Analysis and Interpretation of Music for Dance

Daniel P. Wright

March 27th, 2007

Abstract

The automatic analysis and interpretation of music for dance is a multidisciplinary problem incorporating ideas from the fields of musicology, music information retrieval, artificial intelligence and computational linguistics, as well as requiring an understanding of the dance being simulated. This project aims to collect research from each of these fields, and consider how they could be used to create an automated choreography system, targeting either a paper choreography to be followed by live dancers, or a computer animation of the dance generated using motion capture technology.

1 Introduction

This project aims to explore methods by which a dance could be choreographed automatically, in response to music. The dance that has been chosen for study in this project is Argentine Tango, both for its improvised and structured nature. This is a multi-faceted problem which will need to draw input from a number of disciplines.

- The form of the dance itself
- Musicology
- Music Information Retrieval (MIR)
- Natural Language Processing (NLP)
- Visualisation

Current research in each of these areas was surveyed, compared, and summarised, with reference to the overall goal of building a choreography system. After this, a plan was drawn up describing how they may be used together to create the system.

2 Disciplines explored

2.1 Musicality in Tango

Argentine Tango is an improvised dance form led by the male dancer in response to properties within the music being played. There are many elements within the music that affect which figures the dancer will choose to lead, or which ornaments either dancer will use to lend the dance character. *La Cumparsita* was chosen for study; as the most famous Tango piece internationally, it is widely agreed to represent the character of traditional Tango music. Videos were consulted with reference to the sheet music, and a tool was developed called LISA to annotate audio recordings of the piece.

2.1.1 LISA: LISten and Annotate

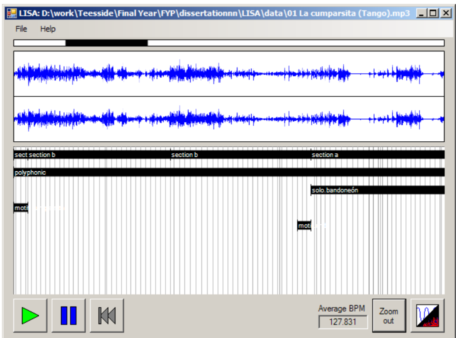


Figure 1: LISA: LISten and Annotate

LISA is an analytical tool which allows annotations to be made across a piece of music. It was intended for preliminary analysis of the music, and was useful for tasks like dividing the piece up into sections and noting the use of cadences and motifs. LISA contains a number of features aimed at facilitating manual annotation of audio files:

- Waveform Display
- Spectral Analysis
- Beat Detection
- Alternative forms of annotation
- Multiple layers
- Zoom
- XML-based file format

2.1.2 The musical framework

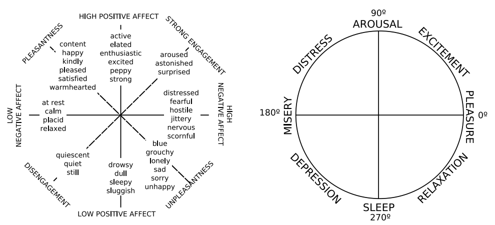
With the help of LISA, a number of musical structures were identified which limit the movements a dancer will choose between.

Tempo is clearly of vital importance in simulating any form of dance. A dancer moving out of time with the music will be immediately noticeable, even to those not trained in music or dance.

There are five main forms of Tango dance, these being *Tango*, *Milonga*, *Vals*, *Nuevo* and *Contemporáneo*. The key musical feature differentiating these forms is time signature.

Much of the complexity of Tango is the result of the layered rhythms it contains. If the bandoneón is playing a different rhythm from the violin, for example, the dancer may choose to dance to either.

There are in Tango a number of clear sections, and the boundaries between these sections offer convenient moments for the dancer to change their style, move into a new figure, or strike a dramatic pose.



