WHOLODANCE

Whole-Body Interaction Learning for Dance Education

Call identifier: H2020-ICT-2015 - **Grant agreement no**: 688865

Topic: ICT-20-2015 - Technologies for better human learning and teaching

Deliverable 3.3

Report on music-dance representation models

Due date of delivery: December 31st, 2017

Actual submission date: December 31st, 2017

Start of the project: 1st January 2016
Ending Date: 31st December 2018

Partner responsible for this deliverable: POLIMI

Version: 1.1

Dissemination Level: Public
Document Classification

Title	Report on music-dance representation models
Deliverable	3.3
Reporting Period	Second
Authors	Massimiliano Zanoni, Michele Buccoli, Augusto Sarti, Fabio Antonacci, Sarah Whatley, Rosemary Cisneros, Pablo Palacio
Work Package	WP3
Security	Public
Nature	Report
Keyword(s)	Music, dance, representation, models

Document History

Name	**Remark**	**Version**	**Date**
Massimiliano Zanoni	TOC and objectives	0.1	12/12/2017
Massimiliano Zanoni	Music Representation model	0.2	20/12/2017
Pablo Palacio	Description of the use-case Piano&Dance	0.3	21/12/2017
Sarah Whatley and Rosemary Cisneros	Description of the relation between music and dance within Flamenco	0.4	20/12/2017
Michele Buccoli	Movement representation model	0.5	24/12/2017
Massimiliano Zanoni, Augusto Sarti and Fabio Antonacci	Review and finalization	0.9	26/12/2017
Massimiliano Zanoni	Finalization after the internal review	1.0	30/12/2017
Anna Rizzo	Final reviewed version	1.1	31/12/2017

List of Contributors

Name	Affiliation
Massimiliano Zanoni	Polimi
Michele Buccoli	Polimi
Augusto Sarti	Polimi
Fabio Antonacci	Polimi
Sarah Whatley	Coventry University
Rosemary Cisneros	Coventry University
Pablo Palacio	Stocos

List of reviewers

Name	Affiliation
Antonella Trezzani	Lynkeus
Anna Rizzo	Lynkeus

1 Executive Summary

This Deliverable will be based on the outcomes of the task T3.1.3 *Joint music-dance representation models*. Dance and music are highly dependent in many dance genres: in some genres, dance performances cannot be even executed without a reference music. For this reason, the analysis of the correlation between music and movement is essential in the WhoLoDancE project.

Music and movement are different in nature, so, to study their correlation, we will define two representation models: a music representation model and a movement representation model. The two models are both based on the extraction of a set of representative features able to capture specific aspects of the relative signals. The two models are then used to study the dependence and the interaction between music and movement in two use-cases: *Piano&Dancer* performance and joint movement-music analysis in Flamenco.

Table of contents

2 Objectives

Dance practice presents a wide diversity across genres and contexts. Across the centuries, music has played a very important role in the dance performance experience for both dancers and the audience in several dance styles and contexts. For some genres, like Ballet, we cannot think about the dance practice disjoint from the music. On the other side, some music genres have been specifically created to be thought as the sound track for dance performances. For all these reasons, we can define the dance practice as multidisciplinary and multimodal by nature.

In order to study and model the interaction and the dependence between music and dance it is important to develop an exhaustive representation model. However, music and dance are very different disciplines in terms of both practice and language. For this reason, there is the need to develop two disjoint representation models: one for music and one for movement. The music representation model developed in this project is described in Section 3, whereas the movement representation model is described in Section 4.

A joint analysis is then needed to build the link between the two models. In dance performances music has several roles. From one side, it provides a melodic and rhythmic structure on top of which movements can be developed; this is the case for Flamenco and the Greek music, considered in this project, where the flow of the melody and the pattern of the beat sequences build the base for the choreography. On the other side, music does not represent only a base for dance practice, but the music performers and dance performers often interact. This is again the case for Flamenco and Greek music in live performances. Thanks to the recent development of digital technology and contemporary electronic music composition, the role of the music performer changed. The music performer can now use a computer to design sounds and live coding the composition. Based on the two aforementioned representation models, the use of sensors and cameras allows the composition to slightly change according to the movements of the dancer and, hence, the music performers and dance performers to interact. A use case is *Piano&Dancer* presented in Section 5.1.

Another aspect to take into account in the project is the relation between a dance genre and the social and cultural context in which the style was created. Dance conveys emotions and messages through a proper language that is specific to the style. Music used in a performance needs to be correlated to the dance experience in order to enhance the power of the message. In Section 5.2, we describe the specific case of Flamenco.

3 Music representation model

Music is strictly dependent on cultural factors and it expresses many anthropological aspects, such as geographical origin, social condition, religion and so on. For this reason, as long as societies evolved, the same happened to the formalisms used to describe musical content. This is particularly evident in the evolution process of verbal expressions used to describe music taken from the natural language. Words like *groovy*, for instance, which have been included in English dictionary since a long time, have not been used in the music context until the '30s. This is the reason why people of different ages tend to assign slightly different meanings to terms. A teenager can relate the term *groovy* to a specific type of electronic music, whereas for a person who was a teenager in the '50s will point to a specific rhythmic pattern that was popular at that time.

The aforementioned social and cultural factors highly impact aspects of sound and music production and aspects of human perception. The way humans perceive and describe musical characteristics has been the focus of many disciplines of the Music Information Retrieval (MIR) for decades: Computer Science, Musicology and Psychology.

Specifically, the MIR community have been studying and investigating this subject for a decade. MIR is a multidisciplinary research area that deals with the retrieval of useful information from music. Music includes a lot of elements and can be perceived under several perspectives, related to the activity we are involved in: Performing/Composing and Listening/Perceiving. The former aims at producing acoustic waveforms by the interpretation of structural information, such as melody, harmony, rhythmic pattern, and so on. The composition and the way performers play invoke highly subjective emotions in the listeners. The latter is much related to the perceptual dimension (emotion, genre, theme, timbre). All the elements involved in music can be formalized and described in a hierarchy going from a lower level, which is related to sound signals, to a higher level, associated with the perception of sound and the structure of music. Those elements are referred as *features* or *descriptors*. For the purposes of the project, we adopt here a two-level representation model where each level captures a different aspect. An explanatory scheme of the two levels is shown in Figure 3-1.

The first level of the representation model comes from the physics area. Many information (descriptors) can be computed directly from the audio signal, such as timbral, loudness, etc. They refer to the audio signal and are highly objective. However, a formal description using such approach totally lacks in semantic and loses each connection to the artistic and structural characteristics of music excerpts. Those descriptors are referred as **low-level features (LLF)**.

Melody, rhythm, such as harmony, are examples of descriptors related to structural information of music and they are particularly important in music composition and execution. Descriptors able to capture these components compose the mid-level of abstraction. We refer to these descriptors as **mid-level features (MLF)**. MLFs introduce a first level of semantics and they intend to fill the gap between audio signals and music annotation and description. They combine acoustics proprieties and musicological knowledge.

Even though the two levels seem to be complementary, as they are devoted to capturing different aspects of music, they are also strictly related to each other due to their interdependency. An in-depth understanding of the relationship between the two levels is far from being well understood and research groups spend much effort on this topic. However, we know that there is a hierarchical dependency among them. From both perceptual and computational perspective, MLFs are built on the composition of low-level ones (Figure 3-1).

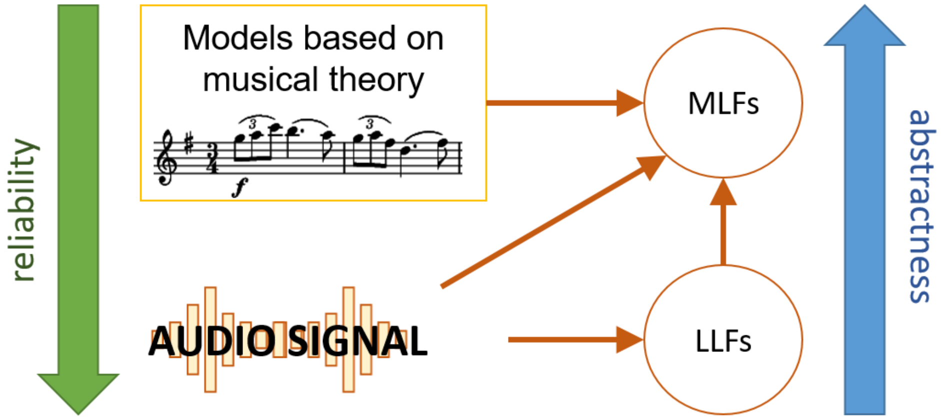

Figure 3-1. General scheme of LLF and MLF extraction

3.1 Low-level features

The ability of human to perceive emotion in music content, the attitude to discriminate and isolate sounds, the listening experience in general, are highly basically related to several acoustic elements that characterise the sound source. The representation model adopted in the project is based on simple descriptors able to capture the aforementioned elements. We call those descriptors low-level features (LLFs).

