

PRAXIS META/HODOS:
THREE LOVE SONGS FOR JAMES TENNEY

Thesis

Submitted in partial fulfillment

of the requirements for the

Degree of

Master of Fine Arts

in Electronic Music and Recorded Media

Mills College, 2009

by

Luke Selden

Approved by:

Chris Brown
Director of Thesis

Fred Frith
Head, Music Department

Sandra C. Greer
Provost and Dean of the Faculty

Reading Committee:

John Bischoff

James Fei

Acknowledgements

I would like to thank the following people for their help in making this thesis possible. Each person contributed some aspect to my understanding of and/or interest in the material contained herein, however tangential, fleeting or life-sustaining. I am profoundly grateful for their influence and assistance.

Thanks to (in randomized order):

Vickie Greene: statistics
Chris Brown: music
Helene Rosenbluth: sound
Kiyomitsu Odai: scale
Samuel Selden: ideas
Ben O'Brien: hierarchy
Anna Seagrave: frame
Harris Ipock: analysis
John Bischoff: encouragement
Ellery Royston: ants and cows
David Bernstein: James Tenney
Casey Selden: fuzzy
Jimmy Johnson: novelty
Students of Mills College Music Department, 2006-2008: improvisation
Christa Martin: spatio-temporal relationships
James Fei: conceptualization
Erin Cody: purpose
James Tenney: ergodicity
Xavier Serra: detrended fluctuation analysis
Mark Robertson: chaos
William Shively: multidimensionality
Allen Anderson: contour
Heather Jovanelli: metaphor
Perry Cook: nonlinearity
Suzanne Thorpe: variation and the now

Contents

Acknowledgements	iv
Introduction	6
Data Preparation	10
Love Song #1: Variation	13
Entropy	13
Ergodicity	15
Experiment	17
Love Song #2: Scaling	21
Scale	22
Trends	25
Experiment	27
Love Song #3: Space is the Place	30
Experiment	31
Conclusion	40
Appendix	42
Bibliography	49

Introduction

Traditional music theory tries to find universal laws of music. But when applied to various styles of music the limitations of these theories become apparent. Each theoretic approach is more or less adapted to the musical style it describes best. Traditional music theory can be used as a guide, but not as a **normative** reference that dictates the kind of processes and structures that operate in the mind of the listener.

--Computational Models of Music Perception and Cognition I (Purwins et al. 4)³

In 1983, referring to the hierarchical segmentation algorithm proposed in Hierarchical

Temporal Gestalt Perception in Music: A Metric Space Model (Tenney & Larry.

Polansky) Larry Polansky writes: "...the algorithm and model even now seem to be of revolutionary importance in the understanding of musical form and perception, and one hopes that others will see fit to continue Tenney's work." (Larry Polansky 267) In his writings on musical form—Meta/Hodos, META Meta/Hodos, Form, and Temporal Gestalt Perception in Music—Tenney presents a persuasive argument for the "theory of hierarchical perceptual gestalt formation."

This thesis is an attempt to continue Tenney's work. In the time since the last paper in the series (Temporal Gestalt Perception in Music) there have been many advances in the automatic processing of musical signals from cognitive models of how the human brain processes audio stimuli (Purwins et al. 1). These models tend to support Tenney's hypothesis that Gestalt Theory provides a useful rubric for the grouping of musical objects (Purwins et al. 14).

A *Temporal Gestalt Unit* (TG) is Tenney's name for a cluster of sound objects in time.

Temporal Gestalts are formed from clustering the base musical material, and then hierarchically clustering the resulting TGs together into higher-level TGs. These Temporal Gestalt Units together describe the hierarchical structure of a piece.

This paper is an attempt to revisit three theoretical principles that Tenney proposes in META Meta/Hodos for this hierarchical grouping of music into Temporal Gestalts: entropy, scaling, and grouping. For each principle I will explain the fundamental concept, provide an algorithm that might help practically apply this concept to musical analysis, and provide an experimental analysis of a piece of music using that algorithm. This is by no means an exhaustive survey of current research, nor a full musical analysis using these measures. Rather, these are the results of initial steps in building a framework for continuing the algorithmic application of Tenney's theories of musical form and perception. To emphasize the tentativeness of the conclusions of these results I title these experiments "Love Songs to James Tenney".

The musical example used for analysis is the first track off of an Audio CD "Skywriters" by Chris Brown and Pauline Oliveros—recording of free improvisation from September 2008—on the request for an arbitrary example of free improvisation. I chose this work for two reasons. First, I may assume that my thesis advisor (Chris Brown) would be an authoritative source in determining whether my characterization of this piece of music is appropriate. The second reason is for its musical properties. As an example of free

improvisation it presents an example of music in which the structure of the piece of music is complex and non-linear. This makes the application of algorithms—adapted from the domain of Nonlinear Dynamic Systems Analysis—appropriate for the task of segmentation. Furthermore, the content of the piece displays a wide range of musical textures and levels of intensity, thereby providing a significant measure of each algorithm’s fitness in using the entire feature space.

The first “Love Song” is on the concept of variation. Tenney proposes the use of Entropy (adapted from the field of Information Theory) and Ergodicity to judge the amount of variation and stability of music at any given time. To reexamine this proposal I apply the Multiscale Entropy measure and discuss the experimental results.

The second “Love Song” concerns the concept of scale. Tenney’s approach to grouping depends on the hierarchical clustering of musical material. The concept of scale applies to the question of how many levels of this hierarchy to use and the ratio of time span over which each level is grouped; this remains an unanswered question in Tenney’s work.

The concept of scale is also important in quantifying the distribution of parametric range over each level of the hierarchy; since there is a finite range of intensities possible for each feature vector, a large range at one level of the hierarchy limits the possible range for the other levels of the hierarchy for that feature vector. For this question I will apply the Detrended Fluctuation Analysis algorithm and discuss its significance to a music analysis of the music example in question.

The third “Love Song” is an attempt to approach the concept of space. In Temporal Gestalt Perception in Music Tenney proposes an algorithm that utilizes a multidimensional metric space (this space will be referred to as “feature space” for the purposes of this thesis) to cluster musical data and determine the clustering of musical material based on their distance in this musical space. An issue to which Tenney admits with his algorithm is that it requires manual weighting of each dimension to their determination of cluster formation. Another issue with his approach that he leaves as an open question is the overlapping of clusters. In his algorithm the grouping of material is strictly delineated by time; transitions occur as strict divisions and do not account for gradual changes between clusters. To address these issues I apply an alternative algorithm for clustering in space—Probabilistic Principal Component Analysis for Time-Series Segmentation (PPCA-TSS)—which attempts to overcome both of these limitations in Tenney’s approach. I then explain the significance of the results of this experiment to the analysis of the musical piece in question.

Data Preparation

DEFINITION 9: A *parameter* will be defined here as any distinctive attribute of sound in terms of which one sound may be perceived as different from another, or a sound may be perceived to change in time.

COMMENT 9.1: This definition refers to "subjective" or *musical* parameters (e.g., pitch, loudness, etc.) as distinct from "objective" or *acoustical* parameters (frequency, amplitude, etc.).

COMMENT 9.2: There is not, in general, a one-to-one correspondence between musical and acoustical parameters. Where there is such a correspondence, the relation is more nearly logarithmic than linear.

(Tenney 103)

For the three analytical methods described below I relied on the same feature set derived from the Audio CD recording of the piece in question. I imported the CD data into jAudio (McEnnis, D. et al. 3), downsampled to a one-channel 22.5k PCM WAVE file^{*}, and extracted the following features with a window size of 4092 and a window overlap of 0.75:

(descriptions quoted from jAudio documentation (McEnnis, D. et al. 3))

Beat Sum	The sum of all bins in the beat histogram. This is a good measure of the importance of regular beats in a signal.
----------	---

^{*} downsampling was necessary for memory limitations; processing the original 203MB file was prohibitively taxing on my computer

Compactness	A measure of the noisiness of a recording. Found by comparing the components of a window's magnitude spectrum with the magnitude spectrum of its neighboring windows.
Fraction Of Low Energy Frames (Amplitude "Peakiness")	The fraction of the last 100 windows that has an RMS less than the mean RMS of the last 100 windows. This can indicate how much of a signal section is quiet relative to the rest of the signal section.
Root Mean Square (Amplitude)	A measure of the power of a signal over a window.
Spectral Centroid	The centre of mass of the power spectrum.
Spectral Flux	A measure of the amount of spectral change in a signal. Found by calculating the change in the magnitude spectrum from frame to frame.
Spectral Rolloff Point	The fraction of bins in the power spectrum at which 85% of the power is at lower frequencies. This is a measure the right-skewedness of the power spectrum. (note: comparable to the concept of "brightness")
Spectral Variability	The standard deviation of the magnitude spectrum. A measure of how varied the magnitude spectrum of a signal is.
Strength Of Strongest Beat	How strong the strongest beat in the beat histogram is compared to other potential beats.

Strongest Beat The strongest beat in a signal, in beats per minute, found by finding the highest bin in the beat histogram.

Zero Crossings The number of times the waveform changed sign in a window. An indication of frequency as well as noisiness.

These feature vectors were then imported into Matlab for data analysis.

