Imperial College London

Interim Report

Turning Music Into Game

Author:
Paulina Koch

Supervisor:
Dr. Iain Phillips

January 2015

Chapter 1

Introduction (1-3)

> Music and games share a fundamental property: both are playable, offering their listeners and operators an expressive experience with the framework of melody and rhythm. [1]

As the quote suggests, both games and music have one thing in common — the act of playing. Just as player's character might die in the attempt to complete a level, causing him to lose the game, the pianist can fail at the attempt of performing a musical piece.

Perhaps this analogy inspired programmers to develop a new genre of games - music games. Music games are games in which players interact with music. Possibly the most commonly known franchises in this genre are Guitar Hero, Rock Band and Dance Dance Revolution. In this type of games user has to follow the indicators on the screen telling him which buttons to hit.

The concept of a music game stormed the industry in 2005, after Guitar Hero was released. The project soon turned into the fastest new video game franchise to reach $1 billion in retail sales in the history of the business, with Guitar Hero III being the first game to reach $1 billion. [3]

However, a limited amount of songs transcribed and adjusted to the gameplay soon caused the popularity of such music video games to decline. Some brave fans of the franchises took it upon themselves to transcribe songs to create new levels. The producers, seeing the tendency, started releasing the in-app purchases to enable the players to extend their library and thus, keep the users.

Due to the time consuming and difficult nature of the process of manually adding new songs, most players usually limit themselves to pre-processed songs provided by the game producers, not really taking advantage of the full capabilities of the games.

This project aims to change the way users look at the music rhythm games. We are creating a game which will allow them to upload any song they would like and automatically generate a Guitar Hero-like level corresponding to it. We will try to achieve that by implementing an algorithm to detect a main melody in a music track and map it to a series of keyboard presses as well as an algorithm for a mood detection, for dynamic rendering of the game surrounding.

With this project we would also like to show that sophisticated academic music analysis techniques can be combined together and applied to real world problems in an efficient and reliable manner.

Finally the project aims to be more than just a research study of feasibility. The result of successful completion will be an application of sufficient reliability and quality that it can be released to, and used by, untrained computer users. This report only lightly touches on this facet of the project, as creating usable polished applications is a reasonably well solved problem, and the least interesting area of this project.

Chapter 2

Background (10-20)

In this section, we go investigate different types of music games, along with a deeper look into Guitar Hero, on which we base our main concept for the gameplay. This is followed by a discussion of the most applicable publications in music analysis, for both defining the mood and finding the main melody in a musical track.

2.1 Music Video Games

A music video game can be defined as a type of game that uses music or rhythm as an integral part of gameplay. This may involve pressing buttons in time with a song, whether on a conventional controller, and instrument controller or some kind of dance mat, singing into a microphone or creating original music. Players can often perform different parts of the same song together in local multiplayer games or over the Internet, providing enjoyable social experiences. [2]

Some games exhibit a sandbox style that encourages a free-form gameplay approach whereas other a hybrid style, which combines musical elements with more traditional genres, for example puzzle games or shooters.

Below we will briefly go over different types of music video games that can be found on the market.

2.1.1 Music Memory Games

The goal of the music memory game is to score player on their musical memory. Music track is presented to the user who then has to provide an appropriate response to each prompt from the game. Games may be based on different primary musical aspect (whether it is the rhythm, pitch or volume). However, a vast majority of the releases available on the market are rhythm-based.

Rhythm Music Games Games of this genre typically focus on dance or the simulated performance of musical instruments, and require players to press buttons in a sequence dictated on the screen. Doing so causes the game's protagonist or avatar to dance or to play their instrument correctly, which increases the player's score. An example of such games could be Guitar Hero or Dance Dance Revolution.

2.1.2 Free Form Music Games

In free form music games, the main task of the user is to create content. This form of music game is often compared to non-game music synthesisers. Free form music games are somewhere between generative hybrid music games and non-game utilities, depending on the degree to which their gameplay relies on a driving underlying plot-line.

2.1.3 Hybrid Music Games

Hybrid music games are characterised by substantial and meaningful interactions between a player and the music game in a game that apparently belongs to a non-musical genre. This type of games can be further split into two sub-types.

Generative music video games make use of user's actions. By monitoring interaction with the surroundings in the game, the mechanism generates sounds that are then integrated into the soundtrack, permitting the player's direct interaction with the score. This encourages the creation of a synesthetic experience — when upon stimulation of one sense others activate, causing an involuntary experience. An example of such game could be Rez, which is a simple rail shooter. However, thanks to integrating sounds generated by player completing the normal task of rail-shooting, the score is dynamic.

Reactive music games, in contrast to generative one, employ music to determine the gameplay. In such games, the player takes cues from soundtrack to devise his gameplay. For example, iS - internal section, uses the music to determine the dynamic of the non-musical components of the game.

