THE UNIVERSITY OF TEESSIDE
SCHOOL OF COMPUTING
MIDDLESBROUGH
TEES VALLEY TS1 3BA

# ANALYSIS AND INTERPRETATION OF MUSIC FOR DANCE

BSc Computer Games Programming
March 27, 2007

Daniel P. Wright
Supervisor: Fred Charles
Second Reader: Julie Turnell

THE UNIVERSITY OF TEESSIDE
SCHOOL OF COMPUTING
MIDDLESBROUGH
TEES VALLEY TS1 3BA

ANALYSIS AND INTERPRETATION OF MUSIC FOR DANCE
BSc Computer Games Programming
March 27, 2007

Daniel P. Wright
Supervisor: Fred Charles
Second Reader: Julie Turnell

**Abstract**

The automatic analysis and interpretation of music for dance is a multi-disciplinary 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.

## Dedication

*This project is dedicated to the memory of my Grandfather, Lindsay Steele. Music and language were his two great passions, and I am sure he would have been fascinated by the relationships that can be drawn between them.*

# Acknowledgments

First and foremost, I would like to thank my project supervisor, **Fred Charles**, for his guidance, for the interest he displayed in my work, and for late-night exchanges of research papers and ramblings.

I also owe the following people my thanks:

**My parents**, for continuing to support me throughout University, for having faith in me, and for always offering invaluable advice.

**Liliana Tolomei**, **Carlos "El Tordo"**, and **Ahmet Ersoz**, for introducing me to dance, the topic of this report; **Mark Llewellyn**, for hours discussing the intricacies of dancing Argentine Tango; and **Franckie Jeandenant**, for accepting bribes.

Finally, to my crack team of distributed proofreaders: **Alex Caithness**; **Steve Lee**; **Kevin Williams**; and **Carolyn Wright**. Thank you!

# Contents

# List of Tables



# List of Figures

# 1 Introduction

This project aims to explore methods by which a dance could be choreographed automatically, in response to music. This is a multi-faceted problem which will need to draw input from a number of disciplines.

- **The form of the dance itself**: There exist a wealth of different dance forms, each of which have expression of movement and flexibility in varying degrees. Some involve little more than stepping in sequence to the beat of the music; others are entirely freestyle. Understanding the range of movements available to the dancers, and how much freedom they have to modify these movements in order to lend the dance a particular feeling or emotion, will be vital in attempting to make decisions about the choreography. In addition, each dance form complements a particular style of music. A solid understanding of how that music inspires a dance is essential in order to understand how the dance is formed. Argentine Tango has been chosen for study in this project. As an improvised dance form, Tango offers great scope for making interesting and varied decisions about how it should be danced. At the same time, movement in Tango is highly structured, which makes it very suitable for machine processing.
- **Musicology** - In order to make decisions based on the music, it must be put into a structured form. Ways of structuring music and understanding it have been investigated in the field of musicology, and will be considered as part of this report.
- **Music Information Retrieval (MIR)**: Before the question of how the dance should be choreographed can be considered, it must be determined what information can be drawn from the music. MIR is a very active field of research, so there are many resources to draw from. Bearing in mind the properties of Tango music as analysed in the previous sections, existing methods of extracting this information from raw music data will be surveyed and considered.
- **Natural Language Processing (NLP)**: Extracting music data from the audio file and structuring it will produce a set of discrete units of information, which must then be tied together into a form from which decisions can be made. This process of taking information at a high level of granularity and linking it together to create an overall meaning is similar to that which occurs in NLP, where the smallest units are words, which form sentences, paragraphs, or passages.

- **Visualisation**: Once all the decisions have been made and the dance has been choreographed, it must be visualised. Visualisation can either take the form of a paper choreography or a computer animation. Both possibilities will be explored as part of this document.

This report will survey existing research in these areas and consider how this could be used to build an automatic choreography system. Each topic will be considered individually, with reference to the overall goal of building a choreography system, after which they will be summarised and a plan will be drawn up describing how they may be used together to create the system.

# 2 Musicality in Argentine 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*(Rodríguez, 1917) was chosen for study; as the most famous Tango piece internationally(Collier et al., 1997), it is widely agreed to represent the character of traditional Tango music. Videos showing different couples dancing to various interpretations of *La Cumparsita* were compared with reference to the sheet music. It was discovered that there were elements within the recordings themselves, that were not or could not be marked on the manuscript, which affected the dance. These elements are considered separately, in section 2.3. A transcription of the manuscript itself can be found in Appendix A, and the URLs of the videos compared are listed in Appendix B.

## 2.1 LISA: LISten and Annotate

To help identify signals within the music, a prototype tool was developed named LISA. LISA is an analytical tool which allows annotations to be made across a piece of music. Its principal features are as follows:

- **Waveform Display** - An amplitude/time graph for the waveform. This provides a frame of reference to help the user determine at which point in time a particular event occurred.
- **Spectral Analysis** - As an alternative to the waveform display, the spectral frequency analysis displays the amplitudes for each frequency sub-band over time. Amplitudes are represented by colour, while frequency is plotted on the $y$-axis and time is plotted on the $x$-axis. This is useful for identifying the density of musical events at different frequency sub-bands.
- **Beat Detection** - Rudimentary beat detection was added when it was found that lining annotations up with specific moments in the music was difficult and frustrating. Beats are marked by vertical grey lines in the annotations frame. Darker lines represent stronger beats. A "snap-to-beat" feature was included, so annotations would line up with the nearest beat. The algorithm used was similar to that described in section 3.1.1 (Patin, 2003).

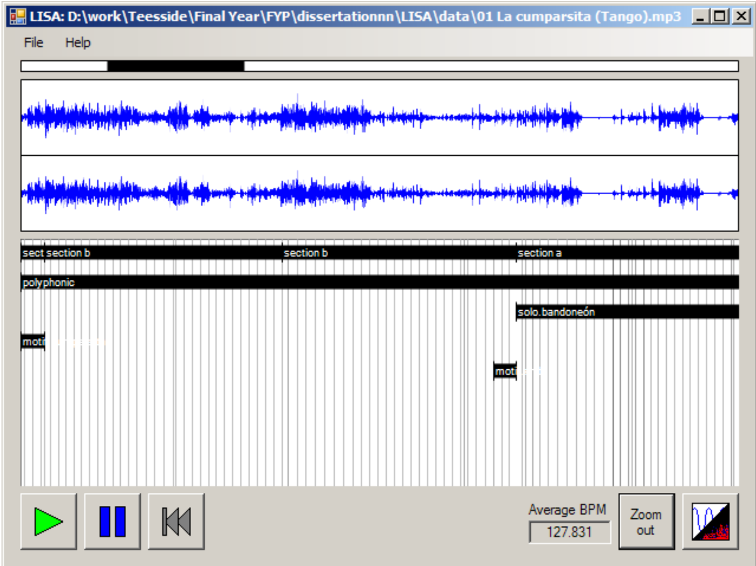



*Figure 2.1: LISA: LISten and Annotate*

- **Alternative forms of annotation** - Annotations could either be "bar annotations", which have a begin and end time and are represented visually as a bar, or "point annotations", which mark a single, precise point in time and are represented by a crosshair. Point annotations were intended to be used for marking out beats that the dancer may want to highlight, however they were found to be far less useful than bar annotations.

- **Multiple layers** - Annotations can be made on multiple layers, allowing overlapping annotations to describe the interleaved properties of the music.

- **Zoom** - It was made possible to magnify certain sections of the waveform for detailed annotation. In addition, a custom scrollbar which incorporated both scrolling and zooming functionality was designed, to integrate zoom with navigation.

- **XML-based file format** - An XML-based format was chosen to load and save annotation data. This made writing input/output code easier and less error-prone.

LISA proved a useful tool for preliminary analysis. For tasks like dividing the piece up into sections and noting the use of cadences and motifs, it was ideal. However, there were some forms of analysis for which it wasn't well-suited, and some more features that would have improved functionality:

- **Lack of standard types** - Currently, annotations are nothing more than text strings. While the flexibility this offers was useful when trying to identify what in the music was being looked for, now that certain properties have been discovered, it would be useful to have a number of standard annotation formats, so that — for example — motifs could be identified and extracted easily from the XML file.

- **Mood representation** - Mood is not represented well in LISA. Neither of the visualisations give a very firm indication of the shifting mood across the song. The addition of a third visualisation, based on Tzanetakis and Cook's (2000) concept of "TimbreGrams", would help identify areas in music involving shifts in mood.

- **Improved beat detection** - The beat detection algorithm employed in LISA was useful for obtaining a rough analysis of musical properties, however it suffered some quite large inaccuracies at certain points in the music, usually when the intensity variance was low. As a result it would not be possible to use these annotations directly in a system hoping to interpret them for dance, as they would quickly fall out of time. If LISA ever needed to be used for a purpose other than that for which it was intended — rough, preliminary analysis — the beat detection would need to be improved.

## 2.2 The musical framework

### 2.2.1 Tempo

While most Argentine Tango music does not have a percussion section, it still follows a strict rhythm which — in traditional Tango — is maintained throughout the length of the piece. This rhythm is usually marked clearly by either the piano or the bandoneón, though in some cases the double bass is used. It is vital that a dancer is able to hear this rhythm and step on the beat.

### 2.2.2 Time Signature

Unlike Ballroom Tango, Argentine Tango covers a wide range of different styles, and this includes a variety of time signatures. A *Vals* or a *Milonga* would usually be danced quite differently to a traditional *Tango*(Turner, 2006). There are some figures which work across the different time signatures; others are limited to a particular style. Some of the key differences between the various styles are summarised in table 2.1, on the following page.

