THE COMPUTATIONAL EXTRACTION OF SPATIO-TEMPORAL FORMAL STRUCTURES IN THE INTERACTIVE DANCE WORK '22'

Vidyarani Dyaberi *Hari Sundaram Thanassis Rikakis* *Jodi James*

Arts Media and Engineering Program

Arizona State University

E-mail: {vidyarani.dyaberi, hari.sundaram, thanassis.rikakis, jodi.james}@asu.edu

ABSTRACT

In this paper we propose a framework for the computational extraction of spatial and time characteristics of a single choreographic work. Computational frameworks can aid in revealing non-salient compositional structures in modern dance. The computational extraction of such features allows for the creation of interactive works where the movement and the digital feedback (graphics, sound etc) are integrally connected at deep level of structures. It also facilitates a better understanding of the choreographic process. There are two key contributions in this paper: (a) a systematic analysis of the observable and non-salient aspects of solo dance form, (b) computational analysis of spatio-temporal phrasing structures guided by critical understanding of observable form. Our analysis results are excellent indicating the presence of rich, latent spatio-temporal organization in specific semi-improvisatory modern dance works that may provide rich structural material for interactivity.

1. INTRODUCTION

In this paper we focus on computational extraction of the choreographic structure in solo semi-improvisatory modern dance. The problem is important as it uncovers rich semantics of movement in modern dance communicated through non-salient middle level structures. These characteristics are often carefully interwoven, selected by the choreographer to provide a framework for the piece, to create meaning through movement phrasing and to create an aura of suspense or familiarity. This can significantly impact embodied multimodal content creation, such as interactive dance performance. Many current audio-visual generative frameworks do not exploit in real-time, some of the complex semi-improvisatory syntactical structures found in modern dance. Extraction and use of less obvious formal elements of movement can increase the organic relationships between the different modes and the computational elements of the work.

We analyze the formal properties of '22' (choreographed by Mr. Bill T. Jones) in two steps – (a) observable, qualitative aspects of time elements of the choreography, and (b) computational extraction of middle level structures in time. Our qualitative analysis reveals the basic structural elements of movement timing used by the dancer and suggests the presence of latent middle level temporal structures. Our computational analysis of speed into a shape, and time spent in a shape and during transition indicates that all three variables show rich temporal structures important to communicating form and associated semantics in movement. The speed shows a sub-phrasing structure for transitions that allows the piece to flow in a continuous manner. The shape times show a convincing two level phrasing structure with the higher level climaxing near the golden mean ratio point. The transition times also contain structured phrasing with a set of fixed climax points. The phrasing of the transitions is organized in counterpoint to the phrasing of shape times creating a balanced mix of variation and hidden repetition. Such mixes are found in many interesting art forms [3]. This kind of analysis has never been attempted for choreographic structures.

We are interested in determining the presence of relative spatial consistency to the movements across expositions and performances. Our approach is to predict the spatial location of the *next pose,* given the current estimates of location, speed, and direction of movement. The predictors are trained using prior training data (motion capture data), and then tested on the current sample using leave one out testing. The successful prediction of the location of the next pose using past training data would indicates that Bill T. Jones exhibits significant spatial consistency, a key ingredient to spatial form.

2. OBSERVABLE ELEMENTS OF FORM

'22' is an interactive, multimodal dance work. It is driven by movement created and performed by Bill T Jones. It is called "22" because the structural movement vocabulary consists of 22 shapes that are inspired by related cultural references (sculptures, movies, sports references, everyday shapes with globally recognized meaning, etc). In between individual shapes or sets or shapes Bill T Jones inserts improvisatory movement. Much of the shapes and movement are accompanied by verbal commentary by Bill T Jones some of which is fixed and some improvised. Some of the sections of the piece include just movement and verbal commentary. Some other sections also include interactive graphics and sound. Such sections involve a custom made engine for the creation, of interactive multimedia works. The engine is able to recognize the 22 shapes, variations or mixes of the shapes and differentiate all those from improvisatory vocabulary not related to the shapes. Visual (Paul Kaiser, Shelly Eskhar and Marc Downie) and sound artists (Roger Reynolds) use the recognition engine to drive visual and audio feedback that comments on, enhances, reinterprets or creates a dialogue with the movement and the story as expressed by BTJ.

