Investigation on Meter in Generative Modeling of Music

Jensen, Kristoffer

Published in:
Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Investigation on Meter in Generative Modeling of Music

Kristoffer Jensen[1]

[1] Aalborg University Esbjerg, Niels Bohr Vej 8,
6700 Esbjerg, Denmark
krist@aaue.dk

Abstract. Generative models of music rhythm are in severe need of performance additions, i.e. inclusions of subtle temporal and dynamic alterations so as to render the music musical. While much of the research is based on music theory, the work presented here is based on the temporal perception, which is divided into three parts, the immediate (subchunk), the short-term memory (chunk), and the superchunk. By review of the relevant temporal perception literature, and analysis of performances of metrical music, the necessary performance elements to add in the metrical generative model, related to the chunk memory, are obtained.

Keywords: Rhythm; human cognition; perception; chunking; music generation.

1 Introduction

Automatic, adaptive music generation has more and more uses in today's media. Be it in computer games, interactive music performances, or in interactive films, the emotional effect of the music is primordial in the appreciation of the media. While traditionally, the music has been generated in pre-recorded loops that is mixed on-the-fly, or recorded in traditional orchestras, the better understanding and models of generative music is believed to push the interactive generative music into the multimedia. Papadopoulos and Wiggins (1999) gave an early overview of the methods of algorithmic composition, deploring "that the music that they produce is meaningless: the computers do not have feelings, moods or intentions". While vast progress has been made in the decade since this statement, there is still room for improvement.

The cognitive understanding of musical time perception is the basis of the work presented here. According to Kühl (2007), this memory can be separated into three time-scales, the short, microtemporal, related to microstructure, the mesotemporal, related to gesture, and the macrotemporal, related to form. These time-scales are named (Kühl and Jensen 2008) subchunk, chunk and superchunk, and subchunks extend from 30 ms to 300 ms; the conscious mesolevel of chunks from 300 ms to 3 sec; and the reflective macrolevel of superchunks from 3 sec to roughly 30–40 sec. The superchunk was analyzed and used for in a generative model in Kühl and Jensen (2008), and the chunks were analyzed in Jensen and Kühl (2009). Further analysis of

ISBN: 978-84-614-0853-5
7th International Symposium on Computer Music
Modeling and Retrieval
Málaga (Spain), June 21-24, 2010

the implications of temporal perception is related to analysis of durations and timing of existing music, and anatomic and perceptual finding from the literature, and this is given in the next section. Section 3 presents current work on the analysis of differences between accented and unaccented notes in 7/8 music, and the section 4 discusses the integration of the metrical rhythm in the superchunk generative music model. Finally, section 5 offers a conclusion.

2 Cognitive and Perceptual aspects of rhythm

According to Snyder (2000), a beat is single point in time, while the pulse is recurring beats. Accent gives salience to beat, and meter is the organization of beats into a cyclical structure. This may or may not be different to the rhythmic grouping, which is generally seen as a phrase bounded by accented notes. Lerdahl & Jackendorff (1983) gives many examples of grouping and meter, and show how this is two independent elements; Grouping – segmentation on different levels is concerned with elements that has duration, and Meter – regular alternation of strong and weak beats is concerned with durationless elements. While grouping and meter are independent, the percept is more stable when they are congruent.
The accentuation of some of the beats gives perceptually salience to the beat (Patel & Peretz 1997). This accenting can be done (Handel 1989) by for instance an intensity rise, by increasing the duration or the interval between the beats, or by increasing the frequency difference between the notes.