*Table 2.1: Styles of Argentine Tango*

| **Style** | **Time Signature** | **Notes** |
|---|---|---|
| *Tango* | 2/4 | This is the most common, traditional Tango style. There are no limits on tempo, though usually *Tango*s are quite slow. A dancer dancing a *Tango* may hold for a number of beats, or step twice in a single beat. |
| *Milonga* | 2/4, stressing the 1st, 4th, 5th, and 7th beats, and occasionally the 2nd | *Milonga* is similar to *Tango*, but typically faster. The music is closer to the African roots from which both *Tango* and *Milonga* evolved. The dance involves more walking, with shorter steps and fewer complicated figures. In *Milonga*, the rhythm is often more marked, and usually a dancer will step on every beat. However, the development of *Milonga con Traspié* allows him to inject variety and dynamics into the dance by using *Traspié*s and *Contrapasos*. |
| *Vals* | 3/4 | *Vals* is the Argentine interpretation of the classic Waltz, and as such is simply a song in 3/4 time, usually using an *Orquesta Típica*, or other typical Tango instrumentation. It is possible to dance *Vals* by simply adjusting *Tango* steps to fit the 3/4 time signature, however *Vals Criollo* is a specific style developed for dancing *Vals*. It is a fluid dance, so there are no stopping moves. |
| *Nuevo / Contemporáneo* | Various | The *Nuevo* and *Contemporáneo* styles of Tango move right away from the traditional roots of the above styles and bring influence from jazz, electronica, and occasionally even rock and rap. This allows for the time signature to change over the course of a song. These wildly ranging influences lead many to question whether these are indeed forms of Tango, suitable for dancing, or whether they are new styles entirely which have evolved out of Tango, and impossible to dance. Those who do dance to these styles often try to break previously maintained rules about how to dance traditional Tango, and as such almost anything is possible. |

### 2.2.3 Interspersed Rhythm

Though the rhythms as laid out in table 2.1 are quite prescribed, often different instruments in the *orquesta* will play around these rhythms, playing faster or slower than the main beat of the song. In these cases the dancer chooses which of the instruments to bring out in his dance. It is even possible for the leader to follow the rhythm of one of the instruments while leading the lady to follow another, as can be seen in video 2 (Appendix B), where the man is dancing to the fast bandoneón and the lady dances to the beat marked out by the double-bass.

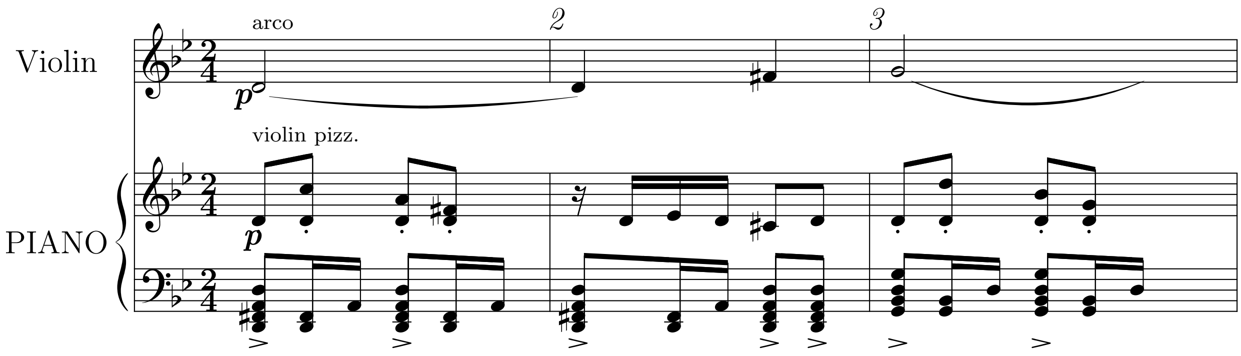



*Figure 2.2: In this extract from "La Cumparsita", one violinist plays slow notes arco, while the pianist and another violinist mark a faster rhythm.* (Rodríguez, 1917)

### 2.2.4 Breakdown of Sections

Newcastle-based Tango teacher Ahmet Ersoz used the idea of *musical sentence*s when trying to develop a beginner's musicality, an allegory which will be considered further in section 4.1. The borders between these *section*s can be used as points where a dancer might change his style of walk, or aim to complete a dramatic figure. In traditional Tango music, *section*s are usually eight bars in length, however music theory identifies many ways we can recognise the beginning and end of sections in a piece of music.

A clear indication to human listeners of what the music will do next is the use of certain *cadence*s, which carry with them particular meanings. Tango specifically uses *perfect cadence* often to signify the end of a section or of the piece of music. Another common technique employed very often by Tango composers is to use *motif*s, which could convey a variety of meanings to the listener.

Figure 2.2 contains an example of motif. The first two bars form an instantly recognisable motif which immediately identifies the piece to listeners. This is

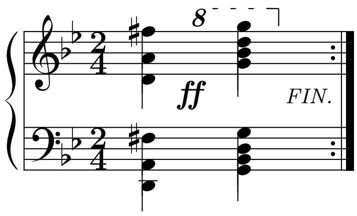



*Figure 2.3: Perfect cadence in "La Cumparsita".*

particularly useful for this piece, as "La Cumparsita" is often the last Tango played at a *milonga*. If it can be recognised easily, patrons hear this motif and know that this will be their last chance of the evening to dance.

Perfect cadence, as demonstrated in figure 2.3, signifies the end of the piece. In the version of the score transcribed in Appendix A, this perfect cadence doesn't feature. The example given above is from Juan D'Arienzo's 1940s interpretation.

## 2.3 The emotional performance

The above-cited musical properties give the dancer the framework within which he can compose the dance. The time signature will govern from which groups of figures he may choose; the tempo will specify the speed at which he must dance, as well as further limiting the movements available to him. The interspersed rhythms give him a variety of options as to what in the music he would like to emphasise, and the breakdown of sections provides particularly convenient moments for the dancer to make changes in style or emphasis.

Once the dancer has utilised this framework to limit his possible movements, he must choose from those possibilities a particular movement, and how he will make it. The distinction between *which* movement and *how* it should be made is important here. For example, a dancer may find himself presented with the option of moving in two directions - a step forward, or a side-step to the left. *Which*ever is chosen, there are then further choices to make regarding *how* this movement is made: his leading leg could move in its direction either smoothly or sharply, whilst his other leg could follow it either quickly or slowly.

These decisions will be made according to the *mood* of the piece at that moment. This presents the system with two problems: how to define mood, and how to draw it from the music. This section will concern itself with the definition and measurement of mood.

### 2.3.1 A brief survey of existing mood models

A variety of models for the measurement and description of mood were explored; in particular those which considered mood in a musical context. Li and Ogihara (2003) consider a series of adjective groups as presented in table 2.2. Similar multifactorial models have been explored in the past, which describe an emotion in terms of 6-10 "primary" emotions (Lorr, 1989).

*Table 2.2: Adjective groups for describing emotion in music as presented by Li and Ogihara*

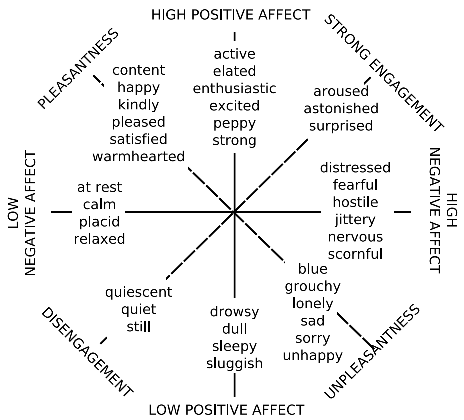
| | | | |
|---|---|---|---|
| A | cheerful, gay, happy | H | dramatic, emphatic |
| B | fanciful, light | I | agitated, exciting |
| C | delicate, graceful | J | frustrated |
| D | dreamy, leisurely | K | mysterious, spooky |
| E | longing, pathetic | L | passionate |
| F | dark, depressing | M | bluesy |
| G | sacred, spiritual | | |

Table taken from Li and Ogihara (2003)

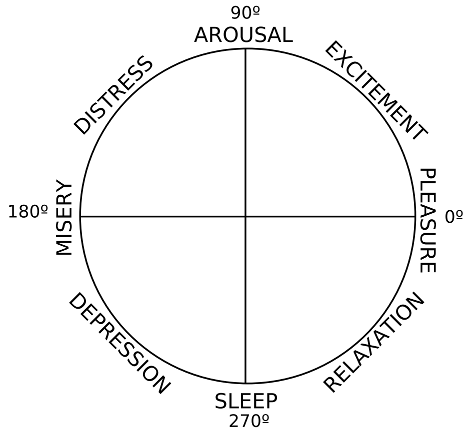
Models like this one are very powerful and can describe a wide range of moods and emotions. Many researchers have found different options for the primary emotions (Lorr, 1989), some of which attempt to describe a broad scope of human feeling, others focused on identifying particular emotional characteristics. The above list is a progression from Farnsworth (1958), which in turn was a refinement of Hevner's (1936) circle of adjective clusters.

A number of mood models use circles to represent the range of emotions and their interrelationships. Some examples are given in figure 2.4. Tellegen and Watson (1985) espouse a two-dimensional model, as expressed in figure 2.4(a), whereby emotions can be described in terms of Positive and Negative Affect. Tellegen and Watson (1999) note that this model "was not designed to be exhaustive or exact ... it is an approximate formulation, and it emphasizes PA and NA as relatively independent dimensions".

The circumplex model shown in figure 2.4(b) uses English words to measure emotion. It is similar to the multifactorial lists of emotions mentioned above, but because the emotions are plotted on a two-dimensional graph, their interrelationships are demonstrated explicitly. As a structural model, the circumplex is extremely useful (Russell, 1989).