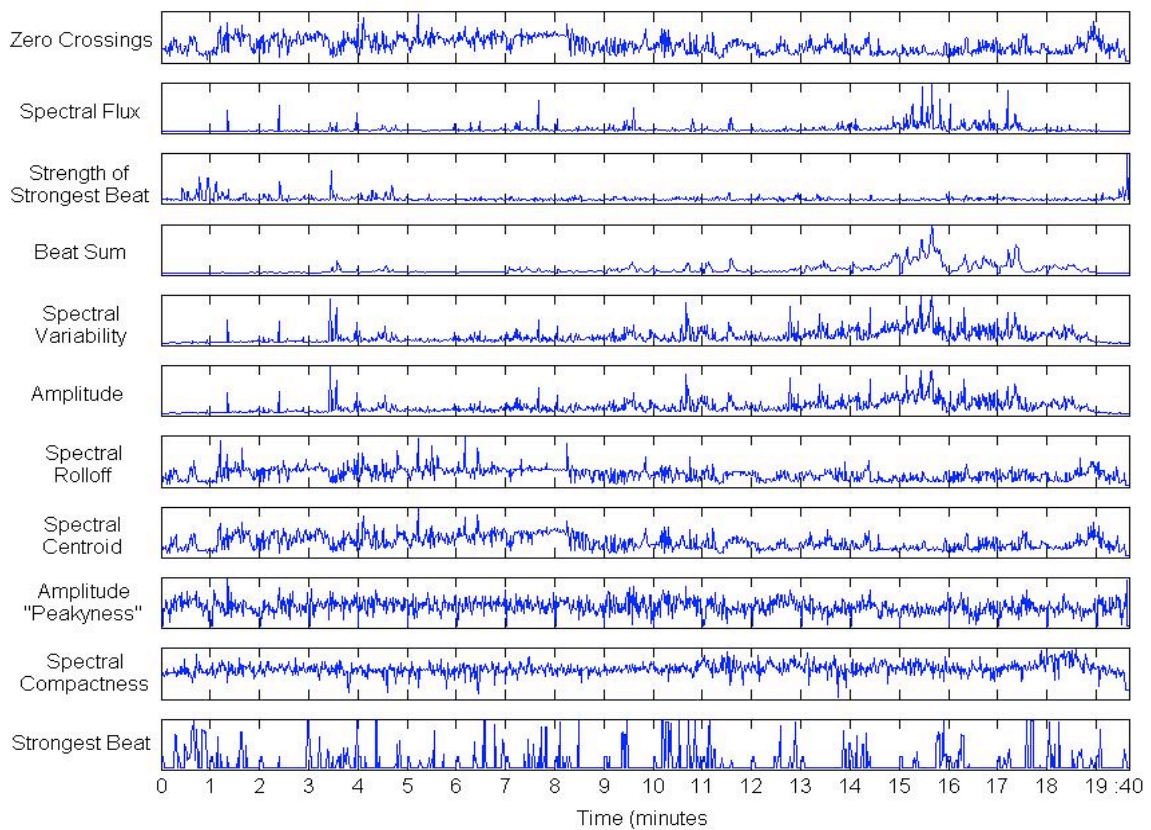


Figure 1: plot Feature vectors over time

Love Song #1: Variation

Entropy

DEFINITION 28: One of the most important aspects of musical experience is the perception of variation, and a useful measure of variation is *entropy*. In information theory, the entropy of a "message" consisting of a series of n discrete "symbols" drawn from an "alphabet" of N equally probable symbols is

$$H = n \log_2 N \text{ (bits per message).}$$

The entropy of each symbol is

$$H = \log_2 N \text{ (bits per symbol).}$$

COMMENT 28.1: The most important variable here is N , the number of symbols available. In the special case where $N = 1$, $H = 0$.

COMMENT 28.2: When the available symbols are not equally probable—i.e., when they do not occur with the same relative frequencies (p_i)—then

$$H = -\sum p_i \log_2 p_i \text{ (bits per message).}$$

(Tenney 114)

Entropy, in the context of Information Theory, is a measure of the uncertainty of a given variable (Moles 22-27). The basic unit of information (again, in the context of Information Theory) is the *bit*, which is either true or false. A coin flip, for example, conveys one bit of information: heads or tails. A dice roll conveys 2.5 bits of information:

Result	1	2	3	4	5	6		
Binary Value	000	001	010	011	100	101	110	111

Table 1: Six-sided dice roll and binary bits used

If each state of a system is equally likely (for example, a coin flip or a “fair” dice roll) then the amount of information that each result gives about the state of the system is equal. The chance that I roll a 4 with a die is the same as the chance of rolling a 6. If the states of a system occur with *differing* frequencies then the less likely states convey *more* information than more likely states. This corresponds with our intuition. For example, in a properly functioning car the check engine light only comes on when there is a problem with the car that needs the driver’s attention—perhaps the oil is low, there is a mechanical error, or the gas cap is loose. If the light is not on then the driver can assume there is no problem with the car. Most of the time this light stays off, so the driver can ignore the light and pay attention to their driving. If the light turns on—which ideally is an uncommon state—it then suddenly conveys original and significantly different information about the state of the engine; something must be wrong.

Compare this to a malfunctioning car, in which the engine light turns on and off for a myriad of reasons; when the car hits bumps, if it’s hot or cold, if there’s actually a problem with the car, if there’s no problem with the car, etc. In this case whether the engine light is on or off conveys much less information to the driver, because there’s no way to tell if the engine light is on because of a real problem or because of some confounding factor.

Entropy is a method of quantifying this uncertainty. The mathematical description of this measure is beyond the scope of this thesis, but can be found applied to music in the works of Tenney and Moles (Tenney 114; Moles 22-27).

Ergodicity

DEFINITION 23: A TG whose component, next-lower-level TG's all have the *same state* in a given parameter will be called *ergodic* with respect to that parameter.

COMMENT 23.1: The shape of an ergodic TG is thus "flat" in that parameter.

COMMENT 23.2: An ergodic TG has the same parametric state as each of its component, next-lower-level TG's.

DEFINITION 24: A TG whose component, next-lower-level TG's have *different states* in a given parameter will be called *non-ergodic* with respect to that parameter.

COMMENT 24.1: The *shape* of a TG may thus be either ergodic or non-ergodic, with respect to a given parameter.

(Tenney 115)

In META Meta/Hodos James Tenney describes a concept he borrowed from dynamic systems analysis: Ergodicity (115). Ergodicity is a measure of whether a system is statistically static at a given time span. For instance, consider the flame of an oil lamp. Even when the air is completely still around it, a candle's flame will flicker, moving and changing shape. Over time, however, the space which the candle occupies never changes outside a certain range – it won't flip horizontally, but rather tend upwards. Since the

probability of where in space the candle is at any given moment is not time dependant (at least until the fuel runs out) this system can be considered ergodic. If, however, there is a breeze that runs through the room or someone blows on the candle that would change the state space in which the candle inhabits—it may flicker to the side, or momentarily shrink in size significantly from its normal perturbations. In this case the probability of the state space changes within the time span in question (until the oil runs out), and the system is not ergodic.

If one were to consider the time span before the breeze or breath and after it, each of these time segments would be ergodic. That is to say, ergodicity is dependant on the time span in question. One can therefore segment a series of data by clustering moments in which the probability distribution functions are statistically equivalent to each other.

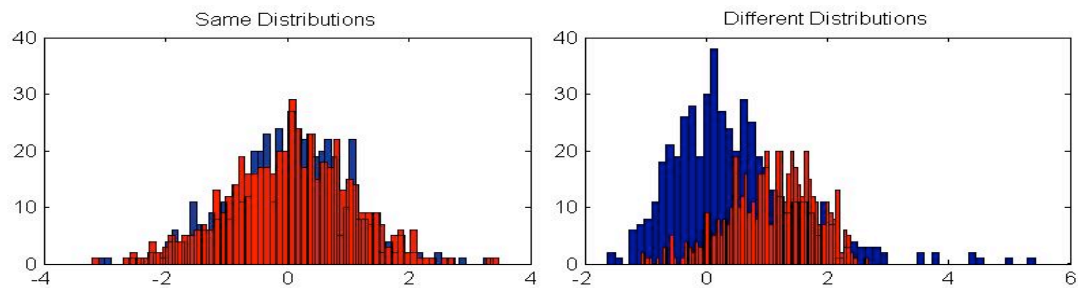


Figure 2: Ergodic and Non-ergodic processes

In the figure above the first histogram shows a time span of a process in which the two time samples (blue and red) are statistically equivalent – their mean, variances, and skewness are functionally equivalent. Therefore, this example is ergodic. The second shows a time span of a process in which the two time samples (blue and red) are statistically dissimilar, so it does not satisfy the criterion for ergodicity.

Experiment

To quantify the amount of originality at each moment of the musical piece in question I utilized the Multiscale Entropy measure. The simple entropy measure of a signal compares the uncertainty of a sample using its immediate neighbors, which disregards possible larger-scale relationships. Multiscale Entropy addresses this problem by chunking the signal at multiple time scales, determining the entropy at each of these time scales and summing the result (Costa M, Goldberger AL, & Peng CK). I applied this measure to 9 of the feature vectors, excluding “Spectral Flux” and “Strength of Strongest Beat” because they resulted in zero entropy in my analysis across the entire time series.

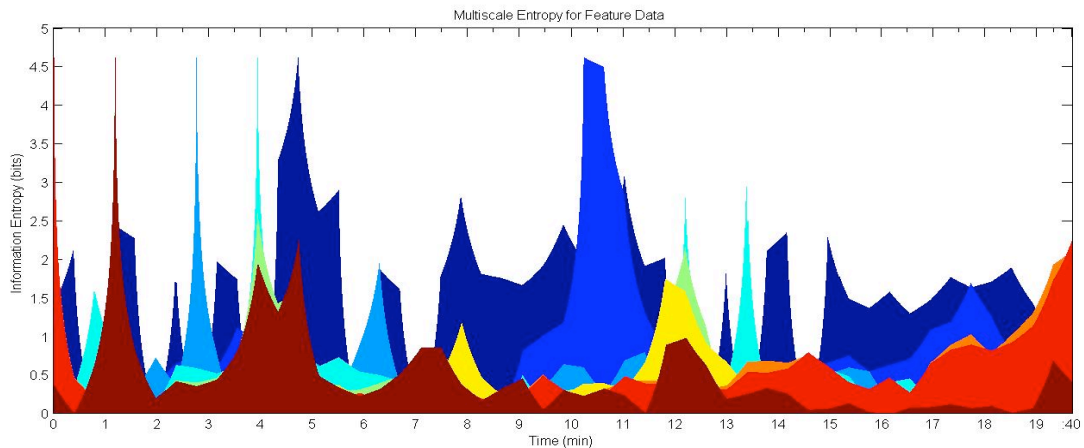


Figure 3: Multiscale Entropy of feature set.
Color corresponds to feature, but feature names were lost in analysis.

As a raw measure this data proved hard to interpret musically. Instead of attempting musical analysis with this raw data I used the Matlab toolbox SOMtoolbox (Vesanto) to create a Self Organizing Map (SOM) that organized the data for clustering. A Self Organizing Map is a method of dimensionality reduction which uses artificial neural networks to iteratively map a high-dimensional space into a regular (in this case two-dimensional) manifold.