Samson *et al* (2000) shows that the left temporal lobe processes rapid auditory sequences, while there are also activities in front lobe. The specialized skills related to rhythm are developed in the early years, for instance Malbrán (2000) show how 8-year-old children can perform precise tapping. However, while the tapping is more precise for high tempo, drifting is ubiquitous. Gordon (1987) has determined that the perceptual attack time (PAT) is most often located at the point of the largest rise of the amplitude of the sound. However, in the experiment, the subjects had problems synchronizing many of the sounds, and Gordon concludes that the PAT is more vague for non-percussive sounds, and spectral cues may also interfere in the determination of the attack. Zwicker and Fastl (1999) introduced the notion of subjective duration, and showed that the subjective duration is longer than the objective durations for durations below 100ms. Even more subjective deviations are found, if pauses are compared to tones or noises. Zwicker and Fastl found that long sounds (above 1 second) has the same subjective durations than pauses, while shorter pauses has significantly longer subjective durations than sounds. Approximately 4 times longer for 3.2kHz tone, while 200Hz tone and white noise have approximately half the subjective duration, as compared to pauses. This is true for durations of around 100-200 ms, while the difference evens out to disappear at 1sec durations. Finally Zwicker and Fastl related the subjective duration to temporal masking, and give indications that musicians would play tones shorter than indicated in order to fulfill the subjective durations of the notated music. Fraisse (1982) give an overview of his important research in rhythm perception. He states the range in which synchronization is possible to be between 200 to 1800 msec (33-300 BPM). Fraisse furthermore has analyzed classical music, and found two main durations that he calls temps longs (>400msec) & temps courts, and two to one ratios only found between temps longs and courts. As for natural tempo, when subjects are asked to reproduce temporal intervals, they tend to overestimate short intervals (making them longer) and under-estimate long intervals (making them shorter). At an interval of about 500 msec to 600 msec, there is little over- or under-estimation. However, there are large differences across individuals, the spontaneous tempo is found to be between 1.1 to 5 taps per second, with 1.7 taps per second most occurring. There are also many spontaneous motor movements that occur at the rate of approximately 2/sec, such as walking, sucking in the newborn, and rocking.

Friberg (1991), and Widmer (2002) give rules to how the dynamics and timing should be changed according to the musical position of the notes. Dynamic changes include 6 db increase (doubling), and up to 100msec deviations to the duration, depending on the musical position of the notes. With these timing changes, Snyder (2000) indicate the categorical perception of beats, measures and patterns. The perception of deviations of the timing is examples of within-category distinctions. Even with large deviations from the nominal score, the notes are recognized as falling on the beats.

3 Analysis of meter performance

Music is typically composed, giving a intended and inherent emotions in the structural aspects, which is then enhanced and altered by the performers, who change the dynamics, articulations, vibrato, and timing to render the music enjoyable and musical. Many of these performance enhancements are done within the meter structure. An important aspect of the meter is the accent. This is mainly consisting of a relative loudness and brightness increase, together with an elongation and possible time-shifting of the note. Early work in music performance analysis, include Bengtsson and Gabrielsson (1983) who found evidence of repeated notes are performed in short-long alternations. Friberg (1991) together with collaborators studied performance in an analysis-by-synthesis approach using symbolic music score as the basis. Widmer (2002) uses machine learning approaches to find similar rules.
While many studies has investigated the deviations in time occurring in music performance, it is more difficult to obtain tangible data from these previous studies.

In order to obtain some empiric data about the accentuation done by musicians, an experiment regarding recording of simple music sequences has been done. The goal of this study is to gain understanding of loudness/brightness, timing and timbre issues that can occur when musicians are playing strong and weak beats. Two instruments, bass clarinet, and electric guitar with semi-professional players, has been investigated. In all, 581 guitar loops have been recorded, and 701 bass clarinet loops. The melody was generally kept very simple, but different octave note alterations are made for the guitar, and other fingering solutions made for the bass clarinet. The main metrical accentuation was 3/2/2, but other accentuation patterns were recorded as well, including some unusual and thus more difficult to play ones. The onsets and offsets are found by calculating perceptual spectral flux (**psf**), and summing the positive **psf**, which becomes the onset feature, and summing the negative **psf**, which becomes the offset feature. The **psf** is calculated by summing the differences between two neighbouring spectrums that has been normalized by the equal loudness contour for average loudness. It was originally presented in Jensen (2005) and favorably compared to other onset detection methods by Collins (2005). No formal evaluation of the **psf** with positive or negative filtering has been performed. An example of the onset and offset features is shown in figure 1, together with the estimated loudness, according to the DIN45631 standard as defined by Zwicker and Fastl (1999). The brightness is assumed to be strongly correlated with the loudness, as shown in Jensen (2002). The features have been normalized. The peaks in the features are denoted with plus signs, and the estimated onset and offset points are denoted with 'o'. The notes denoted with a left-pointing triangle are the accented notes (1, 4 and 6). These notes have a significantly higher loudness. They also seem to start later than the unaccented notes, but this doesn't seem to correspond to the fact that longer notes sound stronger. It could be an effect of preparation for the louder note.