(a) Positive and Negative Affects




(b) Circumplex model

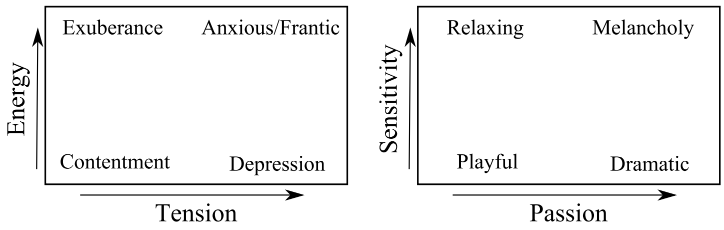
*Figure 2.4: Watson and Tellegen's (1985) two-dimensional map, representing mood in terms of Positive and Negative Affect, and the circumplex model, advocated by Plutchik and Russell.*
Figure 2.4(a) taken from Tellegen and Watson (1999)

### 2.3.2 Thayer's model and its usefulness for music and dance

The mood models that have been presented so far are capable of representing quite a wide range of feelings, and are usually measured through the use of questions designed to assess the subject's mood (Lorr, 1989). It is questionable how directly applicable they would be to the task of identifying mood in Tango music. A dancer need not concern himself over the difference between whether a piece of music is "serene" or "relaxed"; he needs the feelings given from the music in a more abstract sense, to act as a backdrop against the moods he decides to emphasise through his dance.

In their article investigating automatic detection of mood in music, Liu et al. (2003) use Thayer's (1989) two-dimensional mood model, which describes mood in terms of two factors: Energy and Tension. Music mood can then roughly be described relative to four clusters: Contentment, Depression, Exuberance, and Anxiousness (Liu et al., 2003). This allows the mood of a piece of music to be considered in quite simplistic and easily quantifiable terms.

Figure 2.5(a) shows Thayer's mood model, where Energy and Tension are properties within the music. Liu et al. (2003) have linked these properties to measurable quantities obtainable from raw audio data, these being intensity, rhythm, and timbre. It should be noted that these describe the mood we can draw *from* the music.




(a) Thayer's model (b) Moods in dance

*Figure 2.5: Models of mood in music and dance.*

It is hypothesised that similar adjective clusters can be chosen which would apply more directly to Tango music, and that the properties of Energy and Tension within the music will have a direct influence upon the way that piece of music will be danced.

First, alternative adjectives for the four clusters were chosen which are more commonly heard in Tango. Pieces which seem to exemplify these terms were considered. These were:-

- **Relaxing (Contentment)** - *Bahía Blanca*
- **Playful (Exuberance)** - *Así se baila el Tango*
- **Melancholy (Depression)** - *La última copa*
- **Dramatic (Anxious/Frantic)** - *La Yumba*

In fact, usually a piece of music will have variations in its mood as the song progresses. "La Yumba", for example, is mostly dramatic, but features a violin section two-thirds of the way through which give it a melancholy feeling, before building back up to a dramatic climax.

The effect of these different moods upon the dance was considered, by consulting video references for the above-listed songs, and having them played at the local Tango society to view the reaction given by dancers improvising in a social setting. It was noted that, as difference moods correlate with changes in energy and tension in music, they seemed to correlate with the sensitivity and passion with which the dancers performed. Figure 2.5(b) demonstrates a model based on Thayer's, which could be created to represent the way mood affects dancing style.

# 3 Music Information Retrieval(MIR)

The previous section outlined what sort of information would be required from the music in order to make decisions about how to dance to it. This section will survey existing research exploring how to obtain this information automatically from raw audio data.

## 3.1 Rhythm, beat and metre

The basis of any dance simulation must be the rhythm. A step taken out of time with the music will be immediately apparent even to someone not trained in music or dance. Conversely, a figure performed inappropriately to music which does not suit it might not be noticeable except to those with a firm understanding of Tango. It is therefore vital that the beat detection be as accurate as possible, after which other information can be used to refine the choice of movements.

The beat of a piece of music has been described as "a series of pulses, spaced approximately equally in time, relative to which the timing of all musical events can be described" (Dixon, 2001). Each of these pulses may be referred to as a *beat* individually, however the term cannot be used for just any pulse found in a signal; it must be a member of such a series.

### 3.1.1 Simple intensity tracking

One very simple way of tracking these pulses is to maintain a history monitoring the sound intensity over a set length of time. Moments where the intensity is dramatically louder than the mean intensity across this history can be interpreted as beats (Patin, 2003).

This method of beat tracking carries with it two principal flaws:

- In areas where the sound energy variance is low, beats will not be detected easily. This means that when there is a lot of noise surrounding the beat, or if the beat is quiet or muffled, it will be difficult to trace using this system.
- Since this system treats beats individually as dramatic variations in sound intensity, it has no concept of *the beat* as a collection of pulses. It is liable to treat changes in timbre or sometimes even individual notes as beats, and has no way to differentiate these from the actual beat running through the music.

The latter problem is impossible to fix without a more complex technique, or at least some post-processing performed on the "beat" information it results in. The former, though, can be improved slightly, by analysing different frequency sub-bands individually. A common way of separating the intensities by frequency is to use a Fast Fourier Transform (FFT). We can then perform the same comparison by maintaining a history of intensities per-frequency sub-band and comparing against this. This helps the system find some more beats in certain cases - for example when instruments at a higher frequency sub-band combine to create an overall intensity that is fairly invariant, while a low-frequency instrument such as the double-bass marks out the beat.

Simple post-processing that can be done on the information discovered by this algorithm includes monitoring the distance between beats to remove incorrectly detected beats in noisy areas. Once this has been performed, the mode periodicity can be calculated, and beats with multiples of this distance between them can be filled with extra beats. This can effectively "even out" the beats detected, and result in a more regular pulse, however if there are long periods of low intensity variance in which beats are not detected, beats added by this post-processing technique quickly become inaccurate.

### 3.1.2 Tempo tracking on a symbolic representation of music

There are many techniques which try first to move from raw audio to a symbolic representation such as MIDI before performing tempo induction and beat tracking (Takeda et al., 2004; Cemgil et al., 2003; Dixon, 2001).

The conventional method for rhythm extraction from MIDI representation is quantization (Takeda et al., 2004). This simply entails moving note onset times to the nearest beat. While this technique is very simple, with expressive variations in music quantization becomes increasingly difficult (Cemgil et al., 2000). Various attempts have been made to improve the quality of rhythms interpreted from MIDI or MIDI-like data (Takeda et al., 2004; Cemgil et al., 2003; Dixon, 2001; Frieler, 2004; Raphael, 2001). These projects have mostly been aimed at automatic music transcription, and as such require a greater level of detail than that needed simply to work out rhythm and metre.

When a symbolic representation has been found, onset times will have been detected for each note. The time periods between these, called the Inter-Onset Intervals (IOIs), can be used to reveal the main periodicities in the music and filter out spurious onsets (Dixon et al., 2003).

## 3.2 Polyphonic audio

The reference videos (Appendix B) suggest that many of the interesting ways the dancers can interpret the music involve moving between the interspersed rhythms - dancing to the violin one moment, and to the bandoneón the next. Therefore, some way to separate out the instruments and consider their rhythms individually would allow for much more interesting dance to be simulated than would be if they always used the main rhythm driving the song.

Eggink and Brown (2003) have applied Missing Feature Theory to the problem of identifying musical instruments in polyphonic audio, succeeding in identifying up to two instruments played concurrently. Essid et al. (2005) have successfully recognised up to four instruments in a polyphonic audio recording, using as their example a jazz quartet (Drums, Double-bass, Piano, and a monophonic instrument such as Saxophone or Trumpet). Since the Tango *Orquesta Típica* is composed of double-bass, piano, violins, and bandoneón, the number of instruments is similar, though in Tango there is no percussive section. The violin can be polyphonic, usually when playing *pizzicato*, however often it is played monophonically. An *orquesta* will usually have a violin section, to allow for more than one note to be played even while each individual violin is being played monophonically.

### 3.2.1 Extracting major melody lines

Eggink and Brown (2004) focus on "the extraction of a major melody line from polyphonic audio, which we define as the melody played by the solo instrument in an accompanied sonata or concerto". Fundamental frequencies (F0) describe the pitch of a note, so the problem of trying to extract note information for a particular instrument involves estimating the F0 of that instrument. In order to do this, multiple candidate F0s are found, from which the most appropriate can be chosen (Goto, 2001). Goto (2004) works with extracting just the melody and bass lines using a Predominant-F0 Estimation method (PreFEst). These assume particular frequency domains for the bass and melody sections respectively (bass sections will be of lower fundamental frequency). Eggink and Brown (2004) have tried to improve this by analysing melody lines of a solo instrument, which could exist in a number of frequency domains depending on what the instrument is. They use an instrument recognition module in order to choose which fundamental frequency is likely to correspond to the solo instrument sought after.

This technique has achieved reasonable results, however it would be difficult to apply directly to the problem of voice separation in Tango music. The focus on this technique lies with extracting a single solo instrument from a sonata or concerto, rather than extracting each instrument individually from an *Orquesta Típica*. There is possibly some applications for it, however, as Tango will often have sections wherein an instrument is performing a solo over backing from the *orquesta*. Often during these sections the dancer will choose to dance to the most prominent instrument — the solo instrument. Nevertheless, a more flexible system, in which he can dance to any of the instruments, would be desirable.

### 3.2.2 Detecting melody notes

The previous techniques extract "melody lines", which attempt to emulate the musical understanding of an untrained listener, as described in (Goto and Muraoka, 1999; Goto, 2004). Others have tried to extract the exact notes being played, for automatic transcription of music. Paiva (2005) make use of pitch saliences and melodic smoothness.

Chew and Wu (2004) present a Contig Mapping approach to separating voices in symbolic representations of music. This could be used, once onset times and pitches have been calculated, to disentangle the melodies of each instrument in the *orquesta*.

### 3.2.3 Polyphonic tempo tracking

The solutions proposed above approach the problem of finding the melodies within polyphonic music, which is important for applications like automatic transcription and *query-by-humming*. However, for the purpose of dance such specific knowledge is not necessary. Many dancers would not have the musical training to transcribe the music by listening to it, and instead would use mostly the rhythmic and timbral qualities of the music to compose their dance, as well as its intensity. The most important property we need from the polyphonic audio, then, is the rhythmic component, followed by the individual timbral and intensity properties of each instrument.

## 3.3 Motifs, phrases, and section segmentation

Dividing the piece of music into a number of discrete *section*s provides the dancer with convenient boundaries where he can aim to complete a figure, change his style of movement, or otherwise punctuate the music. In addition, *motif*s are often points which the dancer would choose to highlight, since they are usually the more recognisable moments in the piece. Motifs are also used, along with *cadence*s, to divide sections.

### 3.3.1 Pattern learning and recognition

Lartillot (2003) uses Perceptive Heuristics to search for musical patterns, an approach modelled on the cognitive mechanisms involved in music perception. This strategy works with symbolic data, so a preprocessing step to convert the raw audio to a symbolic format such as MIDI would be required. It also works with monophonic data, though similar ideas may be applied to polyphonic audio. Polyphonic patterns can be found within polyphonic music by representing the problem geometrically (Tanur, 2005). This method still requires a symbolic representation of the music, and is intended more for searching within a piece of music for a known pattern than for extracting likely patterns.

Harford (2003) describes another approach to the segmentation and learning of melodic patterns, named SONNET-MAP. A Self-Organising Neural Network (SONNET) is used with an associative map, in order to represent melody as "a learned hierarchy of automatically segmented melodic phrases that are encoded in long-term memory(LTM) by adjustable connection weights". Preliminary simulations with this system are promising, producing segmentations comparable with manually segmented pieces. SONNET-MAP works with symbolic data.

### 3.3.2 Tonal analyses

Section boundaries are sometimes marked by changes in tone, timbre, or key. These are properties that can be used in determining the mood of a piece; indeed, a change in mood often accompanies a section boundary. A cognitive model for tonality perception is compared with various machine-learning models in Gómez and Herrera (2004). Machine learning algorithms gave reasonable results, yielding 84% accuracy in their estimation of mode - a particularly important property in determining the mood of a piece (Hevner, 1935).

The cognitive model has been shown to yield 75.1% accuracy working with raw audio data, however 94.1% can be achieved if exact, relative, dominant, subdominant and parallel keys are assumed to be similar (Pauws, 2004). For this application, it is the *changes* in key which are important, rather than absolutely accurate detection of key, therefore these assumptions would be reasonable.

## 3.4 Timbre and intensity

Timbre and intensity can be used alongside mode and rhythm to determine the mood of a piece of music. Liu et al. (2003) map these musical properties to the axes on Thayer's model of mood, matching intensity to "energy" and timbre and rhythm to "stress" or "tension". As has been hypothesised in section 2.3.2, these axes correspond to "sensitivity" and "passion" within the dance, thus any information we can discover about timbre, intensity, and rhythm will directly affect the choice of movements and how they are performed.

### 3.4.1 Using spectral information to determine timbral qualities

Unlike many other properties of sound, such as pitch, intensity, and duration, timbre is not strictly defined or well understood. The components that make up what we call "timbre" are the subject of psychoacoustic research (Wold et al., 1996). Nevertheless, many of these components have been shown to be determined by the spectral information in different sub-bands (Liu et al., 2003; Zhang and Kuo, 1998). The spectral features affecting timbre as defined in Liu et al. (2003) are described in table 3.1.

*Table 3.1: Definition of timbral features*

| Feature name | | Definition |
|---|---|---|
| Spectral Shape Features | Centroid | Mean of the short-time Fourier amplitude spectrum. |
| | Bandwidth | Amplitude weighted average of the differences between the spectral components and the centroid. |
| | Roll off | $95^{\text{th}}$percentile of the spectral distribution. |
| | Spectral Flux | 2-Norm distance of the frame-to-frame spectral amplitude difference. |
| Spectral Contrast Features | Sub-band Peak | Average value in a small neighbourhood around maximum amplitude values of spectral components in each sub-band. |
| | Sub-band Valley | Average value in a small neighbourhood around minimum amplitude values of spectral components in each sub-band. |
| | Sub-band Average | Average amplitude of all the spectral components in each sub-band. |

Table taken from Liu et al. (2003)

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. Agostini et al. (2003) classify a number of instruments in terms of their timbral qualities.

### 3.4.2 Monitoring intensity

Wold et al. (1996) state that intensity "is approximated by the signal's root mean-square (RMS) level in decibels". 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.

The correlation between intensity in music and energy in the mood that music suggests is clear, as pointed out by Liu et al. (2003). There is therefore a likely negative correlation between intensity in the music and sensitivity in the dance accompanying it.

## 3.5 Summary

Music Information Retrieval is one of the hardest and most important stages for this application. It is vital that the information drawn from the music is accurate, as it will form the input for the system deciding which moves should be made.

These techniques could be added incrementally. Beat detection should be implemented first as all movements are made at a speed relative to the beat of the music, but after that changes in tone, timbre, and mood can be detected to refine the choice of movements the system would make.

# 4 Artificial Intelligence and Music

The application of techniques from Artificial Intelligence (AI) and related fields to music has been approached many times (Meehan, 1979; Roads, 1985; Roads, 1992; McCormack, 1996), however few have discussed the understanding of music from a dance perspective specifically. This section will consider some of the approaches taken toward AI interpretation of music and discuss what relevance they hold - if any - to the problem of composing dance to music.

## 4.1 Music and Language

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 (McCormack, 1996). However, the majority of music attempts to invoke at least some form of emotive response, which is a semantic, rather than syntactic, property. In addition, certain music, especially *programme music*, sets out to represent more specific meanings or even to tell a story. Preliminary neurological studies have found that "both music and language can prime the meaning of a word, and that music can, as language, determine physiological indices of semantic processing" (Koelsch et al., 2004).

It has even been suggested that while music and language are not entirely separate domains, neither is music a subset of language. Rather, "the human brain, at least at an early age…treats language as a special case of music"(Koelsch and Siebel, 2005). Whether music is a subset of language or language a subset of music, it seems natural to consider approaches taken toward language in the field of AI, and to compare how music could be understood using these methods.

## 4.2 Natural Language Processing(NLP)

Natural Language Processing and Computational Linguistics represent large areas of research, so only key concepts will be summarised here, to provide context for an examination of music from an NLP perspective.

### 4.2.1 The basics: forming an hierarchical structure

A simplistic NLP system can be thought of as composed of distinct stages (Lynch, 2007), these being:

- **Morphological processing**. This comprises breaking words up where they have been formed of sub-words, prefixes or suffixes. Examples include "unsteadily" (un·steady·ly) and "keyboard" (key·board). Morphological processing will also recognise inflections, such as "violins" (the plural form of "violin"), "bigger" and "biggest" (comparative and superlative forms of "big").

- **Syntactical analysis**. Here a sentence will be broken down to determine the function of each word within it. Tree structures are a very popular way of representing syntactic relationships.

- **Semantic and Pragmatic analyses**. These stages deal with identifying the meaning of words in the sentence. They are grouped here because the precise distinction between the two is not clearly defined. Russell and Norvig (1995) differentiate them by stating that semantics deal with the words in an isolated sense, while pragmatics deal with using an understanding of the current context to inform the meaning drawn from each word.

In most cases, these stages will not be quite as distinct as presented here. There may be some level of semantic analysis involved in determining the syntactic structure of a sentence, to resolve ambiguities between possible structures. Usually, syntactic and semantic analyses are performed by a single mechanism - the parser (Lynch, 2007).

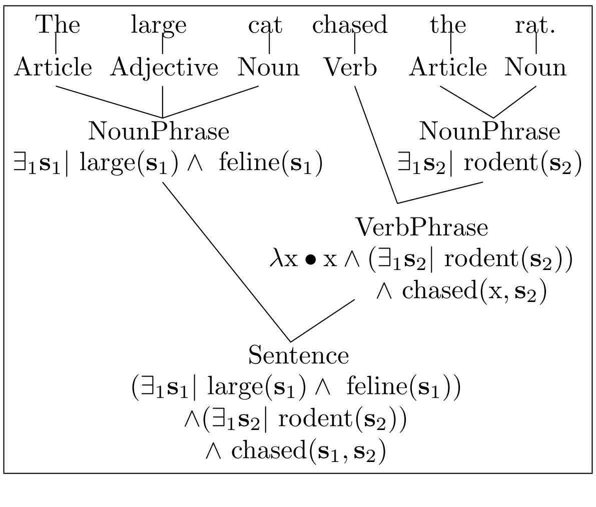



*Figure 4.1: Example hierarchical representation of a simple English sentence.* Figure taken from Lynch (2007)

An example of how the sentence "The large cat chased the rat" could be broken down into a semantically labelled hierarchical structure is given in figure 4.1.

### 4.2.2 The ambiguity problem

The great challenge facing an NLP system is that of solving ambiguities between different possible syntactic and semantic structures. In language, syntactic constraints facilitate this operation. For example, in English, a determiner ("article") must precede the noun with which it will be combined (Bod, 2002). Despite these constraints, ambiguities quickly manifest themselves in quite simple sentences. This can be demonstrated using the sentence, "List the sales of products in 1973".

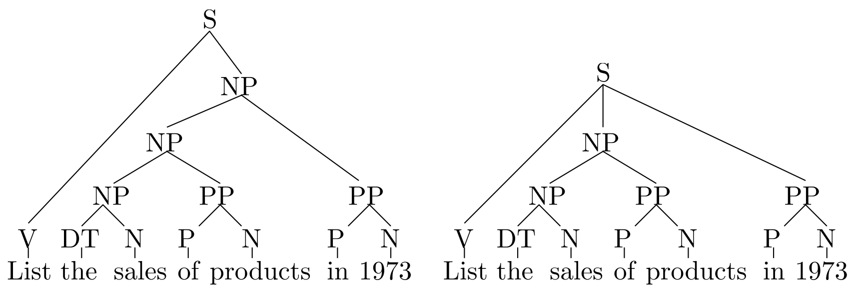



*Figure 4.2: Alternative syntactic hierarchies for a simple sentence.*
Figure taken from Bod (2002)
(originally Martin et al. (1981))

There are two approaches to solving the ambiguity problem: the *likelihood principle* and the *simplicity principle*. The likelihood principle is based around assessing the likelihood of all available tree structures, and choosing the most likely one. It has seen much use in NLP, and has also been applied to music. The simplicity principle favours the simplest tree structure available, regardless of its likelihood. It is very popular in the field of visual perception, but music perception has also been explored from this angle.

Bod (2002) has suggested that it would be possible to combine these principles, using "the simplicity principle as a general cognitive preference for economy, and the likelihood principle as a probabilistic bias due to previous perceptual experiences". This would result in a system which would perform well for structural organisation of language, music, and visual information.

## 4.3 Limitations of NLP applied to music

Many of the arguments for approaching music from an NLP standpoint have been presented in section 4.1. There are definite advantages to using ideas from NLP in analysis of music, however there are problems inherent in trying to bring across these techniques directly. The ambiguity problem is worsened, and semantics — which in language can be used to help resolve certain ambiguities — are less easily defined. This section will explore some of these issues.

### 4.3.1 Infinite ambiguity

Problems with ambiguity become even more complex with musical input than with language, because there are virtually no syntactic constraints. As noted by Longuet-Higgins and Lee (1984), "Any given sequence of note values is in principle infinitely ambiguous, but this ambiguity is seldom apparent to the listener". Two potential hierarchies for the same sequence of notes are shown in figure 4.3.

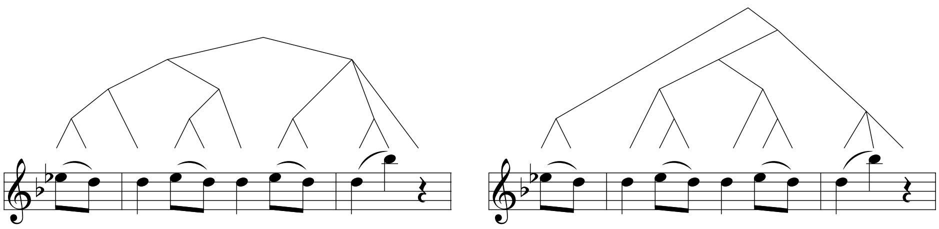


*Figure 4.3: Alternative ways of breaking down music hierarchically.*
Figure taken from Bod (2002)
(originally Lerdahl (1996))

### 4.3.2 The question of meaning in music

Apart from the potentially infinite ambiguity in its structure, treating music from a linguistic standpoint presents another problem: how can "semantics" be defined for musical input? There are certainly arguments for the literal interpretation of "meaning within music", and even some neurological evidence to back these (Koelsch et al., 2004; Koelsch and Siebel, 2005). Wiggins (1998), however, argues that there are "inherent dangers of a naïve, wholesale importation". He proposes a simple thought experiment to warn against any attempt at searching for *precise* semantic meanings in music, as one would in language.

> "If I ask you to read a page of an unfamiliar novel and then ask you 'What does that mean?', you can easily tell me. In particular, the

> request itself is easily understood as one of a small set of possibilities, involving more or less inference from what the text describes... If, on the other hand, I ask you to listen to five minutes of unfamiliar music, and then ask you 'What does that mean?', it is much harder for you to answer."(Wiggins, 1998)

Rather than try and warp the hand of music to fit the glove of Computational Linguistics, Wiggins proposes the term "musical connotation", rather than "semantics" or "meaning". Music can certainly communicate *something* to the listener, but whether it can communicate to the same depth and precision as language — enough to warrant the same terms as used when considering language — is questionable.

Musical connotation, then, would describe the emotional affect achieved in the listener. Wiggins next suggests further sub-terminology: *direct* and *indirect* connotation. Direct connotation directly induces emotion in the listener, in the form of a physical reaction in which the listener shows signs of emotional arousal. Indirect connotation is a feeling in the music which the listener may be able to explain, without having experienced that feeling themselves. It is possible for a connotation to be both direct and indirect.

For the purpose of dance, musical connotations appear to be a very appropriate and useful device. The precise meaning of the song is not strictly important, as demonstrated by the fact that many dance Tango quite successfully who don't understand the lyrics, which are typically in Spanish. What is important is the feeling engendered by the music, in a more generalised sense.

# 5 Visually representing the dance

Performing analysis on music and deciding how it should be danced is of little use if there is no way to visualise these decisions. The system could either output a choreography, which could be printed and followed by live dancers, or it could render an animation on-screen. Which of these the system is targeting will have implications on how much detail it must go into when composing the dance, what sorts of things it can represent, and how it could be used.

## 5.1 Paper Choreography

Since Tango is usually an improvised dance, the idea of printing and following a choreography is a strange one. Nevertheless, there are circumstances in which it would not be so unusual: live Tango shows would usually be choreographed so as to create the most exciting performance for the audience. These shows often incorporate figures that are rarely seen in social Tango, and as a result many refer to the style performed onstage as *Tango Fantasia*.

Apart from this, some form of static, written notation is often useful in testing systems, and experimenting to find the effect of changes either within the music or in the way it is interpreted. It is also useful for communicating results in written form, without having to supply video footage alongside the product or techniques being demonstrated.

### 5.1.1 Benesh Movement Notation

Benesh Movement Notation, devised by Rudolf and Joan Benesh, was first published in 1956 (Benesh Institute, 2005). It is designed to be capable of documenting any movement, and while its main use is in dance, it has been used by physiotherapists and clinicians to analyse other forms of movement.

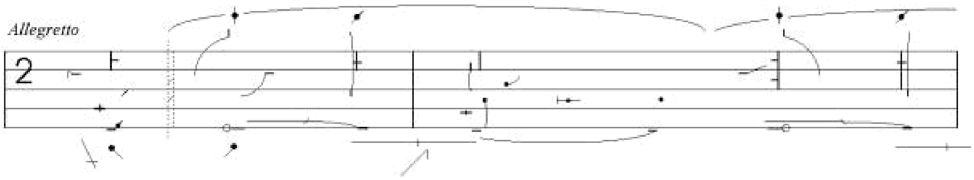
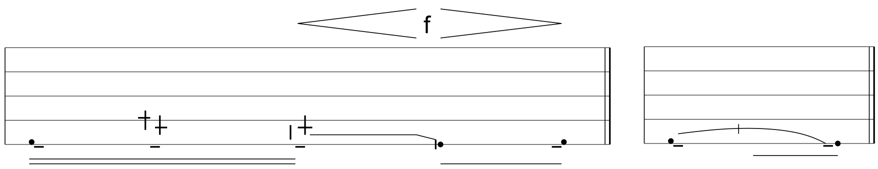



*Figure 5.1: Benesh Movement Notation*
*Giselle Act 1, Peasant Pas de Deux, Male Variation*
*Choreography: Coralli / Perrot / Petipa*
Figure taken from Benesh Institute (2005)

A sample of Benesh Movement Notation applied to ballet, where it is most widely used, can be seen in figure 5.1. It uses a five-line stave, read from left to right and top to bottom, similarly to standard music notation. The positions of the limb extremities are plotted with emboldened lines, while "movement lines" describe the paths they take. In this way both the positions of the dancers at key moments and the quality of the movements taken between those positions can be described.

Benesh Notation has the advantage of being a standard format, well understood by choreographers and professional dancers. It can describe movements in detail, and would be capable of describing *how* the movements should be made, as well as *which* movements they should be. It is quite a lengthy format, however, and the detail it is capable of describing can be cumbersome. A simple forward step of a single person is described in figure 5.2. Even after adopting the shorthand in figure 5.2(b), it is a lot of space to describe such a simple and core movement used in dancing Tango.




(a) Full transcription

(b) Shorthand form

*Figure 5.2: Benesh Movement notation describing a single forward step.*
Figure taken from Bodirsky (2004)

Methods for encoding Benesh notation using the computer have been explored (Neagle et al., 2002; Neagle and Ng, 2003). An interesting thing about these projects is that they include an animation component as well, to visualise the movements on-screen and verify that they've been transcribed correctly.

### 5.1.2 Labanotation

Labanotation is another notation which is very prominent in professional dance choreography. It was originally developed by Rudolf Laban and first published in 1928. Since then it has seen further development, and the Dance Notation Bureau has been founded to promote the use of Labanotation and maintain a library of Labanotation scores.

Labanotation and Benesh Movement Notation have many common properties. Both work by dividing the body and describing the movement of each part of it, and both take into consideration "location, travelling, body action, directions, timing, rhythm and gestures of the body" (Neagle and Ng, 2003). They do this in quite different ways. Labanotation is read from bottom to top, and uses positions on the staff to denote body-parts, while shapes describe the movements those body-parts make. Labanotation does not resemble a musical staff so strongly as Benesh, though both incorporate the concept of bars, and use length to denote time.