2.2 Case Study - Guitar Hero

2.2.1 Overview

Guitar Hero is one of the most popular franchise in the history of music games. The first of the series was published in 2005 by RedOctane and Harmonix. In the games, players instrument-shaped game controllers to simulate playing the instruments across numerous rock music songs. It is widely considered a highly entertaining game fully embracing the rhythm-based music game. [8]

2.2.2 The Controller

Rather than a typical gamepad, Guitar Hero uses an instrument-shaped controller (guitar in the earlier releases, bass, microphone and drums in more recent ones). Playing the game with the guitar controller simulates playing an actual guitar, except it uses five coloured "fret buttons" and a "strum bar" instead of frets and strings, and an analogous mapping for the other instruments. They incorporate most of the real life techniques and motions that an instrumentalist would perform on a real instrument. [7]

2.2.3 The Gameplay

The actual game itself works exactly as many other music titles do. At the bottom of the screen, a number of (varying depending of level of difficulty) buttons is shown. In each attempt, a series of notes moves across the screen and when a note aligns with a button, player is supposed to press a corresponding button, gaining points depending on the accuracy. If the player failed to achieve a certain amount of notes — his performance meter stays low for a longer time, he loses the game.

However, there are a couple minor improvements that Harmonix has made to the general music game formula. By acing certain notes in a song, a player is able to build up Star Power, which when unleashed, doubles up current point multiplier. Star Power also adds a bit of a strategic element —player not only earns more points when it is activated, but he can also raise your performance meter faster, enabling him to last longer when encountering a trickier part of a song.

2.2.4 The Critique

Without a doubt, Guitar Hero features a great selection of music. However, there will always be tracks missing, regardless of how many versions of Guitar Hero are released.

People have different tastes and limiting a game to a set of tracks that everybody is supposed to enjoy is a really hard task.

Some more advanced users familiar with Computer Science attempted to transcribe songs and to create new levels. However, this process is really difficult, especially for non computer scientists, discouraging an average user from fully making use of game's capabilities. The producers, seeing the tendency, started releasing the in-app purchases to enable the players to extend their library and thus, keep the users.

Implementing a feature of uploading some music preferred by the player would definitely improve user satisfaction. However, might have not been implemented yet as the task itself is quite complex. Moreover, enabling the users to load in some music would deprive the company of their income sources.

2.3 Introduction to Melody Extraction

For a long time people were researching ways of estimating the fundamental frequency, be it with monophonic music recording or multi-pitch estimation. Melody extraction differs from both of those problems — unlike monophonic pitch estimation it handles polyphonic tracks and in contrast to multi-pitch estimation, it must also include a mechanism for source identification, to spot the voice carrying the melody within the polyphony.

2.3.1 Melody

The concept of "melody" ultimately relies on the judgement of people listening. This is why it will vary depending on the application context - whether we want to determine symbolic melodic similarity or transcribe a music track.

In order to have a clear framework to work within, the Music Information Retrieval (MIR) community has adopted in recent years the definition proposed by [4], "...the melody is the single (monophonic) pitch sequence that a listener might reproduce if asked to whistle or hum a piece of polyphonic music, and that a listener would recognise as being the 'essence' of that music when heard in comparison".

In practice, research has focused on "single source predominant fundamental frequency estimation" — which means a search for a main melody coming from a single sound source throughout the song analysed. As we can see, the subjective element is still present in this description of a melody as there might not be a definite way of deciding what is predominant. However, it fits well with our project's objective — generating a game level based on changes in the pitch.

2.3.2 Polyphonic Music

Polyphony is a word derived from Greek poluphōnosis meaning more than one sound — a texture consisting of two or more simultaneous lines of independent melody. This can be contrasted with homophony, where musical parts move generally in the same rhythm and one dominant melodic voice is accompanied by chords or monophony, where only one voice is found.

However, in our case, the term polyphonic will simply refer to any type of music in which two or more notes can be played simultaneously. This can be achieved either by playing in different instruments (for example, voice, guitar and bass) or a single instrument capable of playing more than one more at a time (like a piano).

2.3.3 Fundamental Frequency and Pitch

Pitch is the most natural way of ordering sounds on a frequency-related scale. If sounds whose frequency is clear and stable enough to be distinguished from noise, they can be compared among one another as “lower” or “higher”. Pitch is not an objective physical property — it depends on anatomy and physiology of the auditory system, which is a subject of an extensive study called psychoacoustics.

The fundamental frequency f0 is defined as the lowest frequency of a periodic waveform. It can be thought of as the physical property most closely related to perception of pitch. This is why in this context pitch and fundamental frequency can be used interchangeably.

2.3.4 Filter

Any medium through which the music signal passes, whatever its form, can be regarded as a filter. However, we do not usually think of something as a filter unless it can modify the sound in some way.

A digital filter is a filter that operates on digital signals, such as sound represented inside a computer. It is a computation which takes one sequence of numbers (the input signal) and produces a new sequence of numbers (the filtered output signal).