Since "22" is semi-improvisatory and does not follow a set music score or strict choreographic score it *does not* exhibit extensive predetermined regularities in timing or spatial organization. The majority of the opening section of '22', that will be the focus of this paper, is not accompanied by music or sound. The form of that section is primarily determined by the movement choices of the solo performer. In the following paragraphs we will deconstruct the observable spatio-temporal form of the opening section of '22.' We will see how patterns of body movement over space and time help define and communicate the form of that section.

1-4244-0785-0/06/$20.00

2.1 OBSERVABLE CHARACTERISTICS

We now analyze the observable temporal aspects of form of the opening section of '22'. During the analysis we also introduce familiar dance terminology for consistent interpretation.

At the beginning of the choreography Bill T Jones is lying supine in the middle of the stage with his hands forming a circle above him. He remains frozen in that position for a certain amount of time (ranging generally from 2 to 5 seconds depending on the performance). For the purpose of this paper we will define such a static orientation or form of the body as a *Shape*. After holding the opening shape of the work, Mr. Jones slowly rises to standing and moves a small distance and then freezes into a different shape. Movement from one shape to another can be referred to for the purposes of this paper as a *transition*. Mr. Jones follows the second shape with another transition into a third shape. At this point, the viewer starts forming the expectation that shapes and transitions will be basic structural elements of the choreography. BTJ continues to perform different shapes each followed by a transition. This may invite the viewer to group shapes and transitions into pairs and thus form a higher level structural unit, a *sub phrase* (ref Figure 1).

A *Phrase* in dance can be a short sequence of dance steps or body movements linked together by transitions. Discernable events related to body movement can form the punctuation points for the beginning and end of a phrase. Phrases can vary in length depending on the intention of the choreographer and the meaning and significance of that particular phrase. Longer phrases may be comprised of *sub phrases*, shorter sequences of movements. This paper considers, phrases as a middle level feature of dance form.

Since the most discernable rest points in Bill T Jones' movement is the point where he assumes a new discernable shape, the viewer may attempt to consider the beginning of each new shape as the start of a sub phrase. We will see later that this at-first obvious choice of sub-phrasing is challenged by possible different types of sub-phrasing implied by the mover. This purposeful ambiguity at the level of sub phrasing leaves the viewer with shapes and transitions as the only solid observable structural units of body movement for marking time.

It is reasonable to expect that by the beginning of the fourth shape the viewer is beginning to create a custom semiology for following the form of this piece. We can thus also start defining the form semiology we will use in this paper. We will denote each shape by the letter p. We will denote the number of the shape by a subscript under the letter p. We will denote transition by the letters tr and call the transition between p_1 and p_2, tr_1, the transition between p_2 and p_3 tr_2 and so forth. We will denote sub-phrase by the letters s-ph. We will call the grouping of p_1+tr_1 s-ph_1, the grouping of p_2 and tr_2 s-ph_2 and so forth. This terminology has been used consistently in all the figures.

Bill T Jones continues the piece by going through a total of 22 shapes and transitions. After having completed the 22nd shape he performs a fairly long improvisatory section using movement vocabulary that is distinct from the 22 shapes. He thus creates a major delineation point. His improvisation concludes when he returns to the point where he started the first shape, lying down again and beginning again with the performance of the 22 shapes. By the time he repeats the first few shapes in the same order as before the viewer begins to create the expectation that he will repeat all 22 shapes in the same order. The viewer can thus group all 22 shapes and transitions into one higher level section that we will call exposition section 1 (e_1). BTJ performs the set of 22 shapes, in the same order as in e_1 four times in sequence. However each time there are differences creating thus four similar by yet different exposition sections that we will call e_1, e_2, e_3, e_4. These are the key differentiating characteristics of the four exposition sections:

- *First exposition section(e_1):* He performs all the 22 shapes. He is silent and there is no music.
- *Second exposition section(e_2):* He vocalizes each shape number as he performs the shape;
- *Third exposition section(e_3):* He vocalizes each shape number, and gives each shape a related text phrase or label.
- *Fourth exposition section(e_4):* His performance of this exposition section is accompanied by sounds and images that comments on, enhances, reinterpret or creates a dialogue with the movement.