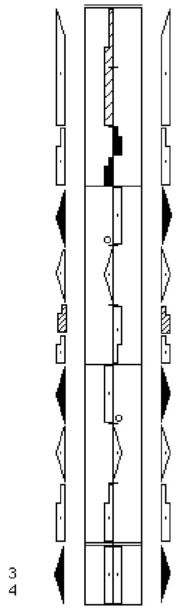



*Figure 5.3: Labanotation*
Figure taken from the Dance Notation Bureau (2005)

Labanotation has been used in a variety of fields. The Dance Notation Bureau a collection of ballet, modern, and jazz scores, as well as folk, historical, and social dances (Dance Notation Bureau, 2005). Loke et al. (2005) have considered using Labanotation when designing movement-based interactive applications, such as games which use the EyeToy™. To a lesser extent, it could also be used when considering applications targeting the Nintendo Wii™, which also depends upon movement-based interactivity for its gameplay. Loke et al. (2005) suggest that Labanotation could be a useful tool in the design of games driven by physical movement, though they comment that "its major disadvantage is the effort required to learn how to use it".

### 5.1.3 Tango-specific notation

Benesh Movement Notation and Labanotation are extremely powerful systems for describing movement. They can describe in detail which movements should be made, how they should be made, and the timing relative to the music. However they both present a significant challenge to learn, and they are designed to be capable of representing any kind of movement. Movement in Tango is quite defined: there is a standard way to take a step, and while dancers will adjust this to suit their style and the feeling of the music, these adjustments will still be variations upon a standard. The disadvantages posed by the difficulty in reading the above forms of notation might outweigh the benefits of using such descriptive notation systems.

In response to this problem, Bodirsky (2004) has developed a notation system specifically for use with Tango, which considers the limitations in movements typically encountered in social Tango and is free to make dramatic simplifications to the format of the notation as a result.

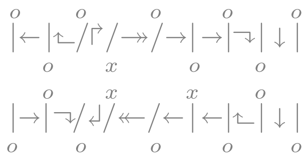


*Figure 5.4: Bodirsky's Tango Notation*
Figure taken from Bodirsky (2004)

Figure 5.4 describes the *Salida*, the basic step in Tango. The arrows denote movements of the feet. The |, \, and / symbols describe the position of the upper body. They don't need to be as descriptive as the symbols describing the movements of the feet, because the upper body will rarely "break posture" in Tango; it will remain upright and face toward that of the partner. The *o* symbols above and below the upper body symbols specify on which foot the dancer is holding their weight. In Tango the weight is always on one foot, and the other foot is free. A *o* above the track means the weight is on the left foot; below means that it is on the right. The leader's track is shown on top, the follower's underneath.

Figure 5.5 shows a choreography taken from the film, "*Tango: no me dejes nunca*" (Saura, 1998), demonstrating how easy and efficient a system this is for describing quite long Tango sequences.

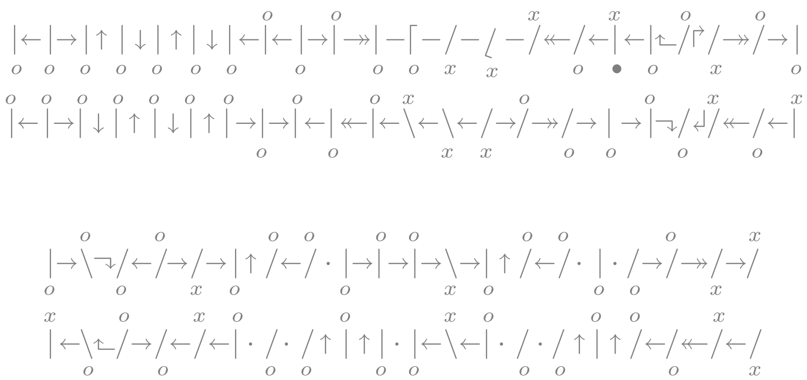


*Figure 5.5: Choreography Extract*
*Juan Carlos Copes, Recuerdo. From Saura (1998)*
Figure taken from Bodirsky (2004)

Bodinsky's system is easy to learn, simple, and space-efficient. It has been designed with Tango specifically in mind. If the system is outputting just the steps the dancer should take, and leaving it up to the dancer to decide *how* they should be taken, it is the logical choice. However, if it is important to determine the quality of the movements as well as exactly which figures are being chosen, a fuller notation such as Benesh or Labanotation would have to be chosen.

## 5.2 Computer Animation

A computer animation might be a useful preview tool, or a more polished animation might be the final target of the system. This might then be used in a game or animated film scene involving dance. Since this system would relieve the animator of having to animate each dance sequence manually, it would enable scenes to be orchestrated involving many couples dancing. In order for this to work, however, the steps and figures must be available separately, and the animation must allow smooth blending between them. In addition, if they are to be available to the system as individual animations, they must either be animated manually (a task which this system aims to alleviate), or captured using Motion Capture (MoCap) techniques.

### 5.2.1 Representing the set of possible movements

There are two schools of thought in the teaching of Argentine Tango (Turner, 2006). One tries to teach a number of figures, then adds variations to those figures, so that dancers build up a repertoire of sequences they have "learned by rote". As they become more experienced they learn to construct new sequences by combining and adjusting the sequences they already know. The other school teaches the fundamentals — how to take steps, to communicate intention, to maintain posture, balance, and axis. From a very early stage, dancers learning this system start to improvise and compose the dance themselves. This might be compared to learning a language: some choose to learn key phrases from a phrasebook, and through use of these phrases and practice through conversation extrapolate a wider understanding of the language they are studying. Others would rather learn the fundamentals of grammar and tables of vocabulary, and construct every sentence themselves.

The designer of an animation system for Tango is presented with a similar choice. The animation must be constructed by linking together a series of smaller animations, but at what level of granularity should they be stored? The "seams" between the animations are a potential problem area. If the animations each have matching *begin* and *end* keyframes, the transition between them will be unnoticeable, however this is an unrealistic constraint to place on the system. There is a danger, if the *begin* and *end* keyframes are liable to be different, of a "snapping" effect occurring between animations. Some sort of crossfading must be applied in order to avoid this, however this can also be problematic and leads to difficulties lining up the animations with the rhythm of the music.

One way to reduce the problems caused by links between animations is simply to build the dance out of fewer, longer animations. This way there will be fewer "seams" and fewer problem areas. The longer each animation is, however, the less flexible the system becomes. Variations on figures must be recorded as entirely new animations, which increases work for the animator setting up the preliminary animations for the system. If a new figure is encountered and the user of the system wants to add it, an entirely new animation must be produced.

An alternative would be to use natural level of granularity as a dancer would consider it; a single step. A number of base positions could be defined which would denote the positions of the dancers relative to each other, and a number of transitions between those positions would describe the steps the dancers can make to move from one to another. A database could be created which would hold a network of positions and transitions, and which could be polled from a particular position to return all the possible transitions that could be made from that position.

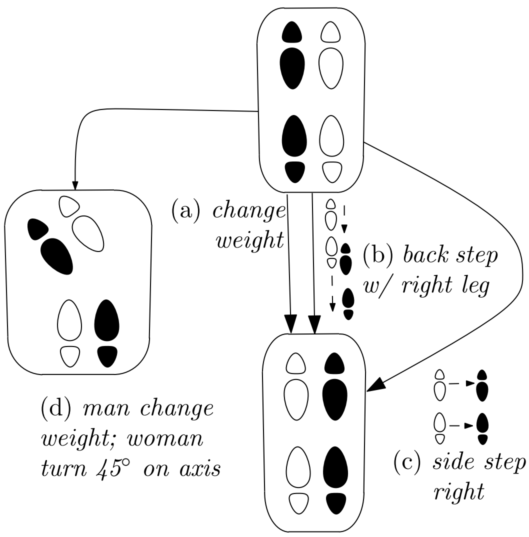



*Figure 5.6: Representing figures using a network of positions and transitions*

This approach offers a number of advantages. The number and length of animations required is reduced, since only transitions must be animated. These transitions will be made to and from distinct positions, so the seams between them won't be so noticeable. Crossfading would still be advisable, so that moving from one transition to another didn't look unnatural, but since the *end* keyframe of the lead-in transition and the *begin* keyframe of the lead-out transition would be guaranteed to be the same, establishing a smooth crossfade between them would not be difficult. If a new sequence needs to be added to the database, no new animations need to be created. Instead, the positions and transitions that make it up can be recorded in a set of known sequences.

The concept of positions and transitions lends itself to computer processing. Similarities can be drawn with techniques used in AI, such as Finite-State Machines, where the positions would be states, which are traversed by means of transitions. Even pathfinding could be explored, whereby if the dancers wished to execute a figure beginning at a certain position in two bars' time, they could find a route of transitions of length two bars that would arrive at that position.

The idea of storing a set of known sequences in a database but also allowing movements that lie outside of those stored in the database is also appealing. As well as pathfinding to a starting position and then performing a hard-coded sequence, this system would offer the possibility for more interesting techniques to be considered. An artificial neural network, for example, could be used to determine which moves should be made. This network would use the database of known sequences, perhaps along with metadata about which music would be appropriate for each sequence, as its training data.

### 5.2.2 Capturing movements using Motion Capture technology

Whether working with long animations depicting entire figures or stitching together individual steps, creating these animations by hand would be a slow, arduous task, and it would be extremely difficult for an animator to capture all the subtleties of movement involved. Motion Capture technology could potentially ease these problems.

In order to explore the possibilities offered by motion capture, some Tango footage was shot using Gipsy5 suits available at the University of Teesside. First, a number of simple figures were recorded individually. After this, a sequence of simple walks and turns were recorded. Finally, an entire dance was improvised.

Motion Capture produces smooth, detailed animations, which having been recorded from human movement naturally look very realistic and humanlike. The very fact of its accuracy, however, leads to some of the difficulties in using it. In order to take advantage of matching *begin* and *end* keyframes, the dancers being recorded would have to end at precisely the same position and angle as they intend to start the next figure. Even disregarding the impossibility of such a feat, floating-point rounding errors quickly lead to slight inaccuracies which would cause small but noticeable snapping between figures. Some crossfading would be necessary to alleviate this.

If the system is to have any notion of *how* the steps should be taken as well as *which* steps, it must be possible to warp captured animations, or blend other animations with them for emotional effect. Various methods of exercising greater control over motion capture data have been explored (Witkin and Popovic, 1995; Pullen and Bregler, 2002). Often these rely on some combination of keyframe-style animation and motion capture data.

There is another, practical difficulty with motion capture. The University of Teesside has two motion capture systems: an optical system, and the mechanical Gipsy5 system. The optical system would be impossible to use for dance purposes, because if any one of the markers cannot be seen, the system breaks down. In dance, where the dancers' bodies will be very close and obscure multiple markers constantly, this would be inevitable. For this reason, Gipsy5 suits were used for the capture.

*Figure 5.7: Motion capture with the Gipsy5 system*

The Gipsy5 system uses a number of rods connected to the body, forming a kind of exoskeleton. This causes significant difficulty when trying to dance a close dance like Tango, where the dancers' legs often brush past one another. During the recording, there were times when the Gipsy5 suits became entangled, almost causing the dancers to fall. Mistakes like these lengthen the time it takes to capture the required data, and pose a small health risk. The dancers were forced to adapt their dance to avoid clashing, which was counterproductive as it means that the captured data was not perfectly accurate. A video was recorded of the dancers improvising *without* Gipsy5 suits, and the difference in their style was significant.

Despite these difficulties, motion capture technology still seems the only option for capturing the amount of animation data that would be required. As newer, more slimline motion capture suits are developed, the problems encountered with the Gipsy5 system may be avoided.

# 6 Conclusions and further work

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. From this a number of conclusions can be drawn:

- **Limit and then refine movements** - Many properties of Tango music limit the moves that can be chosen. After the number of possible moves has been reduced, the mood of the music can be used to determine which should be performed. Following this two-step method means that the mood analysis is in no danger of choosing inappropriate figures.
- **Incrementally add MIR techniques** - A capable prototype of the system could be built which performed only metre and tempo extraction, before adding other MIR techniques such as detection of timbre and intensity for mood, and separating the rhythms of each instrument.
- **Hierarchical structures are useful, but must be used with care** - Syntactic tree structures are convenient methods of storage and offer powerful processing facilities, however potential ambiguities when applying them to music data are difficult to deal with. Often, though, one structure will not be better or worse than another.
- **Choreograph and animate to a granularity of one step** - The base movement in Tango is the step, therefore that is the most natural granularity to work to. Systems which describe movements in finer detail such as Benesh and Labanotation are cumbersome. 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.

The techniques mentioned in this document are enough to start working on the system itself. The following steps are suggested to continue development of this system:

- **Improve beat detection** - Beat detection has already been explored as part of the development of LISA, however the current implementation is highly inaccurate. A more accurate method must be found if it is to be used directly in the simulation. Analysis which uses raw audio as its input would be preferable. Pikrakis et al. (2004) have explored a method which extracts metre and tempo simultaneously, assuming a limited range of possibly time signatures. For traditional styles of Tango, this is a reasonable assumption.

- **Improve functionality in LISA** - New forms of visualisation, such as Tzanetakis and Cook's (2000) "TimbreGrams", would be valuable. After these have been added, LISA could be used to test MIR techniques being explored: rather than entering all annotations in manually, support for those which can be determined automatically could be added to LISA, providing a unified system where MIR techniques being explored can be tested.

- **Explore the generation of Inter-Onset Intervals** - Automatic IOI generation from raw audio data would allow the system to work with symbolic data, which allows for many MIR techniques not possible with raw audio data.

- **Use IOIs to build preliminary hierarchical structures** - After simple trees have been built using IOIs, it will be possible to experiment with more complicated techniques such as those described in Bod (2002).

This would lay the framework for an application to be built which could output either one of the written forms of choreography described in section 5.1, or an animation based on motion capture data.

There are aspects to the system that have not been considered as part of this document. For example, if it is desired that multiple couples should dance simultaneously in the same area, they will need to avoid each other. This presents an interesting problem, as movement is limited to moves available in Tango. This has not been considered here, since the focus of this report has been on the analysis and interpretation of music, rather than dealing with practical limitations imposed by constraints on the dance floor. However, in order to develop a full system which included multiple dancers or any other obstacles, such as tables, chairs, or walls, this would need to be considered.

## References


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

Benesh Institute: 2005, About Benesh, http://www.benesh.org/.

Bod, R.: 2002, A unified model of structural organization in language and music., *J. Artif. Intell. Res. (JAIR)* **17**, 289–308.

Bodirsky, M.: 2004, A notation system for tango, http://www2.informatik.hu-berlin.de/~bodirsky/tango/index.html.

Cemgil, A., Kappen, B. and Desain, P.: 2000, Rhythm Quantization for Transcription, *Computer Music Journal* **24**(2), 60–76.

Cemgil, A. T., Kappen, H. J. and Barber, D.: 2003, Generative Model based Polyphonic Music Transcription, *Proc. of IEEE WASPAA*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY.

Chew, E. and Wu, X.: 2004, Separating Voices in Polyphonic Music: A Contig Mapping Approach, *Computer Music Modeling and Retrieval Symposium - Second International Symposium, CMMR 2004, Esbjerg, Denmark, May 26-29, 2004, Revised Papers*, Vol. 3310 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, pp. 1–20.

Collier, S., Cooper, A., Azzi, M. S. and Martin, R.: 1997, *Tango: The Dance, the Song, the Story*, Thames & Hudson, London.

Dance Notation Bureau: 2005, http://dancenotation.org/.

Dixon, S.: 2001, An empirical comparison of tempo trackers.

Dixon, S., Pampalk, E. and Widmer, G.: 2003, Classification of dance music by periodicity patterns, *ISMIR*, John Hopkins University, Baltimore, MD, pp. 159–165.

Eggink, J. and Brown, G. J.: 2003, Application of missing feature theory to the recognition of musical instruments in polyphonic audio., *ISMIR*.

Eggink, J. and Brown, G. J.: 2004, Extracting melody lines from complex audio., *ISMIR*.

Essid, S., Richard, G. and David, B.: 2005, Instrument recognition in polyphonic music, *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on* **3**.

Farnsworth, P. R.: 1958, *The social psychology of music*, The Dryden Press.

Frieler, K.: 2004, Beat and meter extraction using gaussified onsets., *ISMIR*.

Gómez, E. and Herrera, P.: 2004, Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies., *ISMIR*.

Goto, M.: 2001, A predominant-F0 estimation method for CD recordings: MAP estimation using EM algorithm for adaptive tone models, *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on* **5**.

Goto, M.: 2004, A predominant-F0 estimation method for polyphonic musical audio signals, *Proceedings of the 18th International Congress on Acoustics (ICA 2004)* pp. 1085–1088.

Goto, M. and Muraoka, Y.: 1999, Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions, *Speech Communication* **27**(3-4), 311–335.

Harford, S.: 2003, Automatic segmentation, learning and retrieval of melodies using a self-organizing neural network., *ISMIR*.

Hevner, K.: 1935, Expression in music: a discussion of experimental studies and theories., *Psychological Review* **42(2)**, 186–204.

Hevner, K.: 1936, Experimental studies of the elements of expression in music, *The American Journal of Psychology* **48**(2), 246–268.

Kennedy, M.: 2004, *The Concise Oxford Dictionary of Music*, Oxford University Press, Oxford Oxfordshire.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. and Friederici, A.: 2004, Music, language and meaning: brain signatures of semantic processing, *Nature Neuroscience* **7**(3), 302–307.

Koelsch, S. and Siebel, W.: 2005, Towards a neural basis of music perception, *Trends in Cognitive Sciences* **9**, 578–584.

Lartillot, O.: 2003, Discovering musical pattern through perceptive heuristics., *ISMIR*.

Lerdahl, F.: 1996, *A Generative Theory of Tonal Music*, MIT Press, Cambridge.

Li, T. and Ogihara, M.: 2003, Detecting emotion in music., *ISMIR*.

Liu, D., Lu, L. and Zhang, H.: 2003, Automatic mood detection from acoustic music data., *ISMIR*.

Loke, L., Larssen, A. T. and Robertson, T.: 2005, Labanotation for design of movement-based interaction, *IE2005: Proceedings of the second Australasian conference on Interactive entertainment*, Creativity & Cognition Studios Press, Sydney, Australia, Australia, pp. 113–120.

Longuet-Higgins, H. and Lee, C.: 1984, The rhythmic interpretation of monophonic music, *Music Perception* **1**(4), 424–441.

Lorr, M.: 1989, *Models and methods for measurement of mood*, Emotion: Theory, Research, and Experience. Volume 4: The Measurement of Emotions, Academic Press, Boston, chapter 2, pp. 37–53.

Lynch, S.: 2007, Introduction to NLP. Lecture given at the University of Teesside.

Martin, W. A., Church, K. W. and Patil, R. S.: 1981, Preliminary analysis of a breadth-first parsing algorithm: theoretical and experimental results, *Technical report*, Cambridge, MA, USA.