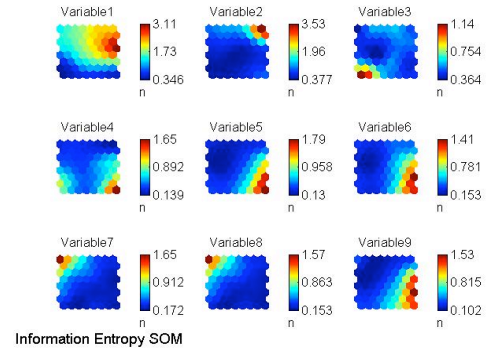


Figure 5: Self Organizing Map of Information Entropy. Feature names were lost in analysis

From this SOM I hierarchically clustered the time series data into 64 separate clusters:

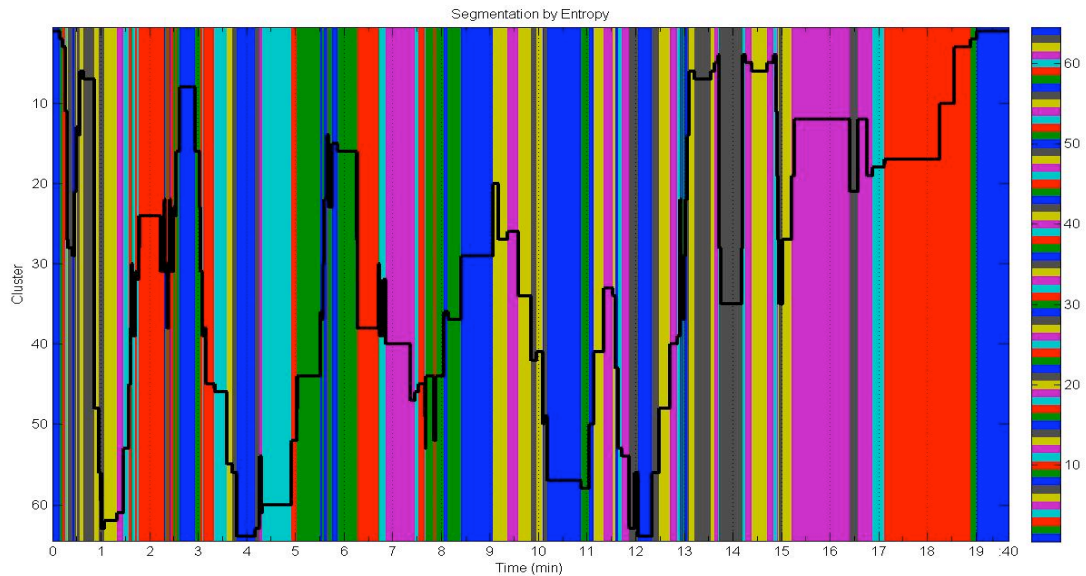


Figure 6: Entropy clustering of time series.

The results of this clustering do not offer much insight into the large-scale structure of the piece, but they do show segments in which the information entropy of the signal is ergodic. Musically this is useful in finding separation points between textures. For

instance, ~10:00~11:00 contains a single segment. This span of music consists of similar textural material—occasional piano notes and background “wind” noises. Right around ~11:00 there is a quick succession of clusters after a long single cluster. This corresponds with a significant change in texture; the accordion comes in with dissonant chords that interrupt the previously static piano texture.

Similarly, right at 15:09 the piano switches into a low rhythmic pulse, and the accordion and piano pick up these pulses, and that pulse continues to inform the onsets of events until around 16:30. This corresponds to a long segment in the entropy clustering data.

The time span ~1:00~3:00, in contrast, shifts clusters relatively often. This is due to the musical texture of fairly constant texture interjected by sudden stabs of various piano sounds. This causes the entropy measure to suddenly shift, because these stabs are unexpected. Like the (properly functioning) car engine light these moments show sudden originality, and so the entropy considers these moments significant in comparison to their surrounding segments. Compare this to the section mentioned earlier (~15:09~16:30) in which similar sudden piano gestures occur, but in this section these events are expected when considering the events around them.

In conclusion, I found that the results from this entropy analysis does give some lead to some musically interesting segmentation, but do not provide a clear method for determining the structural significance of these segmentation points. This method could be improved by weighing the strength of the various feature vectors, but during the

course of analysis I lost the names associated with the information entropy of each feature. Also, another method of dimensionality reduction and clustering other than SOMs, which do not preserve the temporal relationships between samples, may provide better results.

Love Song #2: Scaling



Figure 7: Complete order, chaos, and complete disorder [from (Streich 15)].



Figure 8: Possible inclusions [from (Streich 16)].

Scale

PROPOSITION X: For any parameter with respect to time, the greater the range of variation at a given hierarchical level, the smaller the range of variation possible at the next higher level, and vice versa.

COMMENT X.1: For a given parameter, and under the special condition that the ranges are identical for all TG's at a given hierarchical level, the following relations will hold:

For the first hierarchical level, considered by itself, the maximum range available is $N(1)_{\max} = N_t$, where N_t is the total number of distinguishable values in that parameter. When two hierarchical levels are considered, the maximum range at the second level is

$$N(2)_{\max} = N_t - (N(1)-1).$$

For a third level, the maximum range will be

$$N(3)_{\max} = N_t - (N(1)-1) - (N(2)-1).$$

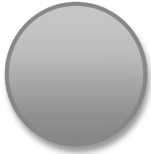

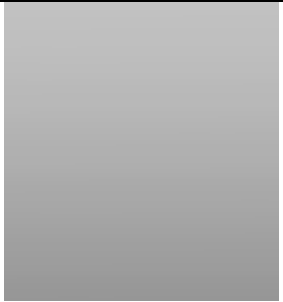
More generally, the maximum range available at a given level (L) is

$$N(L)_{\max} = N_t - (N(1)-1) - (N(2)-1) - \dots - (N(L-1)-1), \text{ or } N(L)_{\max} = N_t - N_L + L-1.$$

Finally, the total available range (N_t) may be distributed equally among some number of levels (L), so that

$$N(1) = N(2) = \dots N(L), \text{ and } N(L+1)_{\max} = 0, \\ \text{By setting each } N \text{ at } N = N_t/L+1.$$

(Tenney 114-115)

To illustrate the concept of scale, consider a sphere:	
If one was to shrink the sphere, or zoom out the frame of reference far enough, the sphere would be perceived as a point:	
Conversely, if one were to zoom in very close to the sphere it would appear as a plane:	

This demonstrates that for scale-variant objects the frame of reference strongly determines how its nature is perceived. A rising tone over one second becomes organized perceptually as one event, but a 40 minute rising tone (*Wavelength*, for instance (Snow)) is perceived very differently. Furthermore, the scale at which one abstracts and groups material strongly affects the inferences one may make about the material.



For a more complex model, imagine an ant crawling on a cow's back.

The ant is able to crawl only a very small amount (in relation to human-scale) per second. For the sake of this thought experiment assume the ant's motion can be simplified as Brownian random motion (a drunk walk). In comparing the ant's position from time X to time $X+t$, the maximum it may move from position X is

$F(X+t) = X + t * \text{crawlRate}$, where t is time and crawlRate is the amount of movement per second. Of course, since this is random motion the position is more likely to fit within a Gaussian distribution around position X .

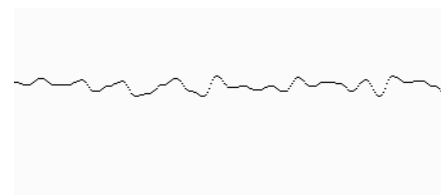


Figure 9: Ant Position over time

[Image from (Elias)]

Nevertheless, this defines the scale of the ant's position probability distribution; let's say roughly the size of the cow's back as a boundary.



But what if the cow were moving also? For simplicity's sake assume the cow is drunk, so wanders about in similar Brownian fashion, but at a larger scale—in this thought experiment, the range of a farmer's flatbed truck.

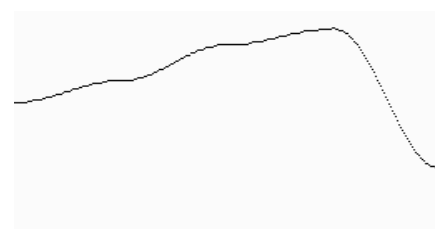


Figure 10: Cow's position over time

[Image from (Elias)]

Now the position of the ant—from the perspective of an observer—becomes more variable, as one must take into account the motion of the cow as well. The position then becomes $\text{antPosition} + \text{cowPosition}$. If the farmer were also drunk and decided to go for a drive with this cow in the back of the truck the motion occurs on an even greater scale, and the position of the Ant then becomes $\text{antPosition} + \text{cowPosition} + \text{truckPosition}$.

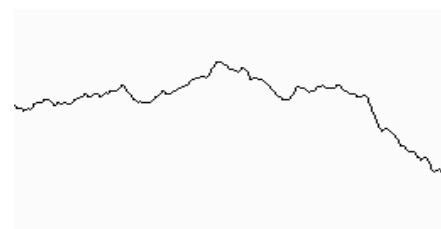


Figure 11: $\text{antPosition} + \text{cowPosition} + \text{truckPosition}$ [Image from (Elias)]

This is the principle behind Perlin Noise (Elias), a multi-scale noise generation algorithm.

Analysis of the power-law scaling behavior of a

signal allows one to decompose such a signal

to find the set of time and space scales that

best describe the motion, and these individual components.

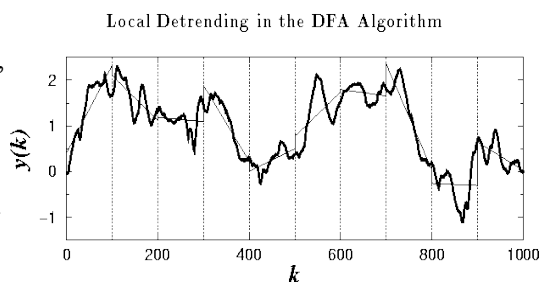


Figure 12: length of linear line is current time scale. For each line the fluctuation around line is calculated [From (Physionet)]

Trends

In analyzing a time series one often needs to account for hidden trends in the data which confound analysis. In economic data, for instance, economists often account for seasonal changes in spending habits. Imagine the CEO of Mattel TM trying to determine if an advertising campaign for Bratz TM dolls has a positive effect on sales. A 20% increase of sales from May through July might be seen as more significant than a 20% increase from October to December, owing to the seasonal increase in consumer goods during the holiday season. If, however, the advertiser detrends the data by accounting for a cyclic trend in spending habits over the year, and still finds a significant increase during the holiday season, then it is more reasonable to infer that the advertising campaign has had an effect.

Detrended Fluctuation Analysis provides a good measure of the scaling behavior of a signal (Peng C-K, Hausdorff JM, & Goldberger AL. 5; Goldberger et al. 5; Peng CK et al. 5). This algorithm works by detrending the time series at multiple window sizes; each

window size describes the time scale of interest for that run of the algorithm. If the amount of variance of the signal after detrending is significantly smaller than before detrending then one may infer that the trend at that time scale (window size) describes some deterministic variation in the signal. If, however, the difference in variance does not vary significantly from pre-detrending to afterwards then one can infer that the signal displays little deterministic activity at that time scale.

DFA is closely related to the Fast Fourier Transform (FFT) of a signal. The Short Time Fourier Transform (STFT) + residual (amount of error between the STFT of signal and the original) is actually a special case of Detrended Fluctuation Analysis in which the contours are sinusoidal (FFT-DFA) (Costa et al. 5).

