Temporal-Gestalt Segmentation—Extensions for Compound Monophonic and Simple Polyphonic Musical Contexts: Applications to Works by Boulez, Cage, Xenakis, and Ligeti

Yao-li Uno and Roland Hübner, College of Music & Department of Computer Science, University of Colorado, Boulder CO 80309-0301
email address: uno@spot.colorado,edu; rolandh@cs.colorado.edu

Abstract: This paper presents extensions of James Tenney and Larry Polansky's temporal-gestalt segmentation algorithm and applies it to selected works by Pierre Boulez, John Cage, Iannis Xenakis, and György Ligeti. The algorithm is based on the Gestalt principle of cohesion and segregation; temporal-gestalt (TG) boundaries are formed on the basis of the disjunction measure (DM) of musical events, the weighted sum of intervallic distances between a given musical event and the preceding one with respect to the musical parameters involved. TG boundaries are formed hierarchically from the lowest to successively higher levels of musical structure. In our extension, we postulate two categories of polyphony, compound monophony and simple polyphony, after Tenney.

I. Introduction. Recent studies offer several different analytical approaches to the segmentation of post-tonal musical repertoires: Jean-Jacques Nattiez's semiotic approach, Le Grand and Jackendoff's generative theory, and its extension to atonal contexts, and most recently, Taeva and Lefkowitz's system for determining segmentation based on the ratios of discontinuities with respect to musical parameters involved. Our approach expands upon a hierarchical model developed by James Tenney and Larry Polansky in the late '70s. Their segmentation algorithm explores the effect of relative weights of musical parameters on segmentation at various hierarchical levels of musical structure. While the application of the original algorithm was limited to monophonic music, our research provides preliminary extensions for application to two types of polyphony in post-tonal music.

Our objective in this research is two-fold: (1) to examine the relative salience of all musical parameters in determining the segmentation of post-tonal music, and (2) to determine the boundaries between syntax and design—the correspondence between the structural boundaries determined by the generative syntax and the formal boundaries artificated by the segmentation algorithm. First, salience refers to the relative strength of a given musical parameter in effecting a break or disjunction in the continuity of musical events. It is important to consider not just the roles of pitch and duration in isolation, but the roles of all other musical parameters in this regard. Second, the syntax that generates a piece, for instance, twelve-tone rows, represents the underlying grammatical organization, while the segmentation algorithm models the change in music or rhetorical emphasis within the musical continuity of a given piece. As in speech, the knowledge of grammar is essential to the comprehension of the underlying structure, but the rhetorical facets of speech, such as intonation, nuance, and vocal quality determine the emotive character as well as shape the internal grouping of musical events within a piece. The segmentation algorithm provides a tool that measures the extent to which the rhetorical or gestural design supports the syntactical structure of a given piece.

II. Fundamental Tenets. Tenney and Polansky's approach to segmentation is based on the Gestalt principles of proximity and similarity. Their fundamental hypothesis is that cohesion and segregation is measured by the interval magnitude of the musical parameters involved. Widely intervallic distances in musical parameters between adjacent time-spans cause a disjunction in the musical continuity, while narrower inter intervallic distances cause the adjacent time-spans to cohere as a temporal-gestalt (TG) unit. Tenney and Polansky define a TG unit as musical event that is internally cohesive and externally segregated from comparable time-spans immediately preceding and following it.

First, the input values of various musical parameters, i.e., pitch, duration, dynamics, timbre, temporal density, are stored as a record in a linked list as shown in Fig. 1a. The values in each parameter are encoded in real number, then scaled to fit the common range between 0 and 1. Each TG at the root level, the lowest level of hierarchy, therefore, contains a set of real numbers that represents the values of participating musical parameters.

Fig. 1a

Second, the disjunction measure (DM) of each node (index n) is computed by taking the weighted sum of intervallic distances to the previous note (index n-1) with respect to number k of musical parameters, P, involved as follows:

\[ DM[n] = \sum_{i=1}^{k} P[i-1] \cdot W_i \]
The weights for musical parameters are exterior variables that can be adjusted by the analyst; everyone the program is run. The analyst calibrates the relative weights for the musical parameters (scaled between 0 and 1) heuristically until the optimum TG segmentation, formal divisions that approximate those based on one's hearing of the piece, is obtained.

Finally, the TG boundaries are determined at successive higher levels by the following procedure: Given $T_G(x,y)$ where $x$ = hierarchical level and the time-span of the TG ranges between $x$ and $y$, then the TG element at the root level, is defined as $T_G(i) = n(i)$, where $n(i)$ is a note in index $i$. Then the higher levels are defined recursively as follows: $T_G(4,i) = T_G(3)T_G(4)\ldots T_G(i)$. This procedure can also be illustrated graphically as shown in Fig. 1b. Here, the first disjunction at the root level occurs at $x$. The hierarchical levels above root are called clang, sequence, segment, section, after Tenney. The program terminates when less than three TGs are formed at the top of the hierarchy.

