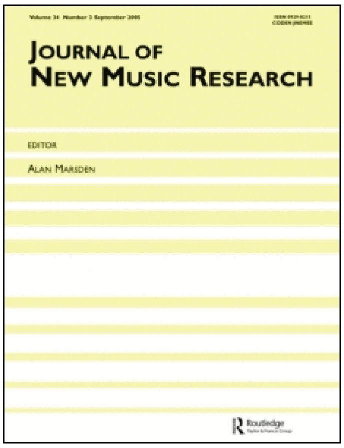



**Journal of New Music Research**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713817838

# Regards on two regards by Messiaen: Post-tonal music segmentation using pitch context distances in the spiral array

Elaine Chew[a]
[a] University of Southern California Viterbi School of Engineering, Los Angeles, California, USA

**To cite this Article** Chew, Elaine(2005) 'Regards on two regards by Messiaen: Post-tonal music segmentation using pitch context distances in the spiral array', Journal of New Music Research, 34: 4, 341 — 354

**To link to this Article: DOI:** 10.1080/09298210600578147

**URL:** http://dx.doi.org/10.1080/09298210600578147

Journal of New Music Research
2005, Vol. 34, No. 4, pp. 341 – 354



# REGARDS ON TWO REGARDS BY MESSIAEN: Post-Tonal Music Segmentation Using Pitch Context Distances in the Spiral Array

Elaine Chew

University of Southern California Viterbi School of Engineering, Los Angeles, California, USA



## Abstract

This paper describes an $O(n)$ algorithm for segmenting music automatically by pitch context using the Spiral Array, a mathematical model for tonality, and applies it to the segmentation of post-tonal music, namely, Olivier Messiaen's *Regards IV* and *XVI* from his *Vingt Regards sur l'Enfant Jésus*. Using the idea of the *centre of effect* (*c.e.*), a summary point in the interior of the Spiral Array, segmentation boundaries map to peaks in the distances between the *c.e.*'s of adjacent segments of music. The best-case computed boundaries are, on average, within 0.94% (for *Regard IV*) and 0.11% (for *Regard XVI*) of their targets.

## 1. Introduction

Segmentation by context is a necessary part of music processing both by humans and by machines. Efficient and accurate algorithms for performing this task are critical to computer analysis of music, to the analysis and rendering of musical performances, and to the indexing and retrieval of music using large-scale datasets. Computational modelling of the segmentation process can also lead to insights into human cognition of music. This paper will focus on the problem of determining boundaries that segment a piece of music into contextually similar sections according to pitch content. In particular, the algorithm will be applied to Olivier Messiaen's (1908 – 1992) *Regard de la Vierge* and *Regard des prophètes, des bergers et des Mages*, the fourth and sixteenth pieces in his *Vingt Regards sur l'Enfant Jésus* (1944).

The $O(n)$ segmentation method uses the Spiral Array model (Chew, 2000), a mathematical model that arranges musical objects in three-dimensional space so that inter-object distances mirror their perceived closeness. The Spiral Array represents tonal objects at all hierarchical levels in the same space and uses spatial points in the model's interior to summarize and represent segments of music. The array of pitch representations in the Spiral Array is akin to Longuet-Higgins' harmonic network (Longuet-Higgins and Steedman, 1971) and the *tonnetz* of neo-Riemannian music theory (Cohn, 1998). The key spirals, generated by mathematical aggregation, bears structural similarity to Krumhansl's network of key relations (Krumhansl, 1990), constructed using multi-dimensional scaling of experimental data, and key representations in Lerdahl's tonal pitch space (Lerdahl, 2000), formulated from a metric based on a hierarchical (chromatic, diatonic and triadic) arrangement of pitch relations.

The present segmentation algorithm, named Argus, is the latest in a series of computational analysis techniques utilizing the Spiral Array. The Spiral Array model has been used in the design of algorithms for key-finding (Chew, 2001), pitch spelling (Chew and Chen, 2005) and off-line determining of key boundaries (Chew, 2002). The

*Correspondence*: Elaine Chew, University of Southern California Viterbi School of Engineering, Epstein Department of Industrial and Systems Engineering, Integrated Media Systems Center, 3715 McClintock Avenue GER240, Los Angeles CA 90089–0193, USA. E-mail: echew@usc.edu

DOI: 10.1080/09298210600578147 

previous model for determining key boundaries is the one most relevant to this paper. This earlier algorithm requires knowledge of the entire piece and the total number of segments, and does not compute in real-time. The present algorithm scans the piece once from beginning to end, and thus can be realized as a real-time method for computing segmentation boundaries. So far, the Spiral Array has only been used in the analysis of tonal music. This paper extends the Spiral Array model's applications to the analysis of post-tonal music. The current Argus method is the first that explicitly uses distances in the Spiral Array as a measure of pitch context difference or separation.

The Spiral Array consists of an array of nested spirals each representing a kind of tonal entity at some hierarchical level. The Argus algorithm for automatic segmentation uses only the outermost pitch spiral in the Spiral Array model and the interior space to compute a distance between the local contexts (captured by a *pair* of sliding windows) immediately before and after each point in time. The distance measure peaks at a segmentation boundary, and the peaks can be used to identify such boundaries. Since the segmentation algorithm detects boundaries between sections employing distinct pitch collections, the procedure does not depend on key context, and can be applied in general to both tonal and atonal music.

The Argus method is highly efficient, computing in $O(n)$ time, and requires only one left-to-right scan of the piece. The algorithm computes in real-time; however, due to the formulation of the problem, there will necessarily be a lag in the assignment of a boundary that is equivalent to the length of the forward window. The algorithm is tested on Messiaen's *Regards IV* and *XVI*, and the results presented for various window sizes. The computational results are compared to manual segmentations of the piece.

Related work includes that on finding local tonal context, for example, Temperley's dynamic programming approach to determining local key context (Temperley, 1999), Shmulevich & Yli-Harja's median filter approach to local key-finding (Shmulevich and Yli-Harja, 2000), and Toiviainen and Krumhansl's self-organizing map approach to determining and visualizing varying key strengths over time (Toiviainen and Krumhansl, 2003). These methods centre around key-finding, which applies only to tonal music; and, the focus of these methods is on the determining of the local key context, rather than on finding segmentation boundaries. Segmentation, in general, is an extremely broad topic, and touches upon motivic patterns, rhythmic structure, metric groupings, phrase groupings, harmonic sequences, and timbral quality, amongst a host of other features. For every recognizable aspect of music for which computational algorithms can be devised, there can be a corresponding study on segmentation by that feature. As a result, a literature review of segmentation in general is beyond the scope of this paper.

The remainder of the paper presents a concise overview of the Spiral Array model, followed by a description of the segmentation algorithm. Then, in Section 3, a descriptive analysis of *Regard IV* is followed by a detailed analysis of the computational results, comparisons of the computed and manually assigned boundaries, and geometric explanations of the effectiveness of the proposed method. A similar treatment of *Regard XVI* is presented in Section 4, followed by discussions in Section 5, and conclusions in Section 6.

## 2. The Spiral Array model

This section provides an overview of the structure of, and underlying concept (namely, the *centre of effect*) behind, the Spiral Array model (Chew, 2000). The description of the proposed segmentation algorithm follows the introduction to the Spiral Array.