DFA is also related to multi-scale Brownian Motion described above (Scafetta & Grigolini 3-5). Theoretically, if one were to perform the DFA on the signal derived from the example AntPosition + CowPosition + TruckPosition one would find the amount of difference of scale between the size of the Ant, Cow, and Truck's positions by examining the fluctuation power spectrum.

Once the signal is detrended and the fluctuations at each time scale are quantified the fluctuations are plotted against their time scales (in a log-log plot because power law scaling exhibits exponential decay) and the slope of the scale (α , or alpha) is calculated.

$\alpha < 0.5$ (Clip Noise) corresponds with anti-correlated data

$\alpha = 0.5$ (White Noise) corresponds with totally uncorrelated data

$\alpha = 1.0$ (Pink Noise) corresponds with signals exhibiting $1/f$ power scaling behavior

$\alpha \geq 1.5$ (Brown Noise) corresponds with signals exhibiting random (drunk) walk behavior

Experiment

I hypothesized that Detrended Fluctuation Analysis would provide a good measure for revealing the underlying structure of a piece of music, by measuring the moments of high and low complexity (Little MA et al. 10). I based this hypothesis from previous tests of applying the DFA to musical signals (Streich 42; Jafari, Pedram, & Hedayatifar 3). To test this hypothesis I applied the FastDFA (Little et al.) algorithm to each feature vector. Here I only include the results from the Amplitude feature, because the results from the other features are difficult to interpret in a musical fashion, and the Amplitude results alone are sufficiently musically salient.

I found the alpha measure to be a poor measure for structural analysis, but the accumulated bin count of the fluctuations of each time scale to be a good predictor of complexity in the music. A low flux bin count demonstrates more stable moments, and a high flux bin count corresponds to complex dynamic, non-ergodic behavior.

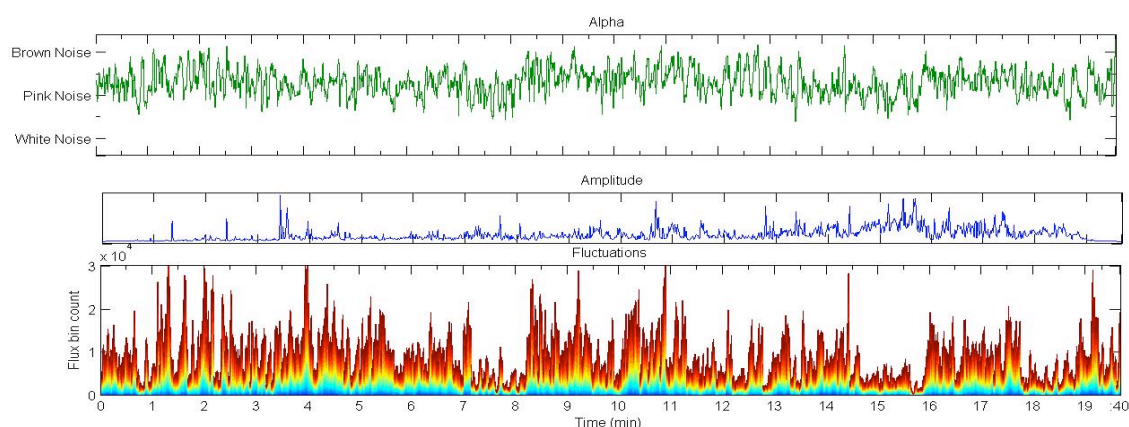


Figure 13: Results of Detrended Fluctuation Analysis. The first graph is a plot of the Alpha value over time, the second graph is of the original Amplitude feature over time, and the last graph is a plot of the aggregate fluctuation bin counts.

In this figure moments with a low flux bin count tend to correspond to moments of timbral stability in the music. For instance, the period of ~7:10-8:00 consists of high frequency filtered noise and what sounds to me like Cowell-esque scrapes on piano strings. The timbre and amplitude stays fairly constant through this section. At ~8:15 a long accordion note occurs followed by a sound somewhat like “moaning”, and there are more changes in the timbral content. Correspondingly the flux bin count jumps up dramatically at that moment, and continues to be high until a brief moment around ~9:30 where the timbre is dominated by a long accordion note, and there is little timbral change.

In contrast, I would characterize spikes in the flux bin count as dramatic gestures which have distinct timbral characteristics from the moments around them. Note, for example, the large peak ~3:56; in the music there is a moment preceding it of relative calmness, then a gesture I would characterize as a sudden reverse reverb to a piano stab. From there until ~5:00 spikes in the flux bin count correspond to quick piano notes.

The period ~16:00-18:00 characterizes very chaotic behavior in the music. The musical material includes a multitude of quick random note runs, and gestures that span the entire extent of the frequency and timbral space. The flux bin count throughout this section is correspondingly high.

From this measure along with a scaling by the original amplitude, a rough musical analysis of form might be four phrases of relative stability (roughly 0:00-1:00, 7:00-8:00, 14:30-16:00 and 18:00-19:40) alternated by three phrases of relative complexity (roughly 1:00-7:00, 8:00-14:30 and 16:00-18:00). Of course a more thorough analysis of form would necessarily include more parameters in determining segmentation, and this is based on manual selection. However, considering this analysis is based on a single measure I conclude that DFA should be a helpful measure in analysis of musical complexity.

Love Song #3: Space is the Place

PROPOSITION II: The perceptual formation of TG's at any hierarchical level is determined by a number of factors of cohesion and segregation, the most important of which are *proximity* and *similarity*; their effects may be described as follows:

PROPOSITION II.1: Relative temporal *proximity* of TG's at a given hierarchical level will tend to group them, perceptually, into a TG at the next higher level.

PROPOSITION II.2: Relative *similarities* of TG's at a given hierarchical level will tend to group them, perceptually, into a TG at the next higher level.

PROPOSITION II.3: Conversely, relative temporal *separation* and/or *differences* between TG's at a given hierarchical level will tend to segregate them into separate TG's at the next higher level.

(Tenney 103)

James Tenney's later work with Larry Polansky Hierarchical Temporal Gestalt

Perception In Music: A Metric Space Model emphasizes a spatio-temporal approach

(Tenney & Larry. Polansky 211) to music segmentation. Taking his example as

inspiration I next attempted to determine structure in this piece via clustering directly

within the feature space, rather than attempt to use information theoretic and complexity measures.

While in theory Tenney’s definition of Temporal Gestalts allows for overlapping TGs—that is, when lower-level clusters are shared by neighboring higher-level clusters, as during a transition point from one phrase to the next—in the practice laid out in Tenney’s papers this principle proved unfeasible using hierarchical clustering. A further limitation of Tenney’s approach is treating each separate feature vector as a separate dimension, because that limits one’s ability to describe the relationships between features. For instance when a trumpet player plays loudly on their instrument the resulting timbre is different from when a trumpet player plays softly. In this example changes in both timbre and amplitude are caused by the same basic component – air pressure from the player.

Experiment

In performing spatio-temporal analysis I utilized a clustering algorithm that overcomes both of these issues—the Probabilistic Principal Component Analysis for Time-Series Segmentation (PPCA-TSS) technique proposed in “Modified Gath–Geva clustering for fuzzy segmentation of multivariate time-series” (Abonyi et al.) and implemented in Matlab (Janos Abonyi & Balazs Feil). The full description of this algorithm is beyond the scope of this paper, but the basic procedure is that it finds the underlying variables that influence the observed feature vectors (via Principal Component Analysis), determines the optimal number of clusters needed to satisfy a certain compatibility criteria (similar to the concept of ergodicity described above), and arranges the size and position of those clusters within the feature space so that they best describe the variances within the underlying variables. It then uses these clusters to segment the original data

into potentially overlapping segments by matching them to the geometrically closest cluster within the feature space. This form of segmentation is termed “fuzzy segmentation” owing to its ability to overlap and interleave segments. This is analogous to the sometimes “fuzzy” distinction between late morning and early afternoon, or the indefinite distinction between river and sea at the point at which the Mississippi River reaches the Caribbean.

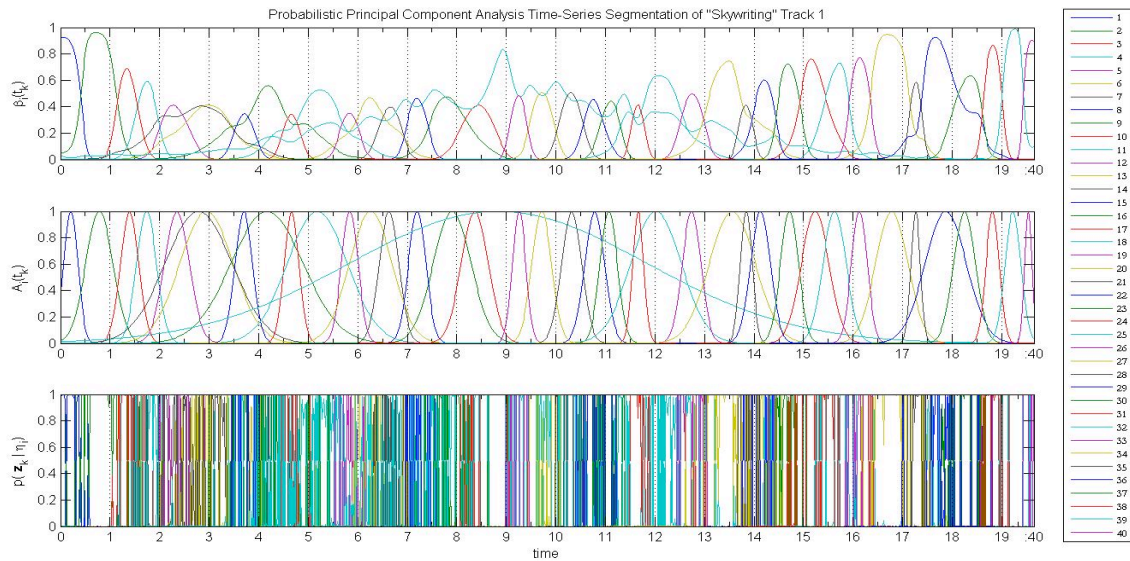


Figure 14: results of fuzzy clustering algorithm

The plot above shows the results of the PPCA-TTS algorithm. The top plot shows each cluster as a line whose height is the amount to which that moment in time belongs to that particular cluster. It is immediately obvious that some time spans are described most by a single cluster, while others are aggregations of long and short time span processes. An example of the former is the sequence of four clusters from ~16:00 to ~18:00. The first, cluster 33 coincides with a period of pointillism in the music. This is replaced at ~16:30 with a much denser spectrum, with a large amount of variability throughout the frequency

space. A small section at ~17:15 is musically dominated by strong 16th note rhythmic activity (cluster 35), but this high energy section dies off by ~17:30, when a period of low energy calm takes over.