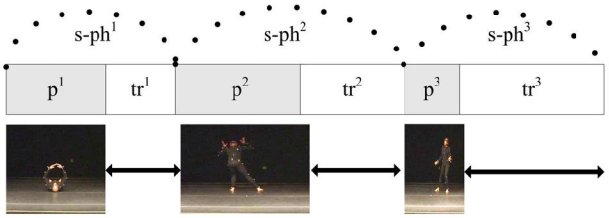

Figure 1: Each sub-phrase (s-ph^1, s-ph^2, s-ph^3) formed by grouping shape (p1,p2,p3) and transition (tr1,tr2 and tr3).

In the section that follows the four expositions (development section), Bill T Jones tells two intricately intertwined stories while also improvising and performing any of the 22 shapes. In this paper we are concentrating solely on the four expositions.

We now briefly discuss the observable aspects of spatial form in '22'. By the time the fourth exposition is performed most viewers are beginning to realize that each complete performance of the 22 poses forms a rough circle in the center part of the stage. Viewers also begin to realize that performance of certain poses might be associated with certain locations on that circle but that those associations are not repeated exactly.

3. NON OBSERVABLE FEATURES OF '22'

We will now discuss key issues raised by the observable analysis of the four exposition sections.

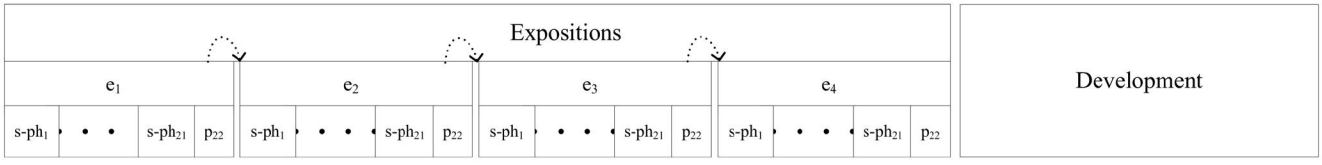

Figure 2: The different exposition sections e_1, e_2, e_3 and e_4, followed by the development.

- **Where are the middle level temporal Structures?**: At the conclusion of the last exposition section the viewer has observed lower level features of choreographic structure (shapes and transitions) and high-level features (exposition sections). *However, the viewer has not been able to observe any middle level formal temporal features.* Considering that each shape or transition is only few seconds long and each exposition section varies from 3 to 4 minutes length can it be that no formal elements connect the micro time structures of the piece (3 to 4 seconds) with the macro time structures of the piece (3 to 4 minutes long)?
- **Is there consistency to the spatial form?**: The people watching the piece agreed that the overall spatial form was roughly circular. Some of the more careful observers also suspected possible consistent correlations between performances of particular poses and spatial location. However neither the spatial form nor the spatial correlations could be defined by the observers with certainty or in detail even after multiple viewings.

Considering that in art communication of meaning relies heavily on form and that the exposition sections felt well organized, were highly communicative and emotionally powerful, we must conclude that there must be latent middle level formal features that exist in the four exposition sections.

In Figure 1 the lengths of *p*s, *tr*s and consequently *s-ph*s are different to denote that the durations of the different shapes and transitions are neither regular nor symmetrical. The viewer is aware that there is repetition of sequences of shapes and transitions but the viewer is also aware that those repetitions are not regular. This does not mean that there is no middle level structure controlling and driving those repetitions or that the viewer does not subconsciously detect that structure. It simply means the structure is not readily observable as it lies in deeper, non surface, levels of organization. The contributors that worked with Bill T Jones on the creation of 22, were aware of a convincing time structure underlying his timing of shapes and transitions but could not readily observe or predict that structure. We shall now use computational analysis to answer the key questions raised by critical analysis and uncover the hidden middle level time structures.

5. DATA ACQUISITION

In this section we describe the 3D marker data acquisition framework. In '22', the time when the dancer gets into a shape and when he gets out of the shape form the key points for extracting duration. Temporal data is of two types: (a) *Shape time*: The amount of time spent in each shape (b) *Transition time*: The time taken for transition from one shape to another shape. The ground truth segmentation of a piece into shapes and transition was performed by a professional dancer (one of the authors). There are total of 15 data sets (each corresponds to one exposition). These data sets (known as takes) were obtained during rehearsals done prior to the actual performance of '22' by Bill T Jones. The rehearsal is very noisy (both in temporal and spatial characteristics), making the computational analysis of the dance form highly challenging.

The data was acquired using the 3D Marker-Based Motion Analysis Corporation motion capture system. We used a twelve infra-red camera system with a capture frame rate of 120Hz. Forty-one 3D markers were placed on the dancer at specific locations, and were tracked. The captured marker data was then cleaned using a robust interpolation technique.

