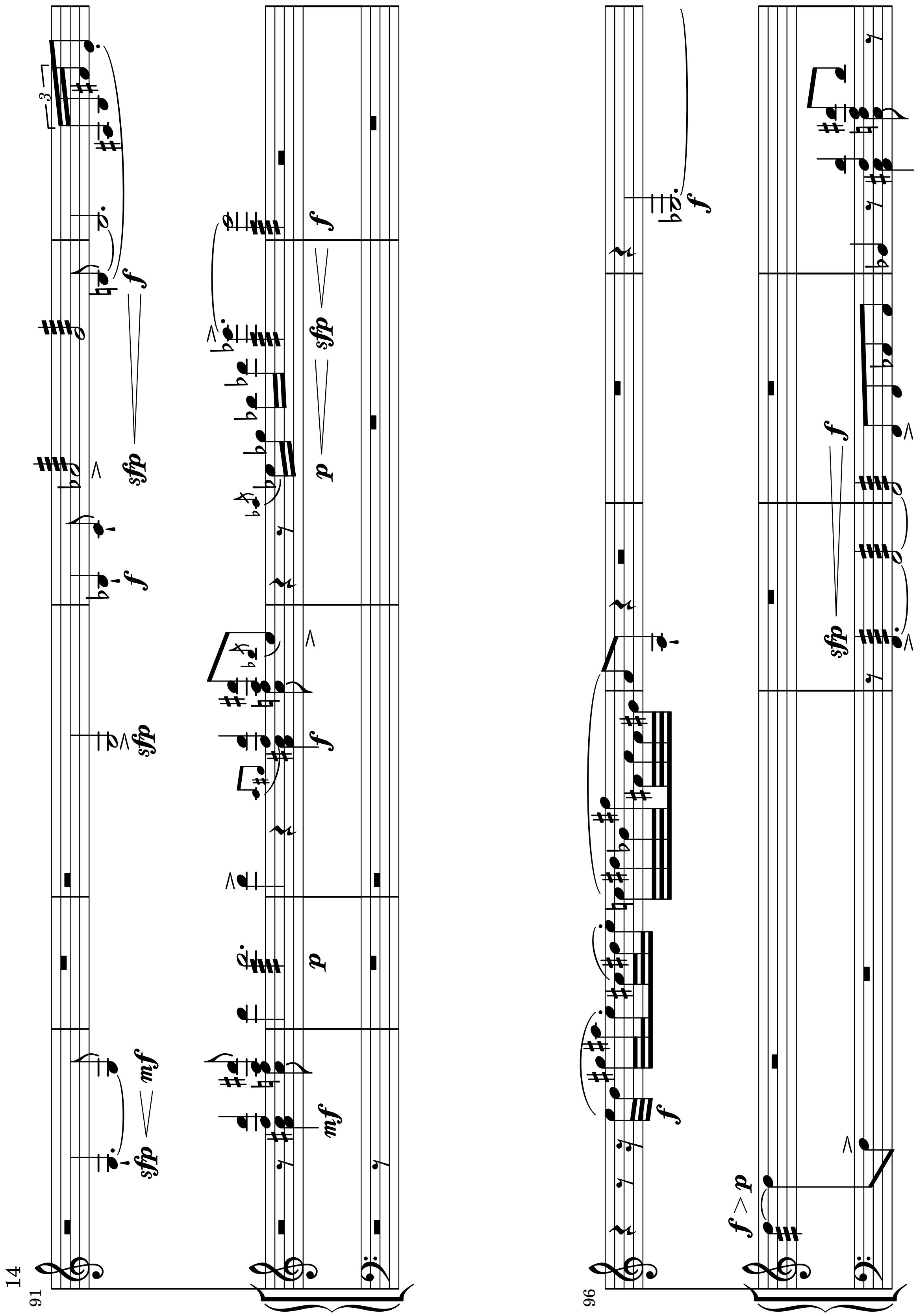

8
50
52
ff
p
f

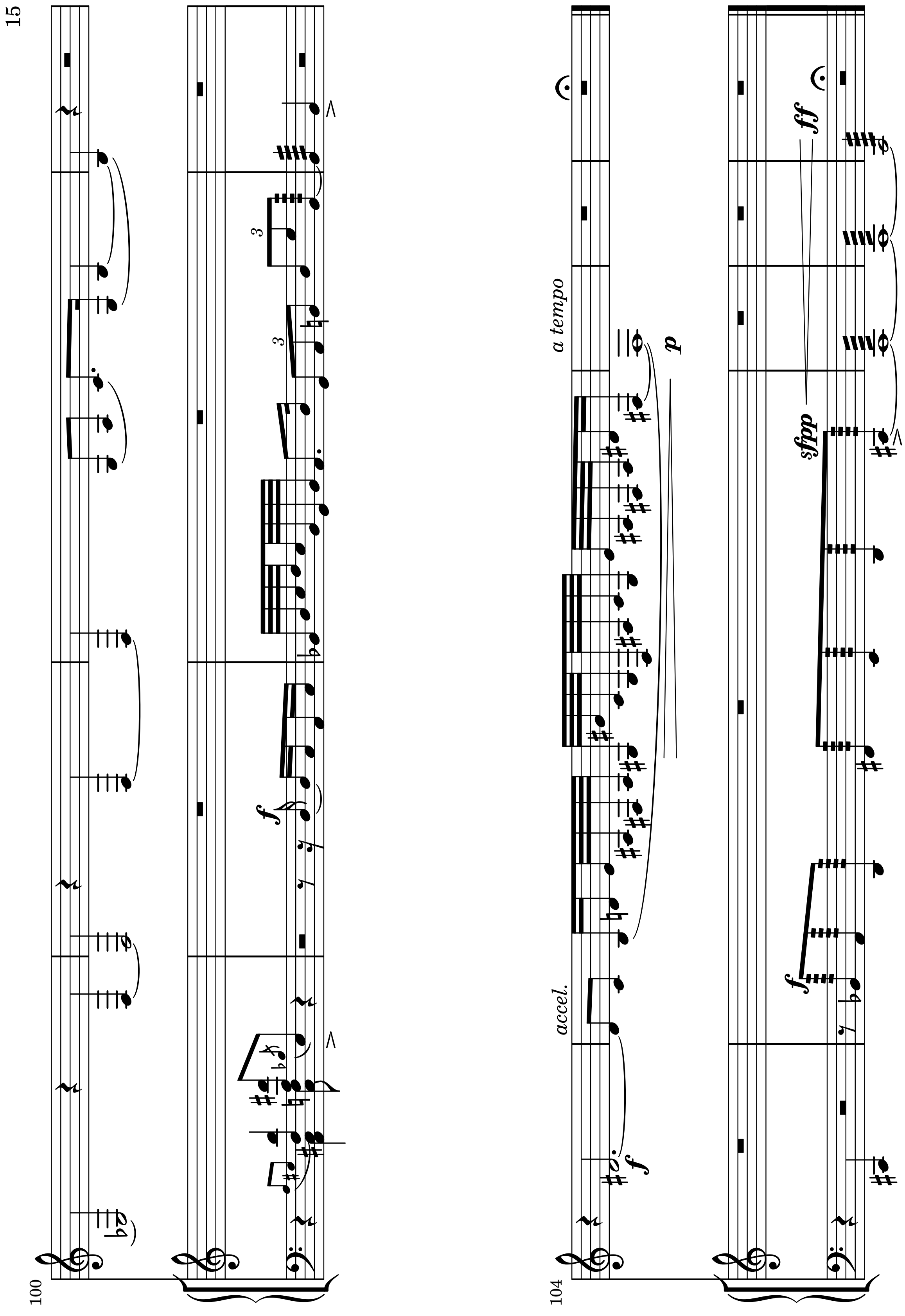





12
79
81


13
84
87
mf
f
ff
p


14
91
96


15
100
104
accel.
a tempo

A.2 Come On, America!