### 2.1 The model's structure

The Spiral Array model represents pitches on a spiral so that spatially close pitch representations form familiar higher-level tonal structures such as triads and keys. The model represents each higher-level object as the convex combination of its lower-level components. The weighted sum of representations of the components result in a spatial point in the interior of the pitch spiral. For example, Figure 1 shows the hierarchical construction of major key representations, from pitches to triads to keys.

Figure 1(a) shows the pitch spiral. Adjacent pitches along the spiral are related by intervals of a Perfect Fifth; each turn of the spiral contains four pitch representations, as a result, vertical neighbours are a Major Third apart. Pitch representations can be generated by the following equation:

$$\mathrm{P}(k) = \begin{bmatrix} \mathrm{r}\sin(\mathrm{k}\pi/2) \\ \mathrm{r}\cos(\mathrm{k}\pi/2) \\ \mathrm{kh} \end{bmatrix}, \qquad (1)$$

where $r$ is the radius of the spiral, and $h$ is the vertical ascent per quarter turn. Each pitch representation is indexed by its number of Perfect Fifths from a reference pitch. For example, C is arbitrarily mapped to the index 0, and represented by $\mathbf{P}(0)$. G, a Perfect Fifth above C, is mapped to the index 1, and represented by $\mathbf{P}(1)$. The model assumes octave equivalence so that all pitches with the same letter name map to the same spatial point on the pitch spiral.

Based on this arrangement of pitches on the pitch spiral, pitches that define a triad form compact clusters. They also outline the vertices of a triangle. Each triad is represented by the convex combination of its component pitches, that is to say, a point in the interior of the triangle. For example, the major triad is defined as:

$$\mathrm{C_M(k)} = w_1\mathrm{P(k)} + w_2\mathrm{P(k+1)} + w_3\mathrm{P(k+4)}, \qquad (2)$$

where $w_1 \geq w_2 \geq w_3 > 0$, and $\sum_{i=1}^{3} w_i = 1$.

The sequence of major triad representations also forms a spiral, as shown by the inner spiral in Figure 1(b).

Three adjacent major triads uniquely define the pitch collection for a major key. They also form the IV, I and V chords of the key. Hence, major keys are defined as:

$$\mathrm{T_M(k)} = \omega_1\mathrm{C(k)} + \omega_2\mathrm{C(k+1)} + \omega_3\mathrm{C(k-1)}, \qquad (3)$$

where $\omega_1 \geq \omega_2 \geq \omega_3 > 0$, and $\sum_{i=1}^{3} \omega_i = 1$.

Again, the sequence of major key representations form a spiral. This major key spiral is shown as the innermost spiral in Figure 1(c). Corresponding definitions exist for the minor triad and key representations (see, for example, Chew 2000).

### 2.2 Parameter selection

The Spiral Array model is calibrated using mathematical constraints to reflect perceived closeness among the different entities. For example, in the definition of the major triad, the weights are constrained so that the weight on the root is no less than the weight on the fifth, which is no less than the weight on the third. Since the segmentation algorithm uses only the pitch representations, we shall describe here only the parameter selection for the pitch spiral, namely, the choice of $r$ and $h$.

The pitch spiral is uniquely defined by its aspect ratio $h/r$. The goal is to constrain the parameters so that the distance between any two pitch representations correspond to their perceived closeness. Suppose the desired rank order of the interval distances is as follows: {(P5/P4), (M3/m6), (m3/M6), (M2/m7), (m2/M7), (d5/A4)}, where P denotes a perfect interval, M a major interval, m a minor interval, d a diminished interval, and A an augmented interval. Figure 2 shows examples of these intervals inside the Spiral Array. Admittedly, this is an approximation of reality as it assumes equivalence within interval classes, and of ascending and descending intervals of the same size.

Algebraic manipulation shows that the mathematical constraints on the aspect ratio, $\sqrt{2/15} \leq \mathrm{h/r} \leq \sqrt{2/7}$, produce the desired distance-based ranking of interval relations. In the remainder of the paper, $h/r$ is always set to $\sqrt{2/15}$, the case in which major triad pitches form equilateral triangles.

### 2.3 The centre of effect

One of the main ideas behind the Spiral Array model is the use of points in the interior of the spiral to represent collections of pitches. More specifically, any pitch collection can generate a *centre of effect* (*c.e.*), a weighted sum of the pitch representations that is a point in the model's interior. This concept was demonstrated in the definition of the major triad and major key in the previous section. Generalizing the idea, any segment of music can also generate a centre of effect.

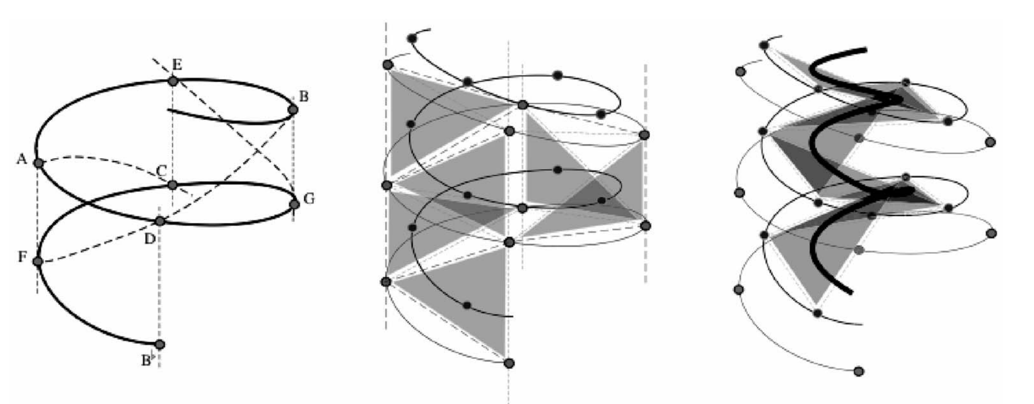
For example, each pitch in a musical segment can be mapped to the pitch spiral, and each pitch position be weighted by the pitch's proportional duration in the segment to generate the c.e. Suppose

$n_j$ = number of active pitches in the $j$-th time period, and

$\mathbf{p}_{i,j}$ = the Spiral Array position of the $i$-th pitch in the $j$-th time period,

then, the formal definition of the duration-based c.e. of the musical segment in the time interval $[a,b]$ is:

$$\mathrm{c}_{a,b} = \sum_{j=a}^{b}\sum_{i=1}^{n_j}\frac{\mathbf{p}_{i,j}}{N_{a,b}}, \quad \text{where } N_{a,b} = \sum_{j=a}^{b} n_j. \qquad (4)$$




(a) pitch spiral (b) major triad representations (c) major key representations

Fig. 1. The Spiral Array model.

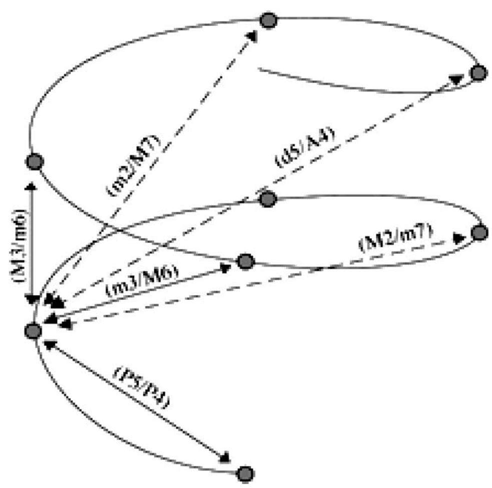



Fig. 2. Intervals in the Spiral Array.