4. TEMPORAL ANALYSIS

In this section we use computational means to analyze the temporal aspects of form. The goal is to determine *middle-level temporal structures* that were not observable.

4.1 ANALYSIS OF SPEED

The rate of change of speed forms a key organization unit in '22'(ref Section 3). The speeds during the shape and during shape transition were calculated. When the speed is analyzed it is found that the duration of a shape is very small. When the dancer starts moving the speed increases reaches a maximum and then again decreases. The speed indicates a sub-phrasing arch-like structure during transition with the shapes acting as anchors on each end of the arch. The movement characteristics of these two structural units (p and tr) are very different. Thus the mover/choreographer might have decided to create separate phrasing structures for the shapes and separate ones for the transitions. Since the constituting elements of these two

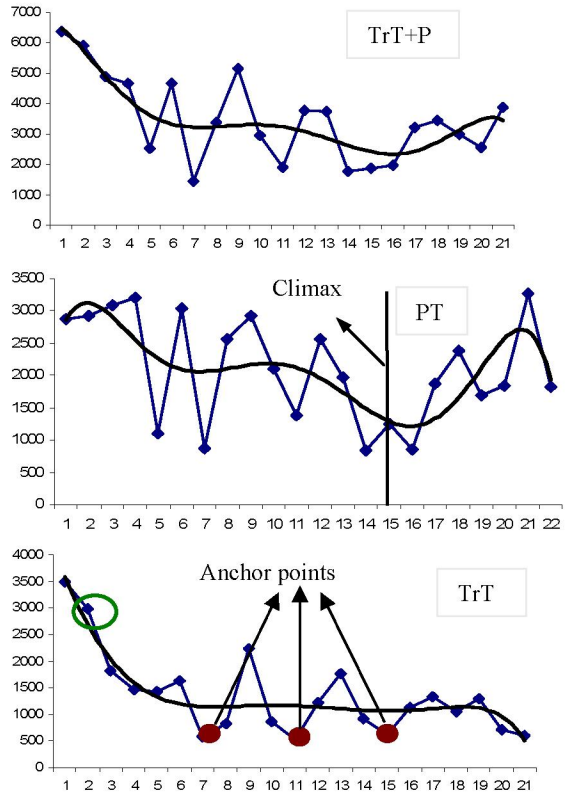

Figure 3: The trend lines of the transition time (TrT), shape time (PT) and total time (TrT+PT) for the 2nd exposition. The trend lines across the expositions are similar.

phrasings would not be seen sequential but in alternation, such phrasing structures would be hard for the viewer to detect. Hence we analyze shape and transition times separately.

4.2 MICRO-TIME ANALYSIS

In this section we look at the micro-time analysis that is the time spent in each shape (shape time) and the time between shapes (transition time). The analysis is done across the 15 data sets for each exposition separately.

In this subsection we discuss the micro-time analysis of shape time. It is found that the actual values of the shape time are different; however the trend of the shape time across different expositions is the same. The mean value of the shape time for each exposition from the 15 data sets is found and the trend is obtained by fitting a polynomial of degree two using least square fit to the actual shape time values. The trend of the shape time for first, second and the third exposition are similar. The trends of shape time show that the dancer is phrasing the sequence of shape durations at two levels. There is an overall arch form that covers each exposition section. The climax of that arch does not come at the middle of the section but near the golden mean ratio point of the duration of the section (approx. at 3/5s of duration). Using the golden section ratio when dealing with time organization of a piece in order to achieve a "natural" feeling is a technique that has been used consciously and/or unconsciously by many artists [2]. There are two to four (depending on the exposition section) smaller arch-like phrases embedded within the overall phrasing of the section. Figure 3 shows the trend lines of shape time for the second exposition.

The overall non-symmetrical arch phrase and the embedded smaller arches can be seen in the first, second and third exposition section but the phrasing changes in the fourth exposition. In the fourth exposition the dancer's movement is constrained by the multimodal feedback (ref Section 3.1). The dancer waits for feedback from the system, and has to synchronize his movement to the music and graphics feedback. The embedded smaller arches can still be seen in the fourth section but they are not as pronounced as they are constrained by some of the regularity dictated by the feedback functions. The last embedded arch with its large peak is dictated by the music. In expositions one, two and three where the dancer is free to move without any constraints, the sense of phrasing achieved just by movement means, is more visible.