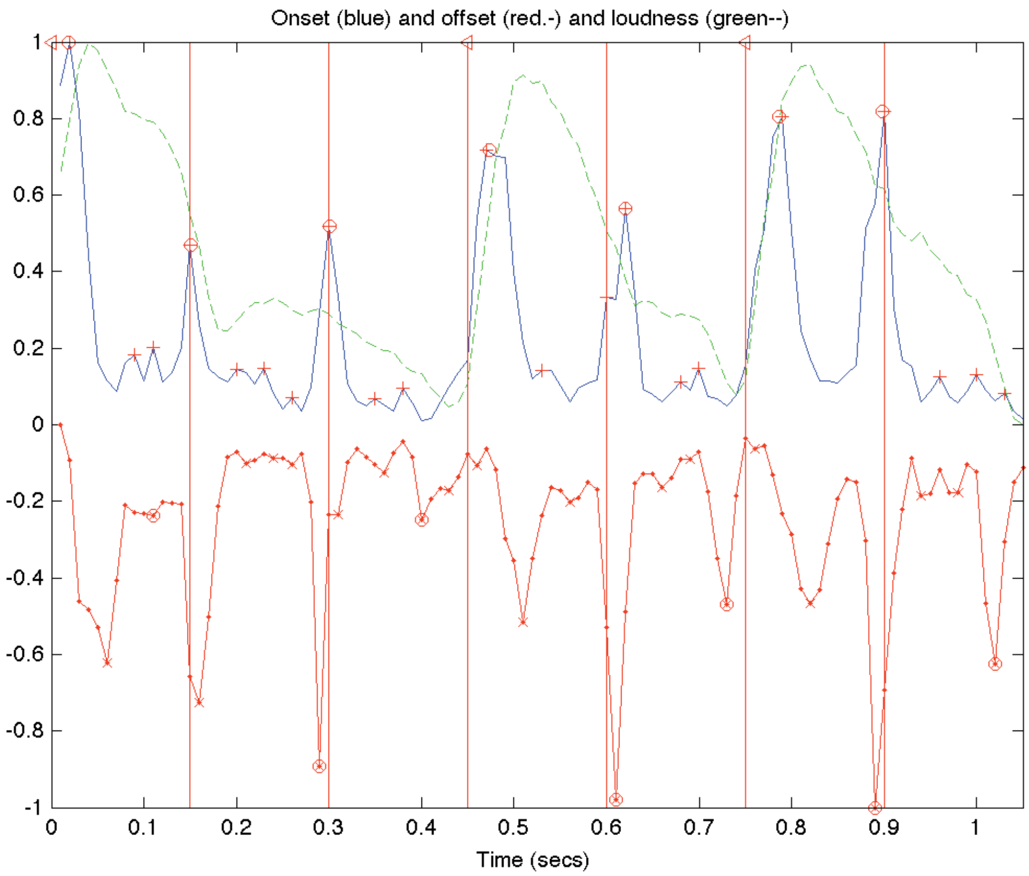

Figure 1. Example of note onset and offset features for a *piano* 7/8 note bass clarinet melody. Onset is blue (solid), offset is red (dash-dotted) and loudness is green (stipled). Accented notes are marked with a right triangle. The notes played are D, B*b*, G, D, B*b*, D, B*b*, chosen as to be easy to perform.

An example of the same features for a similar melody for the electric guitar is shown in figure 2. The guitar notes is performed very close to the nominal onset values. Perhaps the stronger notes a shorter in the case of the guitar too, but again, whether this is due to deliberate musical effect, or it is a bi-product of preparation of the next note is not possible to ascertain. The onset and offset points are also polluted by mechanical noises in the case of the bass clarinet, and pickup noises in the case of the guitar. Certainly, the main effect found for the accented note is a higher loudness.

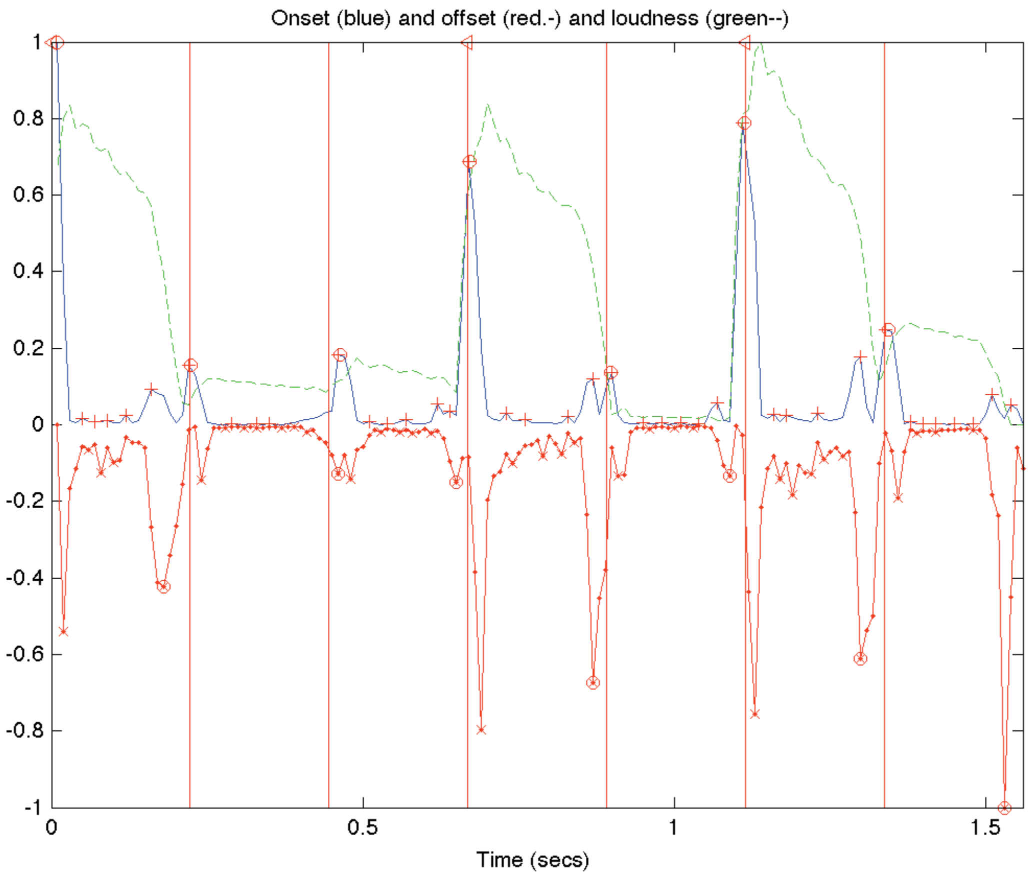

Figure 2. Example of note onset and offset features for a *mezzo-forte* 7/8 note guitar melody. Onset is blue (solid), offset is red (dash-dotted) and loudness is green (stipled). Accented notes are marked with a left triangle.

According to Widmer (2002), only slightly more than 11% of the accented notes have a systematic loudness difference. In the musical case, this is related to the congruency of the grouping and metrical structure (Lerdahl and Jackendorff 1983).

A preliminary statistical analysis of the correlations between the deviations of the onset and offset, the duration of the notes, and the loudness of the notes, as compared to the nominal values, has been undertaken here. The duration is calculated as the onset to offset duration. Only the metrical deviations are investigated here. The statistical method used is the Pearson correlation coefficient. To showcase the issues of onset detection, the largest effects for the guitar are found for offset and next note onset, $r=.82$, $p<.0001$ for several note positions. Similarly, the relationship between same note onset and duration is large, $r=.9$, $p<.0001$. These are effects that have to do with estimation accuracy, mainly, or necessary correlation between onset/offset and duration of notes.

Among the possible relationships, two hypothesis has been verified here, one that a strong accent has a significant effect on some of the features of the same or neighboring notes, and two that the same loop played with and without accents for the same note, and otherwise is the same will show an effect on some of the features. As for hypothesis 1, in the guitar on the fourth position on a 7/8 played 3-2-2 vs. 3-4, i.e. the fourth note is accented or not, a strong relationship is found between accent and duration, r=.5, p<0.01. Furthermore, a medium effect is found between accent and onset, r=.3, p<.05. This signifies, for the guitar (and guitarist) that the accented notes are played longer, and started earlier. No such effect is found for the bass-clarinet, perhaps because it is so expressive in the timbre that the elongation of accented notes is not necessary. For the third note in the same 7/8 played 3-2-2 or 3-1-1-2, i.e. the third note is either accented or not, a similar result is obtained, however only with a medium effect, r=.3, p<.001. In this case, the relationship was found between the accent and offset, r=.2, p<.001. In all cases the relationship between accent, loudness and brightness was large and significant, r=.8, p<.001, as could be expected (Jensen 2002).