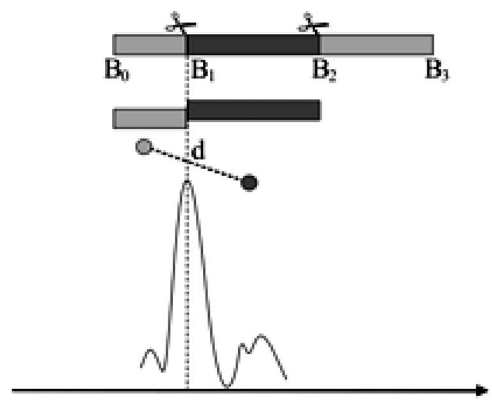



Fig. 3. Idea behind the segmentation algorithm.



Other definitions can be devised to emphasize the metric weight of the note. The naïve duration-based definition is the one used in the current implementation of the segmentation algorithm. This is not an unreasonable choice in the case of Messiaen, given the frequent metric changes in his *Vingt Regards*.

### 2.4 The Argus segmentation algorithm

The intuition behind the segmentation algorithm is that at a boundary point, the distance between the c.e.'s of the local section immediate preceding and succeeding the boundary is at a maximum. Consider the degenerate case when two adjacent segments have identical pitch sets in the same proportional durations. In this case, the distance between the c.e.'s of these two segments of music is zero. If the pitch sets are the same, but the proportional duration differences are small, as may be the case for two segments operating in similar ways in the same tonal context, this c.e. distance is small. As the difference between the two segments increase, the distance between their c.e.'s will also increase. Note that this distance is measured inside the Spiral Array space, and corresponds to tonal context change across the boundary between the two segments. At the point of transition between two pitch contexts, the c.e. distance reaches a local maximum as adjacent segments to either side will necessarily share more pitch material than adjacent segments at this boundary.

The idea behind the segmentation algorithm is illustrated in Figure 3. Suppose the piece consists of three sections that employ distinct pitch sets, with boundaries at $B_1$ and $B_2$ ($B_0$ and $B_3$ represent the beginning and end of the piece) as shown in Figure 3. At $B_1$, the distance between the c.e.'s of the preceding light gray section and the succeeding dark gray section (represented by discs of the corresponding colours) is maximized.

Sliding the double window across the entire piece produces a plot of $d$ over time that contains local maxima at the boundaries as depicted schematically in Figure 4, where the scissors indicate the actual boundaries of the piece. As can be seen in the diagram, the algorithm requires only one left-to-right scan of the piece and computes in $O(n)$ time.

### 2.5 When to take a peak seriously?

There exist many local maxima in the graph of the distance, $d$, between the c.e.'s of the segment pairs over time. One question remains: how can one separate the significant peaks in the distance values from the insignificant ones? Statistical process control theory offers a solution. One objective measure is to only select maxima that are above a given threshold, say, one standard deviation, $\sigma_d$, from the mean $d$ value, $\mu_d$. Assuming that the distances can be suitably approximated by a normal distribution, $N(\mu_d,\sigma_d)$, values more than two standard deviations above the mean have a statistical significance of approximately 97.5%. For a real-time realization of the algorithm, this threshold needs to be predetermined. For the purposes of evaluating the algorithm in this paper, the threshold is assigned after computing all the distances. An advantage of obtaining the distribution for $d$ from the data in this fashion is that the parameters so derived are specific to the piece being analysed.

## 3. Messiaen's *Regard IV*

We first apply the segmentation algorithm described in the earlier part of this paper to *Regard de la Vierge*, the fourth piece in Olivier Messian's (1908 – 1992) *Vingt Regards sur l'Enfant Jésus* (1944). Any discussion of the computational results first requires some ground truth to which to compare the algorithm's outcomes.

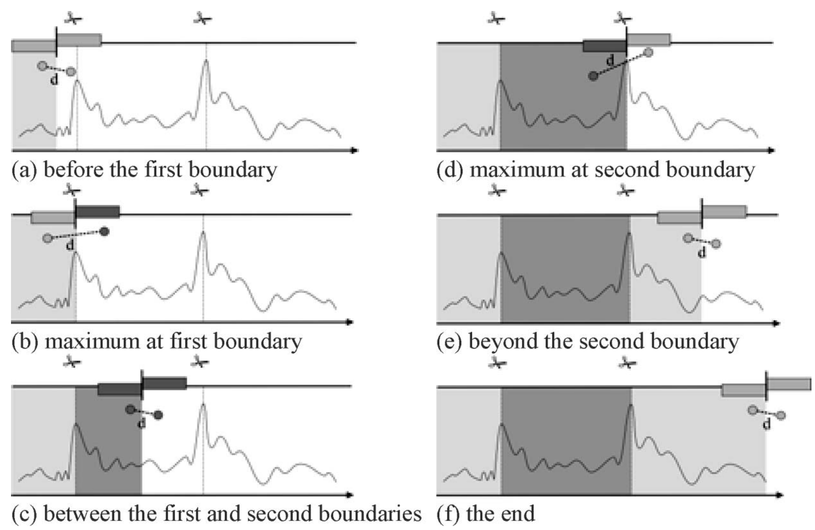



Fig. 4. Progression of segmentation algorithm.



In Section 3.1, a manual analysis of the piece is presented, showing the different sections. Then, the computational results using various forward and backward window sizes are compared to this manual segmentation in Section 3.2.

### 3.1 Manual segmentation of *Regard de la Vierge*

A manual analysis of Messiaen's *Regard de la Vierge* yields three main sections with defining motifs as shown in Figure 5. The first, call it Section A, presents phrases built on an asymmetric $6+7$ sixteenth-note-long motif shown in Figure 5. Section B contains the fluid triplet motif and a second more angular and wide-ranging six-chord sequence. Section C is the most energetic of the three and contains a rhythmic motif whose baseline turns into a second motif with stacked octaves. The square patches on the left show the grayscale tone used to represent that section. The demarcation between sections is clear from the score. In addition to the different motivic material, the composer also writes into the score tempo changes at the beginning of each of the identified sections.

The sequence of sections in the piece, ABACABAC, can be visualized using the bar in Figure 6. The numbers at the boundaries correspond to the number of sixteenth notes from the beginning of the piece. In this particular counting scheme, the sixteenth notes in the triplet motif in Section B are counted as full-valued sixteenth notes and all ornaments are grouped with the pitch set which they embellish.

Some variations occur in the return of each section in the second half of the piece. For example, the A sections in the second half of the piece combine material from Section A (main motif) and Section C (baseline of first motif) – see Figure 7(a). In the return of Section B, the triplet motif is reversed as shown in Figure 7(b). However, the note material is still primarily the same in each case, and the sections are recognizable as being similar to their earlier counterparts.