In these examples the relationship between clusters and musical material is temporally clear. This is due to the musical material's strong cohesiveness in the feature space, which causes the Principle Component Analysis* to strongly align those clusters along the time axis, which most defines the separation between clusters.

Cluster 18 (centered at ~9:00), in contrast, lasts almost the entire piece. This cluster is most related to a upper mid frequency noisiness, which alternately sounds to my ear like it is due to something rattling, a cymbal, a shaker, or computer generated pink noise. Its character is best heard at its highest point around (~8:50), and then skipping through the section ~5:00 - ~7:00.

It is important to note that a contiguous cluster does not correlate with a contiguous musical object. Rather, it is a projection of an auditory stream over that time. Cluster 32, for instance, (~15:30) is strongly defined by the "Strongest Beat" feature. This cluster

* "If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA supplies the user with a lower-dimensional picture, a "shadow" of this object when viewed from its (in some sense) most informative viewpoint." - http://en.wikipedia.org/wiki/Principal_components_analysis

most groups musical material by its rhythmic characteristics, so even though the pitch of the material is changing and both piano and accordion contribute this time span gets grouped together. For Cluster 18 this means music material is grouped by the presence of that mid-range noisiness, which may be due to a cymbal, electronic synthesis, air from the accordion, or any multitude of acoustic sources.

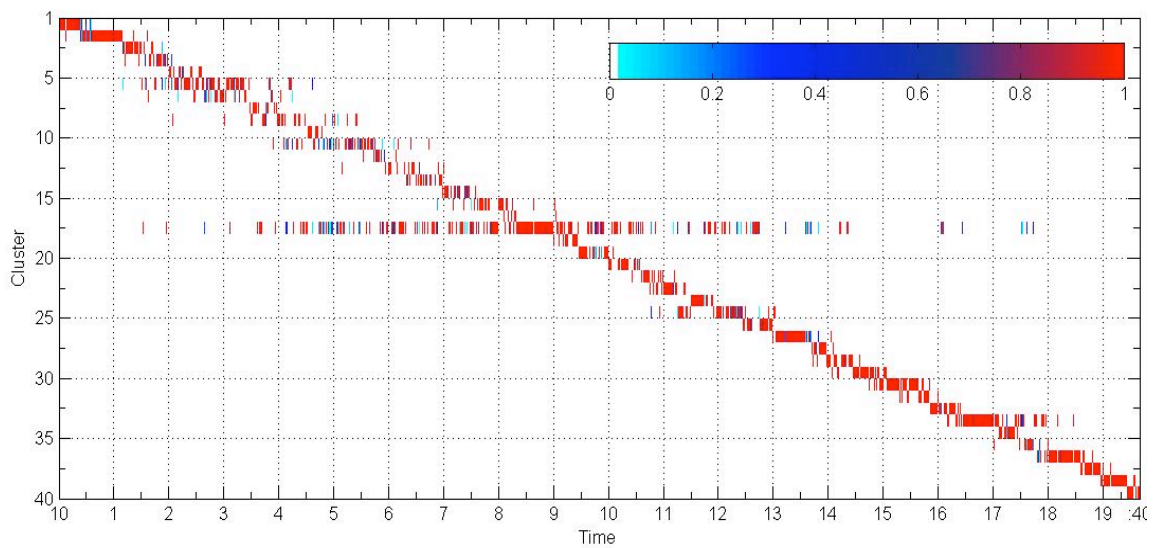


Figure 15: Fuzzy Clusters over time.

The plot above (Figure 15) shows the results of the PPCA-TTS analysis as a sequence of clusters, and their distribution in time. Sections of time that show increased overlap can be characterized as being less episodic – they consist of a mixture of clusters that slowly transition between each other. The time span ~1:30~7:00 exhibits this behavior.

Sections of time that show little overlap are more episodic in nature – for instance the time span ~10:00~13:00.

Finally, weighing the data to the calculated clusters results in a fine-grained fuzzy segmentation of the time series.

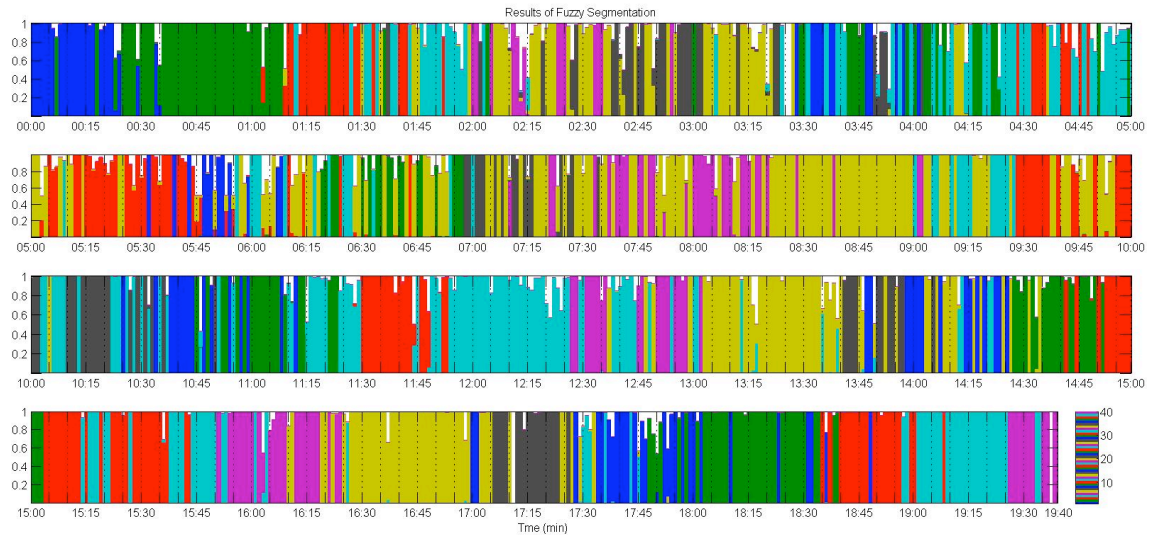


Figure 16: Fuzzy Segmentation of time series from Clusters

In this figure each second of the piece is represented by a block. The color of each block corresponds to the cluster (or clusters) to which a segment belongs (within a blocks' local neighborhood—the colors repeat over the course of the piece). Any white part of a block corresponds to variance of the feature data that is not explained by the clustering model. This plot is a horizontal projection of the same data in Figure 15. Similarly to the Multiscale entropy analysis, the fuzzy segmentation of the time series reveals blocks of statistically similar time spans. This analysis, however, segments the data at musically significant locations.

In examining the piece from directly after the fade in (?) one finds that Cluster 2 begins at 0:24 with a muted piano note^{*}. The two blocks of blue (Cluster 1) at 0:29 and 0:34 are due to moments where the original data again matches Cluster 1 more than Cluster 2. Since the center of Cluster 1 in the feature space was centered in low values in all features (because the audio signal was silence or very quiet) these two blocks correspond to moments of slightly less amplitude than the green (Cluster 2) blocks around them.

Notice that from 0:35-1:00 this algorithm does not create segmentation points at the muted piano strikes. Intuitively these onsets do not signify a new phrase each time they happen. In this musical analysis, onsets only trigger phrase boundaries if the material before and after are sufficiently different. This means that the click at 1:03, though timbrally distant from the muted piano notes, appears as a member of Cluster 3 (red), but 1:04-1:08 remain members of Cluster 2 (green). Cluster 3 is characterized by a high frequency, noise, sleighbell-like sound, and the click at 1:03 is timbrally closer to this musical material. Although sounding on the same piano string as the muted piano notes, the onset at 1:12 appears as a member of Cluster 4, because of the ringing harmonics. And the onset at 1:21 better matches the dark muted piano notes in Cluster 2, so it is marked as such.

The benefit of fuzzy segmentation is that textures need not be assumed to be sequential and exclusive. Intuitively, this matches the listener's real-time experience of

^{*} Refer to the full size fuzzy segmentation figure in the supplementary material.

improvisation; one cannot be sure if the hint of a transition will become a transition to a new section until after it has happened.

To complete the hierarchical analysis of this piece I used the positions of each cluster in the feature space and, with an added weighing for time, created a dendrogram that groups these clusters in terms of their similarity and proximity in time.

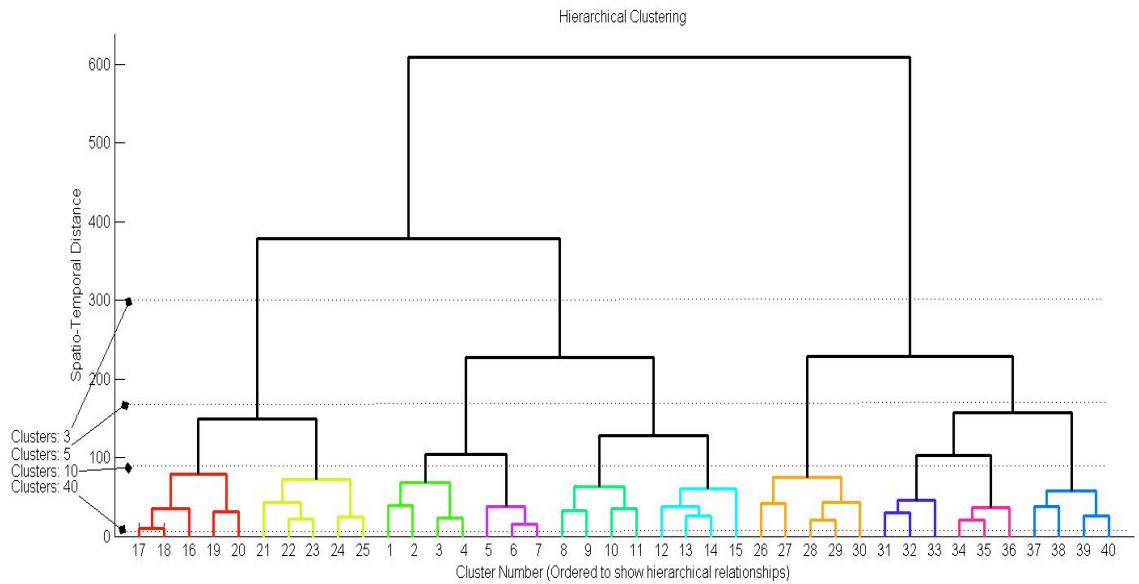


Figure 17: Dendrogram showing hierarchical grouping of clusters

In this diagram (Figure 17) the height of each U shaped stem is equal to the time distance (in seconds) and the distance in the feature space (normalized) between the centers of the two clusters it connects. The clusters are laid out along the horizontal axis for ideal representation of the tree plot instead of in temporal order. As you can see from the plot the first major hierarchical division is between the segments 1-25 (roughly ~0:00-12:00) and 26-40 (roughly ~12:00 to ~19:40).