(a) (b)

Figure 2: Circular mood models.
Figure 2(a) taken from [1]

2.1.3 The emotional performance

After the choice of movements has been limited, the specific movements which should be made are primarily chosen in response to the mood of the music at that moment. This presents the system with two problems: how to define mood, and how to draw it from the music. A number of models of mood were investigated. Including the two circular models in figure 2

The most interesting mood model which was considered was Thayer's model, as utilised by Liu et. al[2]. They identified that properties of "Energy" and "Tension" within the music could be measured by monitoring intensity, timbre, and rhythm. Figure 3 shows how adjective clusters line up with these properties.

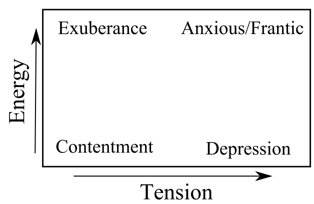


Figure 3: Thayer's model of mood[2].

Alternative adjectives were suggested for this project which are more in line with terminology used to describe Tango; these were: *Relaxing*, *Playful*, *Melancholy*, and *Dramatic*. It was hypothesised that a similar model could be constructed to describe moods in dance, by replacing "Energy" and "Tension" with "Sensitivity" and "Passion".

2.2 Music Information Retrieval

Having identified key features within music affecting the way it should be danced, Music Information Retrieval (MIR) techniques were explored which could extract these features automatically from raw audio data.

2.2.1 Rhythm, beat, and metre

A number of techniques were explored to calculate the tempo and metre of music. The simplest way to discern rhythm is to track moments of high intensity in the waveform. This has the advantage of being very simple to implement, but can be error-prone. In moments of low intensity variance, beats may not be detected, while when the variance is high individual notes might incorrectly be detected as beats.

Other systems move into a symbolic format such as MIDI first, before using Inter-Onset Intervals (IOIs) to determine more detailed information about beat and metre. This is not ideal, as the transition from raw audio to MIDI is non-trivial, and errors at this stage might propagate to the rhythm-detection stage.

An attempt has been made with beat and metre tracking on raw audio with traditional Greek dance music [3]. The limitations this imposes on possible time signatures are reasonable for application in Tango, which also uses quite a limited range of time signatures.

2.2.2 Timbre and intensity

Timbre and intensity can be used alongside mode and rhythm to determine the mood of a piece of music. Along with rhythm, they can be mapped onto Thayer's model of mood [2].

Unlike many other properties of sound, such as pitch, intensity, and duration, timbre is not strictly defined or well understood. Nevertheless, many of these components have been shown to be determined by the spectral information in different sub-bands [2, 4]. These characteristics are very important for mood detection, for example a high centroid is more likely to be part of exuberant/playful music than depressing/melancholy music, due to its higher pitch. A high roll-off might imply less energy, or greater sensitivity. Timbral characteristics might also be used as part of an instrument recognition module [5].

Intensity "is approximated by the signal's root mean-square (RMS) level in decibels"[6]. This calculation can be performed over a series of windowed frames throughout the piece, so that the intensity varies across the length of the song. It would also be possible to calculate the intensity at each sub-band, so that sections with loud bass lines (likely being played by the piano and double-bass), and those where the higher frequencies are louder (which are most likely to be the bandoneón, though possibly also the violin) can be differentiated.

2.3 Artificial Intelligence

A common approach to the structuring and analysis of music for AI processes is to make comparisons with structures developed for Natural Language Processing (NLP). It is widely agreed that music follows syntactic and grammatical principles. The question of whether music can contain semantics is not quite so clearly answered. In the strictly literal sense, a musical expression cannot be said to be true or false [7]. However, the majority of music attempts to invoke at least some form of emotive response, which is a semantic, rather than syntactic, property.

2.3.1 Syntactic tree structures

One of the key stages of NLP is syntactic processing, wherein sentence components are organised into an hierarchical structure. Figure 4 demonstrates one way music might be put into a tree structure.

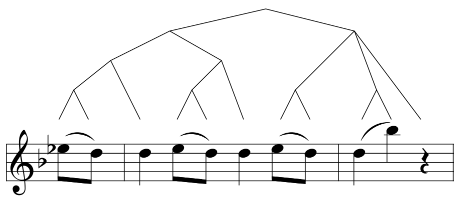


Figure 4: Hierarchical structure for music[8].