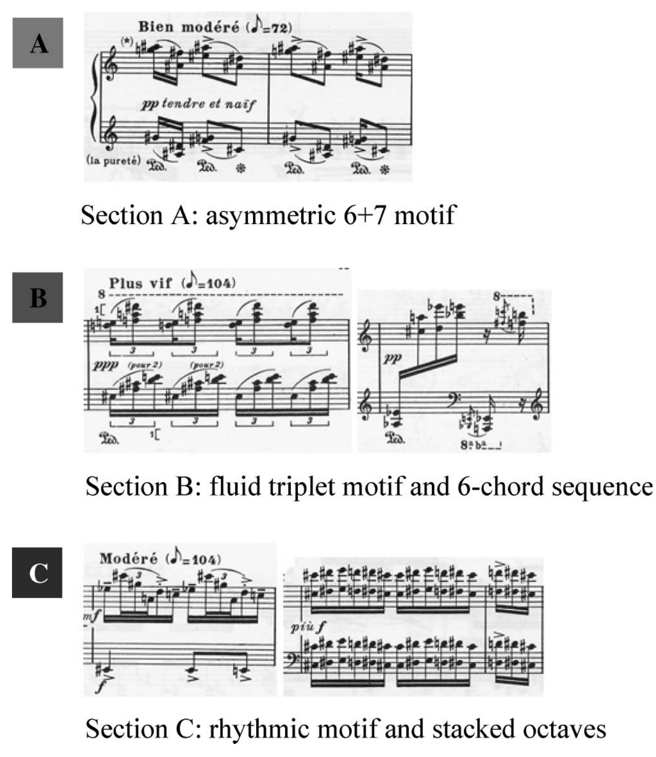
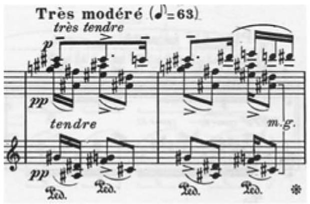



Fig. 5. Sections in the fourth Regard and their defining motifs.

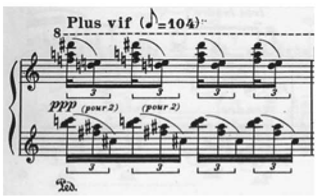
### 3.2 Automatic segmentation

Messiaen's *Regard de la Vierge* was encoded in text format so that at each sixteenth note instance, the names of all pitches present are known. As in the previous


Fig. 6. Manual segmentation of Messiaen's *Regard IV* into component section.




(a) combination of motivic material in A sections in second half of piece




(b) retrograde motif in return of Section B

Fig. 7. Variations in the return of the sections.



section, the sixteenth notes in the triplet figure in Section B's first motif are assumed to be full-valued sixteenth notes in this encoding, and all ornaments are grouped with the notes that they embellish. Note that an encoding that respects the tempo changes (one that would be closer to the performed timing) would be slightly different, but would be expected to produce similar results.

The segmentation algorithm was implemented in Java. Recall that the algorithm requires the user to specify the forward and back window sizes shown in Figure 8. For the analysis of Messiaen's *Regard IV*, we consider the algorithm's results when $f = b = 60$, 40, 20 and 10 sixteenth notes.

### 3.3 Results when $f = b = 60$ sixteenth notes

The segmentation boundaries are computed with forward and backward windows of size 60 (sixteenth notes). The resulting distance values are plotted over time and shown in Figure 9. The horizontal solid line cutting crossing the plot of $d$ marks the average $d$ value, $\mu_d$. The dotted horizontal lines correspond to one and two standard deviations above the mean, $\mu_d, + \sigma_d$ and $\mu_d + 2\sigma_d$, respectively. Maxima above one standard deviation are marked by vertical lines in the plot. Overlaid on the top part of the chart is the manual segmentation of the piece for comparison with the computed segmentation boundaries.

The average absolute discrepancy between the computed and actual boundaries is 8.43 sixteenth notes. Comparing this number with the number of sixteenth notes (893 in total), the boundaries are, on average, off by 0.94%.

Most of the discrepancies are caused by bridge material that either contains pitch collections of the upcoming sections or motivic material borrowed from other sections. Two examples are given in Figure 10, where the outlined arrows indicate the computed boundaries and the solid arrows indicate the actual boundaries. Figure 10(a) explains the discrepancy between the computed boundary at 196 and the actual boundary at 208. In this case, the final chords in Section B are re-spelt enharmonically (with sharps instead of flats) to prepare for the return of Section A. Figure 10(b) explains the largest discrepancy value (corresponding to the peak at 719), which occurs where the Section C motif is appended to the end of Section B.

### 3.4 Results when $f = b = 40$ sixteenth notes

Next, the algorithm is applied with forward and backward context windows that are 40 sixteenth notes wide. The results are documented in Figure 11. At $f = b = 40$, the plot of $d$ over time shows some ambiguity in determining the significant peaks. The three peaks inside Section C at times 385, 423 and 451 are higher than the one for the boundary at 662. In Section C, 385 and 451 correspond approximately to the beginning of the octave motif subsections (at 386 and 448); and, 423 corresponds to another smaller boundary between subsections. Note that the similarity between the ABA sections in the first and second half of the piece are now more apparent.

### 3.5 Results when $f = b = 20$ sixteenth notes

When the window sizes are reduced to 20 sixteenth notes, the peaks at the section boundaries still exist, however, the peaks representing the local boundaries within Section C have become more pronounced


Fig. 8. Parameters in the algorithm: the forward and backward window sizes, *f* and *b*.

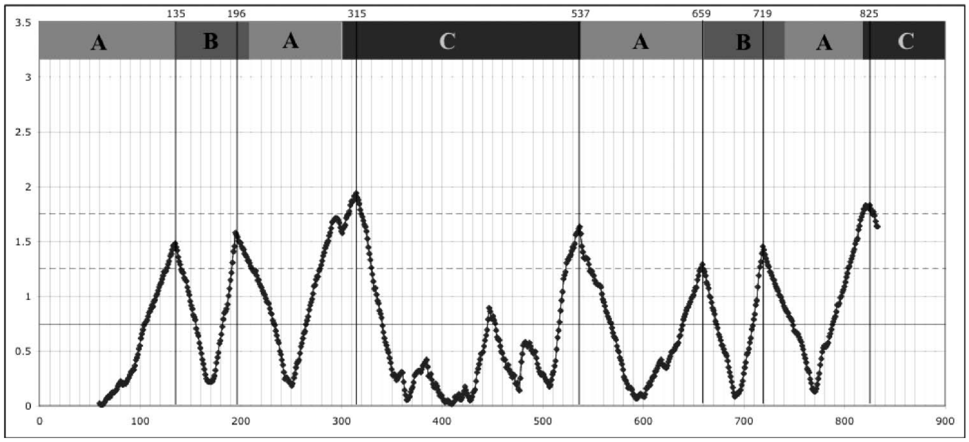



Fig. 9. Plot of *d* over time (in sixteenth notes), $f = b = 60$ [$\mu_d = 0.7418$, $\sigma_d = 0.5083$].


(a) enharmonic spelling of chord pitches in B in preparation for return of A

(b) largest discrepancy caused by motivic material from Section C appended to the end of Section B before the return of Section A

Fig. 10. Analysis of computed-actual boundary differences (computed, actual).

(see Figure 12). In the first Section C, the highest peaks (at 385 and 447) mark the beginnings of the octave motif (exact positions in the score are at 386 and 448). The beginning of the second motif in Section B is marked by a peak at 173, and a corresponding but smaller peak can be seen at 699. In addition to the more pronounced similarity between the two ABA sections, the repeated patterns in the more homogenous Section A are now visible as sequences of low-lying humps.

