JIG: Jazz Improvisation Generator

Maarten Grachten

Artificial Intelligence Research Institute, IIIA
Spanish Council for Scientific Research, CSIC
Campus UAB, 089193 Bellaterra, Catalonia, Spain
`maarten@iiia.csic.es`

Abstract

This article presents JIG, a system for improvisation on jazz ballads. It was described previously in a Master's Thesis [4]. JIG generates so called 'formulaic' improvisations. It uses constraints, in combination with probability-driven randomness to generate note-attributes. In this way, numerous different improvisations can be generated on a single song. JIG has been incorporated in SaxEx, a case based reasoning system for generating expressive performances of jazz ballads.

1 Introduction

The jazz improvisation system JIG generates monophonic improvisations on jazz ballads, using the theme of the song and the accompanying chords to construct a new melody. It was designed as a modular addition to SaxEx, a case based reasoning system for generating expressive performances of jazz ballads, designed at IIIA [1], [2]. SaxEx changes the expressive note attributes, like dynamics and articulation. Therefore it suffices for JIG to generate only the pitch and duration attributes for the notes.

The improvisations are supposed to satisfy three major constraints:

tonality the improvisation must be tonal to the key of the music, and thus be predominantly consonant;

continuity the melodic contour of the improvisation must be mostly smooth; large intervals are sparingly used and registral direction is not too frequently reversed;

structure the improvisation should not be merely a sequence of non-related notes; in some way, interrelated groups of notes should be identifiable.

The first two constraints are relatively easy to satisfy, using the local context of the notes to be generated. The structure constraint requires some more consideration. The structure of an improvisation can take many forms. A cognitive model for improvisation, by Pressing [7], supposes that improvisations can be segmented into adjacent, non-overlapping parts, called 'event clusters'. According to the model, an event cluster is generated either in association, or in contrast to the previous cluster. In this way, the improvisation will clearly satisfy the structure requirement. The model however does not address the question how different types of improvisation can be achieved. Typically, a different kind of improvisation involves different kind of structure. In [8], one kind of jazz improvisation that is recognized, is *formulaic* improvisation. It consists in using *formulae* (small melodic fragments, also called *motifs*) as the building blocks for the improvisation. The formulae are drawn from a personal repertoire, or from the original melody of the song. They need not occur literally, but can be transformed to fit in the current context.

A similar way of improvising is described by David Sudnow, in his autobiographic book [10]. While mastering jazz improvisation on piano, he learned to apply melodic formulae on appropriate places in the chord scheme. As he advanced, he gained control over the formulae (i.e. playing them in the way he wanted) and he learned to play bridging note sequences in between the formulae. Sloboda calls this 'relaxed melodying' [9].

The improvisations that are generated by JIG were aimed to have a structure like those of Sudnow. This implies that there are two alternate kinds of phases within the improvisation: *motif* phase, where the formulae (or motifs) are played and the *melodying* phase, where notes are generated as it were 'out of the blue'. These different phases suggest the use of two different note-generation processes, that are alternately active. JIG realizes these two processes, in order two generate note pitches and durations. However, neither of the processes generate the eventual pitches, only abstract ‘pitch types’ (see section 3.2) instead. From these pitch types, in conjuction with additional contextual information, an actual pitch is derived. This last step is executed equally for both processes.

The global form of the improvisation process, including the regulation of the melodying and motif processes, will be addressed in section 2. After that, the melodying and motif processes will be explained (in section 3 and 4, respectively).

2 Global form of the improvisation process

A first thing that should be noted, is the interdependence of pitch and duration and their depencies on the musical context. The dependence of a pitch (or duration) on context means that it will only sound well under certain conditions. Undoubtedly, these conditions are more strict for pitch than for duration: a duration that is unnatural in its context, is usually not as stunning and repellent as a pitch that is unnatural in its context. For this reason, the focus has been on pitch generation.

To get a grip on the dependencies of pitch on its context, the concept of *pitch tolerance* is introduced. It refers to the perceptual [1] quality of pitches, namely their degree of sounding either satisfying or repulsive. The pitch tolerance of a pitch is primarily derived from the harmonic relation between the pitch and its underlying chord. The more consonant a pitch is relative to the chord, the greater is the pitch tolerance. But within a musical context, other factors can improve or aggrevate the pitch tolerance of a pitch. For example, the pitch tolerance of a dissonant pitch would be decreased if it gained accent. In this way, three prominent factors seem to be involved in the determination of pitch tolerance:

1. The duration of the note
2. The metrical position of the note
3. The (forward) melodic context of the note

Duration and metrical position affect the pitch tolerance in a straight-forward manner. Notes with longer durations will attract more attention, and therefore, if the pitch of that note has a low pitch tolerance, it will be aggrevated. Metrical position has a similar effect: if a note is on a metrical position with strong accent (usually the first beat of the measure), a low pitch tolerance will be decreased even more. The influence of melodic context (i.e. the pitches of proximate notes) is more complex. In general, the dissonance of pitch seems to be more prominent if the pitches of surrounding notes are very different (in terms of melodic intervals). More specifically, the pitch tolerance of a dissonant pitch can be greatly improved if the pitch is part of a chromatic sequence that resolves into consonance.