![Diagram](image)

*Fig. 1b: Extensions for Compound Monophony and Simple Polyphony*

In designing the extensions of the algorithm for polyphonic musical contexts, I posit two categories of polyphony, compound monophony and simple polyphony, after Tenney [Mead-Hodos, 102]. Compound monophony refers to music with multiple parts or voices that are, nonetheless, perceived as one composite musical gesture due to the lack of contrapuntal, regional, or timbral differentiations among them. For instance, Boulez's Structures Ia comprises two piano parts, but the parts cannot be distinguished polyphonically due to the pointillistic, atematic character of the music and the constant registral intersection between the parts. Cape's *Music of Changes* (Book I) and Xenakis's *Hermes* also fall into this category.

Since in compound monophony, musical gestures are heard as a composite unit, we introduce a procedure for averaging and compressing simultaneous into a single musical TG at the root level. As the "before" and "after" conditions in Fig. 2b illustrate, the parametric values of notes that fall on the same time-points (thereby forming a simultaneity) become averaged and compressed into a single strand of TGs. Wherever a simultaneity is followed by another, the DM is averaged as the weighted sum of the DM between all members of the two simultaneities as shown in Fig. 2a.

![Diagram](image)

*Fig. 2a: DM(sin(x),sin(y)) = \sum_{i,j} DM(m_i,n_j)x_iy_j*

In averaging the parametric values of a simultaneity, it is assumed that the outer pitches should be weighted more in relation to the inner pitches of the simultaneity; the perceptual salience of the outer notes, of course, varies according to the range, spacing, and timbre of the instrument. At this preliminary level, the outer notes are weighted more strongly based on the bandwidth of the simultaneity for pitch; that is, where the outer notes lie within the regions defined by the standard deviation and the mean for the pitch range of a given piece. As shown in Fig. 2c, the weighting of the outer notes, $k$, varies according to regions, $R$, delimited by the distance of standard deviation around the mean, so that the further away the range of the simultaneity is from the mean, the outer notes are weighted more strongly in relation to the inner notes, and vice versa.

![Diagram](image)

*Fig. 2c: m*(a(x)) = w*(a(x)) + w*(a(x)) + w*(a(x))*

The second category called simple polyphony refers to music with multiple voices or parts that can be perceived independently of one another due to contrapuntal, regional, and/or timbral differentiations among them. Here the algorithmic extension is applied to the first piece from György Ligeti's *Ten Pieces for Wind Quintet*. While the piece is characterized by chromatic voice-leading within narrow, intersecting pitch ranges, the timbral differences among the instruments, e.g., flute, English horn, clarinet, and horn, lend polyphonic distinction to the five parts.
Temporal-Gestalt Segmentation

In the algorithmic extension, the five parts are stored as independent strands of TGs at the root level. Here, in addition to the weights, the analyst explores the level at which the strands are merged. The analyst determines this level according to the optimum TG segmentation obtained. The strength of the polyphonic independence of the five parts is modeled by the hierarchical level at which the five strands become eventually merged: the higher the level at which the strands are merged, the greater the level of independence of individual parts, and vice versa. In the case of Ligeti’s work, the TG strands are merged at the lowest or root level. This corroborates our intuition that the polyphonic independence of the five parts, aside from their timbral differences, is relatively weak; one can hear gestural connections that cut across different voices frequently.

IV. Application. The relative weights for the four works are shown in Fig. 3a; the parameter with the highest weight for each work is highlighted in bold. Sim refers to the external weight placed on simultaneities wherever they occur.

Fig. 3a  

\[ \begin{align*}
\text{Boulez:} & \quad 0.2 \quad 0.6 \quad 0.4 \quad 0.3 \quad 0.1 \quad 0.8 \\
\text{Cage:} & \quad 0.05 \quad 0.9 \quad 0.1 \quad 0.01 \quad 0.03 \quad 0.3 \\
\text{Xenakis:} & \quad 0.1 \quad 0.8 \quad 0.15 \quad 0.02 \quad \ldots \quad 0.1 \\
\text{Ligeti:} & \quad 0.1 \quad 0.5 \quad 0.2 \quad 0.9 \quad \ldots \quad 0.5
\end{align*} \]

\* \* 

\* \* 

Fig. 3b  

\[ \text{exact correspondence;} \]

\[ \text{ portraying the piece;} \]

\[ \text{approximate correspondence} \]

The differences in the relative weights of the four works parallel the relative salience of musical parameters in regulating the TG boundaries. Furthermore, Fig. 3b shows the TG segmentation of each work that results from the given weights. Here, the vertical axis of each graph displays the fluctuations in DMs at the next-highest TG level, while the horizontal axis depicts elapsed time in seconds. The letters or Roman numerals above the plotting times show the extent to which the TG boundaries correspond with the structural boundaries articulated by the underlying syntax.

