Algorithmic Composition of Popular Music

Anders Elowsson

January 4, 2012

DT2215, PROJECT.

Abstract

This report explores possibilities and techniques regarding algorithmic composition of popular music. Methods in algorithmic composition are reviewed with a focus on Markov models and generative grammar. The report also gives an account of psychological research concerning human music perception. The concept of *Global Joint Accent Structure* is introduced, as a way of understanding how melody and rhythm interact to make "good music". The idea is based on earlier psychological research concerning *Joint Accent Structure* (Jones, 1993), but focuses globally on the interaction between rhythmic accents in the music and the melody. A program (called *PAC*) that composes popular music based on the ideas presented, is described. The important parts are methods for creating tempo, basic rhythms, chord progressions, phrase structures, the rhythm of the melody and the pitches of the melody. Some results are presented and the main conclusion is that the method handles tempo, rhythm, chord progressions and phrase structures well but need further improvements when it comes to the pitches of the melody. Some ideas on how to improve this is suggested.

Contents

1 Introduction

How does one write popular music? A songwriter would perhaps mention feeling as an important aspects whereas an engineer would focus on technical terms. I believe as stated in *Introduction to Computer Music* (Collins, 2010) that it is perfectly possible to capture something important about human music with models devised by humans. But fact is that there are few, if any, studies that cover the topic of popular songwriting from an engineers point of view. This report aims at presenting some ideas of how to algorithmically compose popular music. I will try to extract some rules that can easily be implemented and describe what an implementation of these rules can look like. The resulting program is called PAC, an acronym for *Popular Algorithmic Composer.* The program was implemented in Pure Data (Pure Data, 2011), a graphical programming environment.

Several authors on popular songwriting writes about text, harmony, leitmotif, and title (Cole, 2006; Webb, 1998; Citron, 1985) but other features are covered as well. For example Stephen Citron (1985) claims that a listener's greatest satisfaction comes from form and balance as a song entirely made of choruses will eventually become tiresome for the listener. Jason Blume (Blume, 2004) addresses how a melody of short phrases are easier to remember than a melody with longer phrases. Just as how a number divided into 2782-1227-1210 is easier to remember than a number written 2782122712-10.

There has been a lot of work done on algorithmic composition of classical music (Cope, 2000; Tanaka et al, 2010; Farbood & Schoner, 2001), and there are reasons why this style is studied more. For one there is the formalization of species counterpoint made famous by Johann Joseph Fux. It is a codified description of songwriting for several musical lines that combine to make a harmonic structure. Another reason may be the historically special position that classical music has in music science whereas popular music is not as investigated. Perhaps this is slowly changing, as noted by cognitive psychologist professor Daniel J. Levitin (2006).

Some ideas about melody as a whole (Narveson, 1984; Holm, 1984) will be mentioned in the theory section and discussed further. Studies of music are however often focused on more precise elements. To incorporate all of these small rules into a generative grammar of music seems unfeasible in an algorithmic model. Instead a broader idea would be easier to use. The Gestalt psychologists did focus on music in a broader sense. Grouping of elements and relationships between elements were considered important (see Levitin, 2006). Drake, Dowling and Palmer (1991) have suggested that an accent structure that coincides with the meter makes the segmentation of the melody sequence into smaller pieces as effective as possible. I will try to create an approach to popular music derived from music psychology and I call the idea *Global Joint Accent Structure.* This idea is based on earlier psychological research concerning *Joint Accent Structure* (Jones, 1993).

Some statistical models will be described as well. A statistical approach is based on the observation that humans seem to have an unconscious statistical understanding of music where probable movements in music are experienced as pleasurable (Huron, 2006). This means that a statistical model would involve the same processes and framework as the listeners evaluating the music. The performance of a musical piece is closely intertwined with how the music is perceived (Friberg & Battel, 2002). This subject will however not be covered in this report.

2 Theory

2.1 Melody & Rhythm

Let us start by focusing on the interaction between melody and rhythm in western popular music. Western music has a rich tonal tradition. But the classical music that laid the foundation for the popular music of today has a striking absence of complicated rhythm, at least compared to other cultures. That an absence of complicated rhythm is combined with a strong tonality and vice versa suggests that there is a connection between the two. Let us first look at melody.

2.1.1 Melody

The melodic concept has been used with the same meaning as today since the beginning of the 17th century (Friberg & Ahlbäck, 2009). A good definition being *'A succession of notes, varying in pitch, which have an organized and recognizable shape'* (Kennedy, 1980 as cited in Friberg & Ahlbäck, 2009). The melodic movements, on the surface seemingly random, will reveal multiple patterns at a closer look. The elements (rhythm, melodic intervals, etc.) of some phrases or sub-phrases are often coming back in new shapes throughout a piece of music.

Narveson (1984) attributes great significance to this reoccurring key phrase or sub-phrase and calls it the melodic core (MC). The MC normally consists of 3-6 notes but this of course varies with every song. The MC is here the first appearance of the key phrase and when it appears again in new shapes it is regarded as a core variation.

There are numerous ways to vary the core. Probably the most commonly used variation idea is the alteration variant (A) where the MC is played with the same relative differences but in higher or lower pitch. When the core is not altered in any way this is referred to as the exact repetition (ER) of the MC. The mirror form 'inverted' (I) that plays the MC with inverted pitch intervals is also implemented in PAC. One sometimes encounter multiple MCs that interact in a phrase. Some typical forms are A-B-A-B, A-A-B-B and A-B-B-A (Narveson, 1984). Here the two melodic cores A and B take turns and interact in different patterns. The idea of the MC and multiple MCs will be incorporated in the phrase concept of the program. PAC allows a great variety of lengths for its phrases and they can consist of repeating alterations or a progression of non-connected phrases. In addition, shorter phrases can be joined together in a bigger phrase where the different parts will consist of alternating MCs. Note here that the rhythmic patterns of phrases are also an intrinsic part of the MC. They will too be repeated. Repetition is important as music is very repetitive to its nature:

"...there is probably no other stimulus in common human experience that matches the extreme repetitiveness of music." (Huron, 2006, p. 141)

There are a number of details that need to be explored when considering the melody as a whole. For example, the highest note of a song is remarkably often the tonic or the sixth in sung folk music (Holm, 1984). They occur far more often then the other five alternatives together. For the lowest note, the tonic and the fifth are even more dominating. Another aspect when looking at the bigger picture is to explore if there is a pattern for the tonal direction within the phrase. Data collected by Craig Sapp (2011) from the Tirol folk songs (2011) indicate that phrases seem to start in an upward direction and end in a downward direction. As an example the most common three-note ending is 3-2-1 and the most common start is 1-2-3 followed by 3-4-5.

To examine this pattern further data from 80 000 songs in Themefinder (2011) was examined by the author. The songs are not marked with phrase indicators but the very start of each song can be analyzed. Here 1-2-3 is the start notes for 1770 songs and 3-2-1 is the beginning of 619 songs. This can be related to the total count of the three-note combinations in the Themefinder database. They were extracted by splitting the melodies into all the possible three-note combinations. The results were that 1-2-3 occurs 23732 times and 3-2-1 occurs 32312 times. This means that 7.5 % of every 1-2-3 comes in the beginning but only 1.9 % of every 3-2-1. This making the upward 1-2-3 approximately 4 times more inclined to start a song if their mutual commonality is taken into consideration.

Related statistics have been found by Huron (2006). He examined the phrase contour in a large set of folk songs and classified each phrase into nine different types. What he found was that 40 % of the phrases belonged to a convex arch-shaped type, which confirms that upward movements tend to start and downward movements tend to end phrases. These last details considering melody as a whole were not implemented in the program but it would be desirable to examine them further. I propose some ideas about such implementations in the discussion section.

2.1.2 Rhythm and meter

It is likely that a rich tonal tradition provides a need for a stable rhythm. The reason might be found in the functioning of the working memory. Rhythms are a means for the listener in the quest to decipher the melody, as rhythms help to divide the melody in equal parts, referred to as chunking.

According to Dowling the rhythmic organization of a melody is perceptually more salient than the note pattern (Dowling, 1993). And according to a similar theory, the listener group melodies (without accompaniment) based on the perceptually strongest of the two patterns, the rhythmic pattern (Monahan, 1991). This means that a melody (here just a pitch pattern of a song) played together with the wrong melodic rhythm pattern, makes the song hard to identify.

We will now look at the meter. It has been shown by Caroline Palmer and Carol Krumshansl that listeners rates “the goodness of fit” for tones as higher if these tones coincide with important beats in the metric structure (Palmer et al, 1990, as cited in Huron, 2006). Tempo and meter can be linked to the rhythm. With the help of the rhythmic figures that appear in the music the listener creates an idea of tempo and meter (Monahan, 1993). This is achieved as the rhythmic figures allows us to systematize and predict the meter (Smith, 2000). The part of the rhythm with a maximum amplitude (usually the snare drum and bass drum in western popular music) is the most important part of the framework (Gabrielsson, 1993). We also react to transients with slightly lower amplitude. These are usually, but not necessarily, the beat (pulse) of the music. Implicit in the description of rhythm is some sort of regularity. The rhythmic figures repeat themselves and therefore so does the rhythmic accents (Gabrielsson, 1993). A melody also forms a rhythmic figure. Thus one can conclude that the rhythmic figures of the melody helps to define the meter. Meter becomes a structure that helps us group the melodic pitch pattern. To understand this, one can imagine how difficult a melody would be to listen to (ergo enjoy) if the accompaniment would not play in the same meter as the melody.

2.2 Patterns

When we study music, it is important to understand that humans perceive things like pitch in a different way than an oscilloscope (Monahan, 1991). We perceive it in a logarithmic tonal scale where the notes have relationships to one another: fourth, fifth, etc. If we instead measure the frequencies in hertz (number of periods during one second), the scale is not logarithmic. Our view of the time domain is also logarithmic where we hear whole notes, half notes, etc. The important thing to understand here with regard to pitch and duration of notes is that the listener perceives melodies as part of a whole. What is interesting to the listener is not the pitch in Hz, but how the pitch is related to the notes that have come before. The notes form patterns in the durations of notes and the pitches of the notes. Let us now examine the pattern concepts more closely. Bartlett (Bartlett, 1993) uses "pattern goodness" in order to draw conclusions about melodic patterns. The Gestalt psychologists e.g. Wertheimer (1944), Köhler (1947) and Koffka (1935) (as cited in Bartlett, 1993), studied patterns (including visual patterns) in search of "good pattern" during the 40's. The good pattern would form a whole, a *Gestalt*. Their studies led Garner (Garner & Clement, 1963; Garner, 1970; as cited in Bartlett, 1993), to investigate human perception of patterns. He used patterns formed by dots, and found that the patterns that were ranked best by the subjects were: (a) faster identified (b) easier to remember (c) easier to verbally describe.

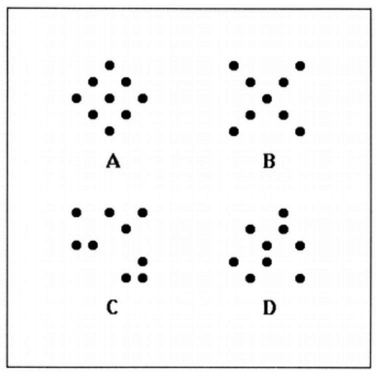

Figure 1: Patterns of dots. A and B are (a) faster identified (b) easier to remember (c) easier to verbally describe.

Most important is that the "good" patterns are perceived to have few alternatives. When people were asked to sort the different patterns into different groups, the patterns that was ranked best was put in groups of fewer patterns. From a statistical point of view Huron (2006) has made the point that probable movements in music is experienced by humans as pleasurable. This as we feel pleasure in finding that we have made accurate predictions about the music to come.

The Gestalt psychologists, and their successors were not only focused on visual patterns. Melodic patterns have long been regarded as a typical example of Gestalts, as they form a distinct, structured whole, and is not just the sum of their parts. Bartlett uses *"pattern goodness"* to draw a conclusion with regard to the scales limitation of the number of alternative tunes: A melody that contains only tones within a scale belong to a smaller group of songs than a melody using tones from several scales. In summary, a melody that moves within a scale is percieved as more unique, as a scale melody is easier to remember than a melody that uses notes from several different scales.

One can easily imagine how, through this thinking, one can group simple melodies in the same scale into smaller groups than the more complex melodies within the same scale. And as we saw earlier, the patterns that are ranked as the best patterns, are the patterns that humans groups into the smallest groups of patterns. However, one must here note that Garner in the experiment on patterns used nine dots. A pattern with only two dots, equivalent to two notes, was not considered. This would minimize the variability so much that the system of grouping patterns had not been particularly useful. A simple melody is instead a melody with the same number of notes as a complex melody. The simple melody would however be easier to predict and describe, and have certain features in its melodic accents. The accent concept is important and we will get back to it. Lets look at human memory functions as it can help us explain the concept of chunking that we earlier touched upon. The concept says that it is easier to decipher the tonal patterns if they are grouped into equal pieces (Monahan, 1993). Another connection between patterns and memory has been put forward by Bob Snyder (2000, as cited in Huron, 2006, p. 197). He sees a group as

"a sequence of events that is held together in short-term memory and processed as a unit before the brain moves on to the next group."

2.3 Memory functions

The short time span in which humans perceive an impression as part of an integrated course of events is called the *psychological present*. Within this time span we perceive the ticks of a metronome or a melody as part of an integrated sequence. When the gap increases to about 3 seconds, each tick is perceived as an isolated event (Encyclopedia Britannica, 2009). Within the psychological present, our *echoic memory* retains melodies and rhythm. They are then processed in working memory.

Working memory is defined as the part of our memory that processes what we can store in our *short-term memory*. We can hold about 5-10 things in our short-term memory at once and keep them for about 30 seconds if they are not repeated (Huron, 2006; Encyclopedia Britannica, 2009). Figures that pretty well match the number of notes in the melodic phrases, the number of melodic phrases that normally interact with each other and the normal time lengths of musical pieces. Estimations of the length of the sound sequences that short-term memory has capacity to store varies between 3-12 seconds (Dowling, 1986; Wittgenstein, 1966, as cited in Huron, 2006). To understand chunking one can picture how smaller pieces of information, as an example four digits, can form separate units in short-term memory and in that way expand the number of digits fourfold. In the same way, longer structures like motives and phrases can be encoded and stored in the immediate memory as "cues" (Rowe, 2001), most easily described as higher level structures. These cues can function as signposts that embody the musical material within. There are also indications that auditory stimuli, in the form of short rhythmic patterns, tend to be processed as mental "atoms" (Huron, 2006).

2.4 Accents

An important tool in understanding how the melody and rhythm fit together is given by Mari Riess Jones (Jones, 1993) in a section on *Joint Accent Structure*. The Joint Accent Structure of a melody is defined as how the melodic and temporal accent coincide over time. Notes with melodic accent define the beginning and, usually, the end of a contiguous sequence of notes. Some examples:

- Large melodic jumps mark the beginning of a new cohesive group.
- Notes where a succession of notes turns from being rising to falling, or falling to rising, represents a melodic accent.
- The notes that constitutes a resolution of a melodic sequence, as an example the last note of a phrase, is an accent.

Note length forms the basis of the time-bound accent. For example, longer notes marks the end of a succession of notes, and is a time-bound accent. By listening tests Jones found that melodies with a Joint Accent Structure where temporal accents and melodic accent coincide with each other over time creates stronger accents. If, as an example, both the temporal accents and the melodic accents repeatedly occur on the third beat we say that they coincide over time. Jones also found that if there is a simple relationship of the distance between the temporal accents and the melodic accent (a consonance), the melody is easier to track. Strong accents with consonance were the easiest to identify.

These ideas of accent structures and how they coincide over time, are also useful when one seeks to explain music in a wider perspective. In Jones experiments, the accent structure of the melody was investigated. If one chooses to see the song as a whole, in which the melody and the rhythm of the accompaniment represents two separate accent structures, it is a reasonable assumption that such an accent structure is useful as well.

2.5 An integrated model of melody and rhythm

We have seen that meter is a framework within which we can easily detect the melody. We have also seen that the listener can extract the music's meter through the rhythmic figures. If you consider that the rhythmic figures define the meter and that the melody is deciphered with the aid of meter it can be concluded that the rhythmic figures help us to decipher the melody.

This confirms that there is connection between rhythm and melody where the former builds a framework for the latter. But there are even more distinct ways in which the two interact. Earlier we saw how Bartlett (1993) drew a conclusion on how we can identify good melodies. He did this with respect to the scales limiting the number of alternative tunes. The same Gestalt reasoning can be applied to how the rhythmic figures limits the number of alternative positions for the notes of the melody. Just consider how humans respond to rhythm. We are moving and dancing (responding) to the rhythm we hear in the music. This illustrates how we predict where rhythmic accents will occur. Another illustration can be provided by the live performance itself (Collins, 2007, p. 181).

"...the human perception of time utilizes prediction rather than reaction, no more so than in musical behaviors like synchronization within ensembles; in order to play together, musicians must anticipate a future point of synchrony, because they would otherwise react too slowly to all perform in union."

Our attention is directed towards the accents so if the notes of the melody falls upon them we perceive the result as a good pattern, a good Gestalt. If the notes occur on the accents we will also be more prepared for them and more easily decipher them. This leads to the following important statement:

The rhythmic figure provides a framework of more or less accentuated beats in which more or less accentuated notes are allowed to form patterns.

I call this idea, or framework, *Global Joint Accent Structure*. Besides of it compliance with Gestalt psychology and pattern goodness it also has some obvious connections to Joint Accent Structure. If the melody and rhythm coincide they form stronger accents. Jones found that a consonance (an agreement, or simple relationship) of the distance between strong accents makes the melody easy to track. We concluded earlier (Gabrielsson, 1993) that the rhythmic accents repeat over time, so the distance between rhythmic accents will repeat. If melody and rhythm coincide we will have a consonance of the distance between strong accents! This means that the song will become easier to decipher. The difference to Joint Accent Structure is that Global Joint Accent Structure does not consider the melody as a closed entity but as part of the accompaniment that surrounds it. A natural approach when you consider how popular music is performed.

The idea of Global Joint Accent Structure is not especially far-off from how people in the music industry regard music. This is a quote from Björn Olsson, renowned Swedish music producer.

"... The goal of basic takes is not about expression first and foremost. It's just a word they (the young musicians) have learned. Listen to Creedence. Wonderfully disciplined. There is nothing in terms of expression there. The accompaniment is there to support the vocal. " (Studio, 2009).

As the lead vocal is almost a synonym for the melody in popular music, the agreement with the Global Joint Accent Structure theory is striking. Huron (2006, p. 187) has examined rhythmic patterns in siciliano (a leisurely dance), and found that these influence a listener's temporal expectations.

"In this case (the siciliano), we can see that it is not simply the strict hierarchical metrical frameworks that influence a listener's temporal expectations. In addition to these metric expectations, listeners also form distinctly rhythmic expectations, which need not employ strictly periodic pulse patterns."

My point is that the temporal expectations of the listener does not have to be induced by genre (here siciliano), but can be induced by just one individual song's rhythmic pattern. It is just necessary with a bar or two of rhythmic patterns, for the listener to form expectations. Here follows a few examples of good Global Joint Accents Structures but there are hundreds. Listen to the snare drum and the kick drum in *Should I Stay or Should I Go* by *The Clash.* At the same time listen to the lead vocal. The vocal is temporally aligned with the drums. The same phenomena can as another example be heard in *The Mamas & The Papas* song *California Dreamin'.* It does not have to be drums that provide the rhythmic accents. Listen to the guitar bass string of *Blowin' In The Wind* by *Bob Dylan* (see Figure 2) to hear an example of where rhythmic accents are present (and coinciding with melodic accents), in songs that do not contain drums. The truth is that this connection is present in almost all hit songs.

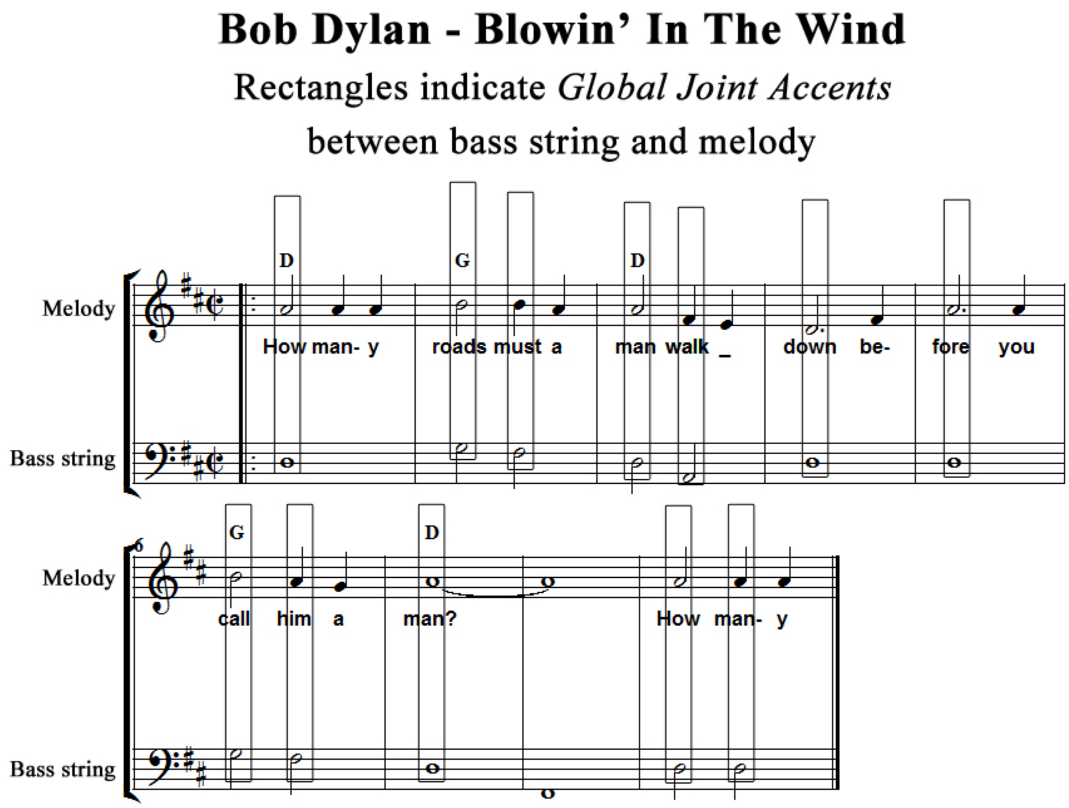

Figure 2: The rectangles indicate *Global Joint Accents* as the pluck of a bass string (rhythmic accents) aligns in time with the melodic accents.

3 Method

3.1 General design

There are numerous methods to compose music algorithmically. With a *generative grammar* the intrinsic parts of the music are described algorithmically. As an example Sundberg & Lindblom (1976) used generative grammar to produce children's songs. Based on the structure of already written songs, rules were derived and new songs produced in a step by step implementation. As mentioned earlier this approach seemed hard to use as the sole basis of the program. However generative grammar is used to some extent where other methods falls short. *Transition Networks* have been successfully used in the past on classical music (Cope, 2000). The idea is to extract information from earlier music examples and mix these pieces of information together in a new musical piece. To use a Transitional Network on popular music would be feasible, but the problem occurs when parts of the music that are not present in the notation are to be used. The rhythm patterns are one example. This would require a clear extraction of these events from the earlier music examples and did not seem feasible for this particular program. *Evolutionary algorithms* are not used as the key to a good evolutionary algorithm is the evaluation function (Nierhaus, 2009). And because of the free form of popular music the evaluation becomes hard to perform. The main method for the program instead became *Markov models*.

The usage of *Markov chains* in music dates back to the 50's (Roads, 1996). They are applicable because music seems to consists of conditional probabilities in many forms (Huron, 2006). The idea behind Markov chains, formulated by the Russian mathematician A. A. Markov, is to determine the probability of future events based on events from the past. A Markov chain can be represented as a transition matrix, where probabilities for future outcomes are listed.

	Rainy	Cloudy	Sunny
Rainy	0.4	0.5	0.1
Cloudy	0.2	0.5	0.3
Sunny	0.1	0.2	0.7

Table 1: A simple Markov chain of order 1. The weather of yesterday provides probabilities for today's weather.

The order of a Markov chain represents the number of past events accounted for. A too low order may cause the output to behave seemingly random and a too high order may just make the output a mirror of the data that created it (Nierhaus, 2009). A higher-order model may therefore not contain any relevant information. PAC uses Markov models for the creation of rhythmic patterns, chord progressions, phrase combinations and melodies. Markov chains can be used to determine other characteristics of the music as well including timbre and instrumentation (Xenakis, 2001). Due to the size of higher order transition tables, Markov models are most commonly used as a way to imitate the style of a data set (Nierhaus, 2009). This because a higher-order Markov chain would be tiresome to create by hand. However PAC is not based on earlier data, but instead analysis-by-synthesis (see below), and the problems of size has been solved by restricting the order and carefully restrict the possible outcomes..

There is a tight connection between melody and harmony in popular music. The harmony is sort of a delineation to the melody, and it is not far away to apply Bartlett's reasoning concerning scales, to harmony as well. Notes of the chord are an extended restriction to the melody in comparison to scales. Harmony, through chord changes, also helps, like the rhythms, to divide the music into bars, etc. If the Markov models does not contain chord information some rules for how chords and melody interact will have to be incorporated within the program. Studies have been done on finding a key given some specific notes, and finding harmonic models by "uncertainty-reducing" predictions (Temperley, 2006). Another solution is to create different Markov models for different chords and chord progressions. A variant of the former idea was used for Markov chains of order 1 and the latter idea was implemented for Markov chains of order 2. (To be able to reach the Markov chain of order n the all the previous states n-1, n-2,...,n-n are of course modeled as well.) The Markov chain was not a data model but based on analysis-by-synthesis. The probabilities were set by the author alone, because of the amount of work that needed to be done. This meant that the results became a subjective mind map of just one persons sense for melody/harmony interaction. As it is a chain based on analysis-by-synthesis it also meant that the Markov chains where not modeled on previous songs.

There are some advantages with this approach. In the case where the Markov chain is modeled on previous songs good is strictly defined as common, leading the program to always give the most common note combinations the highest probability. The problem is that the most common movement do not need to be the best movement. And this is not meant from a strictly artistic point of view. Instead the most common movements are movements that is a proper solution in many circumstances and therefore get used more often. But in more unusual harmonic surroundings or with more unusual MCs these movements would perhaps not have been judged better then some more uncommon movements. Simple put, the most common note combinations are not necessarily the combinations that are the most pleasing to the ear. I would like to illustrate the idea with a metaphor. Consider if a computer were to write a speech based on data from earlier speeches alone. It would gather a large number of speeches (political, family related, etc.) and the most common words, sentences or slogans, would be considered the best. When in fact these words, sentences or slogans, are not common because of any specific value (in most cases) they bring to the speech. *"Freedom for all"* may be a common sentence because it applicable in many different contexts. *"Save the shores of New York from pollution"* may be very uncommon. But in the right context just as good. The same goes for musical phrases. Another way of looking at musical context to understand why the non-data approach may produce better results is to look at the melodic range of pop melodies.

The range between the lowest and the highest note (in this report called *ambitus*) is advantageous to keep track of. A reason of its importance is due to the fact that even well trained singers rarely have a vocal range much above 2 octaves (McKinney, 1994). This sets a definite limit to what ambitus the program should have in its compositions. However it is not very common with an ambitus of 2 octaves. The graphs (Figure 3 and Figure 4) are compiled of data from 285 folk songs, collected by Holm (1984), from Denmark and surrounding countries. The graphs show the data and a normal distribution of the data. The mean ambitus found was 8.9 notes (in major scale steps) and the standard deviation was 1.9. This gives us an idea of what ambitus we should aim at in the program (PAC uses a soft limit as described later).

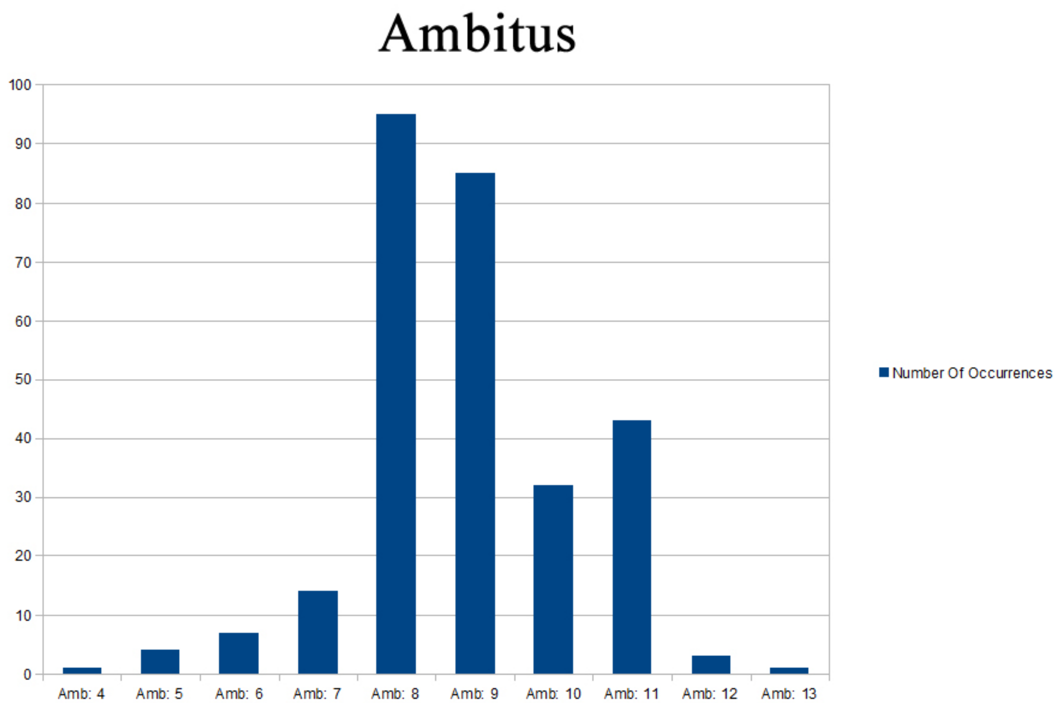

Figure 3: Graph of ambitus, compiled of data from 285 folk songs.

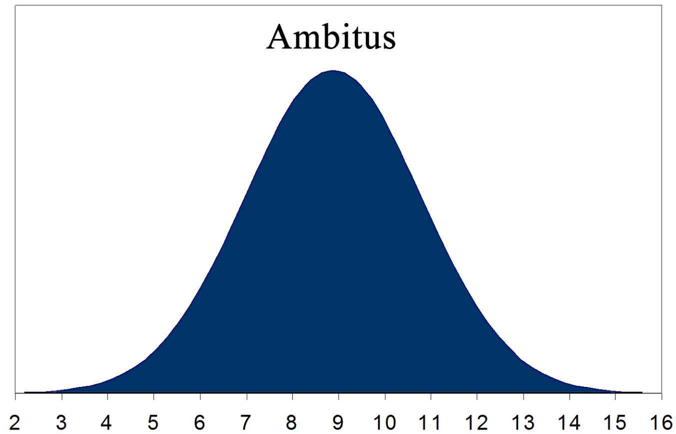

Figure 4: Normal distribution of ambitus in Danish folk songs.

Now let us apply the concept of ambitus using Markov chains that are not based on data but analysis-by-synthesis. An octave leap may sound pleasing to the listener and can receive a fairly high score, but it will never score especially high based on data. This because octave leaps are not very common. At least not if you compare it to a major second. And why would they be? If the allowed ambitus is 8 that means that only one position, the lowest note, will allow an octave leap whereas the seven other positions does not allow it. In case you are positioned at the lowest note the data driven Markov model will still produce a low probability for an octave leap. But analysis-by-synthesis will probably produce a higher probability as octave leaps are not in any way displeasing. And at the lowest note it is now possible to produce the leap as it will not be affected by the ambitus restriction any more. The fact that melodies tend to strive down again after a big jump, so called gap fills (Levitin, 2006) or post-skip reversal (Huron, 2006), also seems to promote ambitus as something an implementation should use (Huron uses the term *tessitura* that represents the comfortable range for a singer, instead of *ambitus* that represents the complete range). But why not just implement a rule that sends the melody downwards after a big leap upwards? It has been shown by Huron (2006) that post-skip reversal depends entirely on the phenomenon of regression to the mean. Melodies only strive down after a big leap upwards if they, after the leap, land above their mean pitch. Of course it most probable, but not definite, that the melody is above its mean pitch after a big jump. This tells us that a correct implementation should keep track of the range in some way instead of keeping track of rules similar to post-skip reversal. Context does not has to stop at the melodic range either as the length of notes could be taken into account as well. An idea would be to make passages of longer notes more inclined to perform longer leaps.

The tempo in popular music usually vary between approximately 70 BPM and 160 BPM. The graphs (Figure 5-6) are based on 123 songs from 10 famous groups/artists in popular music. The data is collected from the BPM Database (2011, BPM Database). Remixes were removed from the data. The mean tempo was 113,7 BPM with a standard deviation of 18,4 BPM.

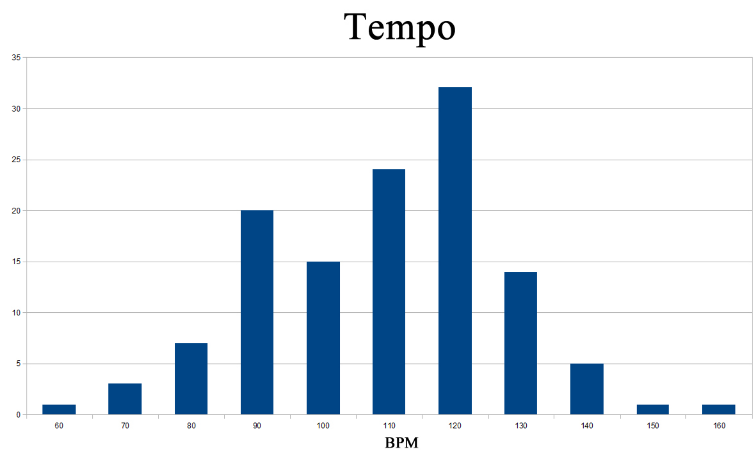

Figure 5: Tempo from 123 songs in popular music. Each song's tempo is rounded down to the nearest ten, creating 11 bars.

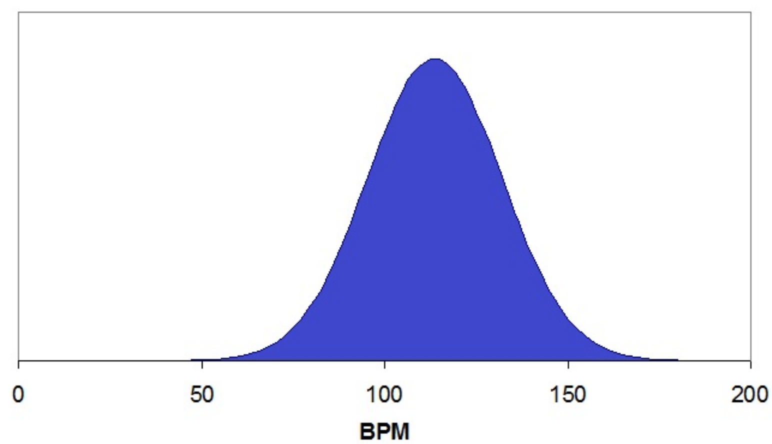

Figure 6: A normal distribution of tempo in popular music. The mean tempo was 113,7 BPM with a standard deviation of 18,4 BPM based on 123 examined songs.

3.2 Implementation

The program works by producing a score through the generation of musical structures. This is the most common approach in algorithmic composition (Nierhaus, 2009). As mentioned earlier, the program was implemented in Pure Data, a graphical programming environment. Presented in Figure 7 is a schematic overview of the songwriting process. The program is throughout non deterministic, meaning that probability and chance always plays a part in the decisions made. Following below is a short description of the different methods implemented.

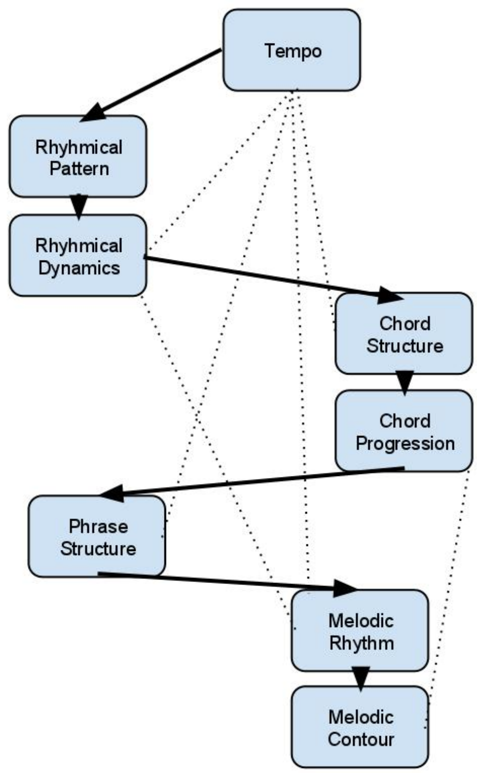

Figure 7: PAC executes in the direction of the arrows. Dotted lines indicate that information is transferred this way as well.

3.2.1 Tempo

Tempo is created first. It is a simple realization of the normal curve for tempo in popular music, shown above. One important aspect of the tempo is how it affects other parameters a songwriting algorithm will deal with. In PAC the tempo influences the length of the notes, chords, phrases and the rhythmic patterns as we will see when these structures are created.

3.2.2 Metrical pattern

PAC only uses drums as a base of the metrical pattern, in this text also referred to as the rhythm (the chords are provided by an organ). First an overall drum pattern is created and then the sound level of each individual part is decided.

The drum pattern consists of three parts. The kick, the snare and the hi-hat/ride. It is constrained to one bar meaning that each bar will have a common rhythmic pattern, something quite ordinary in popular music. We will consider each bar as consisting of 16 discrete time steps where 1 is the first beat of the bar and 16 is the last. A basic rule PAC has adopted is that the kick will hit on the 1st beat of the bar. To simplify things the snare is restricted to beat number 5, 7 and 13 where 5 and 13 are always a snare. The hi-hat can have a main pattern of 16^{th}, 8^{th} or 4^{th}. A ride can have a pattern of 8^{th} or 4^{th}. The algorithm decides how many kicks, snares and hi-hat/ride beats there should be and where they should be placed based on the tempo and probabilities. For the hihat, an increase in tempo means a gradual increase in the probability for a 4^{th} pattern, whereas a decrease in tempo increases the probability for a 16^{th} pattern. It is also more probable with many kicks on each bar if the tempo is low.

The strength of the different rhythm parts is based on the tempo and the rhythmic pattern. As an example 5 kicks in a bar means higher probabilities for a lesser strength in the kick whereas 2 kicks are often stronger.

3.2.3 Chords

The chords available to the program are tonic, subdominant parallel, dominant parallel, subdominant, dominant and tonic parallel. As the key is set to C major this means C, Dm, Em, F, G and Am. The number of chord changes per bar in the music is decided based on tempo with the alternatives being 0.5, 1 or 2 changes per bar. The algorithm assigns a probability in percent for the slower chord change of 0.5 chords per bar by

$$\text{Slow change} = 4 + 0.15\,(tempo - 70)$$

The probability for 2 chords per bar is

$$\text{Fast change} = \frac{70}{6{+}0.1(tempo{-}70)}$$

But a condition for the chords to change in the middle of the bar is that a kick is present there. If not there can not be 2 chords per bar. The probability for 1 chord per bar is given as:

$$\text{Medium fast change} = 100 - (Fast\ change + Slow\ change).$$

Table 2 shows the probabilities as a function of tempo.

Tempo (BPM)	70	100	130	160
0,5 chord/bar	4	8.5	13	17.5
1 chord/bar	84.33	83.72	81.17	79.83
2 chords/bar	11.67	7.78	5.83	4.67

Table 2: The probability (in percent) for the number of chords per bar as a function of four different tempo values.

So how do we build chord patterns? It is quite common that chords repeat in 4-chord patterns. Patterns of 2 chords and 8 chords exists as well and some songs does not contain very much repetition at all. The repetition of chords is in fact more related to the number of bars. Chord combinations often repeat over 2, 4 or 8 bars. If the two last chords in a sequence is twice the speed of the earlier chords this means that there are 5 chords that repeat. To keep things simple enough PAC works with repetition of 4-chord sequences or with 8-chord sequences where each chord has equal length. For four chords this means $6^4 = 1296$ combinations. The number of combinations is small enough to create a Markov chain covering all possible outcomes. This means that probabilities are listed all the way to the completion of the sequence. Table 3 shows the probabilities for the second chord and how they depend on the first chord.

First chord	C	Dm	Em	F	G	Am
Second chord C	2	3	0	20	70	5
Second chord Dm	2	2	5	1	1	5
Second chord Em	2	1	0	1	2	1
Second chord F	39	4	85	1	13	49
Second chord G	20	86	2	76	1	39
Second chord Am	35	4	8	1	14	1

Table 3: The probabilities for the second chord depending on the first chord.

An example of how PAC produces a 4 chord pattern for the first 4 chords of a refrain can be seen in the bar chart (Figure 8). The probabilities for the first chord when the last chord in the verse is G are: C: 60%, Dm: 3%, Em: 1%, F: 28%, G: 0% and Am: 8%. The probabilities are stacked on each other as the chart shows and a random number between 0 and 100 decides which chord to choose.

As an example random number 34 means C and for the next chord the probabilities for C are fetched from the memory. The program then repeats the procedure on the second bar of the bar chart with a first order Markov model. On the third bar a second order Markov model is implemented, etc.

The approach to 8-chord patterns is similar. However the Markov chain does not cover all of the possible outcomes. Instead only the two earlier chords provides probabilities for which chord to choose next. As all songs are written in C major the verse always starts with a C. The 4-chord sequences can repeat exactly or in an altered shape. For the end of the refrain a special function can make changes to the last 3 chords to make sure a resolution from dominant to tonic is present for the melody to dissolve at.

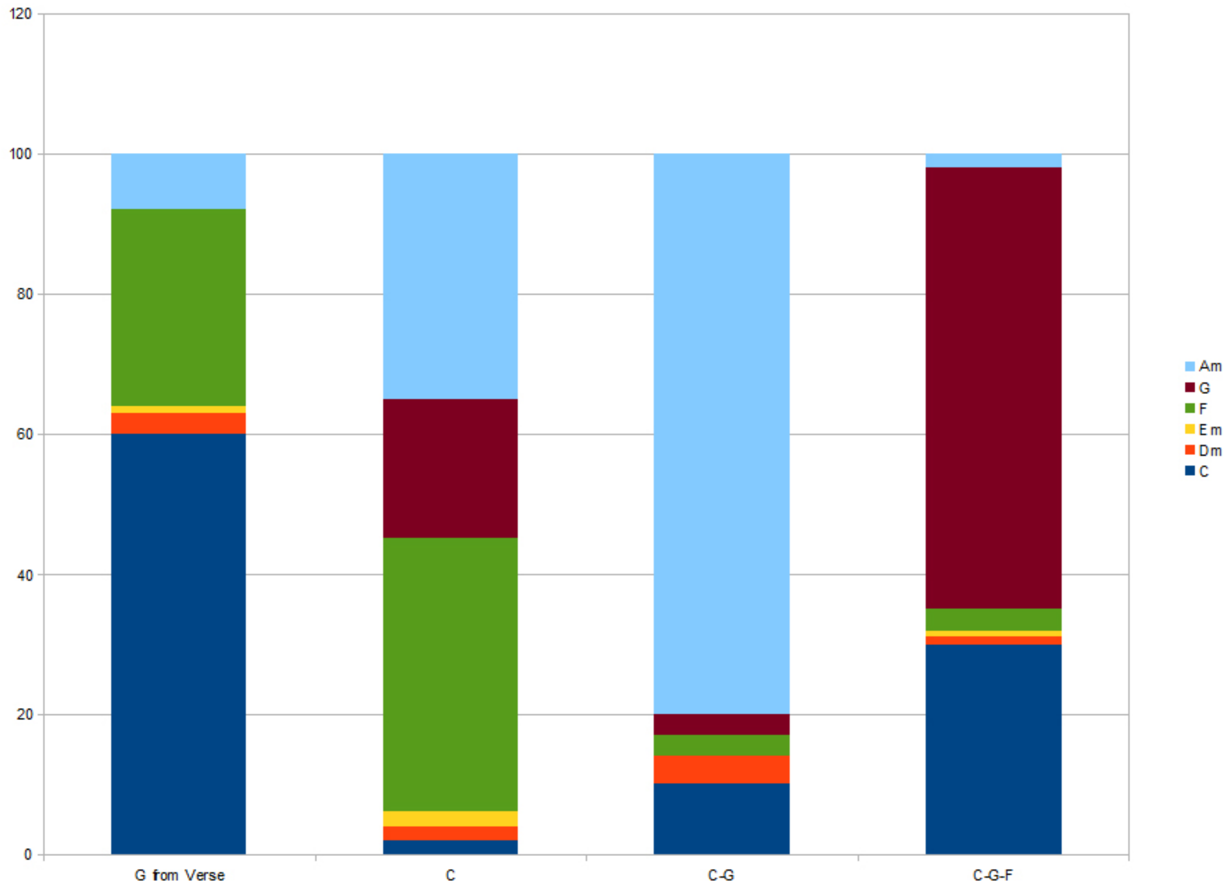

Figure 8: The probabilities as the chorus chord sequence C-G-F-? gets written, represented by bars.

3.2.4 Phrases

The melody in popular music (and most music) is structured into phrases. PAC makes the phrase structure before starting with the melody. The length of each phrase can be between 0.5 bars and 4 bars (8-64 16$^{\text{th}}$ notes). Each phrase can either be followed by a repetition of a earlier phrase or by a new phrase. This is seen in Table 4 where the letters A, B and C represents different patterns and the length of *"A64"* is 64 16$^{\text{th}}$ notes. A repetition means that the note rhythm will be the same and that the probabilities for repeating note intervals (movements) is increased. If two subsequent phrases both have the letter A in Table 4 it means that the second phrase will repeat features of the first phrase. Repeating phrases does not have to be of the same length. As an example a short 0.5 bar phrase can be followed by a 1.5 bar phrase repeating just the first 0.5 bars.

The length of a 1-bar phrase is restricted to 1 bar. This ensures that phrases will not drift away from their given boundaries. In the creation of note rhythm the phrase does not necessarily need to be located in between the bar markers. It can start earlier and end just within the corresponding bar markers for that phrase. A 1-bar phrase (or any phrase for that matter) can however not start earlier than 15 16^{th} notes before the corresponding bar. A 0.5-bar phrase can not start earlier than 7 16^{th} notes. These conditions ensures that phrases will not collide as the note rhythm is established but at the same time it provides flexibility.

Phrases	A8	A16	A24	A32	A64	B8	B16	B24	B32	B64	C8	C16	C24	C32	C64
1^{st} phrase	3	35		55	7										
?															
2^{nd} phrase	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
A8-?	85		6			5		4							
A16-?	5	83		1		2	8		1						
A32-?	1	2		85		1	3		7						
A64-?		2		4	84		1		5	4					
3^{rd} phrase	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
A8-A8-?	33	40				7	20								
A8-A24-?	85	6		1			4		4						
A8-B8-?	42	20				20	13					5			
A8-B24-?	52	10		9		5	12		8					4	
A16-A8-?	75					25									
A16-A16-?	4	75		8					13						
A16-B8-?						100									
A16-B16-?	1	79		10								6		4	
A32-A8-?	80		2			13		5							
A32-A16-?	2	45				3	50								
A32-A32-?		4		91			1		4						
A32-B8-?						70							30		
A32-B16-?						5	91					4			
A32-B32-?	1	4		85		1	2		4			3			
A64-A16-?		80					20								
A64-A32-?		3		77			2		18						
A64-A64	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
A64-B16-?							60						40		
A64-B32-?									60						40
A64-B64	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Table 4: The first rows of the Markov chain for selecting phrase length. Probabilities for the first phrase, the second phrase as depending on the first phrase, and the third phrase as depending on the two earlier phrases. Blanks represents zero probability.

A complete phrase structure for PAC stretches over 8 bars. It is created by combining 2 structures over 4 bars each, or one structure of 8 bars. The restriction is that the Markov chain for creating the last phrase must have an order between 1 and 7. If the Markov chain that created a phrase that completes bar number 4, had an order between 3 and 7 the 4-bar sequence is saved. A new 4-bar sequence is then created partly based on the first sequence. The advantage of this is that a Markov chain can cover all possible outcomes, for short phrases, in 4 bars with a reasonable big data set. But to cover 8 bars with only 0.5 bar phrases would require a Markov chain of order 15. The matrix PAC uses has room for close to 10000 entries (but it is sparse, as seen in Table 4) and that is big enough. Table 4 shows how the 3 first phrases are chosen. Note that as *A64-A64* and *A64-B64* are created they cover 8 bars and are finished. A phrase structure such as *A8-B8-A8-B8-A8-B8-A8-B8* will however have an order of 7 as the last *B8* is chosen. It will by this time only have reached a distance of 4 bars. At this moment there are three alternatives.

- Exact repetition of the first 4 bars.
- Continue at some predeclared position for the next 4 bars. As an example continue from A16-B16 in creating the next 4 bars.
- Do not let the last 4 bars depend on the first 4 bars and start from scratch.

Note also (Table 4) that the phrase-tree grows with a factor of approximately 5 for each added order to the Markov chain in the beginning. As phrases get finished more often at the higher order this factor gets lower at the end.

3.2.5 Melody

Construction of the melody is divided into two main areas. The first is the construction of melodic rhythm and the second concerns melodic movement in the pitch domain.

Creating melodic rhythm PAC checks which phrase combinations and lengths earlier functions decided and creates corresponding melodic rhythms. The melodic rhythm is created separately for each non-repeated phrase. The note lengths implemented are: Half notes, dotted quarter notes, quarter notes, dotted eight notes, eight notes and sixteenth notes. The last note length of the phrase is not explicitly stated and depends on the distance to the next phrase. Therefore other values than the ones mentioned are possible. A probability map (Figure 12) for the positioning of notes acts as an important part of melodic rhythm. The probability map consists of probabilities for the position of the next note. The graphs (Figures 9-12) visualize the creation of a probability map. The tempo of the song is 97 BPM and the two earlier registered note lengths are a dotted 8^{th} note followed by a quarter note. The last note started at position 5 and the rhythm is a simple drum rhythm. The first graph represents the *Rhythm Map*, the second is the *Tempo Map*, the third is the *Markov Chain Map* and the last the resulting *Probability Map*. The smoothing factor is 0.5 and the bar height represents chance in percent to be picked as the next note.

- The rhythm pattern (Figure 9) is the first factor in deciding the Probability Map. The kick, snare and hi-hat pattern affects the Probability Map based on position and strength, where a beat on a particular position increases the probability for the melody to rest on that position. The strength of the beat decides how much the probability is increased, where a stronger beat results in higher probability. Note that this part of the program is a realization of the idea of *Global Joint Accent Structure*. Positions without any rhythm accents receives a probability as well, but they are smaller. Note that the height of the bars is the sum of the drum parts that have an on-set at that moment. The higher the bars are, the stronger the accent is at that position, and the more probable it is that a note will fall there.

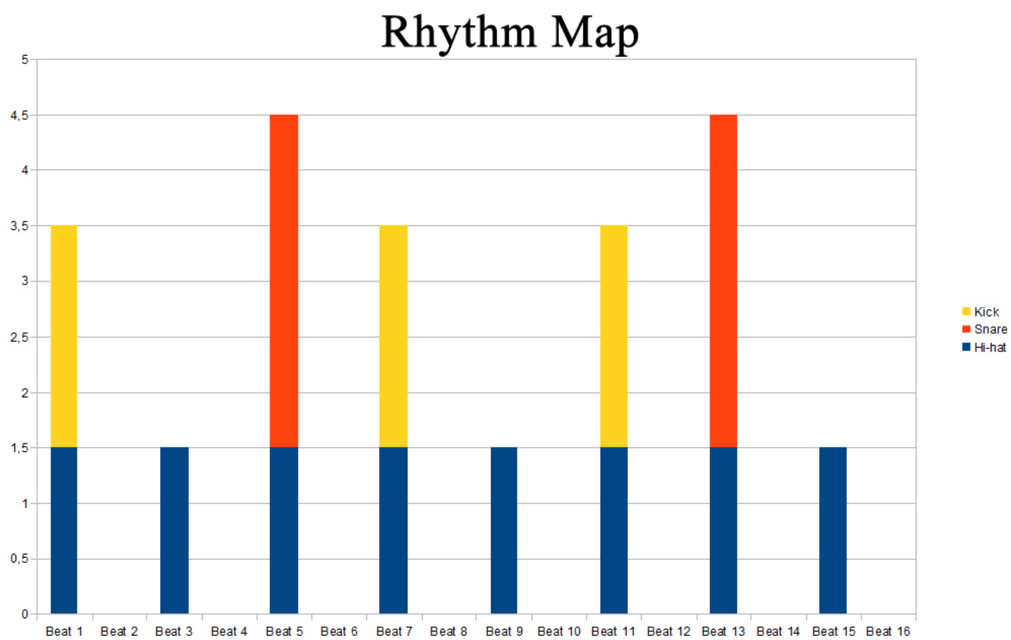

Figure 9: The first step in the creation of a *Probability Map* for melodic rhythm is the *Rhythm Map*. Here a basic rhythm is visualized. The total bar height as the different drum parts are added together represents the strength of the rhythmic accents. Higher bars - stronger accents - means a higher probability of notes occurring, which is consistent with the idea of *Global Joint Accent.*

- The second factor (Figure 10, Table 5) is the tempo of the song. Simply put, a higher tempo makes the probability for longer note values (half notes etc.) higher and a slow tempo makes the probability for shorter note values (sixteenth notes etc.) higher. The equations used to achieve this can be seen in Table 5. The transition is smooth with one setting for each BPM-value. The probabilities for each separate note length is combined with the probabilities given by the third factor, a Markov chain. They are then multiplied with the corresponding positions in the rhythm map, based on the start-position of the previous note.

Note value	Condition	Function
16^{th}	$tempo < 120$	$1.26 - 0.025\,(tempo - 70)$
8^{th}	-	$1.5 - abs(0.012\,(tempo - 115))$
Dotted 8^{th}	$tempo < 124$	$0.181 - 0.002\,(tempo - 70)$
Quarter	-	$0.45 + 0.003\,(tempo - 70)$
Dotted Quarter	-	$0.1 + 0.002\,(tempo - 70)$
Half	-	$0.05 + 0.002\,(tempo - 70)$
Whole	$tempo > 110$	$0.0008\,(tempo - 110)$

Table 5: The functions that determine the probability for each note length. The resulting values are in the range of $0 \leq value \leq 1.5$ with higher values representing more probable.

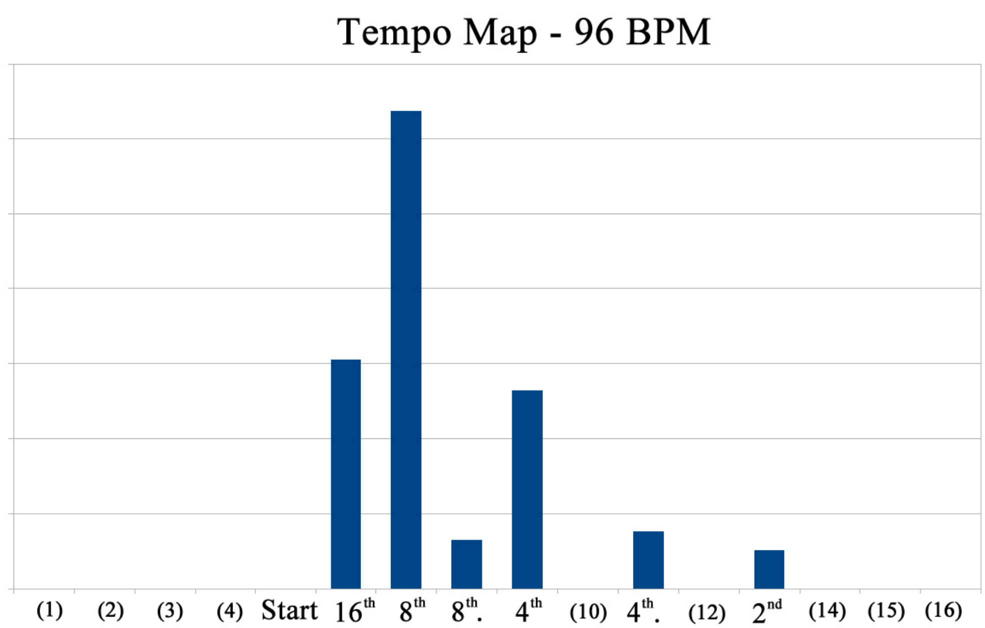

Figure 10: A *Tempo Map* is created for the current note position. Here the *Tempo Map* for a tempo of 96 BPM. The height of the bars visualize the probability for the next note occurring at that position.

- The third factor (Figure 11, Table 6) is a Markov chain of order two, based on earlier note lengths. There are different Markov chains depending on what position in the bar the last note had. The positions are: Position{1}, position{9}, position{5, 13} (Table 6), position{3, 7, 11, 15}, and position{2, 4, 6, 8, 10, 12, 14, 16}. This was necessary, as an important part of the direction of the music, lies in what positions in the bar the music came from. The length of a note is as a consequence of the work flow actually decided when finding the position of the next note. This means that the Markov chain provides probabilities on the decision of the length of the previous note. The probabilities produced by the Markov chain is added as described above.

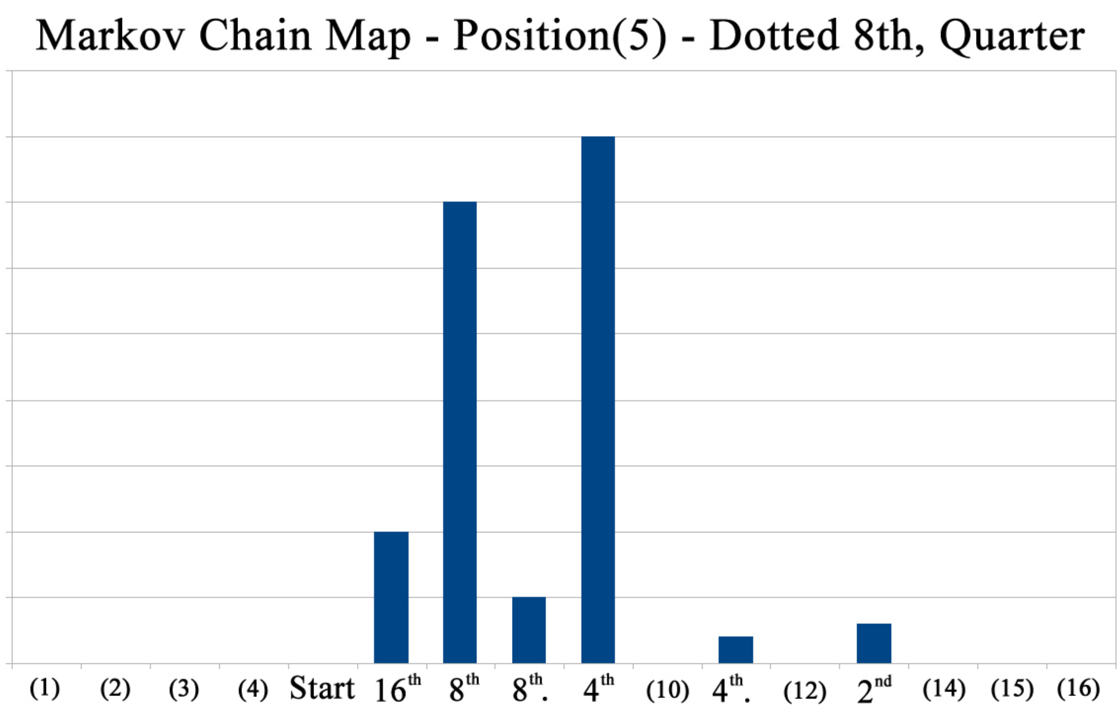

Figure 11: A *Markov Chain Map* for the current note position (position{5}) is displayed. The previous notes have in this example been a dotted 8^{th} note followed by a quarter note. The bar height represents the probability for the next note occurring at that position. A higher bar means a higher probability.

Note Values	16^{th}	8^{th}	Dotted 8^{th}	Quarter	Dotted Quarter	Half
?						
16^{th}-?	75	20		4	1	
8^{th}-?	51	43		6		
8^{th}. -?	50	40		8	2	
4^{th}-?	75	15		8	2	
4^{th}. -?	75	15		8	2	
2^{nd}-?	75	15		8	2	
16^{th}-16^{th}-?	36	56		8		
8^{th}-16^{th}-?	6	78		13	2	1
8^{th}. -16^{th}-?	15	75	1	6	3	
4^{th}-16^{th}-?	5	71		20	2	2
4^{th}. -16^{th}-?	2	88		8	2	
2^{nd}-16^{th}-?	2	84		10	2	2
16^{th}-8^{th}-?	20	48		25	6	1
8^{th}-8^{th}-?	12	41		40	6	1
8^{th}. -8^{th}-?	12	41		40	6	1
4^{th}-8^{th}-?	12	41		40	6	1
4^{th}. -8^{th}-?	12	41		40	6	1
2^{nd}-8^{th}-?	12	41		40	6	1
16^{th}-8^{th}. -?	18	52		20	10	
8^{th}-8^{th}. -?	5	70		20	5	
8^{th}. -8^{th}. -?	15	40		40	5	
4^{th}-8^{th}. -?	3	88		5	4	
4^{th}. -8^{th}. -?	3	25		65	6	1
2^{nd}-8^{th}. -?		24		70	6	
16^{th}-4^{th}-?	5	20		45	30	
8^{th}-4^{th}-?	2	35		33	30	
8^{th}. -4^{th}-?	5	20		45	30	
4^{th}-4^{th}-?	3	22		35	40	1
4^{th}. -4^{th}-?		55		30	15	
2^{nd}-4^{th}-?		26		30	40	4
16^{th}-4^{th}. -?	6	53		25	12	4
8^{th}-4^{th}. -?	2	60		22	15	1
8^{th}. -4^{th}. -?	5	80		10	5	
4^{th}-4^{th}. -?	2	66		26	3	3
4^{th}. -4^{th}. -?	2	75		20	2	1
2^{nd}-4^{th}. -?	2	55		18	23	2
16^{th}-2^{nd}-?	15	71	1	8	3	2
8^{th}-2^{nd}-?	67	30		2	1	
8^{th}. -2^{nd}-?	12	80		15	7	1
4^{th}-2^{nd}-?	40	50		5	5	
4^{th}. -2^{nd}-?	13	65		15	6	1
2^{nd}-2^{nd}-?	13	55		25	6	1

Table 6: A Markov chain for note length at position$\{3, 7, 11, 15\}$. Blanks represents zero probability.

- At last a settable smoothing (the n^{th} power of each probability where n is a number between 0,2 and 1,5) is applied to the map and the probabilities are normalized.

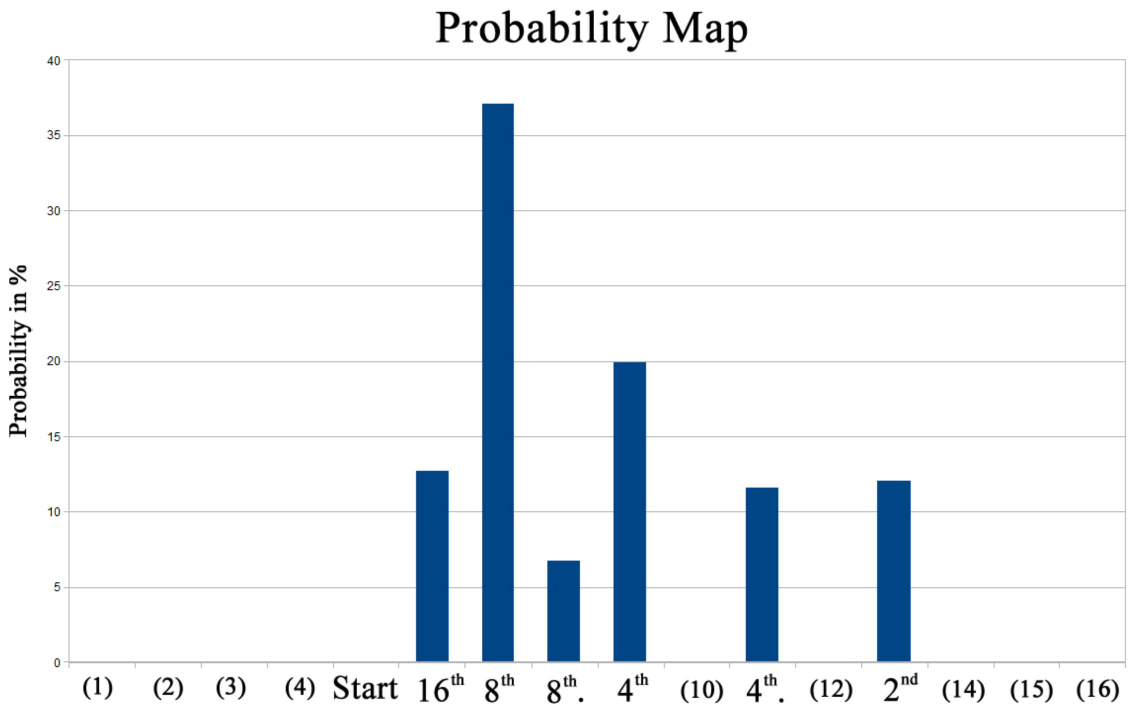

Figure 12: The final result is a *Probability Map* that consists of the probabilities for the next note position. This Probability Map is derived from the *Rhythm Map*, the *Tempo Map* and the *Markov Chain Map*

Procedure The starting point when creating the melodic rhythm is that an approximate start-position for the melody is chosen based on probability. Factors are speed of the chord progression and the length of the phrase that begins. Then the exact position is chosen based on the above mentioned Probability Map, with the difference that the weights based on tempo are not built on previous notes, and the exception that the Markov chain can not be applied. For the rest of the notes the probability map is the base a note is chosen on. The phrase length is decided based on probabilities for the different phrase lengths. When the length is reached or exceeded the phrase has ended. If as an example a phrase is 1.5 bars long and a repetition of that phrase is 0.5 bars long the longer phrase is first created and the shorter phrase is created by splitting of the beginning of the longer phrase.

Creating the pitches

The pitches are created separatedly for each phrase. This means that phrases with rhythmic connections (A32-A32, C16-C8, etc.) are still treated individually. For these particular phrases earlier melodic movements are however taken into consideration. Markov chains of order 1 and 2 are used but they are based on the chords in different ways. The reason is that the multitude of chord combinations made it too time-consuming to design separate Markov chains of order 2 for every combination with the analysis-by-synthesis model.

For the first note of the song a probability table, based on how the notes harmonize with the corresponding chord, is utilized (Table 7).

Scale tones →	c	d	e	f	g	a	b
Am	92	28	85	5	12	91	20
C	94	30	95	16	87	26	22
Dm	20	90	26	92	24	88	8
Em	5	18	87	18	89	24	83
F	85	26	18	87	29	99	9
G	18	92	28	27	95	30	84

Table 7: Notes are given probabilities based on how well they harmonize with each chord.

A Markov chain of order 1 is then used. It handles any allowed chord, or chord combination. The equation for the probability of the next pitch in the Markov chain of order 1 is

$$Probability = \sigma_{prob}\beta\alpha^{\gamma\ell}$$

and the different parts of this equation will be described below. Finally Table 8 shows how the probabilities for the next note in a given situation can be calculated.

The first step of the function is to use the chord based probability table (Table 7) for the two notes, the past note and the present note, to create a total score

$$\alpha = \frac{0.5(past\,note \times present\,note)}{100}$$

for their chord harmonization. Then the interval between the two notes is taken into consideration. This means that instead of representing every possible note combination for every possible chord a probability for different intervals β is instead used. A low score for the harmonization α is also made more important if the interval is large, whereas a smaller interval does not depend as much on harmonization. This is achieved by introducing the power γ. Restrictions are:

$$0 < \alpha \leq 1 \ , \ 0 < \beta \leq 1 \ , \ 1 \leq \gamma \leq 3.$$

The ambitus σ and note length ℓ is also taken into consideration . The ambitus of the whole melody is constantly updated and notes that tries to go beyond the ambitus are awarded lesser and lesser probability the more they reach away. Both the allowed ambitus and the softness of the border is settable. The equations for finding a factor based on ambitus σ_{prob} follow below.

Reduced probability for high notes: $\sigma_{val} = \frac{\big(\big(\eta_{high} - \eta_{prev}\big) + (\sigma_{set} - \sigma_{pres})\big) - \lambda}{\varphi} + 1$

Reduced probability for low notes: $\sigma_{val} = \frac{((\eta_{prev} - \eta_{low}) + (\sigma_{set} - \sigma_{pres})) - \lambda}{\varphi} + 1$

$0 < \sigma_{val} \leq 1 \longrightarrow \sigma_{prob} = \sigma_{val}$

$\sigma_{val} > 1 \longrightarrow \sigma_{prob} = 1 + \frac{\sigma_{val} - 1}{4}$

$\sigma_{val} = 0 \longrightarrow \sigma_{prob} = 0.04$

$\sigma_{val} < 0 \longrightarrow \sigma_{prob} = 0.01$

- η_{high}= Highest note.
- η_{low}= Lowest note (used on falling intervals).
- η_{prev}= Previous note.
- σ_{prob}= The factor derived from the ambitus when deciding probability, based on the current song structure and a proposed new interval.
- σ_{set}= Allowed ambitus.
- σ_{pres}= Present ambitus.

The interval that shall be evaluated is represented by λ and the the softness in how the probabilities shrink when breaking the ambitus is φ with values $1 \leq \varphi \leq 7$.

The length of the note also affects probability. A longer note gains higher probability if it harmonizes well with the chord whereas a shorter note that harmonizes well instead looses some probability. On the other hand for notes with poorer harmonization a longer note looses probability and a shorter note gains. In PAC ℓ represents this change with note length and it is calculated by

$$\ell = \frac{tempo}{115}\left(\frac{note\ value}{16} + 0.5\right)$$

where note value is 16 for a 16^{th} note etc. The idea of a connection between note length and note harmonization is not new. It has been described, among others, by Sundberg & Lindblom (1976) where poor harmonization was referred to as suspension. PAC places great emphasis on writing melodies that harmonizes well with the chords.

Note	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C
Interval	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
α	0.92	0.6	0.885	0.485	0.52	0.915	0.56	0.92	0.6	0.885	0.485	0.52	0.915	0.56	0.92
β	0.01	0.05	0.1	0.3	0.47	0.8	1	1	1	0.8	0.47	0.3	0.25	0.2	0.2
γ	3	2.8	2.6	2.4	2.2	1.35	1	1	1	1.35	2.2	2.4	2.6	2.8	3
σ_{prob}	0.33	0.67	1	1.08	1.17	1.25	1.33	1.25	1.17	1.08	1	0.67	0.33	0.04	0.01
ℓ	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
$\sigma_{prob}\beta\alpha^{\gamma\iota}$	0.003	0.01	0.08	0.07	0.16	0.90	0.81	1.16	0.76	0.75	0.12	0.05	0.07	0.002	0.002
Prob. %	0.1	0.2	1.6	1.4	3.2	18.2	16.4	23.4	15.4	15.2	2.4	1.0	1.4	0.05	0.05

Table 8: The probabilities in a Markov chain of order 1 is calculated. Harmonization, ambitus and note length are taken into consideration. The chord is Am and the previous note was a C in Am as well. The note length is 4 and the tempo is 130 BPM. We also have $\eta_{high} = 12$, $\eta_{low} = 6$, $\eta_{prev} = 10$, $\sigma_{set} = 7$, $\sigma_{pres} = 6$ and $\varphi = 3$.

A Markov chain of order 2 works with some specific note/chord-combinations. It handles all the cases where the 2 previous notes and the present note belong to the same chord. It also handles some chord progressions that are common. This Markov chain works in a more conventional way than the above-mentioned, and outputs probabilities for notes following two specific earlier notes. Below (Table 9-15) is a Markov chain of order 2 for notes that belong to the chords F-F-G. Notes are written as numbers where C{1, 8, 15}, D{2, 9, 16} and so on. A score of 0-100 is given to each note progression over the specific harmony and this scoring has been done by the author as a analysis-by-synthesis model. The probabilities produced by the Markov chain of order 2 are then processed further based on note length and ambitus.

	2	3	4	5	6	7	8	9	10	11	12	13	14
8-4-?		16	5	20	15	20	13	26	14	12	13		
8-5-?		5		10		5							
8-6-?	13	5	5	73	30	65	20	53	24	24	10		
8-7-?	7	4	3	34	26	42	12	20	14	16	12		
8-8-?	18			55	15	70	20	80	30	25	66	20	
8-9-?	13			20	5	54	5	63	10	26	38	10	
8-10-?					7	10	8	38	17	15	15		
8-11-?						30	10	35	16	27	35		
8-12-?							5	14	10	26	24	10	
8-13-?								17	15	16	50	12	30

Table 9: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 8 (C).

	4	5	6	7	8	9	10	11	12	13	14
9-6-?		12		12							
9-7-?											
9-8-?	5	16	24	32	13	37	12	16	19		
9-9-?				5		23	8	12	18		
9-10-?						16	4	10	12		
9-11-?						12			19		
9-12-?									16		
9-13-?									8	4	3

Table 10: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 9 (D).

	5	6	7	8	9	10	11	12	13	14
10-6-?	8		12							
10-7-?			4							
10-8-?			7		16					
10-9-?			8	6	14			18		
10-10-?			2		20			10		
10-11-?			21		28	5	15	35	9	14
10-12-?								4		
10-13-?								18	12	17

Table 11: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 10 (E).

	7	8	9	10	11	12	13	14	15	16
11-8-?	37	5	25	20	18	33				
11-9-?	15		18	10	12	23	13			
11-10-?	20	5	81	20	18	38				
11-11-?	16		43	20	44	46	23	17		
11-12-?	12		24	13	20	34	17	13		
11-13-?	12		35	23	15	46	14	26		
11-14-?								7		
11-15-?								14	3	8

Table 12: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 11 (F).

	7	8	9	10	11	12	13	14	15	16	17	18
12-8-?	5		13			10						
12-9-?			4									
12-10-?			5			4						
12-11-?			18	10	8	15						
12-12-?						9						
12-13-?			23	10	15	43	13	39	15	26	10	5
12-14-?								2				
12-15-?								10		16		

Table 13: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 12 (G).

	7	8	9	10	11	12	13	14	15	16	17	18	19
13-8-?	12		17			25							
13-9-?			6										
13-10-?			16	27	10	18		3					
13-11-?			26	32	10	38		9					
13-12-?		5	38	24	29	74	25	34	10				
13-13-?		8	65	26	23	94	28	57	20	76	20	18	26
13-14-?					4	15	13	40	18	28	8	5	15
13-15-?			65	26	20	58	24	78	27	85	16	15	18
13-16-?			13		8	15	5	26	13	44	17	5	
13-17-?						16	14	35	7	16	10	5	
13-18-?						10	8	37	18	34	27	20	24

Table 14: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 13 (A).

	12	13	14	15	16	17	18
14-12-?	3						
14-13-?	15	2					
14-14-?			1				
14-15-?	20	5	25	4	29		
14-16-?					1		
14-17-?							
14-18-?							1

Table 15: A Markov chain of order 2 with F-F-G as the chords to the 3 notes. This table finds the last note of a three-note combination starting at 14 (B).

If the phrase is not connected to an earlier phrase PAC proceeds to chose the notes based on probabilities. A minimum allowed score/probability can be set to sort out the notes with a too low score. If the phrase is connected to an earlier phrase PAC will first look up the movements from the corresponding notes in the original phrase. It then promotes an alteration (A) of previous movements. It promotes A by awarding higher scores to notes that performs the same, or almost the same note intervals as the previous phrase. The first note of a repeated phrase is a special case as consideration must be taken to how much different starting positions would stretch the ambitus. The highest and lowest notes of the earlier phrase are extracted and compared to the ambitus to provide information. PAC also keeps track of earlier notes (from the last phrase) to find probable starting positions for non-repeated phrases.

4 Results

One example of a typical output is presented as a result (Figure 13-14 and Table 16-18). The tempo for the song was 104 BPM. The rhythm map can be seen in Figure 13. All elements (kick, snare and hi-hat) was in this particular song set to equal strength by the program, which is not especially common. But the rhythm pattern was on the other hand basic.

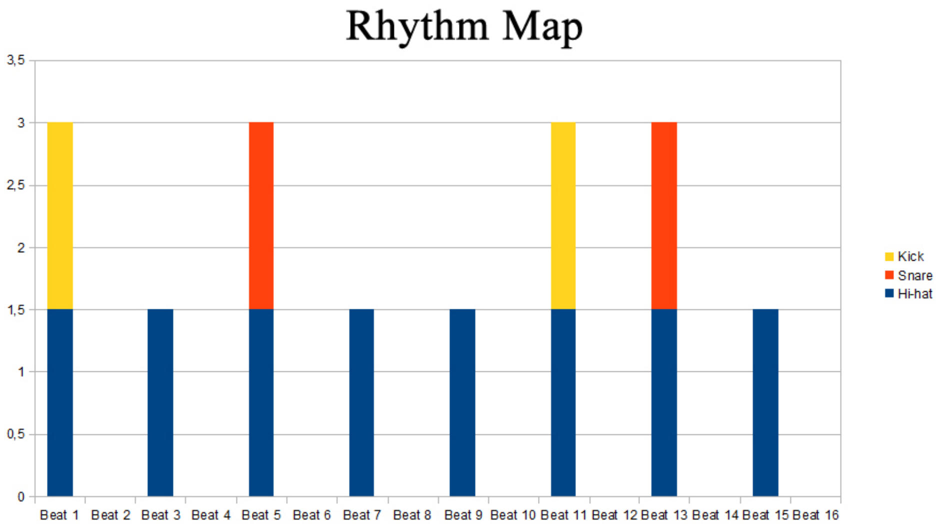

Figure 13: A Rhythm Map with all percussion parts equally strong. The total bar height as the different drum parts are added together represents the strength of the rhythmic accents. Higher bars - stronger accents - means a higher probability of notes occurring, which is consistent with the idea of *Global Joint Accent.*

The chords for the verse was made as an 8-chord pattern and are shown below. The 8-chord pattern is produced by a Markov model of order 2. This means that two earlier chords provide the probabilities for the next chord. The probability for the chords that were chosen is shown below. The last chord of a 8-chord verse is set as G major if the seventh chord is F major.

C	G	Am	Em	F	G	F	G
100%	20%	65%	2%	39%	61%	11%	100%

Table 16: An 8-chord pattern for a verse with their probabilities to get chosen is displayed.

The refrain is built by two 4-chord patterns and is one of the most common patterns. The probabilities again displayed in a table. For the first chord of the refrain (F major), the last chord of the verse is taken into consideration. For the second 4-chord pattern there are two alternatives. It can either be based on the first 4-chord pattern or be a new one. In this song the chords are based on the first 4-chord pattern and the resulting chords are deterministically derived.

F	G	C	Am	F	G	C	C
28%	76%	88%	87%	100%	100%	100%	100%

Table 17: Two 4-chord patterns combined to create the refrain, with resulting probabilities.

The phrase structure for the verse is shown below. A, B and C represent a variation in note rhythms and the number represents the length in 16$^{\text{th}}$ notes.

A32	A8	B8	C16	A32	A8	B8	C16

Table 18: A phrase structure for a verse.

The beginning of the verse is shown below. It starts off with an unusual octave leap. The melody complies with the chords. The first phrase is however rather long and could have benefited from some longer notes or pauses during the G-chord. A smoother melodic contour would have improved the first phrase significantly. Note that the second phrase (A8, three last notes) tries to mimic the melodic contour in the first phrase. It is however restricted by PAC from doing an octave leap as it would have resulted in a too great ambitus.

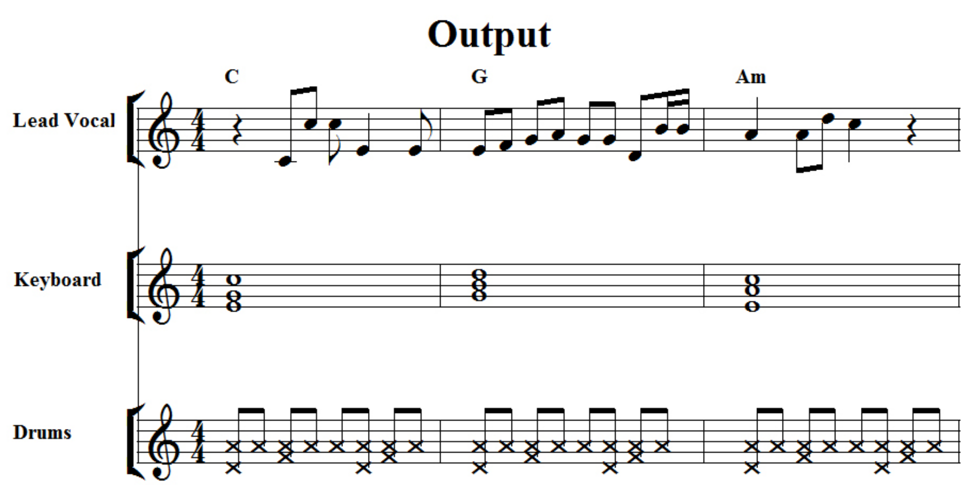

Figure 14: The beginning of a song created by PAC.

5 Discussion

The main conclusion is that PAC handles tempo, rhythm, chord progressions and phrase structures well but need further improvements when it comes to the melody. Here especially the way PAC decides pitches of the melody can be improved. I see three parts of this that can be developed further.

- The order of a Markov chain defines how many of the earlier pitches that are taken into account as the new pitch is created. As the harmonic content was taken into consideration there was a need to create a large set of data for the probabilities. The creation of large data sets gets more time-consuming in the analysis-by-synthesis model than in a model based on machine analysis of popular songs. There are two ways to handle this problem.
  - Use popular songs as a basis for the Markov models. An example would be to use songs that have achieved a certain position on a specific hit chart, as an example the *Billboard Hot 100 Charts.* Much of the work would be to transcribe these songs into a simple data model upon which statistical analysis can be performed. An advantage with this approach would be that if these songs are transcribed into a data model where chords and basic rhythm structures are notated other features of the songs can be extracted as well. As an example the idea of Global Joint Accent Structure could be tested.
  - Expand the equations now performed on the Markov chain of order one. By the equation from section 3.2.5

    $$Probability = \sigma_{prob}\beta\alpha^{\gamma\ell}$$

    a probability can be derived without explicitly defining the harmonic content of each possible movement. Nor is it necessary to define the length explicitly. This means that the Markov chain can easily be expanded to higher orders. An advantage with this approach is that a mathematical model for pitches in popular music will be created. Such a model can easily be modified an it will offer a greater plasticity than the model based on earlier pop songs. A combination of the two ideas will likely bring good results.
- The second problem with the handling of pitches in PAC is the insufficient understanding and control of the melodic contour it has. Ideas regarding the melodic contour is presented in the last sections of 2.1.1. They are not used in PAC as can be seen at 3.2.5 but an implementation of a arch-shapes in the melodic phrases would probably produce good results.

- Tension and release is an important part of music. A link to probability have been expressed by Temperley (2006). *"..a passage of low probability is likely to have an effect of tensions and instability."* Temperley suggests that unexpected events are more difficult to process, and thus cognitively taxing. These ideas are closely related to ideas presented earlier from Huron (2006). A way to achieve tension and release would - from a probabilistic point of view - mean that probable and not so probable movements should be varied. Other aspects of tension and release that should be taken into consideration is to dissolve the melody at the tonic when appropriate.

Note that all instruments that contain transient information can be regarded as rhythmic instruments. Their main contribution to the song may not be rhythmic, but just listen to an acoustic guitar. The strum of the strings presents a clear rhythmic pattern. Ultimately this means that it is always possible to identify a rhythmic pattern in a piece of music. This means that the voice leading in classical music (also rhythmic) can be regarded as the basis of both the Global Joint Accent Structure and as the harmonic basis. It is providing means for Bartlett's (1993) *"good pattern"* in two dimensions at once. And so does the strummed guitar of popular music.

I would like to add that the basic rules of a good composing algorithm, the rules one should strive to find, should be rules that are closely intertwined with human perception of music. Rules based on data can, every now and then, be broken with good results as they do not represent the underlying principles in what we enjoy in music. The idea that rules are made to be broken is widely spread in the music community, here exemplified by a quote from the Brazilian composer and teacher Frederico Richter.

"In music, rules are made to be broken. Good composers are those who manage to break them well." (Miranda, 2001, p. 206)

The idea of Global Joint Accent Structure are in my point of view not a rule to be broken. So are instead rules about the length of a song, or the sound of a song. Global Joint Accent Structure is more a concept about what humans enjoy in music.

Some musicologists and psychologists do not believe that there is much or perhaps any long-term perception of music at all. (Cook 1987, Tillmand & Bigand 2004 as cited in Collins 2010). There is a lot of processing going on in short term memory. But one can also see some clear evidence of where long-term memory is important. It is in the long-term memory we store songs and even styles of music or chord progressions. It is fairly common with musical climaxes towards the end of a piece and when refrains have occurred a number of times they are often varied a little by the end. Another point to be made is that composers almost always re-use melodies for the refrain and the verse on several occasions in the composition.

The discussion of long versus short-term memory has a connection to computational models. If appreciation of music occur both in long- and short-term memory, a program will need to be built on principles that do not only take into account the short term aspects of music. This will mean that the program will have to develop a greater understanding of where the music piece is going on a macro level. It will need to explore variations performed earlier in the piece to build progress. It will also have to take into account what people that listen to the music have heard before, and cannot be disconnected from the world around. This would mean using old music pieces not only to derive what they say about good song writing but also to use them as a palette of note combinations that humans are used to. Another example of macro level thinking can be applied to the end of a song. Pieces of music with a stable rhythm often have a ritardando in the end, resembling how a human stops running (Friberg & Sundberg 1999 as cited in Friberg & Battel 2002).

Even if the algorithmic composer does not perform every step of the composition computers can still be of aid. Perhaps the algorithm can provide inspiration and sketches that a human composer can further develop. With the speed of computers the composer can focus on specific parts as exemplified by Xenakis (2001, p. 144).

"Freed from tedious calculations the composer is able to devote himself (it was the 60's) *to the general problems that the new musical form poses. For example, he may test all instrumental combinations from soloist to large orchestras. With the aid of the computer the composer becomes sort of a pilot."*

6 Acknowledgements

This study was supervised by Anders Friberg, who assisted with ideas for the subject, and material concerning the subject. It was written as a project in the course *"Advanced individual course in musical communication"* held by Anders Friberg. Some parts are based on reports earlier written by the author.

References

[1] Bartlett, J. C. (1993). Tonal Structure of Melody, In Dowling, W. J., Tighe, T. J. (Ed.) *Psychology and Music. The understanding of melody and rhythm.* Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

[2] Blume, J. (2004). *6 Steps to Songwriting Success.* Billboard Books.

[3] Bresin, R., and Friberg, A. 2000. Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal* 24(4): 44-63

[4] Citron, S. (1985). *Song writing. A Complete Guide to the Craft.* London. William Morrow & Company, Inc.

[5] Cole, B. (2006). *The pop composer's handbook - a step-by-step guide to the composition of melody, harmony, rhythm and structure - styles and song writing techniques from rock, reggae and salsa to bhangra, club and steel band.* Mainz, Schott.

[6] Collins, N. (2007). *The Cambridge Companion to Electronic Music.* Cambridge University Press.

[7] Collins, N. (2010). *Introduction to Computer Music.* UK, John Wiley & Sons Ltd.

[8] Cook, N. (1987). *Musical form and the listener.* The journal of Aesthetics and Art Criticism, 46(1):23-9.

[9] Cope, D. (2000). *The Algorithmic Composer.* A-R Editions, Inc., Madison, Wisconsin.

[10] Dowling, W. J. (1986). *Music Cognition.* Richardson, Texas, Academic Press Inc.

[11] Dowling, W. J. (1993). Procedural and Declarative Knowledge in Music Cognition and Education, In Dowling, W. J., Tighe, T. J. (Ed.) *Psychology and Music. The understanding of melody and rhythm.* Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

[12] Drake, C., Dowling, W. J., Palmer, C. (1991). Accent Structures in the Reproduction of Simple Tunes by Children and Adult Pianists. *Music Perception*, Vol. 8, No. 3, p. 315-334. University of California.

[13] Encyclopedia Britannica. (2009). Chicago, Illinois.

[14] Farbood, M & Schoner, B. *Analysis and Synthesis of Palestrina-Style Counterpoint Using Markov Chains.* MIT Media Laboratory.

[15] Francès, R. (1988). *The Perception of Music.* Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

[16] Friberg, A. Ahlbäck, S. *Recognition of the Main Melody in a Polyphonic Symbolic Score using Perceptual Knowledge.* Speech, Music and Hearing, KTH, Sweden, Royal College of Music, Sweden

[17] Friberg, A. & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the acoustical society of America,* 105(3), 1469-1484.

[18] Friberg, A. and Battel, G. U. (2002). *Structural Communication,* Chapter 2.8 of R. Parncutt & G. E. McPherson (Eds., 2002) The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning New York: Oxford University Press.

[19] Friberg, A. and Sundström, A. (1997) Preferred swing ratio in jazz as a function of tempo, *Speech Music and Hearing Quarterly Progress and Status Report*, 4/1997, pp. 19-28.

[20] Gabrielsson. A. (1993). *The complexities of Rhythm - Psychology and Music. The understanding of melody and rhythm.* Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

[21] Garner, W.R. (1970). *Good patterns have few alternatives.* American Scientist, 58, 34-42.

[22] Garner, W.R., and D.E. Clement (1963). *Goodness of pattern and pattern uncertainty. Journal of Verbal Learning and Verbal Behavior*, 2:446-452.

[23] Holm, P. (1984). *Melodianalyse.* Sydjysk Universitetsforlag, Esbjerg.

[24] Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

[25] Jones, M. R. (1993). Dynamics of Musical Patterns: How do Melody and Rhythm Fit Together?, In Dowling, W. J., Tighe, T. J. (Ed.) *Psychology and Music. The understanding of melody and rhythm.* Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

[26] Kennedy, M. (1980). *The Concise Oxford Dictionary of Music.* London: Oxford University Press.

[27] Koffka, K. (1935). *Principles of gestalt psychology.* New York: Harcourt, Brace & World.

[28] Köhler, W. (1947). *Gestalt psychology.* New York: Liverlight.

[29] Levitin, D. J. (2006) *This Is Your Brain On Music: The Science of a Human Obsession.* New York: Dutton Adult (Penguin).

[30] McKinney, J. C. (1994). *The Diagnosis and Correction of Vocal Faults: A Manual for Teachers of Singing and for Choir Directors.* Nashville: Genevox Music Group.

[31] Miranda, E. (2001). *Composing Music with Computers.* Focal Press. Oxford.

[32] Monahan C. B. (1993). Parallels Between Pitch and Time and How They Go Together, In Dowling, W. J., Tighe, T. J. (Ed.) *Psychology and Music. The understanding of melody and rhythm.* Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

[33] Narveson, P. R. (1984). *Theory of melody : a complete general presentation.* University Press of America, New York.

[34] Nierhaus, G. (2009). *Algorithmic Composition, Paradigms of Automated Music Generation*. Springer. Wien.

[35] Olsson, B. (2008). *Studio.* Nr 8 2008. Stockholm.

[36] Palmer, C., Krumhansl, C. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance.* 16 (4): 728-741.

[37] Roads, C. (1996). *The computer music tutorial.* The MIT press. Cambridge, Massachusetts.

[38] Rowe, R. (2001). *Machine Musicianship.* The MIT Press. Cambridge, Massachusetts.

[39] Sapp, Craig. (2011). *Statistics presented in a personal mail communication,* see also Tirol Folks songs.

[40] Smith, L. M. (2000). *A Multiresolution Time-Frequency Analysis and Interpretation of Musical Rhythm. University of Western Australia.* Department of Computer Science.

[41] Snyder, B. (2000). *Music and Memory: An Introduction.* Cambridge, Mass., MIT Press.

[42] Sundberg, J. & Lindblom, B. (1976). *Generative theories in language and music descriptions.* Cognition, 4. s. 99-122. Elsevier Sequoia S.A., Lausanne.

[43] Tanaka, T. & Nishimoto, T. & Ono, N. & Sagayama, S. *Automatic music composition based on counterpoint and imitation using stochastic models.* Graduate School of Information Science and Technology, University of Tokyo.

[44] Temperley, D. (2006). *A Probabilistic Model of Melody Perception, Rochester.* New York, Eastman School of Music.

[45] Tillmann, B. & Bigand, E. (2004). The relative importance of local and global structures in music perception. *The Journal of Aesthetics and Art Criticism.* 62(2):211-22.

[46] Webb, J. (1998). *Tunesmitsh. Inside the art of songwriting.* New York.

[47] Wertheimer, M. (1944). Gestalt theory. *Social Research, 11.*

[48] Wittgenstein, L. (1966). *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief.* Compiled from notes taken by Yorick Symthies, Rush Rhees, and James Taylor. Ed. Cyrill Barett. Oxford: Blackwell.

[49] Xenakis, Iannis. (2001). *Formalized Music: Thought and Mathematics in Composition (Harmonologia Series No.6).* Hillsdale, NY: Pendragon Press.

[50] Figure 1 http://psycnet.apa.org/journals/dev/23/6/images/dev_23_6_839_fig1a.gif

[51] Pure Data http://puredata.info/

[52] Themefinder http://rism.themefinder.org/

[53] Tirol Folk songs: http://kern.ccarh.org/browse?l=essen/europa/tirol