Based on these assumptions, it makes sense to divide pitches into classes with different dissonance. Obvious categories are *chord pitch* which contains all pitches that are contained in the chord; *scale pitch* which contains all pitches that fit to a particular scale (usually a scale that is harmonic to the chord) and are not contained in *chord pitch*; lastly, *chromatic pitch*, which contains all pitches that are neither contained in *chord pitch*, nor in *scale pitch*. These categories represent the *pitch types*, mentioned before. Ideally, further distinctions of relative harmony should be made (especially for *chord pitch*). However this has not been realized in JIG, so far.

To be able to identify which actual pitches are contained in the *scale pitch* set in a given situation, a scale must be defined as the current scale on that moment. To this end, JIG precomputes for each chord in the song an appropriate scale to be played. It does this, by selecting the scale that is harmonic to most chords in the song. Whenever a chord is encountered to which the selected scale is disharmonic, JIG selects another scale, that is maximally similar to the old scale (in terms of common pitches) and is harmonic to the chord.

The melodying and motif processes both determine, each in their own way, the duration for the next note to be played and its pitch type. Both processes are alternately active. So far, the procedure for alternating the processes that has been implemented, is a very simple one. In the current configuration, whenever a note must be generated, a probabilistic selection of either the melodying or the motif process is made. This selection can be controlled by setting a probability parameter for each process, specifying the probability of continuing the current process versus changing to the alternative process. By setting the reflexive probabilities to relatively high values, rapid changes can be avoided, thus preventing improvisations containing a lot of fragmented motifs.

However, one can easily imagine more sophisticated decision methods. A decision method that would resemble the way the improvisations are generated according to Sudnow [10], would be to let the motif process be triggered by the context, particularly the upcoming chord movement. When the motif that was triggered has been finished, the melodying process takes over again. This however, requires a mapping of motifs to appropriate contexts, which is currently not available.

[1] Perceptual qualities of pitches are subjective by nature; nevertheless, it seems reasonable to suppose that the quality of pitch tolerance is fairly persistent within one musical culture.

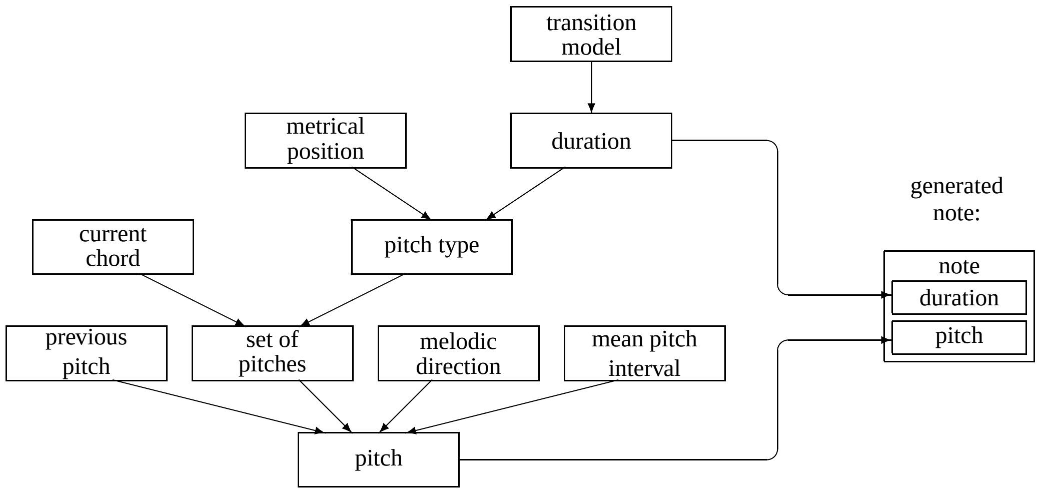

Figure 1: Block diagram of note generation during melodying process

3 The melodying process

The approach of the melodying process can be called a 'temporal' aproach, that is, the constituents of the improvisation are generated in chronological order and have chronological dependencies (as opposed to e.g. a top-down approach, where the constituents have hierarchical dependencies). This approach was derived from Pressing's model. However, because the structure requirement is met by using the alternative motif process, the model was simplified to generating one note at a time, instead of a cluster of notes [2].