In this subsection, we discuss the micro-time analysis of transition time. The analysis of transition time is also done across the 15 data sets . Unlike the shape time, transition times of different expositions not only have a similar trend but the minima also occur at the same place. The minima is found using a window. The mean value of transition time is plotted, and it is found that the minima occur at transition to shapes 8, 12 and 16. These minima can be seen as the climaxes of the smaller embedded arches of the overall phrasing structure of the transition times. Since they are fixed and repeated they can be considered to act as anchor points; the climax points, transition time wise, to which the dancer drives in each exposition section. Figure 3 shows the transition time plots for the second exposition and the anchor points. Additional figures showing similar trends for all exposition sections can be found in [1].

It is seen in Figure 3 that the anchor points (i.e. fixed climax points) exist for the phrasing of the transition times but not for the phrasing of the shape times. As discussed earlier, because shape and transitions alternate, this regularity in the phrasing of the transition times is hard to observe but it never the less provides a sense of hidden fixed structure. Figure 3 also shows that shape time phrases and transition time phrases do not have their peaks at the same points in the section. If shape times and transition times had the same phrasing and the same peaks the phrasing organization would be obvious and predictable. The piece would quickly lose its excitement. *We furthermore mention that the artists working with Bill T Jones could sense regularity in the phrasing but could not clearly detect it.*

Interestingly Figure 3 shows that the shape time plus the transition time (PT+TrT) is nearly flat. This indicates that temporal phrasings structures of the shape times and temporal phrasings structures of the transition times are in a continuous interplay where their peaks and valleys nearly cancel each-other out in the trend line. This interplay maintains variation (durations of shapes and transitions rise and drop continuously and asynchronously) but also regularity since *p* and *tr* times together remain fairly constant. However, since *p* and *tr* continuously alternate and vary this regularity is hard to detect.

5. SPATIAL ANALYSIS

In this section we discuss the computational analysis of spatial form of '22'. In we show that there is significant spatial consistency to the movement across expositions. Experts watching the dance could not identify with accuracy some spatial consistency characteristics. However, *if we are able to successfully predict the location of the next shape using past training data, then we can show that Bill T. Jones exhibits significant spatial consistency, a key ingredient in spatial form.* Our approach is to predict the spatial location of the *next shape,* given the current estimates of location, speed, and direction of movement. The predictors are trained using prior training data (motion capture data), and then tested on the current sample using leave one out testing. In order to estimate the next shape location, we developed models for– (a) transition time between shapes, (b) direction of movement and (c) speed of movement.

5.1 DENSITY ESTIMATION OF TIME

In this subsection, we present our approach to estimate pose time and transition time densities. The estimates are determined separately for each of the four expositions as each exposition involves a different set of communication modes. The aim is to approximate the probability density function of the time data, and determine the parameters of the distribution that fit the data best. Rayleigh distribution [4] is used to find parameters of transition time. The transition times between poses are variable, and can exhibit significant changes. For example, the dancer may decide to slow down the performance during the entire exposition in a consistent manner. This can be accounted for in real time as the overall trend and the relative times remain consistent. We use Linear predictive coding (LPC) [5] to make adjustments to the time estimates. LPC is used to correct for the errors between the estimated transition times from the training data and the actual observed transition time during the performance.

5.2 ESTIMATING DIRECTION OF MOVEMENT

In this section we discuss the estimation of direction. Across expositions and performances, the exact angle of the dancer between any two poses will not be the same. The direction is

found as an angle made with the horizontal between two poses. The number of modes in distribution of direction will help us know the number of directions of movement. The number of modes are variable, and are found by fitting Gaussian mixture models (GMM) using EM algorithm [5] and using minimum description length (MDL). We use a running estimate of direction and check if the dancer is in the last mode from the GMM. If the estimate lies in the last mode in temporal order, then we begin the prediction of the next pose location.