Spectral features, for instance, have a fundamental role in the process to discriminate a piano instrument from a trumpet, or a shout by a scream. Since each of the feature captures a very narrow aspect of the sound, even if the contribute of each is important, the listening experience is related to a combination of them. The investigation of the best set of features to use is still an open issue in audio research field.

LLF descriptors are the basic components of MIR applications, and sound and music analysis and retrieval. They are at the base of several applications like musical genre classification: in the discrimination process between Classical and Hard Rock, the introduction of spectral features, such as Spectral Flatness, would be highly effective, given that it quantifies the presence of noisy components. As an example, in Figure 3-2 Spectral Flatness values extracted from a classical piece (*Mozart Symphony N40in G Minor K 550*) and a rock song (ACDC - *Back in Black*) are shown. Having a descriptive set of LLFs for music genre recognition allows to deal with different music styles and genres. This is important for a project where different dance genres are considered.

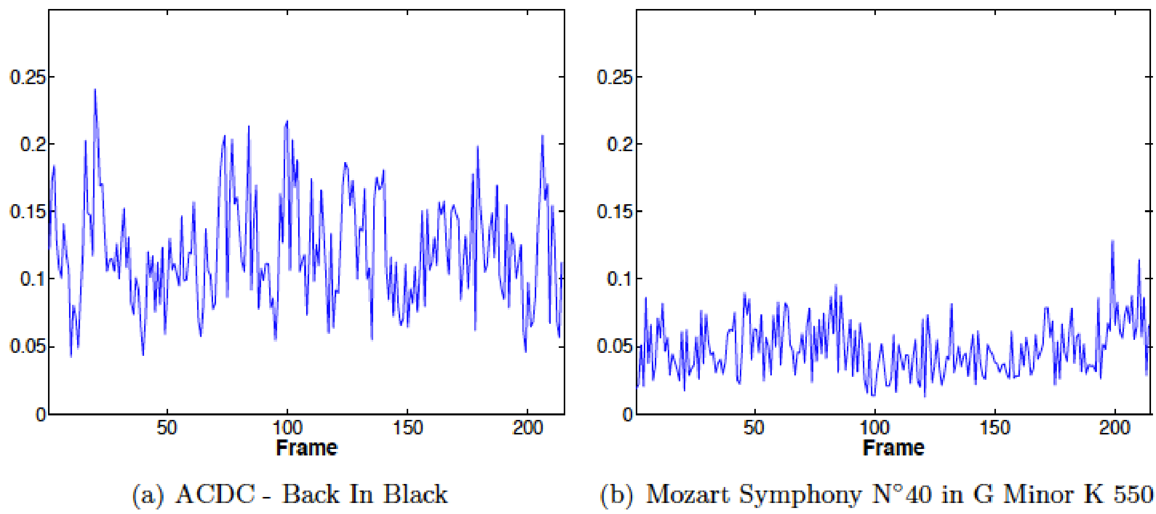

Figure 3-2. Spectral Flatness related to a rock song (a) and to a classical piece (b). The extracted values are related to the 5s long segment from the 100s to the 105s in both pieces.

The choice of the highly representative description model for music is important for the goals of the project in order to have a solid analysis method to combine with dance analysis methods. For this reason, we collected most of the relevant state-of-the-art LLFs.

As far as acoustics elements are considered, our listening experience is based on three main aspects of the sound: **energy**, **tempo**, **timbre**. According to the captured acoustic characteristics, low-level features are categorized as followings: ***energy features***, ***temporal features***, ***spectral features***.

The listening experience seems to be strongly related to the energy distribution and evolution within the musical excerpts. Listening to music at low loudness or at high loudness produces a different perceptual experience. As an example, the variation of the energy values is highly influenced by the type of music. Jazz music tends to be characterized by frequent energy fluctuation along the piece, whereas heavy metal songs, due to the massive use of wave compressors, tend to preserve a uniform higher loudness.

Sound evolves in time and so our perception of sounds. For this reason, it is important to capture the evolution of sound in time by considering a set of time-related (temporal) LLFs.

Spectral features are highly related to the Timbre. Spectral Features are all computed through spectral analysis techniques on Short Time Fourier Transform (STFT) of the audio signal and they represent a set of measures of the shape of the magnitude spectrum.

For all the features, to increase the temporal accuracy of the analysis, audio is decomposed into short pieces (frames) of the signal. The technique is called frame decomposition.

We report in Table 3-1 the set of features that compose the low-level music representation model adopted in this project. Further details can be found in D3.6. Some features described here are not described in D3.6 since they have been included in the model very recently.

Type	Name	Type	Description
LLFs	Root Mean Square Energy (RMSE)	Energy	An indicator of the energy of the audio signal in a frame from time-domain or frequency-domain representation.
LLFs	Low Energy	Energy	Defined as the percentage of frames showing the RMS value minor than the average RMS along the whole piece.
LLFs	Loudness	Energy	Measure of the sound intensity formulated on human perception model; it outlines the frequency spectrum components that highly contribute to loudness.
LLFs	Zero Crossing Rate	Temporal	Rate with which a frame of an audio signal crosses the zero; it is an estimation of the noisiness of the signal.
LLFs	Amplitude Envelope	Temporal	The maximum values of amplitude for the time-domain representation of the signal.
LLFs	(Log) Attack Time	Temporal	The duration of the attach phase of a note onset; it provides information on the timbre.
LLFs	Attack Slope	Temporal	The slope of the attack phase; it is another way to estimate the attack.
LLFs	Attack Leap	Temporal	The amplitude difference between the beginning and the end of the attack phase.
LLFs	Decrease Slope	Temporal	The slope of the deceased phase.
LLFs	Onset Duration	Temporal	Duration of a note, from the attack to the release phase.

LLFs	Spectral Centroid	Spectral	The centre of energy of the distribution of the spectrum over the frequency bins.
LLFs	Spectral Spread, Skewness, Kurtosis, Bandwidth	Spectral	Several statistical moments applied to the distribution of the magnitude of the spectrum over the frequencies.
LLFs	Spectral Flatness	Spectral	The degree of resemblance between the distribution of the magnitude of the spectrum and a flat spectrum; it is an estimator of the noisiness of the sound.
LLFs	Spectral Entropy	Spectral	The amount of information encoded in the distribution of the magnitude of the spectrum; it is another estimator for the noisiness.
LLFs	Spectral Contrast	Spectral	Difference between peaks and valleys in the magnitude spectrum distribution.
LLFs	Spectral Roll-off	Spectral	The frequency below which the 85% of total energy is contained.
LLFs	Spectral Brightness	Spectral	The amount of energy above a certain frequency (given as a parameter).
LLFs	Spectral Roughness	Spectral	An estimation of the roughness depending on the frequency ration of pair of sinusoids closed in frequency (beating phenomenon).
LLFs	Spectral Inharmonicity	Spectral	Amount of energy outside the ideal harmonic series (fundamental frequency and its harmonics).
LLFs	Spectral Irregularity	Spectral	Degree of variation of two successive peaks in the spectrum.
LLFs	Polynomial Features	Spectral	The coefficients of a polynomial fitting over the distribution of the magnitude spectrum.
LLFs	Fluctuations	Spectral	Spectrum computed over the different frequency bins of the real spectrum; it estimates rhythmic periodicities.
LLFs	Mel Spectrogram	Spectral	The spectrogram mapped in the mel scale.
LLFs	Mel-Frequency Cepstral Coefficients (MFCCs)	Spectral	The coefficients computed from the mel-spectrogram by means of the DCT.
LLFs	Similarity Matrix	Spectral	Self-similarity among the representation of different frames of the signal.
LLFs	Spectral Roughness	Spectral	It is a perceptual-based feature and it measure the roughness of the timbre.
LLFs	Tonal Dissonance	Spectral	It is a perceptual-based feature and it computes dissonance using the tonal components of the spectrum.
LLFs	Spectral Dissonance	Spectral	It is a perceptual-based feature and it computes dissonance using all the components of the spectrum.
LLFs	Sharpness	Spectral	Based on psychoacoustical models, it is a subjective measure of sound on a scale extending from dull to sharp.

LLFs	Timbral Width	Spectral	Timbral Width gives a measure of flatness using a psychoacoustical approach.