### 3.6 Results when $f = b = 10$ sixteenth notes

When the window sizes are reduced to 10 sixteenth notes, the local peaks within the large sections (especially C) dominate the picture (see Figure 13). The low-lying humps signifying repeated pitch patterns in Section A at window size 20 have now transformed into more defined patterns that indicate the number of repetitions of the main motif phrase. At window size 10, the beginning of the second motif in Section B is marked by a peak at 169, and a similar but smaller peak can be seen at 698. Section C is visibly the most segmented section of the three. In the first Section C, the other peaks {314, 331, 361, 385, 404, 423, 447, 477, 487} each correspond approximately to the appearance of a motif different from the immediately preceding one. The two dotted vertical lines (at 834 and 851) in the final Section C mark the boundaries of the tremolo bar (in the score at 836 and 855). As can be expected, local details rather than large-scale patterns determine the distance profile of the piece for smaller window sizes.

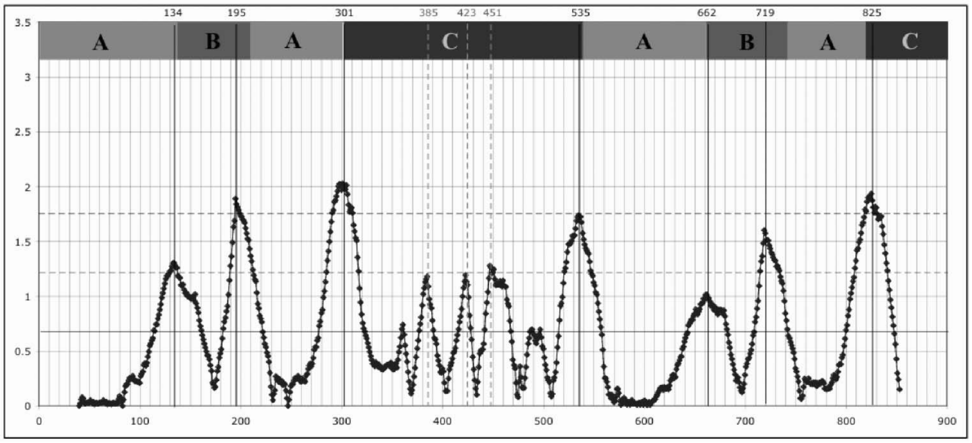



Fig. 11. Plot of $d$ over time, $f=b=40$ [$\mu_d=0.6821$, $\sigma_d=0.5343$].



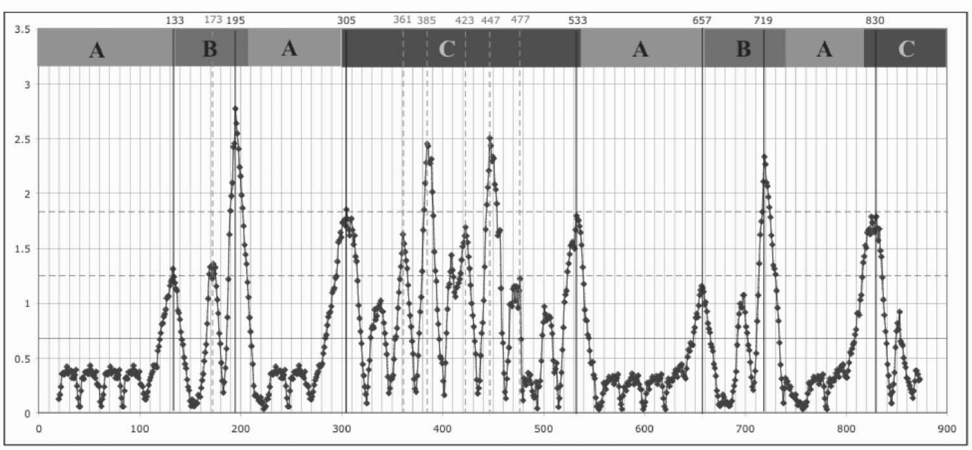



Fig. 12. Plot of $d$ over time, $f=b=20$ [$\mu_d=0.6832$, $\sigma_d=0.5753$].

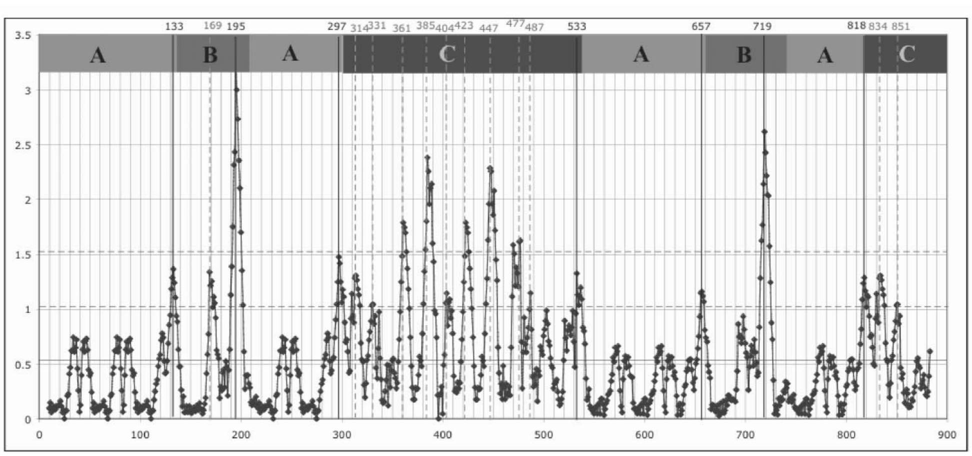



Fig. 13. Plot of $d$ over time, $f=b=10$ [$\mu_d=0.5380$, $\sigma_d=0.4930$].

## 4. Messiaen's *Regard XVI*

To further support the argument for the Argus segmentation algorithm, this section applies the algorithm to another example by Messiaen: the *Regard des prophètes, des bergers et des Mages*, the sixteenth piece in the *Vingt Regards*. Section 4.1 presents a manual analysis of the piece, and Sections 4.2 and 4.3 the computational analyses.

### 4.1 Manual segmentation of *Regard des prophètes, des bergers et des Mages*

Unlike *Regard IV*, the composer did not label all section boundaries in *Regard XVI*. He gives a clue to its structure in the note beneath the title: *Tam-tams et hautbois, concert énorme et nasillard*... The piece can be subdivided into the tam-tams (T), hautbois (H), and nasal concert (C) sections. The defining motifs of these sections are shown in Figure 14, and the sections (obtained by manual analysis) are shown in Figure 15. The numbers correspond to the thirty-second note count from the beginning of the piece.

The presentation of the tam-tams then the hautbois sections are followed by the enormous nasal concert, made up of the Section C motif shown in Figure 14, and joined later by the hautbois theme. Bridge material border the nasal concert section. Nearing the end, the tam-tams return, followed by a re-iteration of the H and C motifs superimposed. The two motifs share many of the same pitches, and are superimposed twice in the piece. The clear boxes with gray edges group the H and C motif sections. Together with the T sections, these boxes delineate the pitch-distinct sections in the piece.

T


Section T: tam-tams

H

Section H: hautbois

C

Section C: concert énorme et nasillard...

Fig. 14. Sections in the sixteenth Regard and their defining motifs.

### 4.2 Automatic segmentation

The text encoding of the piece represents note clusters at the thirty-second note level. All ornaments are grouped with the first thirty-second note slice of the notes they embellish, and the two triplets in bars 52 and 54 are snapped to neighbouring thirty-second note grids according to the proportion 3:2:3.

### 4.3 Results when $f = b = 256$ (32 quarter notes)

When forward and backward context windows are set at 256 thirty-second notes (that is to say, 32 quarter notes), the algorithm found the two major segmentation boundaries within its search range. As shown by the chart in Figure 16, the boundaries at 336 and 1384 were approximated by the two major peaks at 336 and 1380. Considering that the entire piece is 1842 thirty-second notes long, the boundary estimates are, on average, within 0.11% of the correct boundaries.