The musical material in this piece is derived from the 1918 World War I song "Come On, America!" by Kenneth M. Murchison and Edmund Vance Cooke as heard in combination tones. The melody appears in each section but only appears in the correct order in the final section of the piece. The melody is first presented in a random order and then, through the bitonic sorting algorithm, is sorted as the piece is played. Additional musical material is derived from the phonetics of the lyrics of the same piece.

This piece's formed is derived from the standard pop formula of intro, verse, prechorus, chorus, verse, prechorus, chorus, verse, bridge, and verse but the duration of each section has been distorted/modulated by a sine wave. The parameters of both speed and pitch within each formal division are dependent on the length of the distorted division's length.

Come On, America

for the Oregon Bach Festival's Composer Symposium

Stewart Engart (2017)

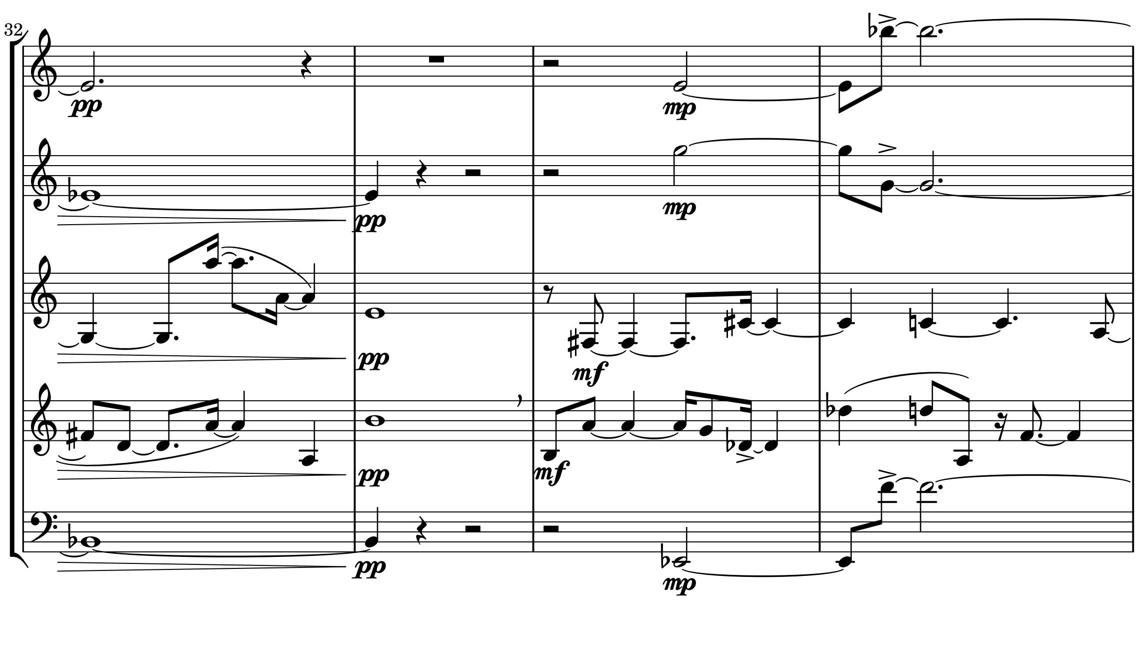

11
mf
mf
mf
mf


16
mf


19


24
Solo


28
mf
mf
mf
mf


32
pp
mp
pp
mp
pp
mf
pp
mf
pp
mp


43
Solo


49


7
52
mp
mp
mp
mp
mp
55


58
mp
mp
mp
mp
mp


61
mf
mf
mf
mf


64
67


70
mp
mp
mp
mp
mp


73


76
with Clarinet
mf
mf
with Flute
mf
mf


79
mf


82
85
mp


88
mp
mp


91


94
f


97
f
f
f
f


100


104
Upper Formant
ff
Lower Formant
ff
Fundamental
ff


108
with Bassoon
mf
with Clarinet
mf


111
mf
mf


115
p
p
p
p
120


124
f
f
f
f
f
127
mf
mf
mf
mf
mf


132
mf
mf
mf
mf
mf
137
f
f
f
f
f

A.3 *Je me suis trouvé*

Je me suis trouvé was inspired by the need to escape, at least mentally, from our COVID confined living spaces.

Reminiscing on a transformational trip to Paris, April's lyrics, which fluctuate throughout the piece between French and English as the singer is transported through nostalgia, reflect both her feelings during the experience and now in hindsight.

As a result of our current COVID situation, all video and audio were recorded on a cell phone. All audio is derived from April's voice, with extensive processing in CREATE's Emission Control 2 and use of concatenative synthesis to organize grains.

I went to find you in Paris
August 2017
I found myself instead on the TGV
Le Train de Grand Vitesse
Depuis dix-huit ans
By le Tour Eiffel,
Le carrousel.
By les jardins du Louvre
I looked for you
Mais j'ai trouvé ma force
Retourner.

This piece was composed for and with April Amante.