Table 3-1. Set of LLFs included in the LLF representation model adopted in the project

3.2 Mid-level features

In the composition process, composers organize sounds of musical instrument to create the idea they have in mind. The final composition is the collection of sounds organized following a set of structural rules. The characteristics of this organization are strictly personal and dependent on many cultural factors that make the piece unique. Therefore, structural information is fundamental in music analysis and description. This is generally classified into **Tonal** and **Rhythmic** categories. In the MIR field, the mid-level descriptors refer to the structural information of songs. Mid-level descriptors are built using low-level features and they present an intermediate semantic layer that combines straightforward signal processing techniques with some *a priori* musical knowledge. In the literature, the problem of retrieving structural elements of a composition is divided into subgroups: pitch detection, rhythmic pattern analysis, harmony detection and so on.

According to the structural characteristics that they capture, MLF are categorized as followings: **rhythmic features** and **tonal features**.

We report in Table 3-2 the set of features that compose the mid-level music representation model adopted in this project. Further details can be found in D3.6. Some features described in the table are not described in D3.6 since they have been included in the model very recently. Within the project, several MLF algorithms have been developed related to the specific issues of beat tracking, chord, key and mode recognition and music complexity analysis. In next sections, we provide a brief description on those methods.

Type	Name	Type	Description
MLFs	Chromagram	Tonal	Distribution of the pitches over time, usually regardless to the original octave (pitch class profile).
MLFs	Tonal centroid (Tonnetz)	Tonal	Projection of the chords along circles of fifths, of minor thirds and of major thirds.
MLFs	Pitch	Tonal	Pitch curves or note events of a melody.
MLFs	Key	Tonal	Harmonic key.
MLFs	Mode	Tonal	Modality of the key, i.e., major vs minor.
MLFs	Chords	Tonal	The chords played during a musical piece.
MLFs	Harmonic Complexity	Tonal	Measure of music complexity concerning the complexity of tonal chord sequences.
MLFs	Tempogram	Rhythmic	Local autocorrelation of the onset strength envelope, to estimate the distribution of onsets periodicities.
MLFs	Beats (beat tracking)	Rhythmic	Instants when the musical beats occur.
MLFs	Pulse Clarity	Rhythmic	Strength of the estimated beats.
MLFs	Tempo	Rhythmic	The speed of a performance in beats per minute.
MLFs	Note onset	Rhythmic	Instants where a musical onset occurs.
MLFs	Event Density (Onset Rate)	Rhythmic	Number of note onsets per second.
MLFs	Metre	Rhythmic	Hierarchical metrical structure of the onsets.

MLFs	Metroid	Rhythmic	Centroid of metre.
MLFs	Rhythm Strength	Rhythmic	Measure of the average intensity of the onsets.
MLFs	Event Density	Rhythmic	Measure of the compactness of the rhythmic texture and is computed as the average frequency of notes per second.
MLFs	Rhythm Regularity	Rhythmic	Measure of the regularity of the onsets over time.

Table 3-2. Set of MLFs included in the MLF representation model adopted in the project

3.2.1 Music beat tracking

Rhythm is one of the fundamental aspects in music as well as in dance. For this reason, capturing the rhythmic characteristics of music is an essential factor in our representation model. Most musical pieces evolve according to an underlying unit of time. This unit of time is called **beat**. The analysis of the progression of beat on a music piece is at the base on a wide range of applications of musical information extraction. Beat instants, in fact, can be used for music segmentation and processing (e.g., structural segmentation, interactive accompaniment, etc.); for beat-synchronous analysis (e.g., drum transcription; chord extraction), and more.

Given the importance of rhythm in several dance genres, beat is also one of the most relevant features for joint music-movement analysis. In many dance genres, in fact, the sequence of beats is used as a grid on which to lay out the choreographies. Dance movements, then, visually highlight the beat to show synchronicity with the background musical piece. In this project, we focus on automatic extraction of the beats (*beat tracking*) from audio signal, starting from a LLF called Onset Detection Function (ODF). ODF highlights the transients of the signal. The musical transients often highlight important events in composition, like, for example, the staring time of a new note, which usually happens with the beat.

In order to perform the beat tracking, our method first analyses the periodicities of the ODF with the purpose of finding an initial estimate of the Inter-Beat Interval (IBI), which is generally time-varying. The IBI is usually also called *tempo* and can be expressed in beats per minute. Successively, a dynamic programming algorithm is employed to efficiently find the sequence of beats through a joint optimization process that maximizes the ODF values of the beat instants and match the estimated IBI.

We focused on this last joint optimization process and proposed a novel strategy based on an efficient generation and joint steering of multiple trackers (paths) [Di Giorgi, 2016]. This solution leads to improved computational efficiency with respect to dynamic programming methods; a computational improvement factor of 2 has been obtained with a negligible accuracy loss.

3.2.2 Chord and key recognition

A second essential aspect of music is harmony. Chord sequences provide a simple and effective representation of harmony. For this reason, we include chord identification and chord sequence analysis in our music representation model. Among the many applications of chord recognition, there are cover identification, interactive music training and automated transcription. The notion of chord and key are well known aspects of the music theory. A chord is a set of harmonically related musical pitches (notes) that sound almost simultaneously. The key can be divided in the *key root*, also called *tonic*, that is defined as the most important pitch class and the *key mode* that is the subset of pitch classes used in a song, relatively to the key root (e.g., major or *Ionian*, minor or *Aeolian*).

The basic step for the chord and key identification is the chromagram, also called pitch class profile (PCP), which is a time-varying 12-dimensional signal obtained from the spectrogram of the audio signal.

In order to recognize the possible sequence of chords, it is important to model the probability of the transition for a sequence of chords to a new one, based on some harmonic rules. The probabilistic formulation we used for the automatic chord recognition system is modelled using the Dynamic Bayesian

Network (DBN), a graphical model that generalizes the Hidden Markov Model (HMM), managing graphs with any number of hidden and observed nodes. In this model, nodes can represent observed random variables, in our case the chromagram, or hidden random variables that can be inferred given the values of the observed variables, in our case the chord, key and bass note. Chord sequence evolves in time, for this reason DBN explicitly model time dependencies; particularly, this is achieved through a set of transition probability distributions, which model the probability of a variable conditioned on the values of a set of related variables at the previous time frame. A natural discretisation of time dimension is obtained using beat instants, which are estimated using our beat extraction algorithm; this process is called beat-synchronization.

Since within the project we deal also with non-pop and rock music, it is important to mention that, differently from other methods in the literature, which only considers the major and minor modes, we can extract a richer description of key by including two diatonic modes not previously considered, i.e., the *Mixolidian* and the *Dorian*. This allows to provide a more complete and expressive model of the harmonic-induced mood, which is more suitable for tonal music languages typically used as the background music in Flamenco, ballet and Greek dance genres.

To achieve this goal, we proposed a set of new parametric distributions. Comparing with similar chord recognition algorithms we found a significant advantage in using such a larger state space for the key variable.

3.2.3 Harmonic complexity of music

The study of the complexity is a very hard task, but at the same time important not only in music field, but also for the analysis of artistic performances where more than an artistic area is involved. The analysis of the complexity of music can, for instance, be correlated to the complexity of movements in dance. Even if everyone can think of relatively complex or simple songs, assigning a precise meaning to the word "complexity" is as hard as describing our mental representation of music. Nevertheless, it is reasonable to address different musical facets separately: within the project we focused on the analysis of harmony complexity [Di Giorgi, 2016].

The method here proposed is based on a data-driven language model of tonal chord sequences. Since we deal with the harmonic complexity of tonal chord sequences, we provide hereafter a definition of tonal harmony and a hierarchy of the main interpretations of harmonic complexity. Tonal Harmony (TH) has been the main theoretical framework for harmony in Western music since the baroque period (around 1650). It implies that there is a certain pitch class, called the tonic, acting as a referential point. The building blocks of this framework are the chords, whose function can be established by the interval between their root and the tonic.

Several studies on harmonic complexity in music are in the literature in research areas such as musicology and psychology. In this project, we design the architecture of a language model of tonal chord sequences that we use to model cognitive expectation. Methods focusing on cognitive harmonic expectations estimate the amount of surprise using rules from music theory or machine learning models.

We show that the data-driven language model of tonal chord sequences has ability to automatically estimate the perceived complexity. Particularly, we show that the probability of a chord sequence, given by a language model trained on a sufficiently large dataset, can be related to cognitive complexity.