### 4.4 Results when $f = b = 128$ (16 quarter notes)

Observe in Figure 17 that, when the window sizes are reduced to 128 thirty-second notes (16 quarter notes), the main peaks above the two standard deviation line are {394, 1380 and 1714}, approximating the three boundaries at {336, 1384 and 1720}. Note that 1714 is at the rightmost edge of the search range and the actual peak could exist to the right of it. Because the period of the cycling pitch patterns in the tam-tams section (16 thirty-second notes) is a divisor of the window size, the plot shows a flat line at zero in the central part of the tam-tams section. The span of this flat line increases as the window size gets reduced to, say, 64 or 16.

## 5. Discussion: Why does it work?

It is reasonable to wonder why a geometric model for tonality such as the Spiral Array that is based on close positioning of pitch classes related by Perfect Fifth and Major Third intervals could be used to segment a post-tonal work, and why the segmentation algorithm worked as well as it did. This section contains reflections on the geometric interpretations of pitch context within the Spiral Array.

The Spiral Array model clusters entities that are perceptually close, which means that pitches in the same key form neat compact clusters in the Spiral Array space.


Fig. 15. Manual segmentation of Messiaen's *Regard XVI* into component sections.

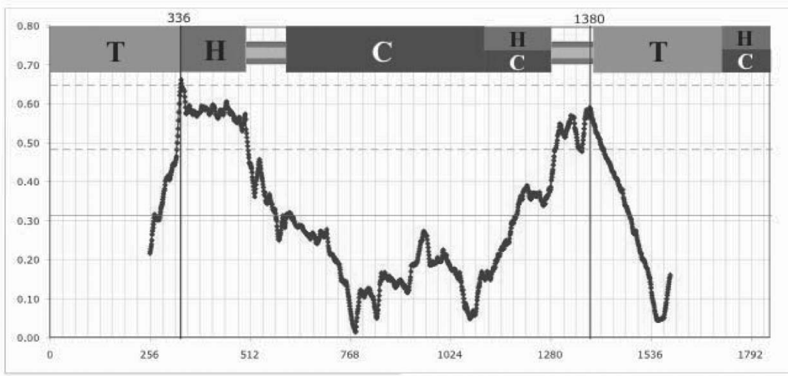



Fig. 16. Plot of $d$ over time, $f = b = 256$ [$\mu_d = 0.3122$, $\sigma_d = 0.1678$].

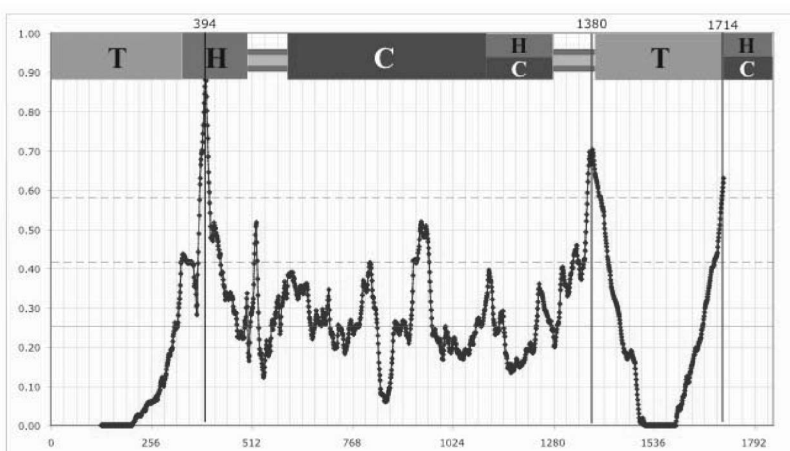



Fig. 17. Plot of $d$ over time, $f = b = 128$ [$\mu_d = 0.2534$, $\sigma_d = 0.1634$].

The c.e. generated by a segment of music in a given key resides inside the convex hull of the pitches of that key, that is to say, the smallest convex volume that contains all spatial representations of pitches in that key. When the context changes, for example when the key modulates, this convex hull shifts either to a neighbouring, or a distant, shape. These shapes are well defined in the tonal context.

What happens when one uses the same geometry for a post-tonal work such as Messiaen's *Vingt Regards*? Are there other, possibly more viable, structures for representing post-tonal pitch collections? Sections 5.1 and 5.2 will address these two questions.

### 5.1 Geometric shapes corresponding to pitch sets in Messiaen's examples

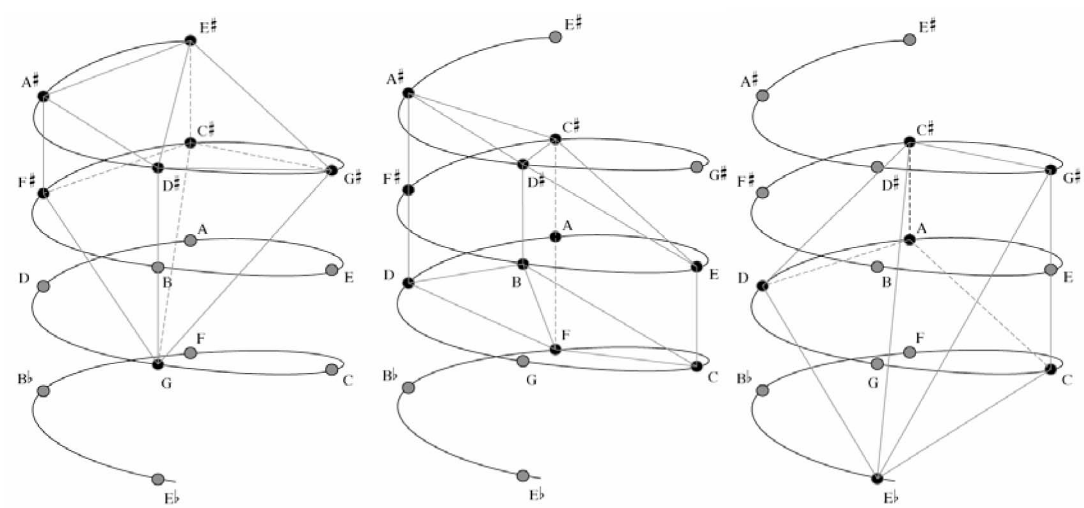
In the case of Messiaen's two *Regards*, the pitch contexts do not necessarily form neat compact shapes. However, they generate distinctive volumes such as those shown in Figures 18 and 19. The segmentation algorithm will work as long as the c.e.'s generated by the notes in each context as represented by these convex hulls are distinguishable (sufficiently distant) from each other. For example, the shapes for the three examples from *Regard IV* (shown in Figure 18) are distinct enough from each other such that c.e.'s generated within each of the three contexts (i.e., inside each of the three volumes) are readily separable one from the other, as evidenced by the automatic segmentation results.

The separation is not as obvious in the samples from Sections H and C of *Regard XVI* (shown in Figures 19 (b) and (c).) In this case, the two convex volumes overlap a fair bit, and the c.e.'s of the two sets are not as distinguishable. This observation is further confirmed by the fact that the composer superimposes the H and C motives on two occasions in the piece.

The position of the c.e. inside each shape is also influenced by the rhythmic structure of the sample. Two samples are distinguishable not only if the pitch sets are different, but also if the distributions of the pitch set are different. The convex hull of the pitch sets, appears to be the dominating factor affecting the success of the segmentation algorithm for these examples.