Souvenir chez moi, 2020

for April Amante

Stewart Engart


S.
de Grand Vi - tes-se De - puis dix - huit
ans
Spoken
By By le Tour Ei - ffel, Le
car - rou - sel.
By By les jar-dins du Lou-
- - vre
Spoken
I
Closed
Stuttering
m - I I looked for you
Spoken
I went to find you in Par - is
Mais j'ai trou vé ma force
Re - tour - ner
Spoken
I went to find
you in Par-is
but, what I found was so much better

A.4 The Blood Red Van

Like all my projects during 2020, this piece started full of hope and optimism, but through the process of writing the dread and anxiety caused by current events stomped out all other emotions.

The title of the piece comes from the main antagonist of Homeward Bound II: Lost in San Francisco, a van that traveled the city rounding up dogs for animal testing.

This piece was composed for HOCKET.

THE BLOOD RED VAN

for HOCKET #what2020SoundsLike

Stewart Engart

August 1, 2020 | Camarillo, CA

Bibliography

[1] Plunderphonics - interviews.

[2] Allolib C/C++ Libraries {#mainpage}, Sept. 2021. original-date: 2016-11-08T09:02:25Z.

[3] Andersen, K., and Knees, P. Conversations with Expert Users in Music Retrieval and Research Challenges for Creative MIR. *New York City* (2016), 7.

[4] Anderson, J. A provisional history of spectral music. *Contemporary Music Review 19*, 2 (Jan. 2000), 7–22.

[5] Berenzweig, A., Logan, B., Ellis, D. P., and Whitman, B. A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures. *Computer Music Journal 28*, 2 (June 2004), 63–76.

[6] Blackburn, M. The Visual Sound-Shapes of Spectromorphology: an illustrative guide to composition. *Organised Sound 16*, 1 (Apr. 2011), 5–13.

[7] Blum, T. L., Keislar, D. F., Wheaton, J. A., and Wold, E. H. Method and article of manufacture for content-based analysis, storage, retrieval, and segmentation of audio information, June 1999.

[8] Bogdanov, D., Wack, N., Gomez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., and Serra, X. Essentia: An Audio Analysis Library for Music Information Retrieval. 6.

[9] Bonet, N. Musical Borrowing in Sonification. *Organised Sound 24*, 02 (Aug. 2019), 184–194.

[10] Burkholder, J. P. The Uses of Existing Music: Musical Borrowing as a Field. *Notes 50*, 3 (Mar. 1994), 851.

[11] Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., and Slaney, M. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE 96* (May 2008), 668–696.