The model is trained on a novel dataset containing approximately half a million chord sequence annotations, therefore being the largest dataset of this type, to the best of our knowledge. The dataset has been collected from the website ultimateguitar.com and contains annotations uploaded by users. Although these data may have questionable accuracy, we believe that the benefits of such a large data set outweigh any effect of the noise in the data. Apart from transcription errors, the differences from high quality annotations may include chord simplification, substitution using similar chord functions and reharmonization, but we cannot quantify them. It is known from music theory that we can formulate much better hypotheses on chords when we know the key of the piece, because of the strong relationship between the two entities. In fact, the key is often interpreted as a latent variable in probabilistic models of chords. Therefore, we estimate the tonic as a pre-processing step, and represent chords using a relative notation (i.e., Cmaj as 0maj when the tonic is C). Estimating only the tonic, but not the specific scale or mode, allows the model to learn common chord progression patterns, such as modal interchange and secondary dominants that exploit notes from different scales or modes. Furthermore, this way we ensure a maximally efficient usage of the models, avoiding a 12-times augmentation of the dataset with the transposed versions of each song.

The machine learning language models used in the project are: Prediction by Partial Matching (PPM), discrete Hidden Markov Model (HMM) and Recurrent Neural Network (RNN). We investigate to find the optimal combination of these three models. In order to quantitatively evaluate our hypothesis, we collected the ratings of complexity and preference obtained from a listening test. Specifically, we performed two perceptual tests, and we assessed that the probability of a chord sequence evaluated by our model is strongly correlated to the high-level concept of harmonic complexity.

4 Movement representation model

The movement can be captured with different devices that generate many signals as output, from video streams to motion signals related to accelerometers up to the complete motion capture, as described in deliverables D2.3 and D2.6. The representation model for movement involves capturing the motion properties by means of the automatic extraction of *motion features*. The taxonomy of the features follows their semantics, i.e., the level of abstractness from the signals, ranging from physical signals related to the kinematics properties, to high-level qualities that require memory and context to be assessed [Camurri2016]. The level of the feature is related with the spatial and temporal scale within which they are computed: low-level features are computed in real time, frame by frame, often locally to single limbs, while high-level qualities require a wider time-span to be assessed (typical within a few seconds) and are global for the whole body or performance. A graphical representation of the different levels of abstractness for motion features in shown in Figure 4-1 and discussed in detail in D3.4, while a comprehensive description of the framework for multimodal analysis of data is provided in D3.3. In the following, we briefly present the models of movement representation for different signals and levels.

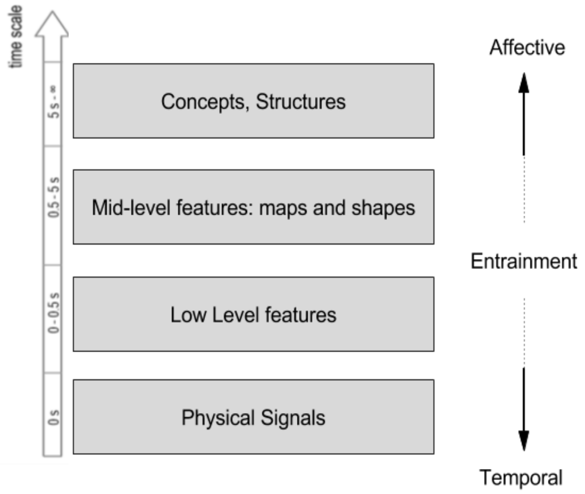

Figure 4-1. The level of abstractness for motion features

4.1 Physical Signals

The physical signals are defined as the properties directly related with the acquisition of the motion captures or the physics of the movement [Camurri2016]. The signals recorded from motion capture provide a skeleton representation that is easily processed to extract the bio-mechanical relationships across the different limbs (see D2.6). The motion capture signal can be represented in many ways, such as the global position of the joints in the space (with regard to an absolute system of reference), or the local transformation of each joint (rotation, translation, scaling) with regard to the transformation of the parent joint (for instance, the wrist transform is a child of the arm transform). From the motion capture signal, it is easy to extract the kinetics of the underlying motion: the velocity of each joints is computed as the first-order derivative of the global position of the joints dimension-wise; the accelerations are the first-order derivatives of the velocities; and so on.

Other devices only extract one of these kinetic properties, or do not provide an explicit description of the motion. For instance, accelerometers are, by definition, sensors that are only able to capture the acceleration of the limb to which they are attached, with regard to a local system of reference (i.e., defined by the device, not by the external world). In order to be reliable for the processing and extraction of higher-level properties, the multimodal signals must be comparable, i.e., able to extract the same physical quantity with similar values. Back to the example of the accelerometers, it is still possible to compute the magnitude of the vector of the acceleration captured from the device and compare it with the acceleration from the motion capture system if considering the same limb. For other devices, however, the comparison may be less immediate.

Video signals capture a stream of images that encode the motion under many layers of abstraction. The image pixels encode a value of Red, Green and Blue (RGB), and proximity of pixels compose shapes, which compose physical entities that can then be interpreted as a dancer. As a matter of fact, the video signal is a 2-dimensional projection of a 3-dimensional scene, which is consequently under-dimensioned and lossy. This is also the reason why even human spectators are not able to perceive all the details of a movement from a video signal, not to mention the video quality and the clarity of the performer with respect to the background. Many techniques, either rule-based and data-driven (learned through artificial intelligence) have been proposed for reconstructing or detecting 3D human action from video signals, but the results are not comparable to the accuracy achieved by state-of-the-art motion capture system.

Within the project, we designed and extracted two simple physical signals from the video signal: the overall horizontal and vertical velocity perceived in the video, i.e., the horizontal and vertical components of the average of the velocity vectors through all the limbs. For the aforementioned reason, the extraction of these features is not easy and require several steps of video signal processing techniques. We first filter the video signal with a Gaussian blur filtering, which removes the possible static background, highlighting foreground elements that are likely to be the moving performers. We then compute the first-order difference over frames and take the absolute values of the resulting signal into consideration. The aim of this processing stage is to automatically extract the silhouette of the dancers with a shortly delayed copy. From this representation, we extract the motion velocity as the Dense Optical Flow, which detects the direction of the apparent motion of subjects present in the scene as an estimate of the local apparent motion of each pixel. This is done by comparing its neighbourhoods over pairs of successive frames. If the background is static enough, so that it has been removed in previous processing stages, and the camera was steady during the recording, the Dense Optical Flow provides a reliable descriptor of the direction and the amount of movement in the scene, i.e., its velocity. We average the velocities over pixels and consider the two components of the vectors, which represent the global amount of movement toward the horizontal and vertical directions. A final smoothing filter is applied to the velocity signals, with the result of removing the spikes that are due to sudden scene or light changes.

If the position of the camera device that captured the video signal is known, it is possible to obtain an analogous 2D projection of the motion capture signal, compute and average the velocities of the joints over the horizontal and vertical direction of the 2D projection. The resulting velocity signals extracted from motion capture recordings are then comparable with those extracted from video recordings. This is the core of an algorithm we developed to analyse and measure the correlation between a video and a motion capture recording, which we ultimately used to assess how much the two recordings are likely to be captured from the same session, and with which time alignment [Buccoli2017]. This is the core of the synchronization algorithm we mentioned in D3.6. In Figure 4-2 we can see an example of two recordings and the corresponding velocity signals, with a high correlation that allows to identify the time-alignment value between the two.

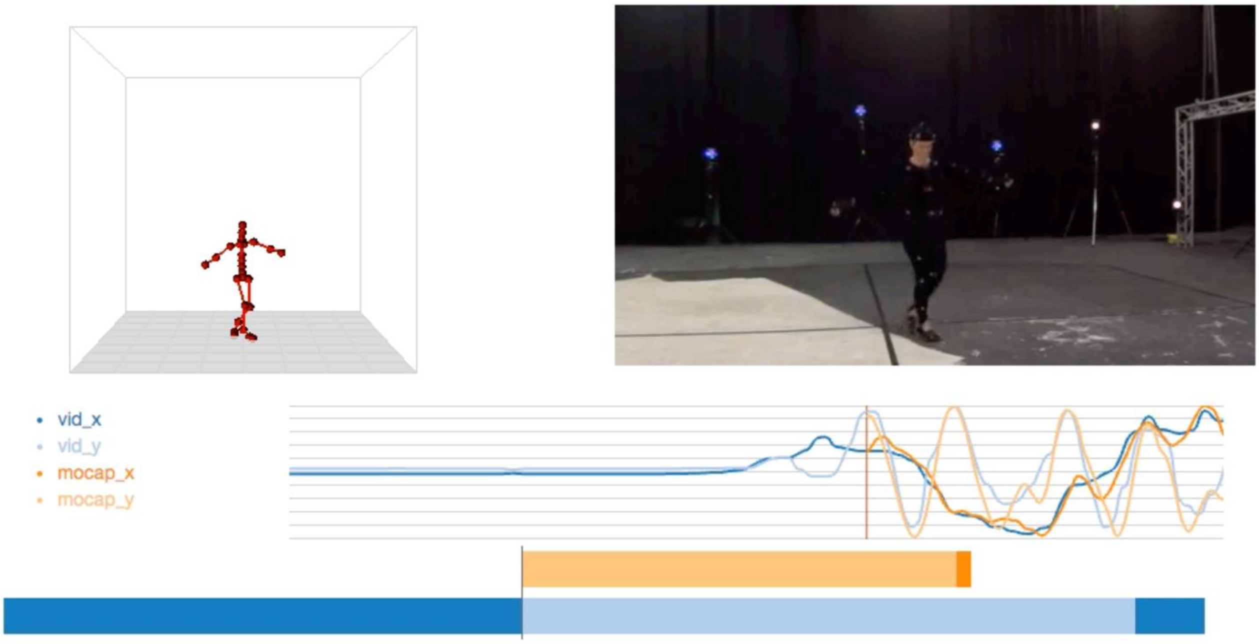

Figure 4-2. An example of the mocap-video synchronization, with the overall motion velocity signals in the horizontal (x) and vertical (y) directions of the two streams

While it is possible to extract many information from different devices, in the WhoLoDancE project we focused on the motion capture recordings, since they provide the greatest amount of information with the highest accuracy.

4.2 Low-Level features

By processing the physical signals with rule-based techniques, we can extract low-level features, which describe complex properties of the motion with a low level of semantics. Low-level features are related with the posture of the dancer, or its balance [Camurri2016]. A list of the main low-level features considered for motion is shown in Table 4-1 for the sake of exemplification. The properties described by low-level features typically occur within a short interval of time or even instantaneously (i.e., one feature for each frame). Low-level features can be multivariate, for example when considering different axis, or univariate, e.g., from the aggregation over directions.

Name	Description
Kinetic Energy	The kinetic energy of the joints, possibly weighted by their masses, using weights from biometric tables.
Motion Index or Quantity Of Motion (QoM)	Area of the difference of the silhouettes' area computed on consecutive frames
Postural Contraction	A measure of the extent at which the body posture is close to its barycentre.
Postural Symmetry	Geometric symmetry of a posture with respect to a plane or an axis.
Smoothness	A joint moving according to the specific laws from biomechanics defining smoothness
Postural and dynamic Balance,	Computed from (i) the measure of the projection to the floor of the barycentre of the body in the area defined by the feet and (ii) the ratio between acceleration of the barycentre of the head and of the barycentre of the body.
Change of Weight between Feet	Computed from feet pressure patterns as measured by a sensitive floor.
Postural Tension	A vector describing the angles between the adjacent lines identifying feet (the line connecting the barycentres of the feet), hip, trunk, shoulders, and head directions. This is inspired by classical paintings and sculptures where such angles are exploited to express postural tension.

Table 4-1. List of the main motion low-level features extracted from motion capture recordings

4.3 Mid-level features

While the physical signals involve the raw motion data or their basic kinetic properties, and the low-level features are directly computable from those, mid-level features capture the properties with a higher degree of semantics, i.e., that are perceptually relevant for the dancers, teachers or spectators in general. There is a gap between the objective signal level and the (usually subjective) perceptual level, which is known in the literature as *semantic gap*. This means that in the current scenario, mid-level features provide a taxonomy for *amodal* descriptors that can be seen as an aggregation of different low-level and physical features, while a clear and closed formulation to extract them may not exist. Example of the Mid-level features include the Movement Qualities, whose extraction through data-driven and model-based techniques is described in D3.5.

Mid-level features capture properties that concern structural aspects, which are related to one single movement unit [Camurri2016]. It is required then that the performance is *segmented*, i.e., divided into the movement units. Since a manual segmentation is cognitively hard for the annotators, some automatic techniques have been developed (see D3.5 for details), with the aim of processing a motion sequence and computing the salient movements. The performance achieved by the automatic techniques is currently under evaluation, but it may result to be not accurate for some situations (e.g., specific dance genres). If the segmentation is not available or reliable, the mid-level features are computed over fixed-length time interval of 0.5-5 seconds, in order to take a sufficiently long dynamic evolution into consideration. A list of the main mid-level features considered for motion is shown in *Table 4-2* for the sake of exemplification.

Name	Description
Contraction	Movement contracting along time.
Dynamic Symmetry	Symmetry of movement features
Directness (Laban's *Space*)	Movement to directly reach a target position (Direct vs. Flexible).

Lightness (Laban's ***Weight***)	How gravity influences a movement, e.g., based on analysis of vertical component of accelerations.
Suddenness (Laban's ***Time***)	Rapid change of velocity (Sudden vs. Sustained) in a movement.
Impulsivity	Movement which is sudden and not prepared by antagonists muscles
Equilibrium	The extent at which a movement is balanced, i.e., the tendency to fall or to keep a stable balance.
Fluidity	A fluid movement is smooth and coordinated (e.g., a wave-like movement propagation through body joints).
Repetitiveness	The extent at which a movement exhibits repetitive patterns.
Tension	The extent at which a movement exhibits rotation of multiple planes, including spirals (computed from Postural Tension).
Cohesion	Whether a movement is made of components exhibiting similar features (e.g., tendency of limbs to move as a single entity in a direction).
Coordination	Whether a movement is made of synchronized components (e.g., synchronization of limbs to operate a body at the unison). This corresponds to temporal entrainment in a group.
Origin	Whether a movement originates at a joint, and at what extent a joint leads the body in the movement. This may correspond to leadership when measured in a group.
Attraction	The degree of influence an external point in space has on movement (e.g., like a magnet attracting or repulsing the dancer).
Slowness	Whether a movement is continuous and at an extremely slow speed.
Stillness	Pause: minimal movements depending on physiology (e.g., respiration), emotions, and attention continuously occur.

Table 4-2 - List of the main motion mid-level features

4.4 Concept and structures

The highest level of features focuses on the *nonverbal communication of movement qualities to an external observer*, which correspond to those properties that cannot be evaluated within a few seconds, but need a longer observation to be properly perceived [Camurri2016]. They involve *memory*, i.e., the history of previous movement qualities, and *context*. A typical example for this level of features is the emotion [Piana2016] that is conveyed to an observer: it is impossible to be assessed by observing a few seconds of the performance, and may be affected by elements from context, including the music that accompanies the performance, the setup of the stage, the costumes, the dancer's size, the generic cultural context shared between the performer, the choreographer and the spectators. A broad categorization of these descriptors is listed in the *Table 4-3*.

Name	Description
Predictability/expectancy	The extent at which an external observer can predict a dancer's movement
Hesitation	When an external observer cannot clearly perceive a movement intention.
Attraction / Repulsion	The extent at which an external observer is attracted/repulsed.
Groove	The extent at which dance elicits movement in an external observer.
Saliency	Saliency: a movement which is perceived as salient with respect to others occurring at the same time

Emotions	The emotions, expressed by full-body movement and posture, which are conveyed to an external observer.
Nonverbal social signals	Entrainment in its temporal and affective components, Leadership, and so on.

Table 4-3. List of the main motion concept-related and structure-related features.

5 Joint music-dance representation model: use cases

In this section, we will illustrate two scenarios that represent an evolution of the music representation model, the movement representation model and dependencies and interaction between the two described in previous sections. The first scenario concerns contemporary dance and contemporary music; the second is related to Flamenco dance and music.

5.1 Piano&Dancer

5.1.1 Introduction

Piano&Dancer is an interactive piece for a dancer and an electromechanical acoustic piano. The piece has been realised by Instituto Stocos as a collaboration between Pablo Palacio, Muriel Romero and Daniel Bisig with the support of UNIGE. The piece presents the dancer and the piano as two performers whose bodily movements are mutually interdependent. This interdependence reveals a close relationship between physical and musical gestures. Accordingly, the realisation of the piece has been based on creative processes that merge choreographic and compositional methods. In order to relate the expressive movement qualities of a dancer to the creation of musical material, the piece employs a variety of techniques. These include methods for movement tracking and feature analysis, generative algorithms for creating musical structures, and the application of non-conventional scales and chord transformations to shape the modal characteristics of the music. In this contribution, we will describe the interesting opportunities for creative cross-fertilisation between dance choreography and musical composition we had to deal with, and the technical and aesthetic principles that were used to connect expressive qualities of movements to the creation of music structures. Being WhoLoDancE a project focused on developing new technologies for dance education, in *Piano&Dancer* we aimed at creating a peculiar and innovative context for the two main elements that are present in the traditional classic ballet dance class, the dancer and the acoustic piano, creating a context that would correlate their activities in a novel manner.

In *Piano&Daner* the music of the performance is produced through the piano mechanical movements. These movements are generated in real-time through a combination of compositional algorithms, stochastic functions, swarm simulations and modal mappings, all of which are influenced by the dancer’s bodily movements. As a result, the piano and dancer are connected with each other through three levels of relationships: they are both physically present on stage and exhibit bodily movements, their respective movements are correlated through a generative intermediate layer, and they exhibit in the musical and bodily domain a clear correspondence in expressivity.

One of the central motivations for the realisation of *Piano&Dancer* is the assumption that a pianist’s bodily movements during a musical performance represents a choreography. This choreography is at the same time very sophisticated but also highly constrained. Furthermore, the experiential aspects of this choreography are revealed through the resulting musical forms. These forms are shaped by the intrinsic physical properties of the piano and by the characteristics of the gestures performed by the player.

Starting from these considerations, the piece *Piano&Dancer* undertakes an artistic investigation into the choreographic qualities of instrumental gestures and how these qualities can be creatively extended while at the same time maintaining their inherent music producing functionality. Accordingly, some of the research questions we tried to answer are: Can musical structures convey acoustically some of the qualitative and expressive characteristics of bodily movement? Which intimate aspects of a dancer's bodily activities that are normally hidden can be conveyed through acoustic feedback, and how this feedback can serve as a tool to complement dance learning? What formal and aesthetic mechanisms need to be established to mediate between the bodily and musical domains of gestural movements? How can the important role of kinaesthetic body awareness be taken advantage of for the control and perception of interactive music? How can compositional and choreographic approaches inform each other in the creation of performance pieces for dancing instrumentalists?

5.1.2 Music representation model

The entire composition is implemented in the Supercollider programming environment and is generated live during the performance. The music of the piece builds upon an algorithmic composition layer consisting on several abstractions that mediate between the dancer's movements and several pre-defined harmonic fields and modal systems. There exist multiple algorithms that create itineraries, rotations, control densities, speeds, rhythms, motifs and relationships within a composed layer of predefined musical entities. These algorithmic abstractions are perturbed, controlled or affected in different ways by the bodily movements of the dancer. From a musical point of view, these layers condition the composition of the work.

Some of these compositional approaches explore the properties of finite groups. This for instance is the case for the automated creation of inverted transpositions on the same fundamental note which are controlled by the rotation of the dancer's wrists. For each joint rotation, the algorithm returns a new version of an array of numbers corresponding to an inverted transposition with the same fundamental note. Other approaches make use of probability distributions employed to shape the density of events and also to distribute them across different harmonic fields. Each distribution has its own specific characteristics and can be associated with different musical aspects. For example, Cauchy and Gaussian distributions are combined and superposed and their means are assigned to harmonic nodes or tonal centres that are associated to specific body joints. The levels of movement qualities of a particular joint may be used to perturb the spread or deviation from the mean that defines these distributions. This causes the musical output to oscillate between harmonic and more dissonant states. A similar application may be developed in the temporal domain to organise durations that move between predictable and unpredictable or random rhythms.

A more sophisticated algorithmic layer is provided by a swarm simulation that has been implemented using the ISO programming library. By integrating swarm simulations as intermediary level between movement analysis and piano control, the interactive systems gain the capability to create complex musical material in a self-organised manner. In this setup, the dancer is no longer able to directly and fully control the behaviour of the algorithmic layer but rather assumes the role of an improvisation partner with an artificial and autonomous musical agency. The repertory of swarm behaviours that are provided by the programming library has been expanded with several additional behaviours. These behaviours have been developed to meet the specific requirements to provide control data for the creation of discrete and note based musical forms (see Figure 5-1).

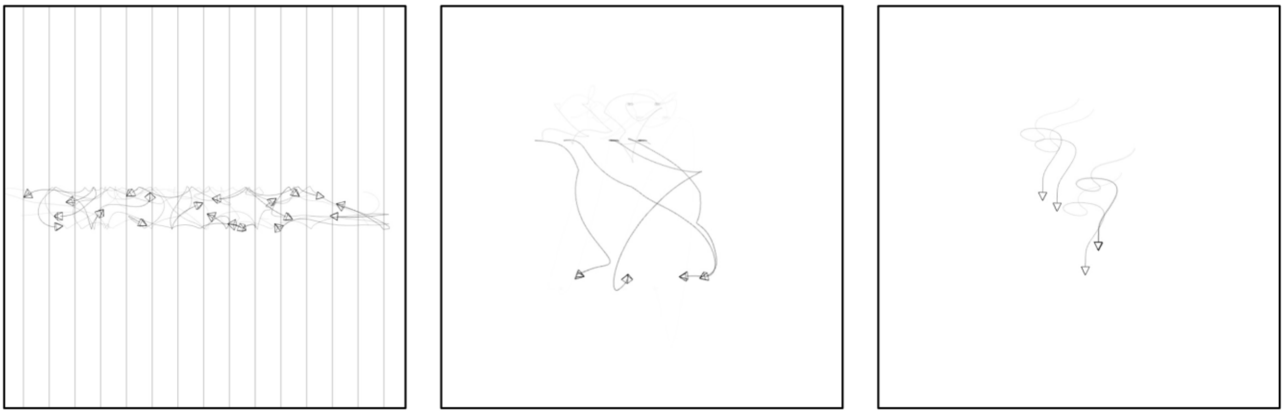

Figure 5-1. Custom Agent Behaviours for Swarm Simulations. From left to right: discretisation behaviour, axis aligned offset behaviour, sequencing behaviour.

The most important of these behaviours are: a discretisation behaviour that maps any continuous agent parameter such as position or velocity to a predefined set of discretised values; a cohesion behaviour that permits the specification of axis aligned offsets among the positions of neighbouring agents; this effect permits the realising of chord-like groupings within a swarm; a neighbourhood behaviour that encodes the positions of neighbouring agents in spherical coordinates in order to simplify the distinction between distance and orientation relationships within a swarm; the distance can be used for instance to control the intervallic relationships between notes whereas the orientation can control the permutations of chords; a sequencing behaviour that triggers a timed series of modifications to a particular agent parameter; the purpose of the sequencing behaviour is to generate control data that exhibits a motivic form.

On top of these algorithmic layers we find different types of predefined modal systems or key chords that shape the harmonic content of the piece. Some of these structures have been calculated algorithmically while others are obtained by hand. Among these structures we find non octavating scales or modes of limited transposition among other colour palettes.

5.1.3 Movement representation model

For almost all the scenes in *Piano&Dancer*, interactivity is based on sensing the dancer's movements with inertial measurement units (IMU) that are attached to the dancer's body. This technique is complemented only during particular moments by a camera-based tracking system. The camera based tracking systems provides an allocentric and absolute frame of reference whereas IMU sensors provide an egocentric and therefore relative frame of reference. The IMU devices that are used for the piece are named Xosc and are provided by the company x-io Technologies. Four of these devices are employed to track the movements of four joints on the dancer's body (two wrists and two ankles). Interactivity is based on both the acquisition of raw sensory data as well as on the analysis of higher level movement features such as Smoothness, Weight, Energy and Dynamic Symmetry. By integrating methods for higher level feature analysis into an interactive system, this system becomes capable of detecting and subsequently responding to movement qualities that are also salient for the dancer and the human audience. This helps to alleviate one of the problems when applying interactive technology for dance: the constraining of dance movements through technological prerequisites and the shifting of the dancer's attention away from intentionality and expressivity towards the purely functional execution of movement. The analysis of higher level movement features is implemented in the EyesWeb programming environment developed by UNIGE. A custom EyesWeb-based software application has been developed. This application processes the raw IMU sensor data and extracts low and higher-level movement features. These features are sent via the open sound communication protocol (OSC) to the composition and piano control software. Figure 5-2 depicts a schematic representation of this technical setup.

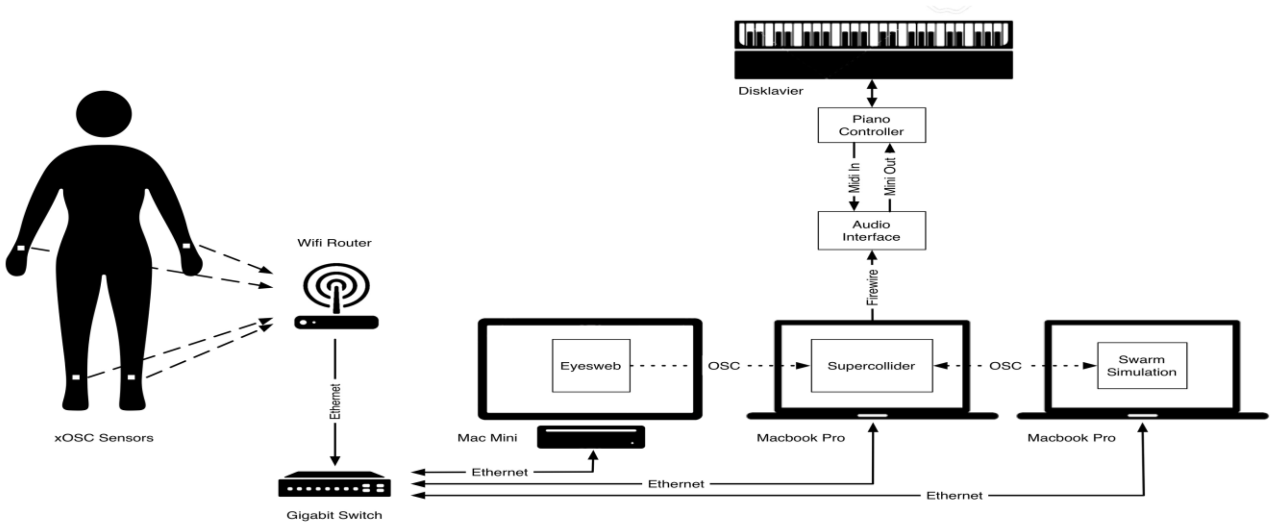

Figure 5-2. Schematic Depiction of Technical Integration of the Sensing, Communication, Computation and Piano Control

5.1.4 Interaction between the music representation model and the movement representation model

For the realisation of *Piano&Dancer*, an interactive setting was chosen that abolishes the necessity for a direct tactile manipulation of the acoustic instrument. In addition, a direct causality between the physical aspects of the dancer's body movements and piano actuation was complemented with interaction techniques that take the expressive aspects of bodily movement into account and that integrate compositional algorithms as part of their mapping mechanisms. By allowing the performer to control the piano through other means than direct tangible interaction, the functional constraints of sound producing gestures and the immediacy of their effects on the musical result are dissolved. This provides the opportunity to invent novel and diversified relationships between physical and musical gestures. As a result, the dancer's bodily movements can be shaped according to choreographic criteria.

Furthermore, through algorithmic means, the relationship between movement and music can be expanded to involve interactive control over the compositional process itself. By integrating algorithmic and generative methods as mediating layers between kinaesthetic and visual domain of body movement on one hand and the acoustic domain of algorithmic composition and piano playback on the other hand, the dancer becomes able to transfer articulated gestural expressions into music without the necessity for her to consciously pay attention to and plan the musical consequences of her movements. Rather, it is up to the music composer to specify the compositional and sonic principles of the music, the automation of which is incorporated and finally delegated to the algorithmic and mechanical mechanisms of the computer and piano, respectively. As a result, the dancer can focus on the creative and aesthetic principles that lie fully within her own area of expertise while relying on the compositional expertise embedded in the interactive musical system to respond in a musically meaningful way to her own performance. The incorporation of algorithmic and generative methods into the mediating layers between movement and music allows the choreographic and compositional elements of the performance to preserve their intrinsic aesthetic principles while at the same time remain connected through a strong causal relationship. These algorithmic abstractions are influenced or perturbed in different ways by the bodily movements of the dancer. The design of these algorithmic abstractions has been carried out in such a way that the qualities of the dancer's movements are transferred into compositional structures and sonic morphologies that convey aesthetically and metaphorically similarly qualities in the acoustic domain.

As a result of these relationships, the choreographic and music compositional principles are tightly interconnected. The composition of *Piano&Dancer* emerges from the integration of movement gestures and corresponding sonic gestures that mirror the dancer's expressivity.

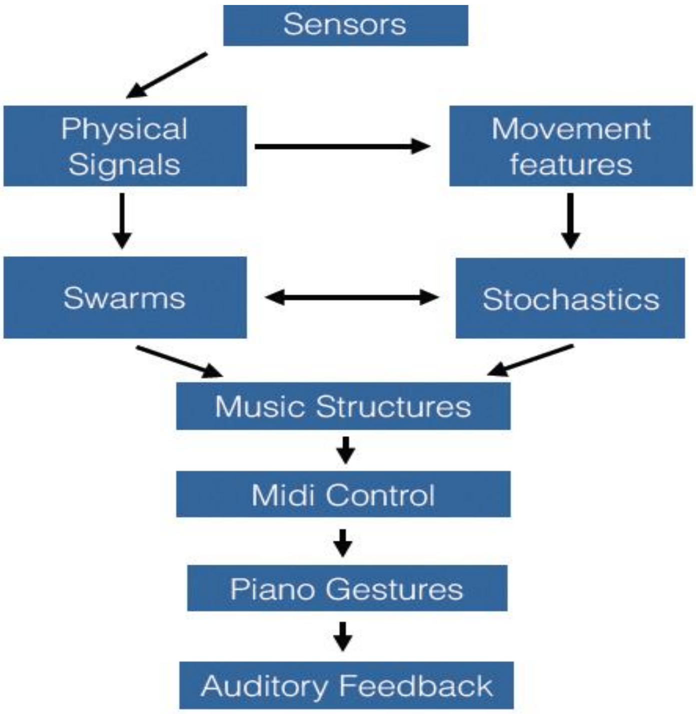

Figure 5-3. Technical integration of all the functional elements in Piano&Dancer

In *Piano&Dancer* each of the composed scenes of the piece develops a particular movement quality. These movement qualities are always linked to a particular musical material or algorithmic approach to composition. Moreover, the choreographic structure highlights these relationships using choreographic strategies for relating movement qualities and music: these have been developed within the framework of the project WhoLoDancE. Of particular importance in the piece are the movement qualities: Energy, Weight, Smoothness, and Dynamic Symmetry of Smoothness as defined within the framework of the project. In the publication presented at MOCO 17 [PALACIO2017] a detailed description of how these qualities are developed and connected to the music models of the piece, explaining the sonification and choreographic strategies that have been developed in order to exploit this connection. We provide the following links to videos of *Piano&Dancer*: a *making off* video with rehearsal footage and a compilation with some moments of the performance piece.
https://vimeo.com/182424752
https://vimeo.com/191080304

5.2 Relationship between dance and music within Flamenco

Within flamenco there are many elements that come together to form the musical and artistic traditions. To understand the relationship between dance and music, a series of terms are needed to be defined to help illustrate the interrelationship between the two. To simplify the vocabulary and in an attempt to make a cohesive story, it is best to separate the terms into primary and secondary groups. The terms in the primary group are the rudimentary basics that would have to be present for flamenco to exist. The secondary group encompasses words that can be seen as jargon or that serve as decoration to the art from. The terminology in the secondary group is important to flamenco and plays a crucial role, but is not critical in understanding the overall history and the relationship between dance and music within Flamenco.

The three pillars of flamenco are the *cante,* the singing, the *toque,* guitar playing, and the *baile*, the dancing. These main elements serve as the foundation of the primary group. Everything is built from the *cante, toque*, and *baile.* Within flamenco, there is a term known as *compás* which is a twelve-count rhythm that underlies everything happening within the flamenco art form. It means slightly different things according to the context. In Spanish music, it is the bar of measure and within flamenco it also means the rhythmic unit of the song. *Compás* also means to stick accurately to that unit while making rhythmic variations within it. The *compás* is a structure that allows for personal freedom and allows the dancer, singer or musician to creatively decorate the rhythms. Because flamenco encompasses complex rhythms, the *compás* of each style must be clear, strong, and consistent. Within flamenco there are up to 50 styles, also known as *palos*[1]. These *palos* can have a 12 count *compás* but are not obliged to be 12 count rhythms. Each *palo* has specific verses, also known as *letras*[2], which adhere to certain syllabic rules. *Letras* fit into *tercios*[3], also known as sections of a song. Sometimes *tercios* are changed and a new *palo or* style, is introduced within the same *compás*, rhythm. However, this combination of *palos* is not always used by performers. *Palmas*[4] are another term that are often used to describe the hand clapping which hold the *compás*. All of these components are part of the primary group of the flamenco art form.

The secondary components are equally important but again, stem from the primary elements. To illustrate this point here is an example within dance, *baile*. Within the primary group we established that dance is a key component to the flamenco art world. Out of *baile* a secondary term would be *braceo*[5], arm movements during a dance. Although *braceo* is important to flamenco, the actual term is simply jargon. The following analysis of the flamenco *palos,* will include vocabulary from the secondary group.

Cante jondo[6], *cante intermedio*[7], and *cante chico*[8] are the three sections that divide flamenco, and the singer, dancer, and guitar player all abide to the breakdown. Each variant conveys a mood that is specific to that category and within that category there are branches, called *palos. Cante jondo* is the oldest form of flamenco and the intense and sad form of *cante* which deals with anguish, pain, suffering, death, and religious sentiment, is our first category. Under this unit, the t*oná, martinete, carceleras,* and the *debla* palos are registered. Our second category *cante intermedio,* meaning intermediate, is a less profound yet moving style that can have an oriental overtone to its character. The *cante intermedio* is between the *cante grande* and the *cante chico*. It can carry a heavier mood, but this too depends on who the singer, dancer or guitar player is and how they are choosing to interpret the verses within the style. The last category is the *cante chico*,

[1] *Palos*- song styles within flamenco; also Spanish word for stick.
[2] *Letra*- verses within the flamenco *cante*.
[3] *Tercio*- sections of the songs.
[4] *Palmas*- hand-clapping which holds the *compás*.
[5] *Braceo*- Spanish word for arm movements during a dance
[6] *Cante Jondo*- category within flamenco that is known to contain the deepest and saddest song styles.
[7] *Cante Intermedio*- category within flamenco that is known to contain the songs that are not sad nor happy; forms that are in the middle.
[8] *Cante Chico*- category within flamenco that is known to contain the happier and lighter song styles.

which literally means "little song". This style of *cante* primarily deals with themes of love, humour, and happiness, often the *palos* of *alegrías, bulerías,* and *tangos.*

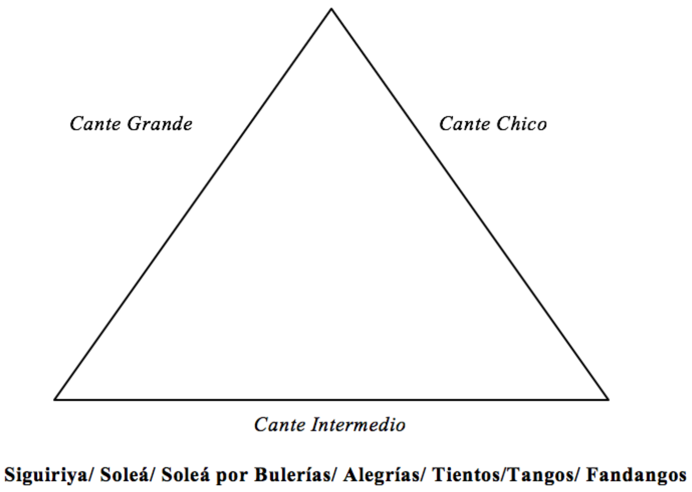

Figure 5-4. Figure three categories which divide flamenco: Cante Jondo, Cante Intermedio, and Cante Chico.

The guitar is the second component of our primary group. All guitar playing is known as *toque*. The word encompasses different elements but means that all flamenco is played on the guitar. Similar to the *cante* within flamenco, *toque* is also divided into categories: There is *toque jondo* and *toque of the soleáres*. Guitar playing in flamenco is just as complex as the *cante*. Within the hierarchy of flamenco, the *cante* will always be above the guitar playing, which means that the musician must accommodate to the singer and dancer. *Toque* within flamenco also adheres to the *compás* of the style. It is a crucial part of playing and the goal of the guitarist is to make the *compás* a subconscious element, which will allow for improvisation and a greater sense of freedom. As we have seen with the *cante*, the singer has the freedom to veer off the framework if they so please, this holds true for the guitarist as well.

The final primary element is the flamenco *baile* which is the exploration of the rhythmical structures and the blending of the individuals' personal expression, within their own interpretation of the *cante* and *toque*. The key to dancing is the combination of the upper and lower body all while honouring the *compas* and the *cante* and *toque*. Within the hierarchy of flamenco, the dancer is the lowest on the totem pole, but is considered a valid component of the overall entity.

There is clearly an Arabic, Persian, Greek, Jewish, and Spanish influence within flamenco, which has also affected its dance. Indian dance elements are interwoven into the *baile* and when comparing Indian Bharatanatyum and Kathak with flamenco, the similarities are clear. Both Indian and flamenco use an extraordinary amount of footwork, spins, and hand gestures, which link the two forms. Apart from the Indian influence there is also a direct correlation with the Arabic culture.

Similar to the *cante* and *toque,* the dancer also uses secondary terms to describe their discipline. The upper and lower body use specific vocabulary that aid in the understanding of the dance. I will start with the lower body and work my way up to the head. Footwork, also known as *zapateado*[9]*,* within flamenco, is an important element. There are different steps and combinations of sounds that make the *zapateado* effective. Each part of the foot creates a unique sound. The *media planta*[10] is used to describe the ball of the foot, where the

[9] *Zapateado*- Spanish word for footwork in flamenco.

[10] *Media Planta*- used to describe the ball of the foot in flamenco dance technique.

tacon[11] is known as the heel. The *planta*[12] is the entire sole of the foot and the jab is also known as a *tacon*. These elements come together to make phrases that fit within the *compás.* However, in the past the footwork was divided by gender roles. Originally the male was the only dancer that could use the *zapateado.* Women were primarily consumed by and encouraged to speak solely through their upper body.

Footwork for a flamenco dancer is viewed as a connection to the earth. The lower part of the leg, from the knee down, should be used as a hammer which does not stomp on the floor but rather dig into it to produce a rounder sound. A *plie*[13]*,* bent knees, is also important as it allows the body to have a grounded feeling that can produce a weighty quality. The hips, las *caderas*[14], are evenly placed and swing from right to left, aiding in the use of the *zapateado.* The stomach is always engaged. The torso is long yet a bit forward and once it is positioned, should be arched back. This opens up the chest and creates a bold quality. The arms, *braceo*, follow the classical dance pathway, known as the gateway. This path requires the elbows to be bent and the shoulders to be turned inward. The hand movements, *floreo*[15]*,* are extensions of the arms and are seen as the final expression of the core. The neck is elongated, and the chin is down. All in all, the combination of the *zapteado*, *braceo*, *floreo,* and the head complete the analysis of the secondary terminology within dance.

It is important to note that flamenco dancing encompasses *compás,* technique, footwork, arm work as well as a dialogue between all of the other performers. Flamenco steps do not have symbolic meaning but could be viewed as a metaphor for the suffering of its ancestors. Just like with the *cante,* flamenco dancing brings its own experiences to the forefront, while honouring the past. It is a tradition that is always changing. *Baile* resembles the inner self while the body serves as a vessel to express its deep thoughts, emotions, and ideas. It is not only what you do but how you do it that is key to flamenco dancing.

In conclusion, the singer has four tools, the *melisma, ayeo*, rhythm and repetition that they use to navigate the three categories which divide flamenco: *Cante Jondo, cante intermedio, and cante chico*. The guitarist uses right- and left-hand techniques to support the *cantaor,* yet has freedom when using, *falsetas*, his guitar solos. Finally, the dancer joins the overall composition by combining upper and lower body movements which include the *zapateado, braceo, floreo,* and *compás.* Each primary element has its own secondary counterparts that are used to bring together everyone involved to form a unit. The singer will be followed by the guitarist, and the dancer will trail both the *cantaor* and *guitarista*[16]*.* All in all, they come together and reflect the relationship between the flamenco music and dancing. Within the WhoLoDancE project, at least one dance from each category was captured and annotated.

---

[11] *Tacon*-jab or heel
[12] *Planta*- entire sole of the foot strikes the floor.
[13] *Plie*- a bending of the knees outward by a ballet dancer with the back held straight
[14] *Caderas*- Spanish word for hips.
[15] *Floreo*- hand movement used during flamenco dancing.
[16] Guitarista- Spanish word for guitarist.

6 References

- Buccoli M., Di Giorgi B., Zanoni M., Antonacci F., Sarti A. (2017) Using multi-dimensional correlation for matching and alignment of MoCap and video signals. *IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*
- Bruno Di Giorgi, Massimiliano Zanoni, Sebastian Böck, Augusto Sarti, Multipath Beat Tracking, in Special Issue on Intelligent Audio Processing, Semantics, and Interaction, Journal of the Audio Engineering Society, vol.64, no.7/8, pp.493-502, 2016
- Piana S., Staglianò A., Odone F., Camurri A. (2016). Adaptive body gesture representation for automatic emotion recognition. *ACM Transactions on Interactive Intelligent Systems (TiiS)*
- Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., Canepa, C. (2016) The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. *Proceedings of the 3rd International Symposium on Movement and Computing (MOCO '16)*
- Palacio P., Bisig D. (2017) Piano&Dancer: Interaction Between a Dancer and an Acoustic Instrument. *Proceedings of the 4rd International Symposium on Movement and Computing (MOCO '17)*