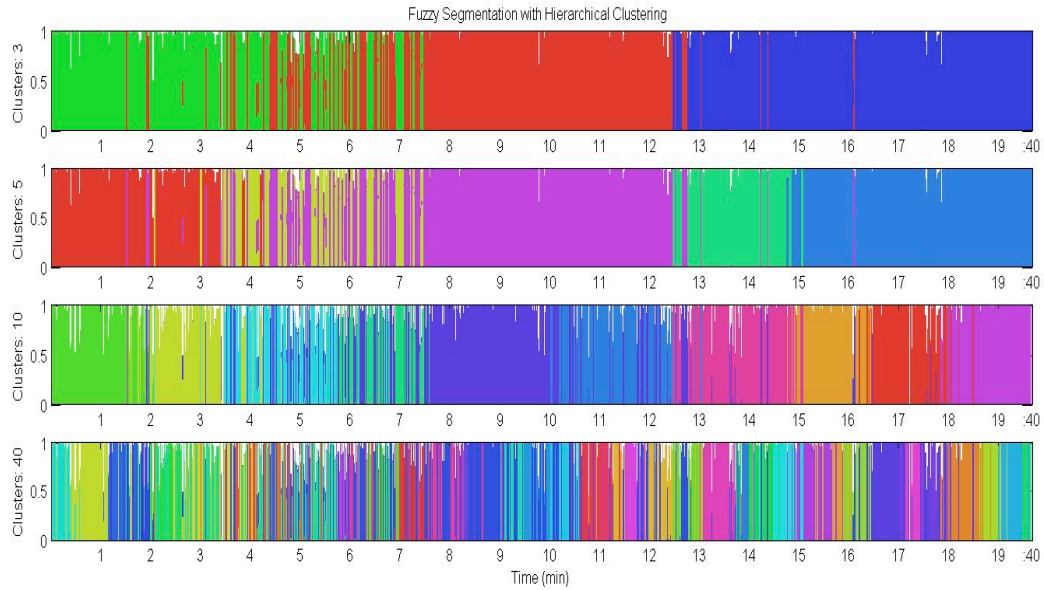


Figure 18: Fuzzy Segmentation at four hierarchical levels

The figure above (Figure 18) shows the results of fuzzy segmentation using different hierarchical levels of clustering. The relative heights at which these hierarchical levels were generated is noted as dashed lines in Figure 16. The first graph shows the piece split into roughly three segments, with the transition between the first and second segments (roughly ~3:00-7:00) to be significantly more gradual and “fuzzy” than the transition between the second and third sections (roughly ~12:30). At the second hierarchical level sampled (Clusters: 5) one can note that this level adds a separate section for the first transition, and splits the last section into two parts (roughly ~12:30-19:40).

A full analysis of these hierarchical relationships and their relation to alternate methods of analyzing this piece is beyond the scope of this thesis. The initial results from this algorithm prove very promising for future research, but modifications to the procedure

are required for any qualitative statement of musical form for the entire piece. These results are from a single run of the algorithm with arbitrarily selected parameters. A more rigorous approach would be to apply the algorithm with multiple initial cluster sizes and tolerance matches for clustering. This method might further be improved by adding pattern matching capabilities and groupings for temporally distant clusters to determine nonlinear structural relationships.

Conclusion

In this paper I proposed three algorithms, each of which approached a specific domain of concern in Tenney's theory of musical organization.

The first, Multiscale Entropy, provided inconclusive results. Though it showed some correlation with structural changes in the music these correlations were inconsistent and provided no specific rubric for which to apply the results to musical organization.

The second, Detrended Fluctuation Analysis, did provide an analytically useful measure of the relative complexity of different sections. I found that the fluctuation bin counts that resulted from this analysis correlated well with what could be roughly described in musical terms as stability or complexity. Further work is required to determine the relationship between this measure and the number of hierarchical levels required to cluster a given section of music. Nevertheless, from these results I am confident that measures of time series complexity provides a musically significant descriptor.

The third and final algorithm, Probabilistic Principal Component Analysis for Time-Series Segmentation, provided the most significant results with regards to Tenney's theories of hierarchical gestalt grouping. The results show that it is possible to algorithmically group clusters of similar musical material together with musically significant results. This initial experiment only applied this procedure to one level of the

musical structure, and further research is required for fully quantifying the correlations between the experimental results and the musical material.

These three “Love Songs” do not conclusively reach the goal of applying Tenney’s theories of Temporal Gestalt formation to computational practice, but rather provide a basis for further research. I would argue that the results of the second two experiments provide evidence that computational models of musical organization can provide musically meaningful results.

Appendix: Full Size Figures

Figure 1: plot Feature vectors over time

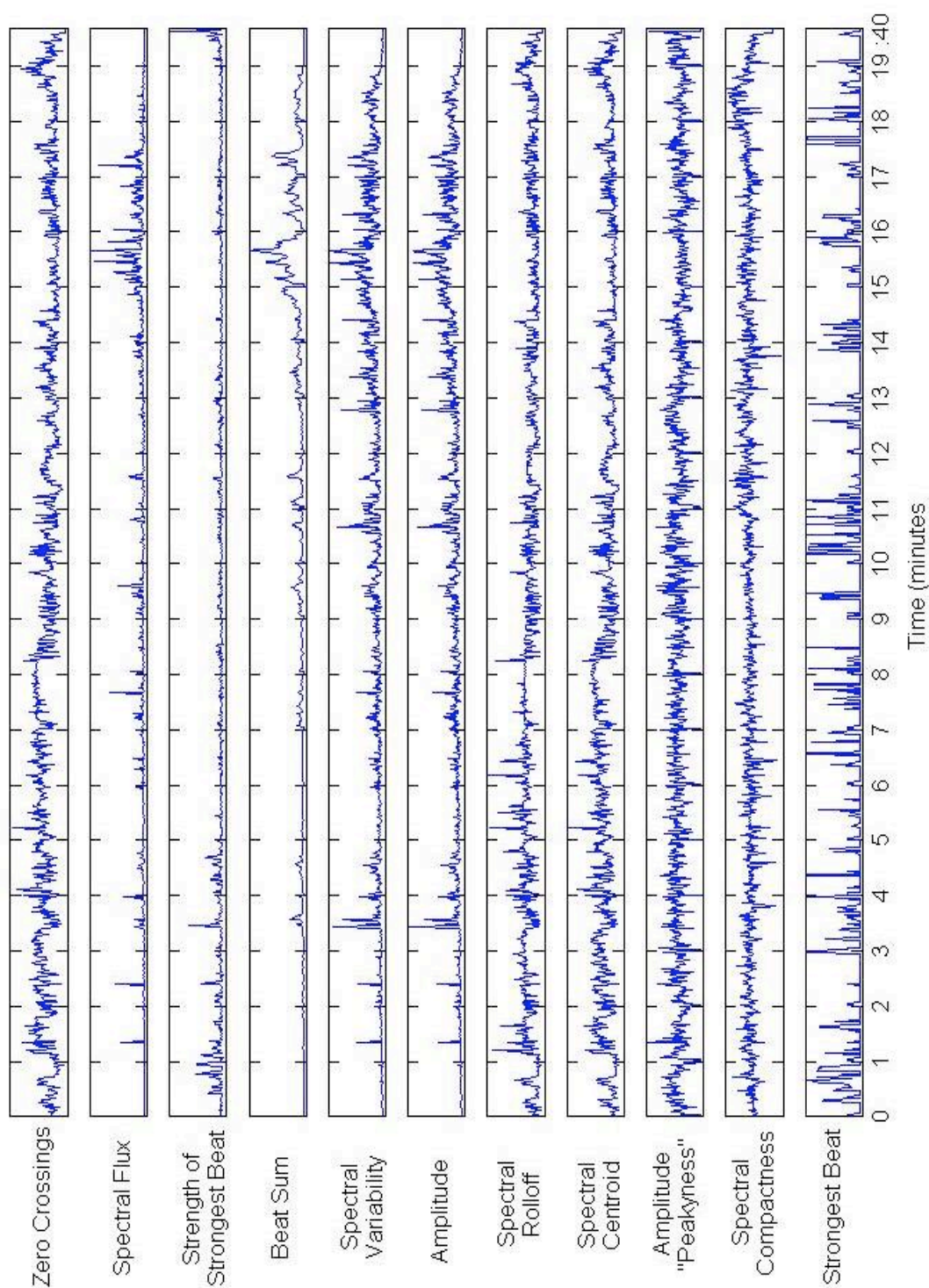


Figure 6: Entropy clustering of time series.



Figure 13: Results of Detrended Fluctuation Analysis. The first graph is a plot of the Alpha value over time, the second graph is of the original Amplitude feature over time, and the last graph is a plot of the aggregate fluctuation bin counts.