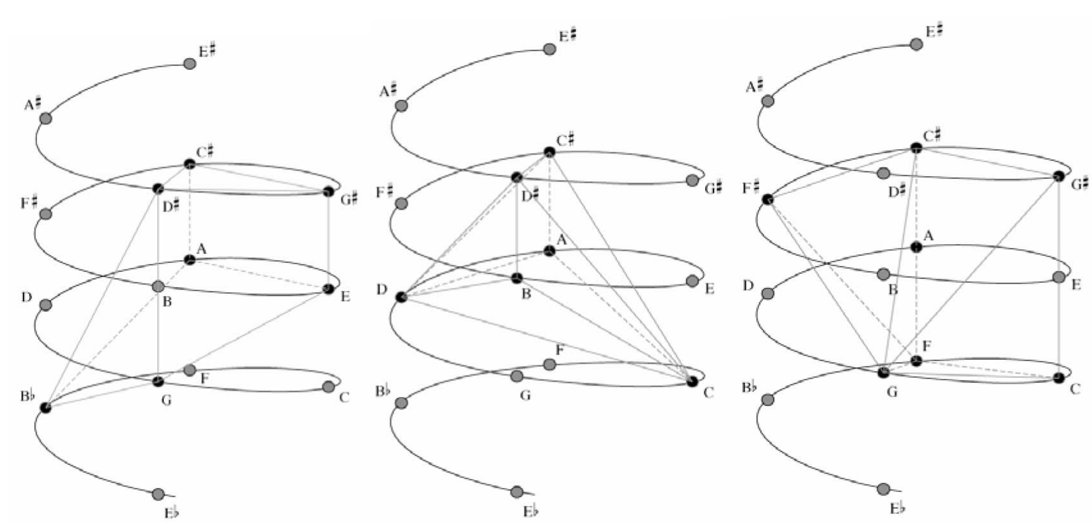
### 5.2 What are the structural alternatives?

The question arises as to whether the Spiral Array space is the best one to create the separate c.e.'s. Conceivably, one could construct a model with the best pitch set separation for segmenting a specific piece of music. It is unlikely that such a model tailored to one piece would generalize to many others. Since the Spiral Array exhibits




(a) Section A (b) Section B: triplet motif (c) Section C: rhythmic motif

Fig. 18. Convex hulls of pitch sets in sections of *Regard IV* shown in Figure 5.




(a) Section T (b) Section H (c) Section C

Fig. 19. Convex hulls of pitch sets in sections of *Regard XVI* shown in Figure 14.

strong correspondence to models rooted in theory and psychology, one might argue that it is a reasonable one to use to model perceived distances between pitch clusters designed to sound distinct one from another. However, readers may still wonder, as one reviewer did, if and what other structural representations for musical pitch might be more appropriate for segmenting the Messiaen examples by pitch content. I shall first paraphrase the reviewer's questions, then expand on some answers and explanations.

According to the one reviewer, one might argue that other non-tonal models may be more appropriate from a perceptual point of view when dealing with non-tonal music. From a practical computational point of view, the question remains open. What happens if one replaces the fifth-related pitches of the pitch spiral in the Spiral Array, for instance, (a) by pitches separated by semitones, (b) by pitches organized in thirds, (c) by random pitch name classes, and so on, and run the algorithm on the same pieces for the same parameter settings.

In answer to part (c) of the question, one can almost always find a more appropriate positioning of pitch classes to maximize the separation between pitch sets in different sections of a piece. Let us consider the extreme and simplistic case illustrated in Figure 20. In a piece that can be segmented into sections described by two pitch class sets, one strategy is to simply cluster pitch classes common to both section types (shown as black dots) in the centre of a space, and to place non-overlapping pitch class sets (shown as the gray set and the clear set) as far as possible in the given space. This structure would be more effective than the Spiral Array, and optimized for this particular example.

Next, we consider generalizable (positions fixed for all pieces) and regular pitch class representations, specifically, helical representations that exhibit a fixed interval between consecutive pitch class representations. The line of fifths is a one-dimensional structure that fits this description. Another constraint inherent in the Spiral Array that may not be immediately obvious is that, although the model assumes octave equivalence, enharmonic pitches are not grouped into the same class, i.e. Ab $\neq$ G#, hence the the helix is not folded back on itself into a torus. Pitch spelling is preserved so as to distinguish between diminished, minor, major, and augmented intervals and chords, and voice leading implications.

Let us now constrain the discussion to helical structures that cover all pitch class names (the line of fifths can be considered a special class of these structures, when the aspect ratio, height-increase-per-step/radius, goes to infinity). Then, assuming enharmonic equivalence, replacing the fifth-related pitches of the pitch spiral by pitches separated by intervals other than major or minor seconds, or perfect fifths would cover only a subset of all pitch classes, assuming enharmonic equivalence.

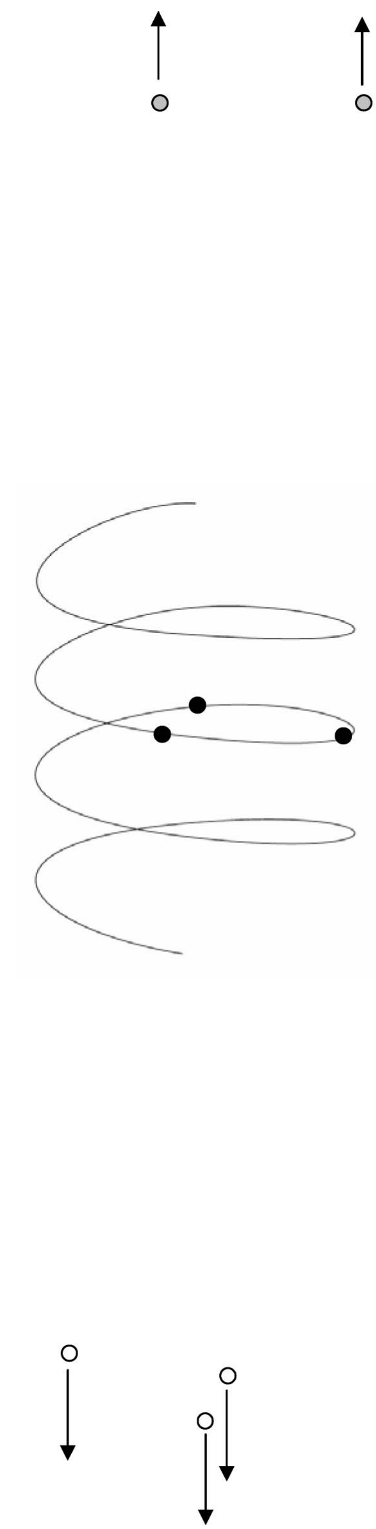


Fig. 20. Best separation of pitch class sets from two distinct sections.