5.3 ESTIMATION OF SPEED

The ground truth estimates of speed of the seven motion capture data sets during pose and transition, suggest a parabolic fit for the speed during transition. From our prior work in pose recognition we know when a pose begins and ends. The speed is estimated from that instant of time when the dancer leaves the pose. The mean and variance of the speed during transition estimated from training data, used to find when the dancer begins to move from his pose. A quadratic fit is used to predict the speed. The estimated speed $\hat{s}_t$ at time t is found as:

$$s_e(t) = c_0 + c_1 t + c_2 t^2, \qquad <1>$$

where the constants c_0 is the constant speed, c_1 is the rate of change of speed and c_2 is the rate of change of acceleration. These constants are found using a simple least square fit. Our prediction accuracy for the speed improves significantly after the peak in the speed. At this point in time, we can be confident of the spatial location of the next pose.

The estimates of the expected transition time, direction of movement and speed allow of us to calculate the location of the next pose in a straightforward manner.

6. EXPERIMENTAL RESULTS

In this section we discuss the experiments where the pose position of the dancer is predicted during the performance using the estimates of time, speed, direction. The temporal data set was obtained from the videos and the spatial data set was obtained from 3D motion captured system, for each of the four expositions. There are 15 data sets for time (i.e. 15 videos), where as for the spatial data there are 7 data sets (seven motion capture data sets). The data was scarce since the performer would frequently stop midway, as we were capturing the rehearsal, and only rarely did he perform the complete expositions. The multimedia exposition is treated differently from the others since the multimodal feedback introduces additional constraints in the dancer movement.

We developed a simple figure of merit to evaluate the system. For estimating the accuracy of the system we determined the distance of the estimated pose location to the observed pose location, *after the dancer reached the next pose*.

The graph in Figure 4 shows the average d / σ of the four data sets of the first, second and third exposition. The results obtained using LPC and quadratic fit are significantly better compared to case of linear speed (by an order of magnitude). However, while LPC does improve the results, the differences are not substantial from using quadratic speed without LPC.

Our results indicate that we can model the spatial accuracy of the dancer per pose, within one standard deviation for most poses, provided we use a quadratic fit for speed. These are excellent results thus justifying the conjecture that there exists

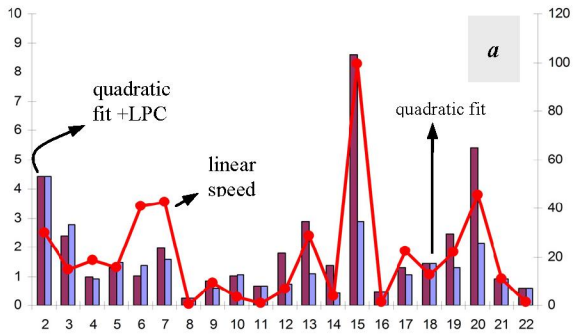

Figure 4: Figure (a) shows the plot of the average d/σ figure of merit per pose, for three different estimates of speed. This averaged over the first three expositions. The red line shows the estimate using linear estimates (right axis). The two bars indicate quadratic estimates to speed with LPC (brown bar) and without LPC (blue bar).

tremendous spatial consistency in the form. Note that since the estimates of spatial location are *relative to the previous pose*, this implies that the spatial consistency is relative. When the dancer changes the location of pose 1, the relationships amongst the pose locations are maintained. This determination of spatial consistency is hard for the critics and audience members, as they would need to create highly accurate spatial maps for each expositions to make a comparison.

7. CONCLUSION

In this paper, we focused on the computational extraction of form in '22', a new multimodal, interactive dance work choreographed by Bill T. Jones. The framework involved the following: (a) analysis of the observable movement form, and its use in guiding the extraction of computational form. (b) Detection of middle level temporal structures of movement form (c) determining the consistency of spatial organization of movement. The experimental results are excellent. In future work we plan to focus on the development section of the composition as that contain interesting semi-structured characteristics.

8. REFERENCES

[1] V. M. DYABERI, H. SUNDARAM, T. RIKAKIS and J. JAMES (2006). *The Computational Extraction Of Spatio-Temporal Formal Structures in the Interactive Dance Work '22'*. Arts Media and Engineering, Arizona State University, AME-TR-2006-09, Summer 2006.

[2] E. LENDVAI (1971). Bela Bartok: An Analysis of His Music. London: Kahn & Avrill.

[3] L. B. MEYER (1956). Emotion and meaning in music. University of Chicago Press Chicago.

[4] A. PAPOULIS (1991). Probability, random variables, and stochastic processes. McGraw-Hill 0070484775 New York.

[5] L. R. RABINER and B. H. JUANG (1993). Fundamentals of speech recognition. Prentice Hall 0130151572 Englewood Cliffs, N.J.