4 Recreating meter in form

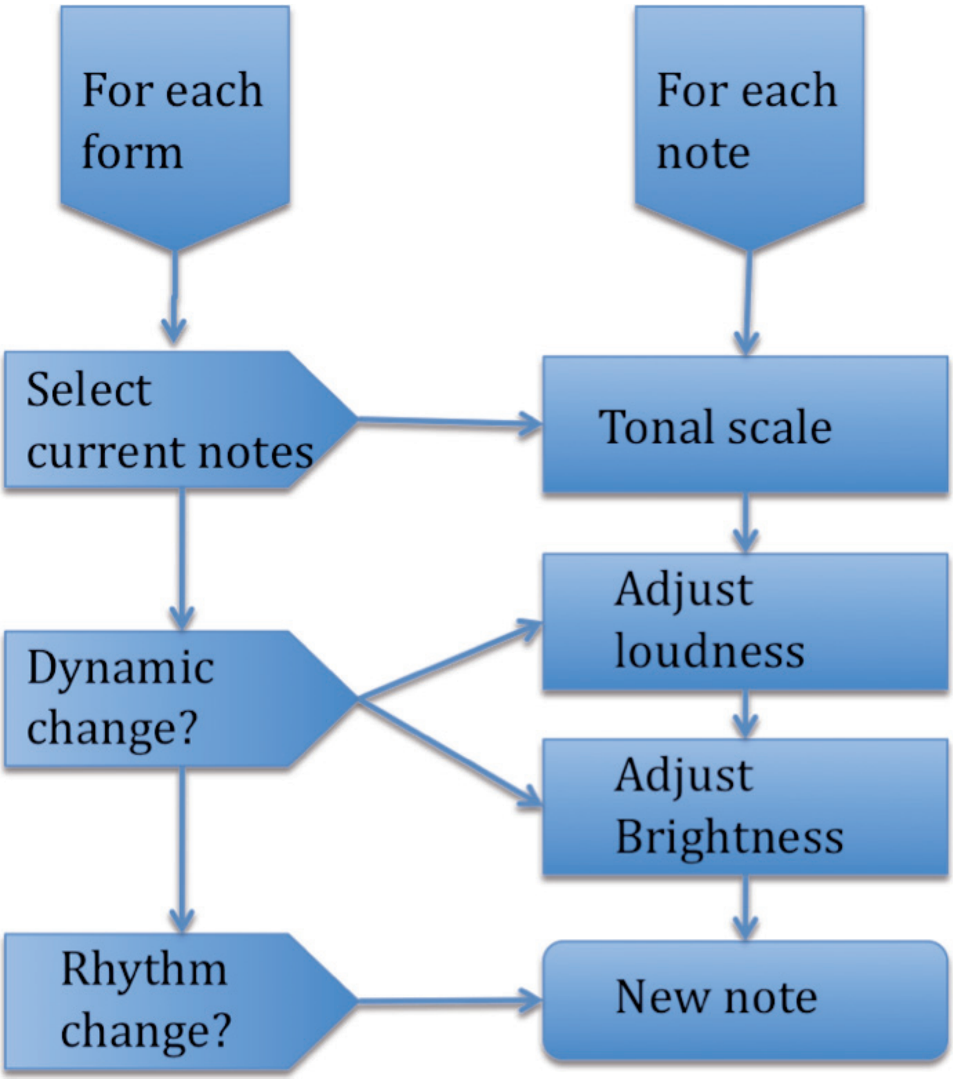

Figure 3. The Form model. Structural changes on the note values, the intensity and the rhythm is made every 30 seconds, approximately.

In previous work (Kühl & Jensen 2008), a generative model that produces tonal music with structures changes was presented. This model, that creates note values based on a statistical model, also introduces changes at the structural level (each 30 seconds, approximately). These changes are introduced, based on analysis of music using the musigram visualization tools (Kühl & Jensen 2008).

With respect to chroma, an observation was made that only a subset of the full scale notes were used at each structural element. This subset was modified, by removing and inserting new notes at each structural boundary. The timbre changes include varying the loudness and brightness between loud/bright and soft/dark structural elements. The main rhythm changes were based on the identification of short elements (10 seconds) with no discernable rhythm. A tempo drift of up-to 10% and insertion of faster rhythmic elements (Tatum) at structural boundaries were also identified. These structural changes were implemented in a generative model, which flowchart can be seen in figure 3. While the structural elements certainly were beneficial for the long-term interest of the music, the lack of short-term changes (chunks) and a rhythm model impeded on the quality of the music.

The meter is included in this generative model by adjusting the loudness and brightness of each tone according to its accent. This is illustrated in figure 4.

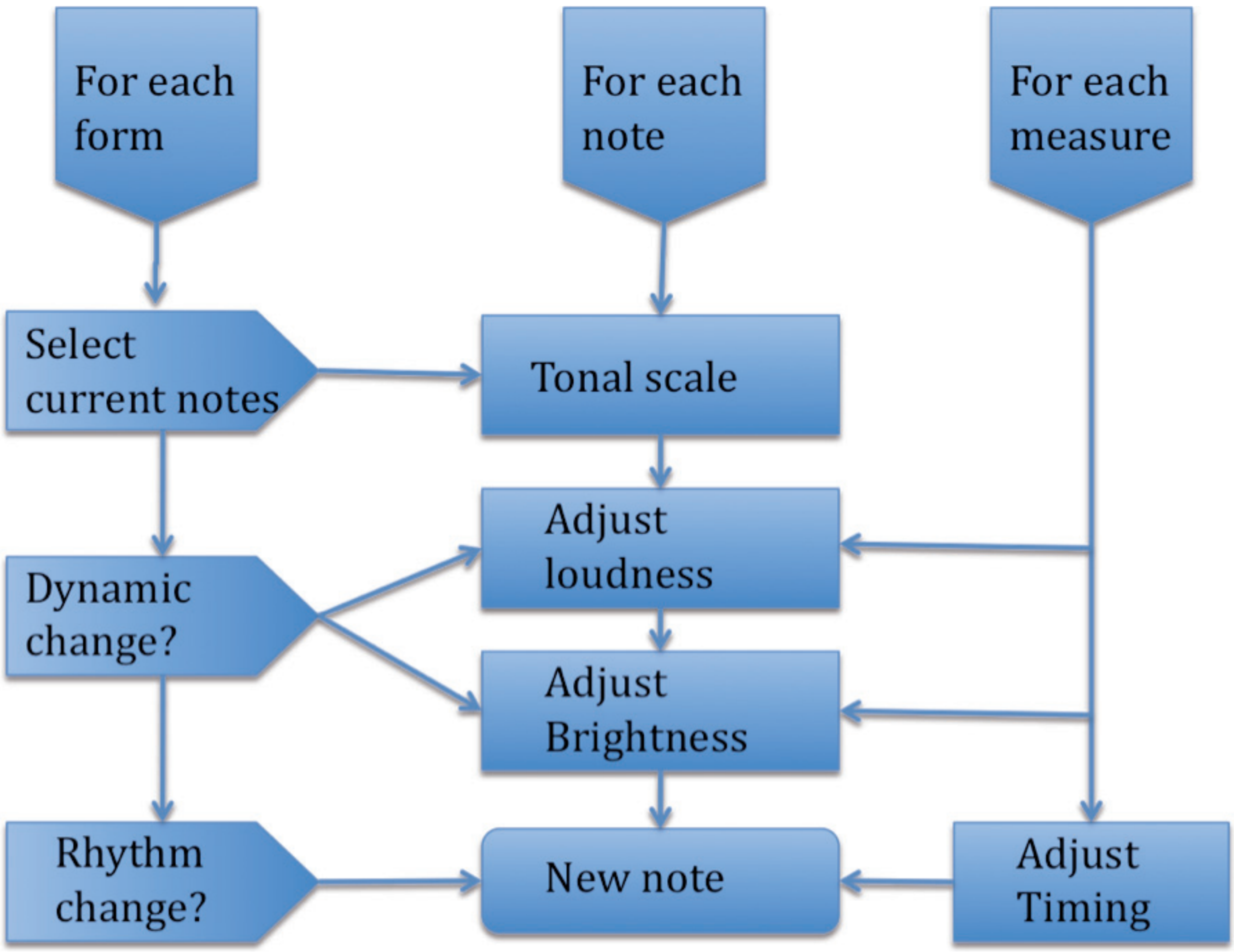

Figure 4. The meter and form model. In addition to the form on the note values, the each note dynamics and timing is adjusted according to the accent of the note.

The notes are created using the envelope model and the synthesis method dubbed brightness creation function (**bcf**, Jensen 1999) that creates a sound with exponentially decreasing amplitudes. The envelope model is not adjustable, but the **bcf** allows the continuous control of the brightness. The accent affects the note, so that the loudness brightness is doubled, and the duration is increased by 25 %, with 75% of the elongation made by advancing the start of the note. The inclusion of rhythmical meter in the form generative model certainly renders a more musical expression, but as there is not much interaction between the form and the meter it is difficult to evaluate the results further.

These findings are put into a generative model of tonal music. A subset of notes (3-5) is chosen at each new form (superchunk), together with a new dynamic level. At the measure (chunk) level, new notes are created in a metrical loop, and at the microtemporal (subchunk) level, expressive deviations are added in order to render the loops musical.

5 Conclusion

The automatic generation of music is in need of model to render the music expressive. This model is found using knowledge from time perception of music studies, and further studies of the cognitive and perceptual aspects of rhythm. Indeed, the generative model consists of three sources, corresponding to the immediate microtemporal, the present mesotemporal and the long-term memory macroterminal. While a single stream in each of the source may not be sufficient, so far the model incorporates the macrotemporal superchunk, the metrical mesotemporal chunk and the microtemporal expressive enhancements.