Without enharmonic equivalence, stepping by perfect fifths is a necessary requirement to cover all pitch classes. As an illustration using the Spiral Array as the reference structure, stepping by major seconds and by major and minor thirds are shown in the diagram sequence in

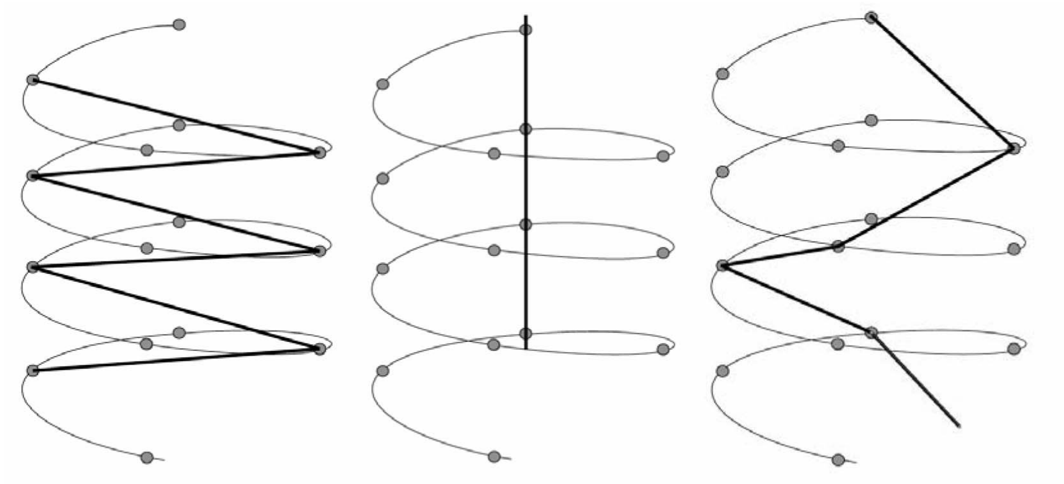



Fig. 21. Illustrations of pitch class sequences based on other intervals.



Figure 21. Assuming that an only thirds (or only seconds) representation covers all the pitch classes in the piece, the result would not differ from that using the Spiral Array. This answers questions (a) and (b).

There remains one more question that could be asked. Why not assume enharmonic equivalence and use the corresponding toroid structures? A reasonable question, especially with reference to post-tonal music. The argument can be further supported by Messiaen's use of his modes of limited transposition, scales that map back to the same pitch classes after only a small number of transpositions; the most "limited" being the whole tone and the octatonic scales, which repeat after only two transpositions. The extent to which Messiaen's pitch spelling is meaningful then comes into question, a topic for another paper. For now, I have assumed that the composer meant the spelling that he chose, and that each notated pitch is not lightly inter-changeable with its enharmonic equivalent. Without further experimentation, this assumption seems reasonable given that the composer, with few exceptions (and for these, only for a very limited time), uses the same pitch spellings whenever the same motivic material recurs.

## 6. Conclusion

In this paper, I have presented an algorithm for segmenting music according to pitch context based on a technique for tracking and assigning a distance measure to pitch context differences between consecutive segments of music, and applied it to the automated analysis of Messiaen's *Regards IV* and *XVI*. The computational results have shown that large search windows identify section boundaries and small search windows find local section and pattern boundaries. Some future work remains to identify the instantaneous optimal window size, if one exists, and to determine the best statistical model for the distance time series, and a robust method for setting the threshold for categorization of a peak as being significant. The use of dynamically varying window sizes may produce better results; however, studying the distance time series for each window size can also offer insights into the pitch context variations at multiple time scales. The method is also amenable to a dynamic programming formulation and implementation. However, an algorithm with a parsimonious description is preferred, and is the one described in this paper.

The strength of the algorithm lies in its use of the interior space in a 3D model to summarize pitch information, making the method highly efficient, and easy to visualize and explain. By using an interior point to summarize pitch context, one gains computational efficiency but loses much information, such as the convex hull of the pitch set, and the actual pitches themselves and the sequence in which they are sounded. In the case when a section is transposed, which did not occur in the Messiaen, the local patterns are preserved, and the relative distances within the section remain unchanged. In the case where the same thematic material is repeated in sequence, the second time some interval higher than the first, the Argus method could still detect a pitch context change, even though the pitch patterns remain unchanged. The Argus algorithm is designed to separate a piece by pitch context, and not by thematic material. In the Messiaen examples, the thematic material corresponded to distinct pitch contexts. Hence, the automatic segmentation not only separated sections with distinct pitch contexts, but also different thematic material.

The algorithm works well on the two Messiaen examples because the pitch collection in each section generates a c.e. that is sufficiently distant from that of its neighbouring sections. This is aided in part by the fact that pitch classes with the same letter name but differing by an accidental, although close in frequency, are far apart and have a distinctive span in the Spiral Array model. In general, the greater the separation of the c.e.'s of the distinct pitch sets, the better the accuracy of the Argus algorithm. Future investigations will include the comparison of models assuming enharmonic equivalence with the current one to determine the effect, or lack thereof, of pitch spelling and other arrangements of pitch representations (when assuming enharmonic equivalence) on the relative positions of the c.e.'s and the resulting segmentation boundaries.

## Acknowledgements

This paper grew out of a talk given at the MaMuX seminar organized by Moreno Andreatta on March 21–22, 2004, at IRCAM, whereby all participants were encouraged to concentrate on a few selected pieces from Messiaen's *Vingt Regards*. I thank Moreno for sparking my interest in analysing selections from Messiaen's *Vingt Regards*, and the attendees of the MaMuX seminar for their valuable feedback at that initial analysis of *Regard IV*. I also thank Gérard Assayag for his encouragement to submit the paper to the First International Conference on Sound and Music Computing, and the opportunity to expand on the paper in this special issue. The reviewers' detailed comments that have helped shape the current version of the paper are much appreciated.

This material is based upon work supported by the National Science Foundation under Grant No. 0347988. The work made use of the Integrated Media Systems Center Shared Facilities supported by the National Science Foundation under Cooperative Agreement No. EEC-9529152. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Science Foundation.



## References

Chew, E. (2000). Towards a mathematical model of tonality. PhD dissertation. Operations Research Center, MIT, Cambridge, MA.

Chew, E. (2001). Modelling tonality: applications to music cognition. In: Johanna D. Moore & Keith Stenning (Eds.), *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*. Edinburgh, UK.

Chew, E. (2002). The Spiral Array: An algorithm for determining key boundaries. In: C. Anagnostopoulou, M. Ferrand, A. Smaill (Eds.), *Music and Artificial Intelligence – Proceedings of the Second International Conference on Music and Artificial Intelligence*. Edinburgh, Scotland, UK: Springer LNCS/LNAI, Vol. 2445.

Chew, E. & Chen, Y.-C. (2005). Real-time pitch spelling using the Spiral Array. *Computer Music Journal*, *29*(2), 61–76.

Cohn, R. (1998). Introduction to neo-Riemannian theory: A survey and historical perspective, *Journal of Music Theory*, *42*(2), 167–180.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Lerdahl, F. (2000). *Tonal pitch space*. Oxford: Oxford University Press.

Longuet-Higgins, H.C. & Steedman, M.J. (1971). On interpreting Bach, *Machine Intelligence*, *6*.

Messiaen, O. (1944). *Vingt Regards sur l'Enfant Jésus*. Durand Editions Musicales: Paris, France.

Shmulevich, I., Yli-Harja, O. (2000). Localized key-finding: Algorithms and applications, *Music Perception*, *17*(4), 531–544.

Temperley, D. (1999). What's key for key? The Krumhansl–Schmuckler key-finding algorithm reconsidered, *Music Perception*, *17*(1), 65–100.

Toiviainen, P. & Krumhansl, C.L. (2003). Measuring and modelling real-time responses to music: The dynamics of tonality induction, *Perception*, *32*(6), 741–766.