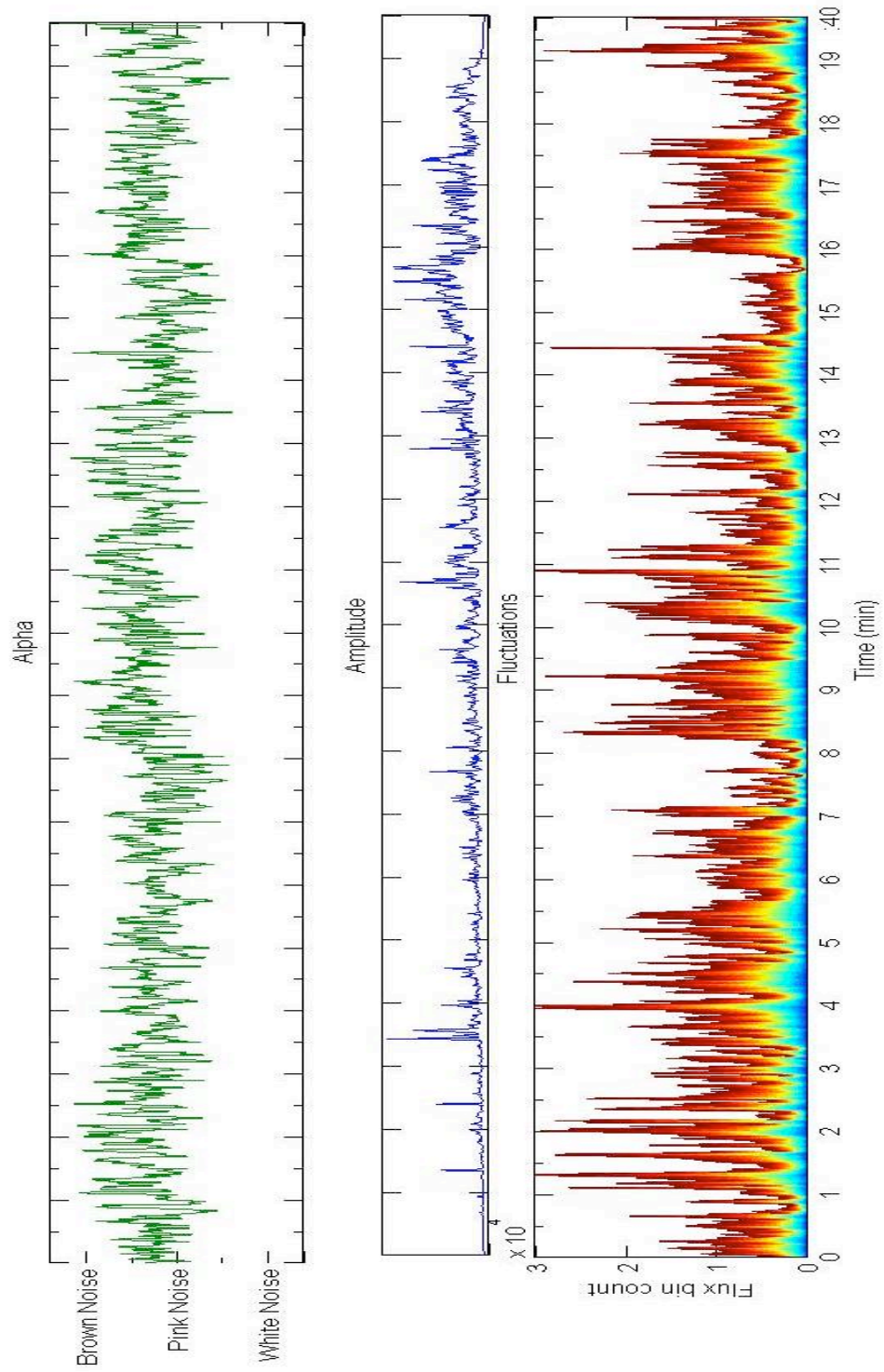


Figure 14: results of fuzzy clustering algorithm

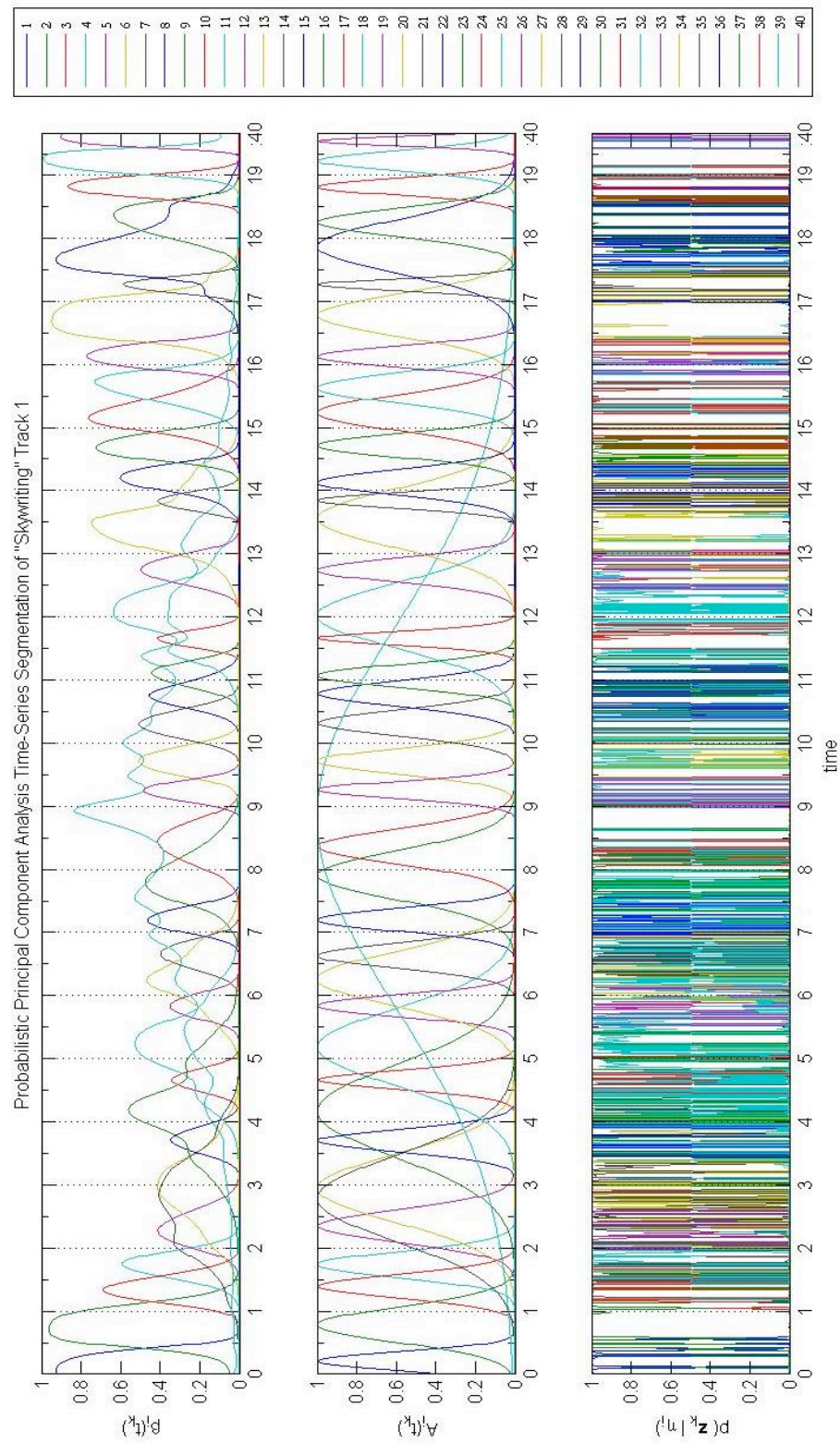


Figure 11: Fuzzy Segmentation of time series from Clusters

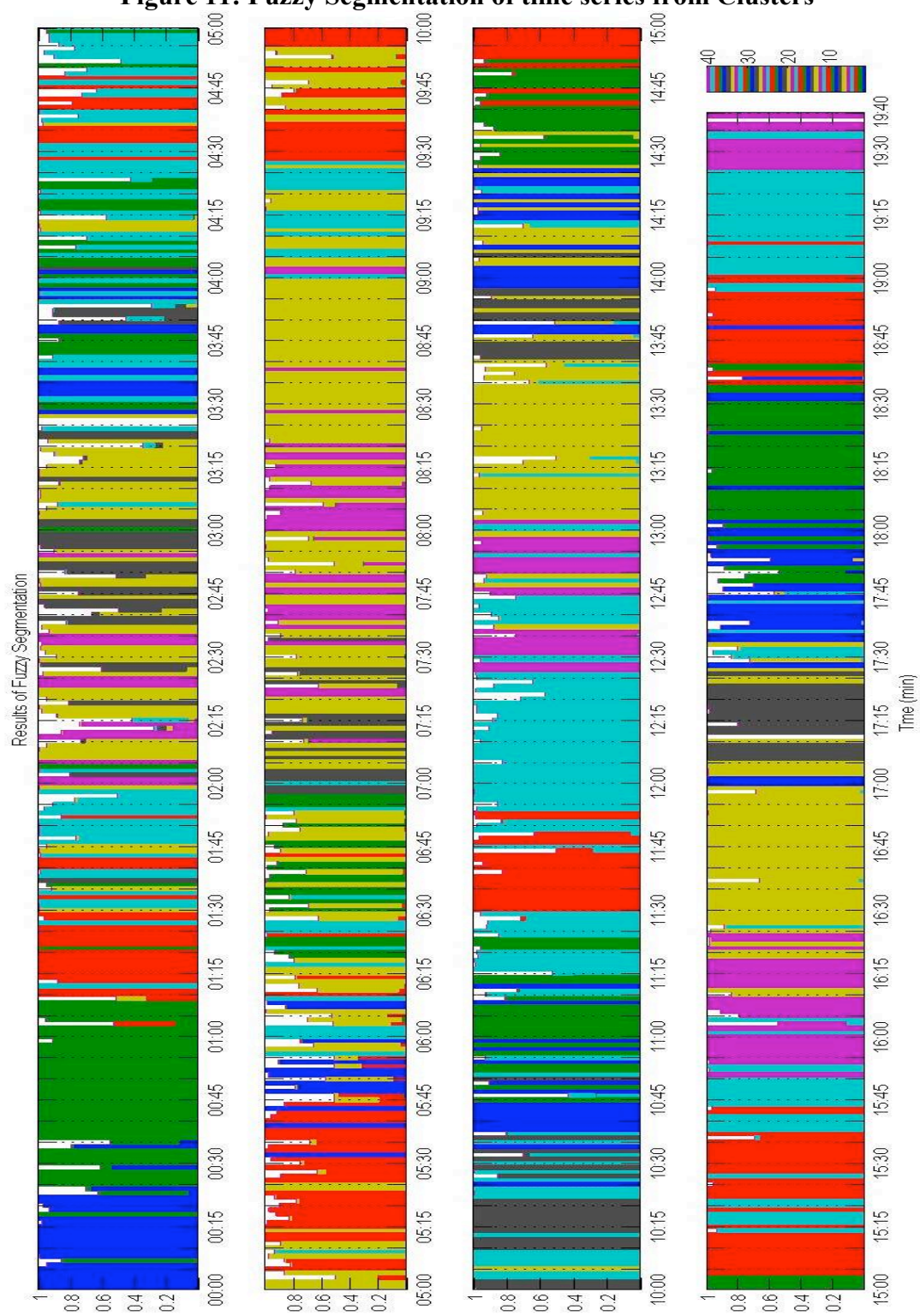


Figure 12: Dendrogram showing hierarchical grouping of clusters

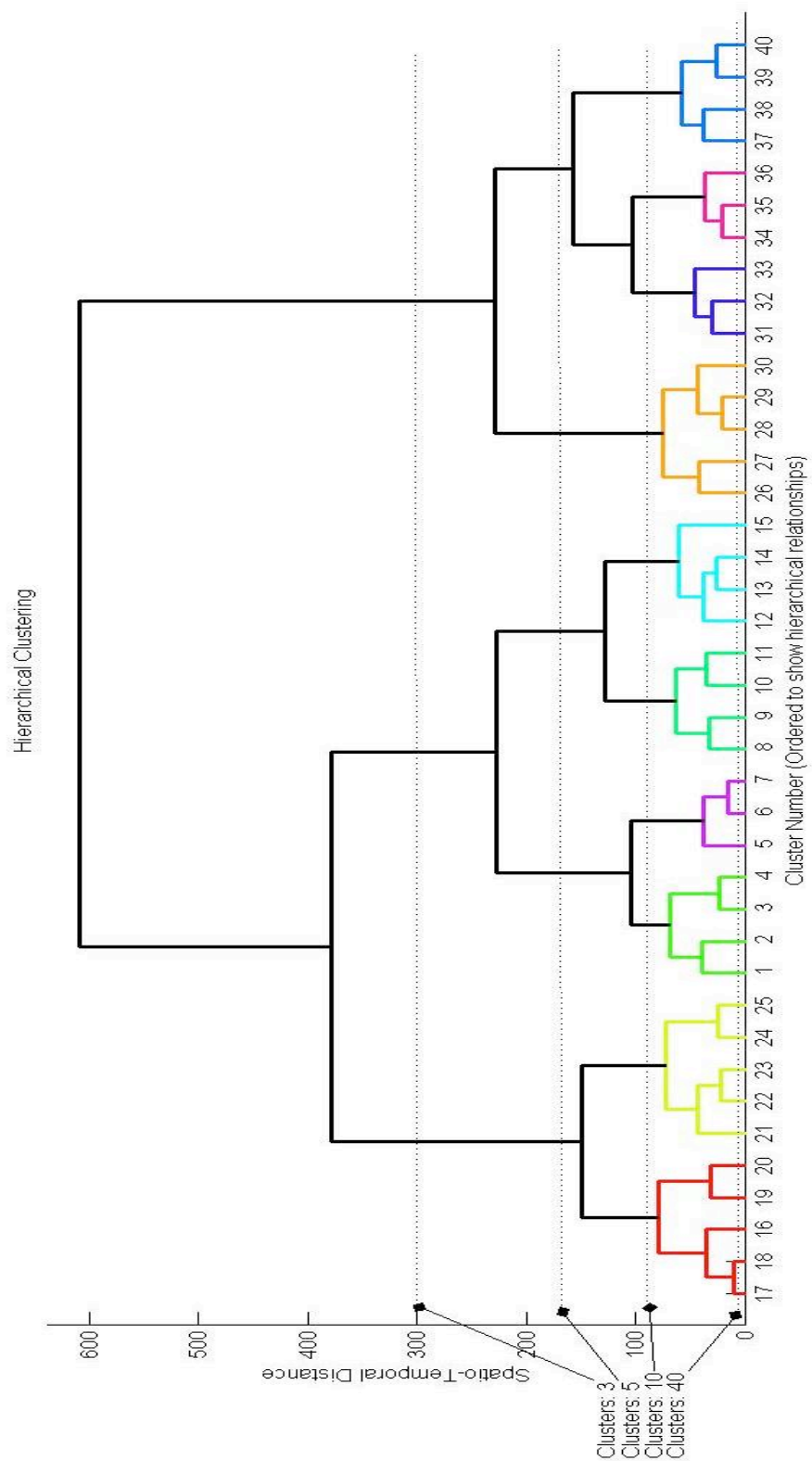
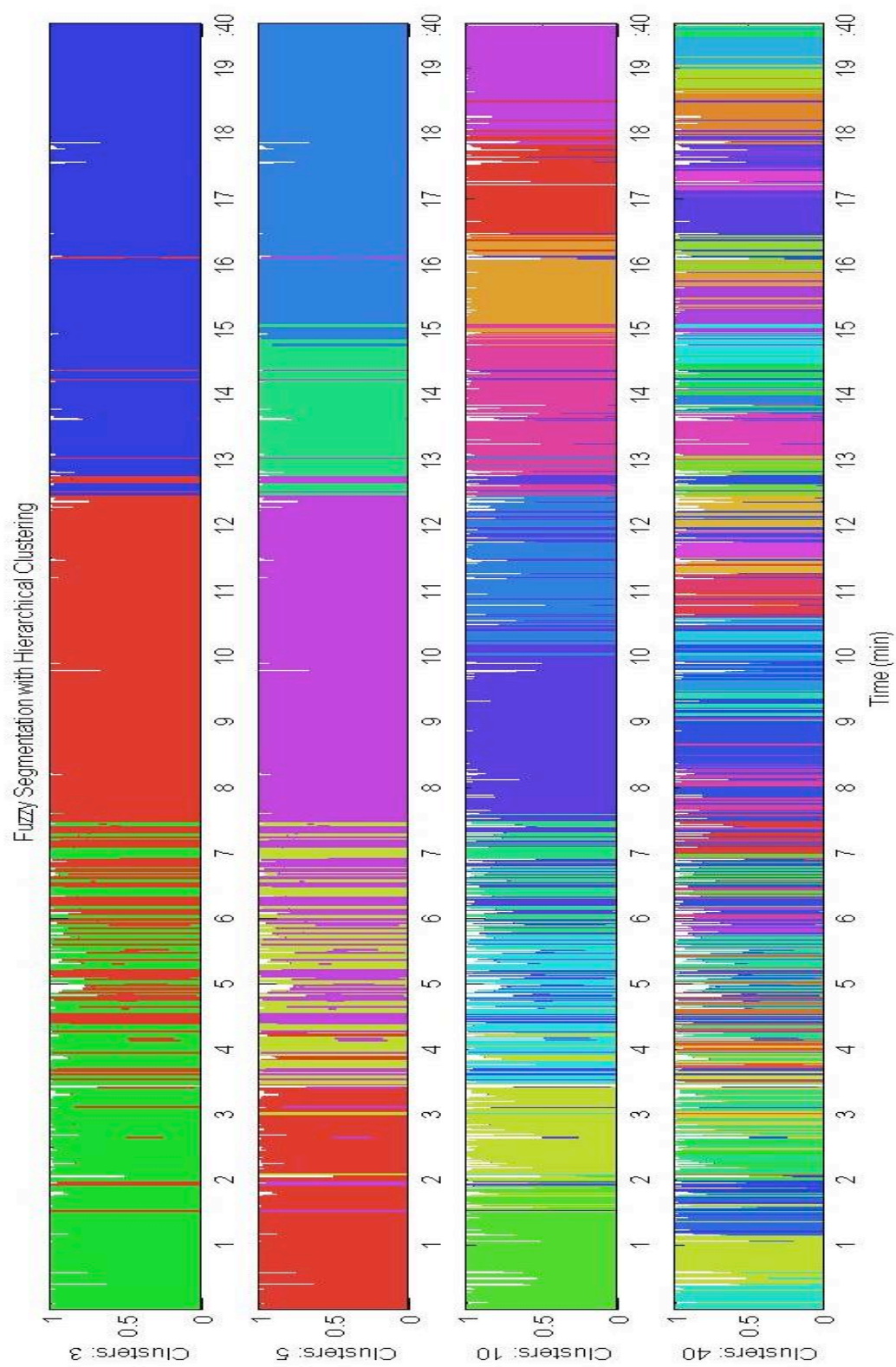


Figure 13: Fuzzy Segmentation at four hierarchical levels



Bibliography

- Abonyi, J. et al. "Modified Gath-Geva clustering for fuzzy segmentation of multivariate time-series." FUZZY SETS AND SYSTEMS 149.1 (2005): 39-56.
- Abonyi, Janos, and Balazs Feil. "Time Series Segmentation." 15 Dec 2008
<<http://www.fmt.vein.hu/softcomp/segment/index.html>>.
- Costa M, Goldberger AL, and Peng CK. "Multiscale entropy analysis of complex physiologic time series.." Physical review letters 89.6 (2002).
- Costa, Luciano da Fontoura et al. "Spectral Detrended Fluctuation Analysis and Its Application to Heart Rate Variability Assessment." q-bio/0507016 (2005). 15 Dec 2008 <<http://arxiv.org/abs/q-bio/0507016>>.
- Elias, Hugo. "Perlin Noise." Perlin Noise. 15 Dec 2008
<http://freespace.virgin.net/hugo.elias/models/m_perlin.htm>.
- Goldberger, Ary L. et al. "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals." Circulation 101.23 (2000): e215-220.
- Jafari, G. R, P. Pedram, and L. Hedayatifar. "Long-range correlation and multifractality in Bach's Inventions pitches." 0704.0726 (2007). 15 Dec 2008
<<http://arxiv.org/abs/0704.0726>>.
- Little MA et al. "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection.." Biomedical engineering online 6 (2007).
- Little, M. et al. "Nonlinear, Biophysically-Informed Speech Pathology Detection." 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. Toulouse, France, 2006. II-1080-II-1083.