2.3.2 Semantic understanding

A common problem in NLP which is worsened when using these techniques for music is that of *ambiguity*. Grammars can help reduce this problem, but ambiguities still arise. One way of clarifying some of them is to use semantic and pragmatic analyses. The former concerns the meanings of individual words, while the latter uses context to discard parse trees which are impossible.

How far these ideas can be transported across to music is debatable. Wiggins warns against trying, and suggests use of "connotations" instead [9]. These would describe the emotional affect the music can engender in the listener.

2.4 Visual Representation

2.4.1 Paper choreography

Three choreography systems were explored: Benesh Movement Notation, Labanotation, and a Tango-specific notation [10]. It was found that Benesh and Labanotation are very powerful, but quite cumbersome if all that needs to be expressed is the steps that must be taken. In those cases, the Tango-specific notation is more appropriate. However, if more detail about *how* the steps should be taken is desired, a more complex system such as Benesh or Labanotation would have to be used.

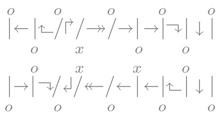


Figure 5: Bodirsky's Tango Notation[10]

2.4.2 Computer animation

Motion capture techniques were explored, as well as ways in which the animations could be divided up and later combined to form new figures. It was decided that dividing animations to a granularity of one step is likely to be optimal. Animating to coarser levels of detail, by storing entire figures as single animations, is inflexible and the number of requisite animations would quickly build up. Care would have to be taken, however, to ensure that the animations crossfade correctly and remain in time with the music.

3 Conclusions

This project has defined many of the disciplines that would be involved in creating a system to choreograph dance performances automatically. It has investigated current research in these fields, and compared a variety of techniques.

The techniques mentioned in this document are enough to start working on the system itself. The first step should be an improvement in the beat detection algorithm used in LISA. New forms of visualisation, such as "TimbreGrams"[11], would help understand what is happening in the music. The generation of Inter-Onset Intervals (IOIs) should be explored, from which preliminary hierarchical structures describing the music could be formed.

References

[1] Auke Tellegen and David Watson. On the dimensional and hierarchical structure of affect. *Psychological Science*, 10(4):297, June 1999.

[2] Dan Liu, Lie Lu, and HongJiang Zhang. Automatic mood detection from acoustic music data. In *ISMIR*, 2003.

[3] Aggelos Pikrakis, Iasonas Antonopoulos, and Sergios Theodoridis. Music meter and tempo tracking from raw polyphonic audio. In *ISMIR*, 2004.

[4] T. Zhang and C.-C. J. Kuo. Hierarchical system for content-based audio classification and retrieval. In C.-C. J. Kuo, S.-F. Chang, and S. Panchanathan, editors, *Proc. SPIE Vol. 3527, p. 398-409, Multimedia Storage and Archiving Systems III, C.-C. Jay Kuo; Shih-Fu Chang; Sethuraman Panchanathan; Eds.*, pages 398–409, October 1998.

[5] G. Agostini, M. Longari, and E. Pollastri. Musical instrument timbres classification with spectral features. *EURASIP Journal on Applied Signal Processing*, 2003(1):5–14, 2003.

[6] Erling Wold, Thom Blum, Douglas Keislar, and James Wheaton. Content-based classification, search, and retrieval of audio. *IEEE MultiMedia*, 3(3):27–36, 1996.

[7] J. McCormack. Grammar based music composition. *Complex Systems*, 1996.

[8] Fred Lerdahl. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, 1996.

[9] G. A. Wiggins. Music, syntax, and the meaning of "meaning". In *Proceedings of the First Symposium on Music and Computers*, pages 18–23, Corfu, Greece, 1998.

[10] Manuel Bodirsky. A notation system for tango. http://www2.informatik.hu-berlin.de/ bodirsky/tango/index.html, 2004.

[11] G. Tzanetakis and P. Cook. Audio Information Retrieval (AIR) Tools. *Proc. Int. Symposium on Music Information Retrieval (ISMIR), Plymouth, MA*, 2000.