McCormack, J.: 1996, Grammar based music composition, *Complex Systems* .

Meehan, J. R.: 1979, An artificial intelligence approach to tonal music theory, *ACM 79: Proceedings of the 1979 annual conference*, ACM Press, New York, NY, USA, pp. 116–120.

Neagle, R. and Ng, K.: 2003, Machine-representation and Visualisation of a Dance Notation, *Proc. Electronic Imaging and the Visual Arts* .

Neagle, R., Ng, K. and Ruddle, R.: 2002, Notation and 3D Animation of Dance Movement, *Proceedings of the International Computer Music Conference (ICMC2002)* pp. 459–462.

Paiva, R. P.: 2005, On the detection of melody notes in polyphonic audio., *ISMIR*, pp. 175–182.

Patin, F.: 2003, Beat detection algorithms, http://www.yov408.com/beat/BeatDetectionAlgorithms.pdf.

Pauws, S.: 2004, Musical key extraction from audio., *ISMIR*.

Pikrakis, A., Antonopoulos, I. and Theodoridis, S.: 2004, Music meter and tempo tracking from raw polyphonic audio., *ISMIR*.

Pullen, K. and Bregler, C.: 2002, Motion capture assisted animation: texturing and synthesis, *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, pp. 501–508.

Raphael, C.: 2001, Automated rhythm transcription., *ISMIR*.

Roads, C.: 1985, Research in music and artificial intelligence, *ACM Comput. Surv.* **17**(2), 163–190.

Roads, C.: 1992, *The Music Machine*, The MIT Press, Cambridge.

Rodríguez, G. H. M.: 1917, *La Cumparsita*, http://www.todotango.com/spanish/biblioteca/partituras/partitura.asp?id=25.

Russell, J. A.: 1989, *Measures of emotion*, Emotion: Theory, Research, and Experience. Volume 4: The Measurement of Emotions, Academic Press, Boston, chapter 4, pp. 83–111.

Russell, S. and Norvig, P.: 1995, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs.

Saura, C.: 1998, *Tango: No me dejes nunca*. Film, Actors: Miguel Ángel Solá, Cecilia Narova; Music: Lalo Schifrin.

Takeda, H., Nishimoto, T. and Sagayama, S.: 2004, Rhythm and tempo recognition of music performance from a probabilistic approach., *ISMIR*.

Tanur, L. A.: 2005, *A geometric approach to pattern matching in polyphonic music*, Master's thesis, University of Waterloo, Waterloo, Ontario, Canada.

Tellegen, A. and Watson, D.: 1985, Toward a consensual structure of mood., *Psychol Bull* **98**(2), 219–35.

Tellegen, A. and Watson, D.: 1999, On the dimensional and hierarchical structure of affect, *Psychological Science* **10**(4), 297.

Thayer, R.: 1989, *The Biopsychology of Mood and Arousal*, Oxford University Press, Oxford Oxfordshire.

Turner, D.: 2006, *A Passion for Tango*, Dingley Press, Dingley.

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

Wiggins, G. A.: 1998, Music, syntax, and the meaning of "meaning", *Proceedings of the First Symposium on Music and Computers*, Corfu, Greece, pp. 18–23.

Witkin, A. and Popovic, Z.: 1995, Motion warping, *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, pp. 105–108.

Wold, E., Blum, T., Keislar, D. and Wheaton, J.: 1996, Content-based classification, search, and retrieval of audio, *IEEE MultiMedia* **3**(3), 27–36.

Zhang, T. and Kuo, C.-C. J.: 1998, Hierarchical system for content-based audio classification and retrieval, *in* C.-C. J. Kuo, S.-F. Chang and S. Panchanathan (eds), *Proc. SPIE Vol. 3527, p. 398-409, Multimedia Storage and Archiving Systems III, C.-C. Jay Kuo; Shih-Fu Chang; Sethuraman Panchanathan; Eds.*, pp. 398–409.

# Appendix A Score: *La Cumparsita*

*La Cumparsita* was composed by G.H. Matos Rodríguez (1917), a Uruguayan student, and sold to Roberto Firpo, who was at the time performing at a Montevideo café (Collier et al., 1997). This version has been taken from the freely available manuscript at TodoTango.com


Edicion Breyer Hnos.

# Appendix B Video References: *La Cumparsita*

The following video references were gathered to study the way dancers used the music of *La Cumparsita* to compose their dance.

1. http://www.youtube.com/watch?v=YpLqCth7DrY
   Juan Carlos Copes and Cecilia Narova. From the film, "Tango" (Saura, 1998). Fairly traditional.

2. http://www.youtube.com/watch?v=CRg7oDVXWaE
   Claudio Hauffman and Pilar Alvarez. This very short clip shows how the faster sections of the song could be used. It also gives an example of the interspersed rhythms - throughout this extract the man dances to the bandoneón, while he leads the woman in time to the double-bass at first, then later with the violins, before finally bringing her with him in time with the bandoneón.

3. http://www.youtube.com/watch?v=12eEf6IOuqE
   Vladimir Estrin and Irene Tumanyan. Another quite traditional clip.

4. http://www.youtube.com/watch?v=eHNz3vEnhUM
   Gustavo Naveira y Giselle Anne. Stage tango interpretation.

5. http://www.youtube.com/watch?v=72U1OjfaPBI
   Anastasia and Nicolás. Quite a showy tango, this video is interesting because the dancers often end sections by striking a dramatic pose, marking the structure of the music explicitly with their dance.

6. http://www.youtube.com/watch?v=Q5PTNIUud8k
   Oleh and Charla. This Tango doesn't use particularly fast or dramatic movements, but is danced sensitively. It demonstrates also that Tango is a dance for the most part about walking.

7. http://www.youtube.com/watch?v=lV6Fgptthy4
   Eduardo Cappussi and Mariana Flores demonstrate how differently it's possible to dance, even to a traditional Tango like *La Cumparsita*. They are well known for giving a very theatrical performance.

8. http://www.youtube.com/watch?v=kRNVBHxgidg
   http://www.youtube.com/watch?v=OzRbPquy3G4
   Social dancing. The dancers here are improvising, and having to be careful to avoid other dancers. Sometimes this very practical limitation has a greater affect on the dancers' choice of moves than musicality.

# Appendix C Specification

The title for this project is, "Analysis and Interpretation of Music for Dance". This project forms a part of a larger project, and is best explained in the context of the full application. This larger project would consist of a system whereby the user could enter music data in some raw form (for example the .wav or .mp3 file formats), and animated on-screen characters would dance to the music. This would have to be composed of a number of systems:-

- First the music would have to be analysed, using a combination of offline and realtime musical analysis techniques. This analysis would return information about the music, such as Tempo, Time Signature, Key, Timbre, Genre, Instrumentation, and so forth. This information would be passed on to the next component.
- Taking the high-level information about the music given by the first section, Artificial Intelligence techniques would be used to make decisions about which steps (from a given database of steps and figures) should be chosen.
- Having set up which steps to dance and determined the speed at which they should be performed, a choreography or sequence of animations would be passed to an animation component, which would take the animations and play them, performing blending and crossfading to give the impression of a smooth, single dance.

This project concerns itself with the preliminary research required to build such a system. Many fields will have to be looked into: what information is required from the music, how to get that information, how to use that information to make decisions, and how to represent those decisions on screen. Rather than trying to implement one of these systems in full, this project will explore all of them to assess the full scope of the application.

# Glossary

**Cadence**

"Any melodic or harmonic progression which has come to possess a conventional association with the ending of a comp., a section, or a phrase." (Kennedy, 2004)

**Contrapasos**

Syncopated steps.

**Milonga**

Simple form of Tango characterised by its lively tempo and irregular rhythm. Also a venue for dancing Tango.

**Milonga con Traspié**

Style of dancing for Milonga involving Traspiés and Contrapasos to liven the dance.

**Motif**

"The shortest intelligible and self-exitant melodic or rhythmic figure. Every 'theme' or 'subject' perhaps has several motifs, and almost every mus. passage will be found to be a development og some motif. But the word has, in mus.analysis, been used as a synonym for 'theme'." (Kennedy, 2004)

**Musical Sentence**

Term coined by some Tango teachers to refer to *sections* in music.

**Orquesta**

Lit. Orchestra; in this document the word is used interchangeably with *orquesta típica*

**Orquesta Típica**

A Tango Orchestra. Typically composed of a string section (usually violins), a bandoneón section, and a rhythmic section (piano and double bass).

**Perfect Cadence**

"Chord of the dominant followed by that of the tonic." (Kennedy, 2004)

**Programme music**

"Instr. music which tells a story, illustrates literary ideas, or evokes pictorial scenes." (Kennedy, 2004)

**Salida**

The basic step in Tango. Usually this is the first series of steps a dancer will learn.

**Section**

A way of partitioning a piece of music.

**Tango**

Argentine style of music originating in the 19th century, set to a 2x4 beat. Tango is also the umbrella term covering Tango, Vals, Milonga, and modern styles, and also covers the various styles of dance followed for each of these different kinds of music.

**Tango Contemporáneo**

Contemporary style of Tango music. Often tango contemporáneo will feature modern synthesised instruments and will not follow one of the traditional Tango, Vals, or Milonga styles.

**Tango Nuevo**

Modern style of Tango music as pioneered by Ástor Piazzolla. Tango Nuevo incorporates influences from jazz and classical music.

**Traspié**

A step stressing a particular beat, usually syncopated.

**Vals**

Lit. Waltz; however in the context of this report it refers to the Argentine interpretation of the classic Waltz. A subset of Tango, played in 3/4 time.

**Vals Criollo**

Another name for *Vals*.