[12] de Almeida, G. B. Composing Music by Selection Content-Based Algorithmic-Assisted Audio Composition. 231.

[13] Diakopoulos, D., Vallis, O., Hochenbaum, J., Murphy, J., and Kapur, A. 21st Century Electronica: MIR Techniques for Classification and Performance. *Poster Session* (2009), 5.

[14] Evanstein, M. Composing with Interfaces: Analog and Digital Feedback Loops in the Compositional Process. 31.

[15] Fineberg, J. Spectral music. *Contemporary Music Review 19*, 2 (Jan. 2000), 1–5.

[16] Foote, J. An overview of audio information retrieval. *Multimedia Systems 7*, 1 (Jan. 1999), 2–10.

[17] Grisey, G., and Fineberg, J. Did you say spectral? *Contemporary Music Review 19*, 3 (Jan. 2000), 1–3.

[18] Hackbarth, B. AudioGuide: A Framework for Creative Exploration of Concatenative Sound Synthesis. 4.

[19] Hackbarth, B., Schnell, N., Esling, P., and Schwarz, D. Composing Morphology: Concatenative Synthesis as an Intuitive Medium for Prescribing Sound in Time. *Contemporary Music Review 32*, 1 (Feb. 2013), 49–59.

[20] Harvey, J. Spectralism. *Contemporary Music Review 19*, 3 (Jan. 2000), 11–14.

[21] Holm-Hudson, K. Quotation and Context: Sampling and John Oswald's Plunderphonics. *Leonardo Music Journal 7* (1997), 17.

[22] Kim, H.-G., Moreau, N., and Sikora, T. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval.* John Wiley & Sons, Ltd, Chichester, UK, Oct. 2005.

[23] Manning, C. D., Raghavan, P., and Schütze, H. Introduction to Information Retrieval. 506.

[24] Michael Casey. Soundspotting: A New Kind of Process?

[25] Nisbet, R., Miner, G., Yale, K., Elder, J. F., and Peterson, A. F. *Handbook of statistical analysis and data mining applications*, second edition ed. Academic Press, London, 2018. OCLC: on1012917853.

[26] Roads, C. *Microsound.* MIT Press, Cambridge, Mass, 2001.

[27] Schwarz, D. A System for Data-Driven Concatenative Sound Synthesis. 7.

[28] Schwarz, D. Concatenative sound synthesis: The early years. *Journal of New Music Research 35*, 1 (Mar. 2006), 3–22.

[29] Smalley, D. Spectromorphology: explaining sound-shapes. *Organised Sound 2*, 2 (Aug. 1997), 107–126.

[30] Stark, A. M. Sound Analyser: A Plug-In For Real-Time Audio Analysis In Live Performances And Installations. 4.

[31] Stoll, T. M. Beyond Concatentation: Some Ideas for the Creative Use of Corpus-Based Sonic Material. 4.

[32] Sturm, B. L. Adaptive Concatenative Sound Synthesis and Its Application to Micromontage Composition. *Computer Music Journal 30*, 4 (Dec. 2006), 46–66.

[33] Tutschku, H. Hans Tutschku - Interactivity, Apr. 2021.

[34] Tversky, A. Features of similarity. *Psychological Review 84*, 4 (1977), 327–352.

[35] Typke, R., Wiering, F., and Veltkamp, R. C. A Survey of Music Information Retrieval Systems. *Proceedings of the International Conference on Music Information Retrieval* (2005), 153–160.

[36] von Kaenel, N. nvk_create.

[37] Wishart, T. CEC — eContact! 15.2 — Encounters in the Republic of Heaven by Trevor Wishart.

[38] Xambo, A., Lerch, A., and Freeman, J. Learning to Code Through MIR. 3.

[39] Xambó, A., Lerch, A., and Freeman, J. Music Information Retrieval in Live Coding: A Theoretical Framework. *Computer Music Journal 42*, 4 (May 2019), 9–25.