In Boulez’s Structures Ia, the TG boundaries are articulated primarily by the weights of simultaneities (0.8) and attack-point duration (0.6). Simultaneities create audible juxtapositions in the course of the piece since the rhythmic alignment of rows allow them to occur only at the beginning of new rows boundaries. While this work is frequently cited as an example of integral serialization in the early ’60s, note that the gestural design of the piece is shaped not by the serialized parameters, e.g., pitch, duration, dynamics, and attacks, but by the changes in polyphonic density—a factor that was determined on an ad hoc basis. In fact, Boulez’s serial method proved to be largely ineffective since the syntactical relationships between pitch and duration become randomized and obscured at the musical foreground. The smaller weights associated with pitch and duration show the relative insignificance of the serial structure in articulating the global design of the piece.
The assigned weights for Boulez, nonetheless, produce a good fit between row and TG boundaries. Notice how the structural mid-point of the piece coincides with the maximum disjunction occurs.

For comparison, the attack-point duration plays a prominent role in regulating the segmentation of Cage's and Xenakis's works. Most importantly, the salience of dorons resonates with Cage's aesthetic view of the primacy of duration in establishing a foundation for music-in-sound. While Cage used the 1 Chute to derive the musical content of this piece, the musical realization attests to a careful interplay between chance and choice; for instance, it is his choice in determining the probability of sound and silence to be fifty-fifty: this factor allowed for abrupt temporal discontinuities, e.g., prolonged silence between ants that contribute to the peak in DM at odd boundaries. As illustrated by the graph, it shows the most abrupt changes in the fluctuation of DMDC; it reaches a peak at the beginning of units 2 and 3, as the temporal density increases suddenly.

Xenakis's Hermà is also dependent primarily on the salience of attack-point duration; it is the temporal discontinuities, abrupt changes between activity and silence, that punctuate the internal junctures in the piece. The pitch organization of Xenakis's Hermà is based on the sets A, B, and C, and the transformation of these sets based on the logical operations of union, intersection, and negation (or complementation). The letters inside Xenakis's graph, indicate correspondences between set boundaries and TG boundaries; the statistically homogeneous dispersion of pitches (owing to the stochastic operations) makes it otherwise impossible to distinguish where transitions occur with respect to the underlying set transformation.

Finally, Ligeti's work is distinguished from the other two in that the primacy of dynamics is articulating the TG boundaries. This relatively short piece is characterized by two contrasting textures. The first part features micropolyphonic, meandering chromaticism in the five instruments that becomes more compressed and condensed. The texture to the second part (m. 16/2 to end) is characterized by sudden shifts in the dynamic and regional levels; the texture presents as chromatic as the instruments enter in pairs of two or three to sustain long held notes at the registral and dynamic extremes (44). The TG segmentation captures the abrupt contrast between the two sections; the first section shows gradual decline in DM as the micro-polyphonic texture is sustained, followed by the sudden peak (at elapsed time 93.38 sec.) that parallels the abrupt textural change in the second section.

V. Future Considerations: The segmentation algorithm recognizes a "neutral" level of organization, where the disjunction is based on the composite changes in intervallic magnitude of musical parameters; the TG segmentation is, however, not influenced by memory of events, knowledge of syntax, familiarity with the piece, all those factors that vary according to the listener's background [what psychologists call a set]. Further refinements need to be introduced as follows: 1) include context-sensitive criteria, i.e., pattern recognition, motivic recurrences, in regulating the segmentation; 2) allow flexibility in the disjunction criteria, so that it is not based solely on the condition that the DM must be greater than the previous and following TGs; 3) develop a "dynamic" as opposed to the "statistical" system of weights presently used so that the weights can be altered in the course of a piece as deemed necessary; and 4) refine the criteria and procedure for measuring disjunction in simple polyphonic contexts.

The last category merits special attention. According to Hartmann and Johnson's study of stream segregation, our ears tend to shift in and out of linear and vertical gestural connections in addition. Therefore, the disjunction for simple polyphony may be more accurately modeled by weighting the DMs formed linearly within each polyphonic strand with the DMs formed vertically between polyphonic strands. This can be expressed as follows. Given a, b, and c, the DM for n and m parts, dm = time[n1]-time[m1] and time[m1]-time[n1], dp = pitch difference p[n]-p[m], k = scaling factor to change the influence of the vertical distance, exp is exponential function and T changes the change of α′ sigmoidal from a step to a very smooth function. Then the composite DM can be measured as follows:

\[
DM(n, m) = \frac{\sum\text{DM}[n, m]}{\text{pitch difference} \cdot (1-\exp(-k \cdot \text{distance between voices}))}.
\]

The value of k is inversely proportional to the linear degree of polyphonic independence between voices: the lesser the value of k, the more strongly the DM weighted on the linear DMs within each strand, and the greater the value of k, the more strongly the DM weighted on the vertical relationship between voices. The segmentation algorithm can go expanded, in this way, to test different criteria involved in our perception of polyphony.

References:


