Algorithmic Composition of Popular Music

Anders Elowsson,[*1] Anders Friberg[*1]

[*1] *Speech, Music and Hearing, KTH Royal Institute of Technology, Sweden*
[1]elov@kth.se, [2]afriberg@kth.se

ABSTRACT

Human composers have used formal rules for centuries to compose music, and an algorithmic composer – composing without the aid of human intervention – can be seen as an extension of this technique. An algorithmic composer of popular music (a computer program) has been created with the aim to get a better understanding of how the composition process can be formalized and at the same time to get a better understanding of popular music in general. With the aid of statistical findings a theoretical framework for relevant methods are presented. The concept of Global Joint Accent Structure is introduced, as a way of understanding how melody and rhythm interact to help the listener form expectations about future events. Methods of the program are presented with references to supporting statistical findings. The algorithmic composer creates a rhythmic foundation (drums), a chord progression, a phrase structure and at last the melody. The main focus has been the composition of the melody. The melodic generation is based on ten different musical aspects which are described. The resulting output was evaluated in a formal listening test where 14 computer compositions were compared with 21 human compositions. Results indicate a slightly lower score for the computer compositions but the differences were statistically insignificant.

I. INTRODUCTION

Human composers have used formal rules for centuries to compose music. With the introduction of computers it became possible to execute these formalized set of instructions without the direct involvement of the human composer, thus, music could be entirely created by algorithms. In this study popular music has been the musical style of focus. How can we formalize popular music? This is an essential question if an implementation of an algorithmic composer of popular music is to be successful.

In which way may a composition program help us better understand music? As pointed out by Rothgeb (1993) a computer program can contribute by exposing deficiencies in theories about composition. If the computer is unable to compose given specific rules then those rules are incorrect, or they insufficiently describe the composition process. Which algorithmic approach is best suited to the task if we at the same time wish to gain an improved understanding of music? If the approach is to use *Markov chains* or *neural networks*, the achievement of a composition program will be merely that. A generative model where probabilities may interact at many levels can instead provide a deeper understanding. By using such an approach, we intrinsically answer questions about music (Ahlbäck, 2004), while at the same time providing possibilities for easy user interaction. On a similar theme Nierhaus (2012) has also used algorithmic composition to formalize composition techniques used by human composers.

Authors on popular songwriting (Citron, 1985; Webb, 1998; Blume, 2004; Cole, 2006) stress aspects such as lyrics, harmony, refrains and hooks, which of course are relevant to every composer of popular music. But perhaps the most important lesson, often neglected in scientific research, is the emphasis on the composer as a listener. In this project, a constant evaluation of the sounding results has been used to shape the songs into music that better corresponds to the genre. Let us take a look at the more scientific approaches to music composition. A lot of work has been done on classical music (Cope, 2000; Tanaka et al, 2010; Farbood & Schoner, 2001). A reason for the interest in this type of music may perhaps be an already established formalization such as species counterpoint made famous by Johann Joseph Fux. Another reason may be the special position that classical music has reached within music science where contemporary popular music is more or less disregarded (Levitin, 2006).

If the output of the program is to be in accordance with the structure of sung music it seems feasible to use data from sung music as a model for the program. Statistical analyzes have been performed on such data to extract probabilities which were employed in the music generation. The statistical approach to music is supported by the observation that humans seem to have an unconscious statistical understanding of music, where predictable events in music are experienced as pleasurable (Huron, 2006). A statistical understanding has been important in this study and a statistical analysis has been performed on the Essen Folksong Collection (Schaffrath, 1995). Results will be presented to support the methods that are described.

II. THEORY

A. Music & Patterns

Humans sees pattern everywhere. IQ tests are mainly built on finding patterns in a sequence of letters, numbers or pictures. Even where there are no patterns to be found, confronted with random noise input, humans still sees input that breaks small sequences of order as exceptional (Simon & Sumner, 1993). Patterns in music are almost always multidimensional with a constant formation of patterns at many simultaneous levels (Parncutt, 1994b). How should these patterns be arranged? The Gestalt psychologists e.g. Wertheimer (1944), Köhler (1947) and Koffka (1935) (as cited in Bartlett, 1993), studied patterns in search of *"good patterns"* during the 40's. A good pattern would form a whole, a *Gestalt*, when perceived by the observer.

It turns out that the Gestalt psychologists, and their successors were not only focused on visual patterns. They regarded melodic patterns as a typical example of Gestalts, as they form a distinct, structured whole. Bartlett (1993) uses Gestalt psychology in this manner for the limitation a scale becomes to the number of alternative tunes: A melody that moves within a scale will belong to a smaller group of songs than a melody using tones that do not belong to any particular scale. Thus, a melody that moves within a scale will facilitate the perceptual prediction of future events. Beyond scales the melody is imposed several "restrictions" such as harmony,

meter and rhythm. All of these restrictions make the melody easier to decipher, and various ways to ensure that the melody is bound by them will be presented.

B. Music & Memory

Chunking is the process where smaller pieces of information, as an example four digits, can form a single unit or element in short-term memory. This expands our capacity to store information. The process is important in music as it is easier to decipher tonal patterns when they are grouped into equal pieces (Monahan, 1993). It has been pointed out that it is hard for people to chunk information in groups consisting of more than three or four elements (Estes, 1972 as cited in West et al, 1985) and we can see a connection to music as the number of beats per measure often varies between two and four.

In a similar way as chunking, longer structures like phrases can be encoded and stored in the immediate memory as "cues" (Rowe, 2001; Ahlbäck 2004), representing higher level structures. The cues are signposts that embody the musical material within. There are also indications that short rhythmic patterns tend to be processed as mental "atoms" (Huron, 2006). If the melodic structure of the phrases is repeated throughout the song, a sense of cohesiveness can be achieved. Figure 1 shows the probabilities for phrase repetition in the German songs of the Essen Folksong Collection (Elowsson, 2012). Notice that according to the figure, rhythmic phrase repetition is very common as nearly 40 % of the analysed phrases repeat the rhythm of an earlier phrase. Notice also that repetition of contour almost never occurs without repetition of rhythm.
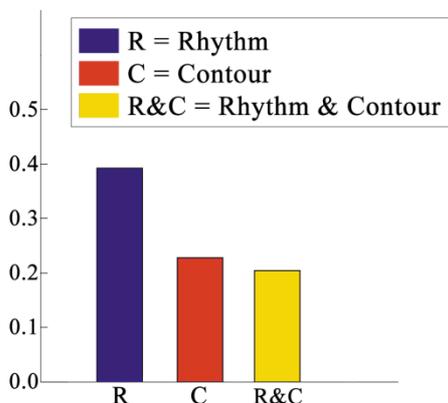
Figure 1. Probabilities for phrase repetition in German folksongs. The probabilities represents how likely it is for any given phrase of a song to be a repetition of an earlier phrase.

Huron (2006, p. 141) concludes:

"...there is probably no other stimulus in common human experience that matches the extreme repetitiveness of music."

Notice that any musical meaningful event leads to expectations of new musical events. Every note that has come to the listener's attention forms new expectations about where the piece will evolve. From a computational perspective this means that the probabilities for future events constantly change.

C. Melody

A good definition of melody is 'A succession of notes, varying in pitch, which have an organized and recognizable shape' (Kennedy, 1980).

When algorithmically processing pitch intervals the size of the intervals are not the only relevant aspect. Listeners recognize repetition of contour (West et al, 1985) so a small falling interval is perceived as very different from a small rising interval, whereas a small rising interval and a somewhat bigger rising leap are not perceived as that different. This is accounted for in the program.

Another interesting aspect is patterns of tonal direction at different positions in the phrase. Huron (2006) has examined the phrase contour in a large set of folksongs, classifying each phrase into one of nine different types. He found that 40 % of the phrases belonged to a convex arch-shaped type, which indicates that an upward movement tend to start and a downward movement tend to end phrases.

As can be seen in Figure 2 (Elowsson, 2012) there is a relationship between note length and interval size. If the notes are shorter the probability for smaller intervals will be higher and if the notes are longer the interval size will on average become larger.
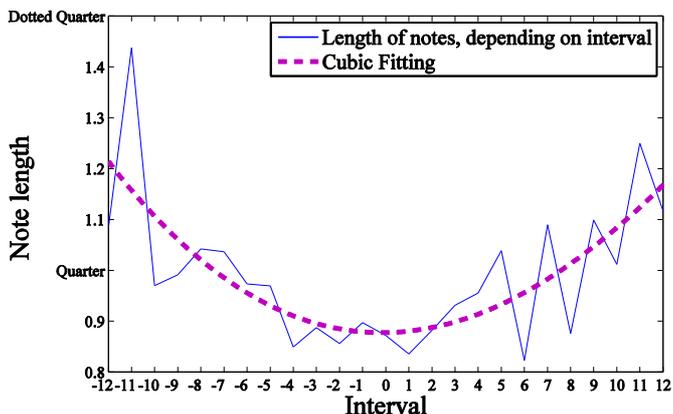
Figure 2. The relationship between interval size and note length in over 1000 German folksongs.

As the melodies will be crafted for the singing voice, vocal range sets a definite limit to the ambitus of our compositions. In this report ambitus is used to describe the total tonal range of the composition and tessitura describes the range within which the singer is comfortable. In what can be considered a regression to the mean, the listener expects the melody to strive towards the mean pitch between the lowest and the highest note (Huron, 2006). An implementation of regression to the mean is used in the program and is covered more extensively in the method section. It features a soft regression within the range which the singer is comfortable (tessitura), a harder regression as the range becomes more expanded and a definite limit so that the ambitus is not greater than what the singer can handle.

Figure 3 shows the ambitus of over 4000 German folk songs in major mode (Elowsson, 2012). Notice that the songs of the Essen Folksong Collections are rather short excerpts. As the songs become longer the ambitus will probably rise a little.
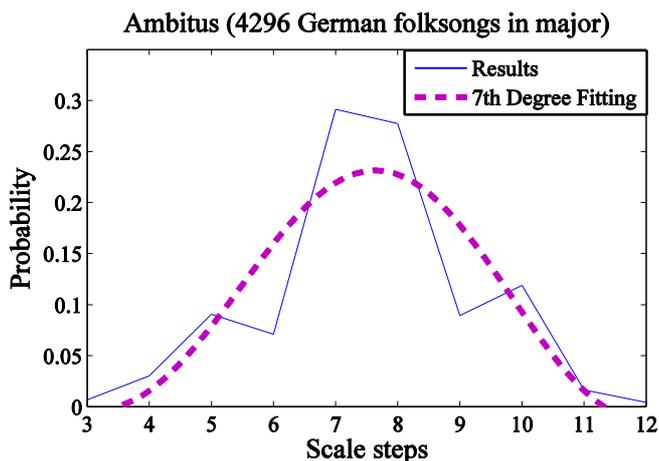
Figure 3. Ambitus of more than 4000 German folksongs.

D. Rhythm, Meter & Tempo

Let us look at a simple drum rhythm in popular music. Pulse sensation is used by Parncutt (1994a) to describe all the different levels evoked in the listener's mind. The parts with the highest amplitudes (usually the snare drum and bass drum) become the most important parts of the perceived rhythmic structure (Gabrielsson, 1993), but we also react to transients with lower amplitude. The rhythmic figures often has some regularity and as they repeat so does the rhythmic accents which they produce (Gabrielsson, 1993).

The listener uses rhythm to create a perception of tempo and meter (Monahan, 1993). This is possible because the rhythmic figure allows us to systematize and predict the meter (Smith, 2000). Just as rhythm, meter becomes a structure that helps us group the melodic pitch pattern. Just imagine how difficult a melody would be to listen to if the accompaniment plays in a different meter than the melody. When notes coincide with important beats in the metric structure listeners perceive a high "goodness of fit" (Palmer et al, 1990, as cited in Huron, 2006).

The tempo in popular music varies between approximately 60 BPM and 160 BPM as can be seen in Figure 4 (Elowsson, 2011). The graphs are based on the tempo of 123 songs from 10 famous groups or artists in popular music.
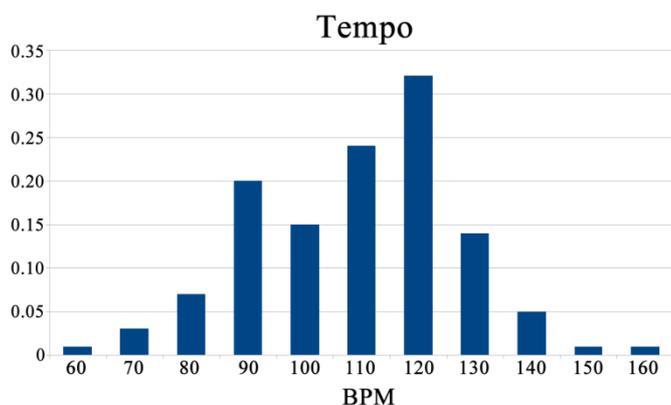

Figure 4. Tempo in popular music.

E. Melodic Accents

Our discussion regarding accents will begin with the Joint Accent Structure theory as described by Mari Riess Jones

(1993). It is an important tool in establishing a connection between melodic accents formed by pitch and melodic accents formed by note lengths. The Joint Accent Structure is defined as how these accents coincide over time.

The accents in the pitch-domain (melodic accents) are formed by large melodic jumps, positions where a succession of notes turns from being rising to falling, or falling to rising and notes that constitute a resolution of a melodic sequence such as the last note of a phrase. Accents in the time-domain (temporal accents) are for example, longer notes that mark the end of a succession of notes.

Jones found that when melodic accents and temporal accents coincide over time the accents becomes stronger. It was also found by listening tests that when there is a simple relationship in the distance between melodic and temporal accents the melody becomes easier to track. These studies considered only the melody. However rhythm also possesses an accents structure as seen in Section D. In the same way as the different accents of the melody are perceptually connected it seems reasonable to assume that an investigation of melodic accents combined with rhythmical accents can be useful as well.

F. An Integrated Model of Melody and Rhythm

Ahlbäck (2004) concludes that the beats can be conceived as points of temporal expectation in music. This assertion seems valid but will be nuanced with a theory of accents strength based on the actual rhythmical content within the music. There is a wide variety of drum patterns in popular music, and it will be shown below that they probably arise as a reaction to the melody. The drum patterns are crafted - often unconsciously – as a support for the melody in the time domain.

Consider dancing, you will find dancing in connection to music everywhere; at a rock concert, a prom or a folk dance. Dancing illustrates how the listeners predict where rhythmic accents will occur. Another illustration can be provided by the live performance (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."

Now remember how Bartlett (1993) used Gestalt psychology to identify good patterns in melodies. The use of a scale limited the number of alternative pitches and made the melody easier to predict. With the same reasoning the rhythmic accents guides our attention towards specific metrical positions, thus limiting the number of alternative positions for the notes to fall upon. In conclusion our attention is directed towards the rhythmic accents and if the notes of the melody fall upon them we perceive the result as a good pattern, a good Gestalt. As the notes occur on accents we will be more prepared for them and more easily decipher them. To summarize:

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

This idea for how the melody is connected to rhythmic accents is called *Global Joint Accent Structure* (Elowsson, 2011). The theory has, besides its name, an obvious connection to Joint Accent Structure. The alignment of important information both in the time-domain and pitch-domain results in stronger salience. Also notice that with this alignment and a repeating rhythm a simple relationship in the distance between strong accents may easily develop which according to Jones (1993) makes the melody easier to track. This as rhythmic accents tend to repeat over time (Gabrielsson, 1993). Perhaps it is so because rhythmically organized patterns are easy to reproduce for the listener (Parncutt, 1994a). A difference to Joint Accent Structure is however that Global Joint Accent Structure does not view the melody as a closed entity. Instead the accompaniment becomes more relevant which seems natural if the performance of popular music is considered.

One listening test that supports the theory has been performed by Ahlbäck (2004). He found that listeners preferred an asymmetrical grouping of rhythm, with good alignment between drums and fiddle over a symmetrical grouping with less alignment. An interesting aspect, which perhaps was what Ahlbäck primarily wanted to show, occurred when grave accents and acute accents in the rhythm had been reversed. This probably shifted the listeners' perception of measure starts and resulted in a poor scoring.

Huron (2006, p. 187) has performed a statistical analysis of the rhythmic patterns in siciliano (a leisurely dance), and he has come to the conclusion that the listeners' temporal expectations are formed by the same recurring rhythm.

"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."

Here genre induces temporal expectations. In popular music there is not one genre specific rhythm pattern but instead a multitude of different patterns can be used. However it is common that one song has a specific pattern and the next song another. The point of Global Joint Accents Structure is that as a specific rhythm repeats several times in a song, temporal expectations are formed for that specific song. To get an example listen to the snare drum and the kick drum in *Should I Stay or Should I Go* by *The Clash* (Jones, 1982) and compare it with the melody as sung by the lead vocal. You will find a temporal alignment with the drums. Another example with the same features is *California Dreamin'* by *The Mamas & the Papas* (Phillips & Phillips, 1965). This phenomenon is not merely found in popular music but in most music with strong rhythmic emphasis. The music styles with the most salient Global Joint Accent Structures are perhaps Hip Hop and Rap. The rhythmic accents do not necessarily need to be provided by the drums. In *Blowin' in the Wind* by *Bob Dylan* (Dylan, 1962) the bass string of the guitar provides important rhythmic accents which aligns with the melody, see Figure 5.

Figure 5. The Global Joint Accent Structure of *Blowin' in the Wind*. Rectangles highlight where rhythmic accents are aligned with the melody.

III. METHOD

The program was written in Java and an outline of the structure is presented in Figure 6. As shown the user controls settings via the GUI and initializes the composition process. The thick arrows indicate the order of execution. First a tempo is created, followed by rhythm, harmony and phrase structuring. Finally the melody is created and playback can commence at the user's command.
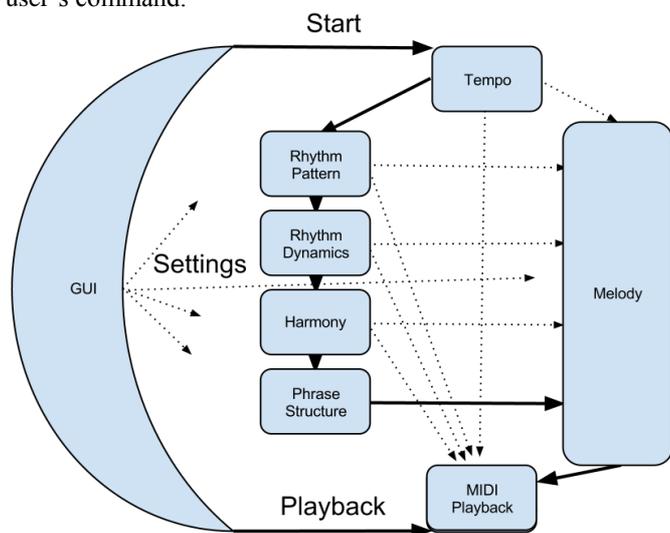

Figure 6. The structure of the program. Arrows represent the composition process and dotted lines represent dependencies.

In the Method section a 16th note constitutes the length 1. The compositions are at the moment restricted to a 4/4 meter. The 16th notes of the measure are referred to as position 1-16. One verse and one refrain are created in each composition.

G. Tempo

Tempo is created from a normal distribution where the user can set a preferred mean of the distribution.

H. Rhythm

The program creates a basic drum rhythm with tempo as input parameter. For the kick it is first decided how many additional kick hits each measure should contain (the kick is always present at the first beat). The number is decided from a normal distribution with a standard deviation of 0.45 where the mean is calculated by:

$$1 + \frac{|tempo - 160|}{40}$$

The result is rounded to the nearest integer and the number of kick hits is only allowed to vary between 1 and 4. Probabilities for different positions in the measure are used to distribute the kick hits.

The hihat can play quarter notes, 8th notes or 16th notes whereas the ride can play quarter notes or 8th notes. The probabilities depend on tempo where a high tempo makes the longer notes more probable.

The dynamic level of the kick, snare, hihat and ride depend on the tempo. For the kick, the number of kick hits is also involved. Below is an example of how a mean dynamic level (internally the dynamic level is described as a number between 1 and 3.3) is calculated for the kick. The variable *num* represents the number of kick hits per measure.

$$\frac{5}{num} + \frac{|tempo - 160|}{110}$$

When the rhythmic foundation is established the total level of the different parts at every 16th note of the measure is calculated. The rhythmic weights that are established will provide probabilities for note positions in accordance with a Global Joint Accent Structure.

The results are added together with a metrical weight representing metrical salience. The default settings for the metrical salience can be observed in Table 1 and they can be adjusted by the user.

Table 1. Weights based on metrical salience.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Weight	10	1	3	1	6	1	3	1	8	1	3	1	6	1	3	1

I. Harmony

The chords are created by a Markov chain where earlier chords provide probabilities for subsequent chords. Only the most common chords in major mode are used and these chords are the tonic, subdominant parallel, dominant parallel, subdominant, dominant and tonic parallel. This means C, Dm, Em, F, G and Am as the key is set to C major during the composition.

The program uses 4-chord sequences or 8-chord sequences where each chord has equal length. For a four chords sequence $6^4 = 1296$ combinations are possible and they are all covered by a Markov chain of order 3. The Markov chain for 8-chord sequences is instead of order 2, where 2 earlier chords affect probabilities for the next chord. As an example Table 2 shows probabilities for the second chord of a 4-chord sequence and how they depend on the first chord.

Table 2. The probabilities for the second chord, depending on the first chord.

First chord	C	Dm	Em	F	G	Am
2nd C	24	35	0	20	70	5
2nd Dm	2	2	5	1	1	5
2nd Em	2	1	0	1	2	1
2nd F	39	4	85	1	13	49
2nd G	20	86	2	76	1	39
2nd Am	35	4	8	1	14	1

The 4-chord sequences can be repeated exactly or in altered shapes. For the end of the refrain a function makes changes to the last 3 chords to make sure a resolution from dominant to tonic is present.

J. Phrasing

A statistical background to phrase repetition can be seen in Figure 1, Section B. To make repetition occur at several levels many different techniques are used. The probabilities for different phrase lengths depend on tempo with relatively longer phrases (number of 16th notes) for faster tempos.

Blocks of length 32 are used to make sure that groups of phrases repeat over time. When a phrase repeats another phrase the note rhythm will be identical in the program. It is also possible that a phrase only repeats a part of another phrase and then only those parts will have identical note rhythms. The repetition in the pitch domain is described in Section K (Repetition). Repetition may also occur for smaller formations than at the phrase level. If a half measure have a note rhythm which is not homogeneous (only 8th notes or only quarter notes etc.) it may be repeated.

The start position of each phrase is altered by a global offset in terms of the number of 16th notes, chosen by a normal distribution. By using an offset we can get phrases of length 16 to go between position (-5-10,11-27,...) instead of position (1-16,17-32,...). How much of each phrase that will consist of pause and how much that will consist of the melody is calculated by:

$$w - 0.0015(tempo - 70)$$

The parameter *w* is a weight with the default value of 0.62. At last, boundaries that constitute where the melody will start and end within each phrase are created.

K. Melody

The melody is generated by iteratively creating one note at the time. First a number of notes of varying pitch and length are suggested. For each of these notes a score is calculated by multiplication of different sub-scores. Ten aspects that all provide such sub-scores are described below. The final note is selected by the probability distribution created by all score values.

The user has the freedom to decide how many of the initial note suggestions, sorted by their score, that the program will finally choose between. The user can also set a power *p*, which if below 1 will lower the difference between the scores *S* for the different note suggestions, and if above 1 will instead increase the difference. This will affect the perceived randomness of the

melody as the probability distribution is altered in favour of higher scoring notes, and the equation is simply:

$$S^p$$

When a note is chosen, which is positioned at or beyond the boundary of a phrase, that phrase is finished.

1) Ambitus. A regression towards the mean pitch is achieved by establishing an allowed ambitus (see Figure 3, Section C). The user specifies a preferred range (default 8) and a maximum range (default 12) as well as an inner and an outer drop-off. With the inner drop off the user can specify how much the score is lowered for each step that the pitch deviates from the median pitch within the preferred range (default 4 %). With the outer drop-off the user can specify how much the score is lowered as the pitch moves towards the maximum range outside of the preferred range (default 15 %). The preferred range represents tessitura, a comfortable range for the singer in which most of the pitches should reside.

2) Harmonic Compliance. How well a given note harmonizes with the chord is an important aspect and is in this paper referred to as *harmonic compliance*. Each of the pitches have been given a value between 0-1 for how well they harmonize with the different chords (Table 3), values that can be edited by the user as well.

Table 3. Default setting for the harmonization between scale notes and common chords.

	C	D	E	F	G	A	B
Am	0.92	0.28	0.85	0.03	0.25	0.91	0.20
C	0.94	0.30	0.95	0.16	0.87	0.26	0.15
Dm	0.20	0.90	0.26	0.86	0.24	0.88	0.02
Em	0.01	0.18	0.87	0.09	0.89	0.24	0.83
F	0.90	0.26	0.18	0.82	0.29	0.99	0.01
G	0.28	0.92	0.28	0.27	0.95	0.30	0.75

3) Intervals & Harmonic Compliance. The size of the interval between two notes has a close connection to harmony. For larger intervals the dependency on good harmonization is much bigger than for small intervals. The harmonization of the notes from Table 3 is used in combination with the size of the intervals to calculate a harmonic compliance of the intervals.

Larger upward intervals are favoured over larger downward intervals, and smaller downward intervals are favoured over smaller upward intervals. Unusual intervals are awarded a lower score even if the two pitches of the interval have a perfect harmonic compliance. Some more rules have been applied as well.

- An interval of at least one pitch step where none of the two notes belongs to the chord is awarded a lower score.
- An interval of at least two pitch steps where one of the notes does not belong to the chord is awarded a lowered score. If both of the notes are pentatonic the lowering is small. If at least one of the notes is non-pentatonic the lowering is bigger.
- An interval of at least two pitch steps where one of the notes is not pentatonic is awarded a slightly lower score, regardless of harmonic compliance.

4) Note Length. As a new note is suggested the previous note will get its length determined based on the onset of the new note. One part of the evaluation of the new note position is therefore an evaluation of the length of the previous note. Here a Markov chain based on statistics would have been a good solution but the complexity that comes with such a solution would have made it harder to overview. Also a connection to tempo was seen as harder to integrate into a Markov chain. In Table 4 the probability for different note lengths and their dependency on tempo can be observed.

Table 4. Calculations of probabilities for different note lengths.

Note Length	Equation
16th	$0.65-0.014(tempo-70)$
8th	$1-\|0.016(tempo-100)\|$
Dotted 8th	$0.151-0.002(tempo-70)$
Quarter	$0.5+0.007(tempo-70)$
Dotted Quarter	$0.35+0.0085(tempo-70)$
Half	$0.15+0.0075(tempo-70)$

A normalization is applied so that the highest scoring note receives the score 1. Negative values constitute a zero probability. Consideration is also taken to positions in the measure for a more musical result. As an example, a note falling on an uneven position is not allowed to have an even length.

5) Note Length & Harmonic Compliance. The program tries to create melodies where longer notes in general have better harmonization than shorter notes. This means that longer notes with poor harmonization are awarded a lower score and that longer notes with good harmonization are awarded a higher score. It also means that shorter notes with good harmonization are awarded a slightly lower score and that shorter notes with poor harmonization are awarded a slightly higher score.

6) Note Length & Interval Size. As can be seen in Figure 2, Section C there is a relationship between note length and interval size. This is implemented in the program so that the probability for a small interval size is higher between shorter notes and the probability for a large interval size is higher between longer notes.

7) Phrase Arch. Compliance with Huron's (2006) findings of convex phrase arches can be ensured if the user choses to.

8) Tonal Resolution. At the end of the refrain the melody will resolve at a tonic. This may happen in the verse as well if there is a position where a dominant V chord is followed by the tonic I. In Figure 7 we see statistical findings for tonal resolution at the end of songs in the Essen Folksong Collection (Elowsson, 2012). The gradually narrowing distance to the tonic, as symbolized by arrows is achieved in the program by a narrowing window.
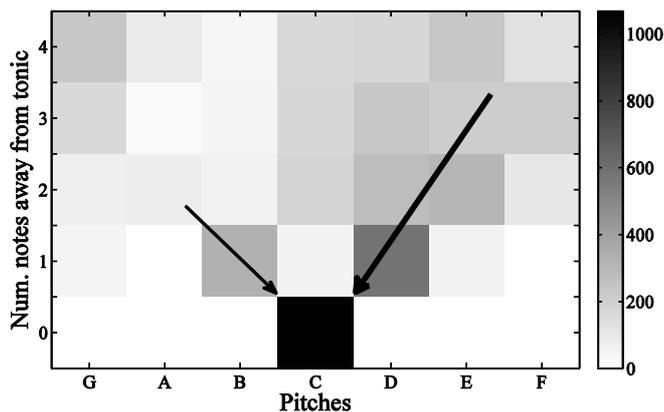
Figure 7. Tonal resolution at the end of approximately 1000 German folk songs.

9) Repetition. Patterns in the melody repeat themselves over and over again both at a rhythmical level and concerning pitch intervals. As we have seen in Figure 1 much repetition comes at a phrase level and the program uses earlier phrases as "mirrors" for the following phrases. If the phrase is to repeat an earlier phrase, consideration is taken to how the intervals between the notes in the phrase correspond to the intervals of the notes in the mirror phrase. Consideration is also taken separately to the difference in contour. The score is given by:

$$c \cdot I$$

The differences in contour determine c. The default setting is
- Same contour = 1.2
- Not same, not opposite contour = 0.9
- Opposite contour = 0.7

The values can be tuned by the user. The value of I is determined by the interval I2 of the mirror phrase and the interval I1 of the current phrase by the equation

$$k^{|I_2 - I_1|}$$

The constant k can be tuned by the user.

10) Good Continuation. To give the melody a sense of direction a higher score is awarded to melodies that continue in a newly established direction. A statistical foundation can be seen in Figure 8 (Elowsson, 2012).
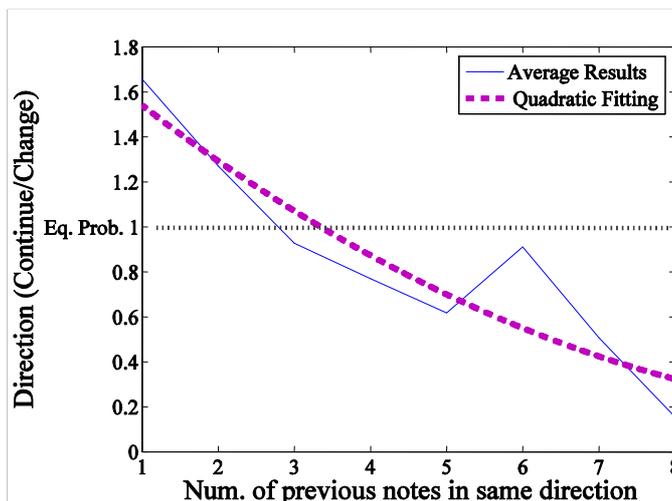

Figure 8. The probabilities for intervals to continue in the same direction after x number of intervals of one pitch step in that direction, based on over 1000 German folksongs.

L. Program GUI

The GUI was developed in the Eclipse (2012) environment for Java and WindowBuilder (2012) was used in combination with Swing. Though the program is not primarily created for commercial use, a friendly user experience was still an object.

Figure 9 is an overview of the GUI. Here we see settings for *harmonic compliance* (top left) described in Table 3 and user settings for *note length* and *phrase length* (top right). Notice that the user only affects their relative probability; tempo affects probabilities in general. In the bottom of Figure 9 are the settings for metrical salience (Table 1).
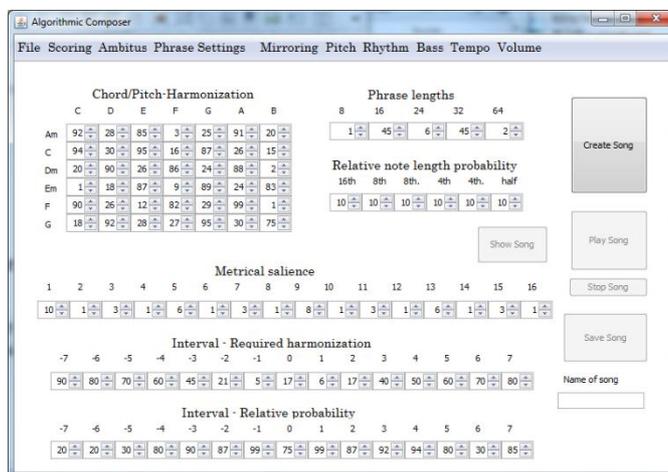

Figure 9. The GUI of the program.

The user can control more advanced settings of the composition with the aid of the drop-down menus at the top of Figure 9. Figure 10 displays three drop-down menus from the GUI for Phrase Settings, Pitch and Rhythm.
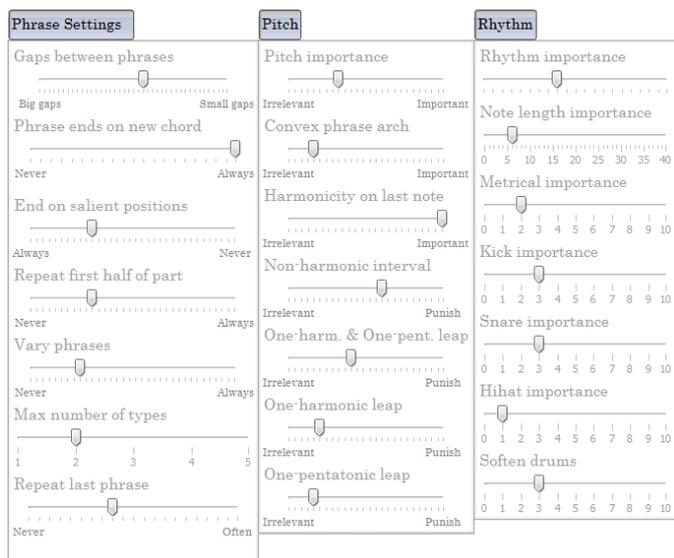
Figure 10. Drop-down menus for Phrase Settings, Pitch and Rhythm.

IV. RESULTS

M. Melodies

Figures 11-12 shows two compositions by the computer program. The rhythm section is omitted.

Figure 11. A composition by the computer program.

Figure 12. A composition by the computer program.

V. LISTENING TEST

A listening test has been conducted on the resulting songs. The set of songs that was used in the test is shown in Table 5.

Table 5. Compositions used in the listening test.

| No. Songs	Type	
7	Selected Computer Compositions	*A*
7	Randomly selected Computer Compositions	*AR*
7	Winner of Norwegian *Melodi Grand Prix* 76'-82'	*N*
7	Author's compositions (Elowsson) - Good GJAS	*GJ*
7	Author's compositions (Elowsson) - Poor GJAS	*GN*

The first group (*A*) is a selection of computer generated compositions. They were selected based on perceived quality by the first author. The second group (*AR*) consists of randomly selected computer compositions. The third group (*N*) consists of the winners in the Norwegian annual music competition *"Melodi Grand Prix"* between the years of 1976 and 1982. The fourth and fifth group (*GJ* and *GN*) are songs with identical melodies composed by the first author. The difference between the two groups is that the arrangements are altered, changing the Global Joint Accents Structure (GJAS). But as the arrangements of the computer music only consisted of simple drum rhythms, organ, bass and a piano melody, and as conformity in the arrangements was desired to make the evaluation fair, the theory of Global Joint Accent Structure was mainly tested by altering the position of the kick drum. That is, for good GJAS the kick drum was temporarily aligned with the melody and for poor GJAS the kick drum was not aligned with the melody. A kick drum did however always appear at the first beat of each measure.

All musical excerpts were rendered with a similar instrumentation using MIDI. Songs were played in random order but the distance between the repeated melody in the songs with a good GJAS and a poor GJAS was always at least 20 songs. Different aspects of the songs were rated on a seven-point scale by 18 participating listeners. The listeners were between 20 and 30 years of age and they rated their own experience as musicians as 4.2 on a seven-point scale. The listeners were among other things asked to rate the quality of the compositions where "Bad" was represented by 1 and "Good" was represented by 7 (Figure 13).

**Figure 13. Rating of the quality of a composition in the listening test.**

The perceived rated quality of the compositions is shown in Figure 14. Here only the first of the two occurrences of songs with or without a GJAS was included, to not let repeated listening distort the results. The results indicate small differences in the ratings of the groups. As can be seen in Figure 14 all five subsets fall more or less within the same confidence intervals.
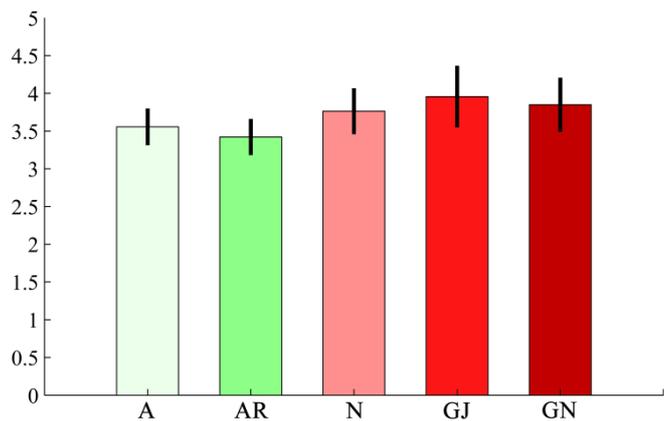
Figure 14. Perceived quality of compositions. Error bars indicate 95% confidence intervals

The small difference between groups *A* and *AR* suggests that the algorithmic composer does not depend on human selection of its output; it seems to compose at a stable level on its own.

The perceived rated quality for each individual composition is presented in Figure 15. Here the repeated occurrences for the compositions with a GJAS were included. The songs within each group are displayed next to each other. Notice that there is a great variation within each group. Evidently the variation within the groups is much greater than the variation between the groups. This suggests that a high perceived quality can be achieved with algorithmic composition.
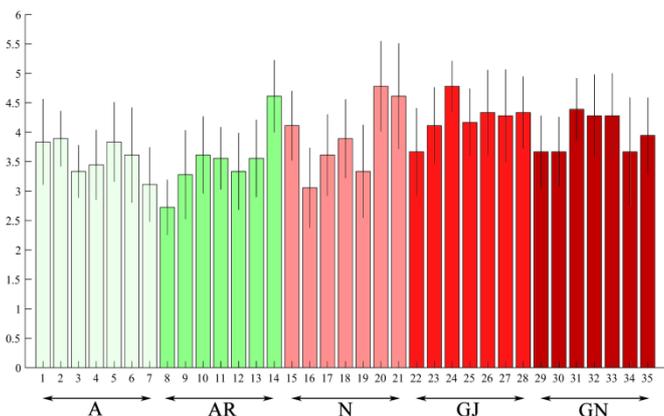

Figure 15. Perceived quality of each individual composition. Error bars indicate 95% confidence intervals

Two of the highest rated compositions (songs 20 and 21) were the only compositions that contained altered chords. These were two human written songs from group *N*. If these would have been removed from group N the mean score for that group would have been lowered from 3.76 to 3.53 making the score of group *N* lower than that of the algorithmically composed songs in group *A*.

The group of songs without a GJAS (*GN*) and the group with a GJAS (*GJ*) are further compared below. The repeated occurrences for them are included in this comparison. Presented in Figure 16 is the score for four parameters that the listeners assessed. The first (*"Good"*) is the rated quality as presented in Figures 14-15. The second is the perceived *"groove"* of the songs, where a higher score means more groove. For the third parameter (*"Human"*) the listeners were asked to assess if the

composer was a human. Here "Computer" was represented by 1 and "Human" was represented by 7. The fourth parameter is the listeners' sense of stress while listening to the song. Here a higher rating means that the listeners felt more stressed. The word "Calm" represented 1 and the word "Stressed" represented 7.
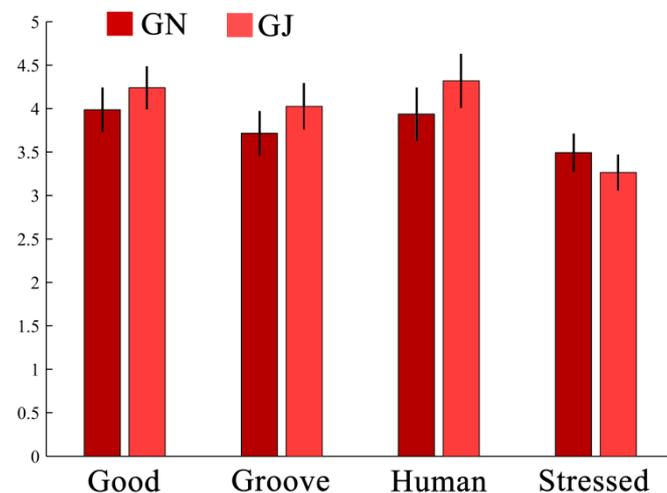
Figure 16. Perceived impression of the songs from group *GJ* and *GN* for four different parameters. Error bars indicate 95% confidence intervals

The compositions with a good GJAS were rated as having a higher quality and more groove. They were perceived to be more likely to be written by a human and they made the listeners calmer. The differences were however rather small and not likely to be statistically significant.

VI. CONCLUSION

The techniques developed in this study are built on a statistical foundation and the statistical approach seems to be feasible for modeling the generation of popular music. The results indicate that algorithmic composition of popular music can be valuable for composers in the field.

Listening tests indicate that melodies composed by the program are relatively close to human compositions concerning the perceived quality. A difference could not be statistically determined in the test as the groups have overlapping confidence intervals. Listeners seem to rate the more complex human written songs (altered chords) higher whereas the simpler human written songs receive a similar rating as computer written songs. This suggests that one aspect that needs further development is the harmony. As altered chords would be easy to integrate within the theoretical framework, it seems reasonable to add these in the future. The user can interact with the program to develop compositions of his or her style. With further development and added features this interaction can hopefully be expanded. It is feasible to let the user compose parts of the music and to let the program compose other parts. The perceived rated quality was however very similar between the randomly selected compositions (*AR* in Figure 14) and the compositions selected by the author (*A* in Figure 14). This indicates that the program performs at a similar level independently of if a human participates in the selection process or not.

A theoretical framework for Global Joint Accent Structure has been proposed and integrated into the program. Evaluation of the theory suggests that the compositions with a good GJAS were rated as having a higher quality and more groove. They were perceived to be more likely written by a human and they made the listeners calmer. The differences were however rather small and further studies are needed to evaluate the validity of the theory.

As the melodies have close resemblance to popular music some of the aspects in human composition has hopefully been captured and can be formally analyzed.

VII.   ACKNOWLEDGMENT

REFERENCES

Ahlbäck, S. (2004). *Melody Beyond Notes: A study of melody cognition.* Göteborgs Universitet: Humanistiska fakulteten.

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.

Blume, J. (2004). *6 Steps to Songwriting Success.* New York: Billboard Books.

Citron, S. (1985). Song Writing: A Complete Guide to the Craft. London: William Morrow & Company, Inc.

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.

Collins, N. (2007). *Musical robots and listening machines.* In N. Collins & J. d'Escriván (Eds.), The Cambridge Companion to Electronic Music (pp. 171-184). Cambridge: Cambridge University Press.

Cope, D. (2000). *The Algorithmic Composer.* Madison, Wisconsin: A-R Editions.

Dylan, B. (1962). Blowin' in the Wind. On *The Freewheelin' Bob Dylan* [Vinyl]. New York: Columbia Records.

Elowsson, A. (2011). *Algorithmic Composition of Popular Music.* Stockholm: Royal Institute of Technology.

Elowsson, A. (2012). *Statistical Analysis of Vocal Folk Music.* Stockholm: Royal Institute of Technology.

Farbood, M & Schoner, B. (2001). *Analysis and Synthesis of Palestrina-Style Counterpoint Using Markov Chains.* International Computer Music Conference. Havana.

Friberg, A., & Ahlbäck, S. (2009). *Recognition of the main melody in a polyphonic symbolic score using perceptual knowledge.* Journal of New Music Research, 38(2), 155-169.

Gabrielsson, A. (1993). The complexities of Rhythm. In W. J. Dowling & T. J. Tighe (Eds.), *Psychology and Music: The understanding of melody and rhythm.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Jones, M. (1982). Should I Stay or Should I Go [The Clash]. On *Combat Rock* [Vinyl]. New York: Epic Records.

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

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

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

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

Nierhaus, G. (2012, February 10). Lecture, The Royal College of Music in Stockholm, unpublished.

Parncutt, R. (1994a). *A perceptual model of pulse salience and metrical accent in musical rhythms.* Music Perception, 11(04), 409-464.

Parncutt, R. (1994b). *Template-matching models of musical pitch and rhythm perception.* Journal of New Music Research, 23, 145-167.

Phillips, J., & Phillips, M. (1965). California Dreamin' [The Mamas & the Papas]. On *If You Can Believe Your Eyes and Ears* [Vinyl]. US: Dunhill Records.

Rothgeb, J. (1993). Simulating Musical Skills by Digital Computer. In S. M. Schwanauer & D. A. Levitt (Eds.), *Machine Models of Music* (pp. 157-164). Cambridge, MA: MIT Press.

Rowe, R. (2001). Machine Musicianship. Cambridge, MA: MIT Press.

Schaffrath, H. (1995). *The Essen Folksong Collection in Kern Format*, [computer database] D. Huron (ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Simon, H. A., & Sumner, R. K. (1993). Patterns in Music. In S. M. Schwanauer & D. A. Levitt (Eds.), *Machine Models of Music* (pp. 157-164). Cambridge, MA: MIT Press.

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

Tanaka, T. & Nishimoto, T. & Ono, N. & Sagayama, S. (2010). *Automatic music composition based on counterpoint and imitation using stochastic models.* Tokyo: University of Tokyo, Graduate School of Information Science and Technology.

Webb, J. (1998). *Tunesmith: Inside the art of songwriting.* New York: Hyperion.

West, R., Howell, P., & Cross I. (1985). Modeling perceived musical structure, In P. Howell, I. Cross & R. West (Eds.), *Musical structure and cognition,* (pp. 21-.52). London: Academic Press.