<http://www.eng.ox.ac.uk/samp/dfa_soft.html>.
- McEnnis, D. et al. "jAudio: A feature extraction library." Proceedings of the International Conference on Music Information Retrieval. 2005. 600-3.
<http://www.music.mcgill.ca/~cmckay/papers/musictech/jAudio_ISMIR_2005.pdf>.

- Moles, Abraham A. Information theory and esthetic perception. Urbana: University of Illinois Press, 1968.
- Peng CK et al. "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series.." Chaos (Woodbury, N.Y.) 5.1 (1995): 82-7.
- Peng C-K, Hausdorff JM, and Goldberger AL. "Detrended Fluctuation Analysis (DFA)." Fractal Mechanisms in Neural Control: Human Heartbeat and Gait Dynamics in Health and Disease. 15 Dec 2008
<<http://www.physionet.org/tutorials/fmnc/node5.html>>.
- Polansky, Larry. "The Early Works of James Tenney." Soundings 13 : the music of James Tenney. Ed. Peter Garland. Santa Fe, NM.: Soundings Press, 1984. 119–294.
<http://eamusic.dartmouth.edu/~larry/published_articles/tenney_monograph_soundings/19_META_HODOS.pdf>.
- Purwins, H. et al. "Computational models of music perception and cognition I: The perceptual and cognitive processing chain." PHYSICS OF LIFE REVIEWS 5.3 (2008): 151-168.
- Scafetta, Nicola , and Paolo Grigolini. "Scaling detection in time series: Diffusion entropy analysis." Physical Review E 66.3 (2002): 036130.
- Snow, Michael. Wavelength. Toronto: Joyce Wieland & Michael Snow Ltd., 1967.
- Streich, S. "Music Complexity a multi-faceted description of audio content." 2007.
- Tenney, James. META+HODOS and META Meta+Hodos. Oakland, CA.: Frog Peak Music, 1988.
- Tenney, James., and Larry. Polansky. Hierarchical temporal gestalt perception in music : a metric space model. Toronto: York University, 1978.
- Vesanto, Juha. SOM toolbox for Matlab 5. Espoo: Helsinki University of Technology, 2000.