The normal beat as is given by different researchers to be approximately 100 BPM, and Fraisse (1982) furthermore shows the existence of two main note durations, one above and one below 0.4 secs, with a ratio of two. Indications as to subjective time, given by Zwicker and Fastl (1999) are yet to be investigated, but this may well be creating uneven temporal intervals in conflict with the pulse.

An experiment with recorded 7/8 loops with different accent patterns reveals that accents and loudness and brightness a strongly related, and accents and durations are also strongly related, the durations increase being seen more as an earlier start of note than an later end of note.

The inclusion of the metrical and expressive models certainly, in the author opinion, renders the music more enjoyable, but more work remains before the generative model is ripe for general-purpose uses.

References

1. Bengtsson, I., and Gabrielsson, A. (1983). *Analysis and synthesis of musical rhythm. In* J. Sundberg (ed.), Studies of music performance. pp. 27-60).

2. Collins, N. (2005). *A Comparison of Sound Onset Detection Algorithms with Emphasis on Psychoacoustically Motivated Detection Functions*. Proceedings of AES 118th Convention, Barcelona, Spain, 2005 May 28–31.
3. Fraisse, P. (1982). *Rhythm and Tempo. In* D. Deutsch (ed.) The Psychology of Music, first edition. New York: Academic Press, pp. 149-180,
4. Friberg, A. (1991) *Performance Rules for Computer-Controlled Contemporary Keyboard Music*. Computer Music Journal 15(2). pp 49-55
5. Gordon, J. W. (1987) *The perceptual attack time of musical tones*, Journal of the Acoustical Society of America, pp 88-105
6. Handel, S. (1989) *Listening*. Mit Press
7. Jensen, K. (1999) *Timbre Models of Musical Sounds*, PhD Dissertation, DIKU Report 99/7.
8. Jensen, K. (2002) *Musical Instruments Parametric Evolution*, Proceedings of the ISMA, Mexico City, Mexico, np (CD).
9. Jensen, K. (2005) *A Causal Rhythm Grouping*. Lecture Notes in Computer Science, Volume 3310. ISBN 3-540-24458-1, pp. 83-95.
10. Jensen, K., O. Kühl, (2009) *Towards a model of musical chunks,* Lectures Notes in Computer Science, Springer-Verlag, LNCS5493 pp. 81-92.
11. Kühl, O., K. Jensen (2008) *Retrieving and recreating Musical Form,* Lectures Notes in Computer Science, Springer-Verlag, LNCS 4969. pp. 270-282.
12. Kühl, O., *Musical Semantics*, Bern: Peter Lang (2007).
13. Lerdahl, F. & Jackendoff, R. (1983) *A Generative Theory of Tonal Music*. Cambridge, Mass.: The MIT Press.
14. Malbrán S. (2000) *Phases in Children's Rhythmic Development. In* R. Zatorre and I. Peretz (eds.) The biological foundations of music. Annals of the New York Academy of Sciences.
15. Papadopoulos G., and G. Wiggins (1999) *AI methods for algorithmic composition: a survey, a critical view and future prospects*. AISB Symposium on Musical Creativity, pp 110-117
16. Patel, A. & I. Peretz (1997) *Is music autonomous from language? A neuropsychological appraisal. In* I. Deliège & J. Sloboda (eds.) Perception and cognition of music. Hove: Psychology Press, pp. 191-215
17. Samson S., N. Ehrlé, M. Baulac (2000). *Cerebral Substrates for Musical Temporal Processes. In* R. Zatorre and I. Peretz (eds.) The biological foundations of music. Annals of the New York Academy of Sciences.
18. Snyder, B. (2000) *Music and Memory. An Introduction*. Cambridge, Mass.: The MIT Press.
19. Widmer, G. (2002). *Machine discoveries: A few simple, robust local expression principles*. Journal of New Music Research, 31, 37-50.
20. Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: facts and models*. Springer series in information sciences. Berlin, 2nd updated edition.