2.3.5 Short Time Fourier Transform

Short-time Fourier transform (STFT), is a signal processing method which is used in analysis of non-stationary signals with statistic characteristics varying with time. In particular, STFT extracts several frames of the signal to be analysed with a window that moves with time. If we set the window size to be narrow enough, each frame extracted can be viewed as stationary so that Fourier transform can be used. With the window moving along the time axis, the relation between the variance of frequency and time can be identified. [6]

2.4 Main Melody Extraction from Polyphonic Music

2.4.1 Source Separation Based Approach

In polyphonic tracks the main melody can be represented by a specific source/filter model. In case of the leading vocal part, the vocal cords are treated as a source and the voice tract as a linear acoustic filter.

In their paper from 2011 [5], authors presented an algorithm in which they assume that at any given time the signal observed is a mixture of two elementary signals - one corresponding to the main source and one to the background music. Therefore, the signal can be represented in an equation x = v + m, which also holds for the short time Fourier transform (STFT) X, V and M respectively: X = V + M.

The background music signal M can be thought of as a mixture of R independent Gaussian sources Mr. Each of the sources is centred and characterised by its power spectral density (PSD), which describes how the power of a signal or time series is distributed over the different frequencies. This together with STFT and an amplitude coefficient associated with each source is used to calculate the likelihood for each of the frequency bins.

The likelihood of the vocal part V is then calculated using two different frameworks.

The first submission uses the source/filter Gaussian scaled mixture model (GSMM). In this model the source element refers to the excitation of the vocal folds and is therefore linked to the fundamental frequency of the sound f0, while the filter part is characteristic of the vocal tract shape. This space of possibilities is then discretised so that we consider one possible filter frequency response, which is then used to calculate the likelihood of the vocal part knowing the filter and f0.

The second model was derived from the first one to find a solution that would be more efficient to compute. The authors came up with a formulation that keep the source/filter model within an instantaneous mixture framework (IMM). In essence, it represents the lead voice as the instantaneous mixture of all possible notes.

Once the parameters are estimated using the maximum likelihood criterion for each of the model, the Viterbi smoothing of the melody line is applied, obtaining a trade-off between the smoothness of the melody and its global energy in the signal.

The authors then parametrise the transitions between the possible main melody without disabling jumps from one note to the other. The source separation framework used by them allows, through Wiener filtering, to obtain the separated signals. Computing the energy for each frame of the separated main melody and thereafter thresholding allowed to discriminate between spurious notes and true positives.

2.4.2 Salience Based Approaches

This approach has been the most popular so far, with majority of algorithms evaluated at MIREX implementing it and can be split into several smaller stages. In particular, a method implemented in paper [9] seems to be quite promising.

As a first step, some sort of preprocessing is applied to the audio signal, usually to enhance the frequency content where we expect to find the melody. In particular, Salamon and Gomez apply an equal loudness filter, which enhances the frequencies to which the human ear is more perceptually sensitive by taking a representative average of the equal loudness curves and filtering the signal by its inverse.

This stage is followed by spectral transform — the signal is chopped into time frames and a transform function is applied to obtain a spectral representation of each frame. This is achieved by applying the Short-Time Fourier Transform with a window length of 46.4ms. Thanks to that we achieve sufficient frequency resolution to identify different notes while maintaining adequate time resolution to track pitch changes in the melody over a short time.

Having done this, we move to frequency/amplitude correction, where the spectral peaks are detected and used for further processing. To avoid a relatively large error in the estimation of the peak frequency caused by binning them in the process of FFT, peak's instantaneous frequency and amplitude are calculated.

Those three steps constitute the spectral processing. But at the core of the salience based algorithms lies the multi pitch representation, i.e. the salience function — a representation of pitch salience over time. In the algorithm described by Salamon and Gomez, this computation is based on harmonic summation, where the salience of a given frequency is computed as a sum of the weighted energies found at harmonics (integer multiples) of that frequency. Using only the peaks for the summation allows us to discard less reliable values and apply further frequency corrections.

// Creating Pitch Contours

// Pitch Contour Characterisation

// Note grouping

Given the peaks of the salience function, we now have to determine which pitch values belong to the melody. This process is initiated by grouping peaks into continuous pitch contours, out of which a melody is selected later.

2.4.3 Comparison of both approaches

2.5 Mood Detection

Chapter 3

Project Plan (1-2)

Chapter 4

Evaluation (1-2)

Game developers use personal preferences and creative programming techniques and tools to develop games with the hopes of successful market penetration. Often, in the course of development, the needs of the end user are lost.

Evaluation can occur during various times during the design and development life cycle of a game – early, in the middle, late, and at the end. However, not all types of evaluation methods can be applied during all phases of design and development.

4.1 Formative

4.1.1 Single-Condition Study

- students
- measure learning, affect, liking
- good for evaluation while implementing, not really for the end product

Were the mechanics easy to understand?
How many time did you need to learn how to play the game?
Did you find the game challenging?
Did you finish the game?
How long did it take before you got bored of playing the game?
Did you find the user interface intuitive and easy to use?
Did you empathise with the main character?

4.1.2 Comparison to Original Songs

4.2 Summative

4.2.1 Comparison to Original Songs

4.2.2 time spent tracking

4.2.3 Evaluation Framework

4.2.4 Release

4.2.2. interviews, questionnaires, surveys

Appendix A

Appendix Title Here

Write your Appendix content here.