In the melodying process, the duration is generated before the pitch, because of the relatively restricted choice of pitch. Once a duration has been found, the context is specific enough to generate a pitch-type, and from that, an actual pitch is derived. A block diagram of the generation process is shown in figure 1.

3.1 Generation of durations

Durations are generated by a probabilistic transition model. This provides a very flexible way of imposing a certain degree of regularity in the rhythm of the improvised melody, while allowing for rhythmic variation within and across improvisations. For each note that is generated, the transition model is queried, and returns the duration value that corresponds to its current state. It then makes a transition to one of the next states, where each transition has a preset probability of being selected. Currently, a fairly simple transition model (but with great generative power) is used. It has transitions going from every duration to every other duration, including reflexive transitions. The probabilities of arriving at half, quarter and quaver duration are greatest. Also, the reflexive probabilities for the quarter and quaver durations are relatively great, whereas reflexive probabilities for long durations are small. This configuration usually results in quite regular rhythms, with sequences of quarter or quaver durations, at times alternated with a half or whole note.

By adding an extra attribute to the transition model states, that discerns durations for notes from durations for rests, improvisations with rests can also be generated. When a state of the rest type is returned, the transition model is queried again, until a state of the note type is retrieved. Every time a rest state is encountered, the only action is the addition of corresponding rest duration to the variable that stores the position of the next note.

An extensive study on the use of transition models in the context of jazz improvisation (for rhythm as well as harmony and melody) is found in [5].

3.2 Generation of pitch types

When the position and duration of a note have been established, an estimation can be made of how much dissonance the current context allows (for instance, if the note is a whole note that starts at the beginning of a measure, it is obvious that the pitch should be quite consonant). Depending on this information, an appropriate probability distribution is assigned to the different pitch types. Based on this distribution, a random selection is made. If the two conditions, of duration, and position, act as a fuzzy *conjunction* (that is, the most restrictive condition is prevalent), a more consonant improvisation will result.

[2]From the model point of view, this is admittedly a rather devastating modification. However, the aim was not to implement the cognitive model, but to build a convincing improvisation mechanism.

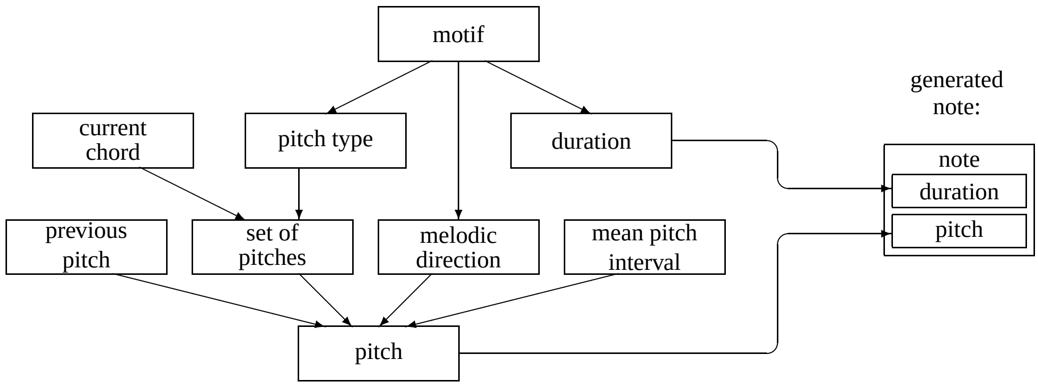

Figure 2: Block diagram of note generation during motif process

For example in the case of a very short note on the first beat of a measure, a chord pitch is likely to be chosen, because the position is very restrictive. If on the other hand, the conditions act as a fuzzy *disjunction*, the improvisations will generally be less consonant, occasionally having dissonant notes with long durations or on metrically stressed positions. Another possible configuration is to bias one of the two conditions.

Because the incorporation of chromatic pitches in improvisations requires a more extensive interaction of the note generation with the generation of surrounding notes then currently is possible, the chromatic pitches are for now excluded from selection.

3.3 Inferring the actual pitch

To derive the actual pitch from the previously selected pitch type, three factors are taken into account, as can be seen in figure 1. Firstly, the pitch of the previous is retrieved. Secondly, a variable is used for storing the desired melodic (or registral) direction for the next chord. Depending on its value (*up* or *down*), the next pitch will be either higher or lower than the previous pitch. Lastly, the mean interval of previous pitches is calculated (possibly a weighted mean, discarding less recent intervals).

These three factors together form the information on which a probability distribution is assigned to the possible pitches. The distribution is such that probabilities are highest for the pitches nearest to the previous pitch, and drop as the distance towards the previous pitch increases (either above or below the previous pitch, depending on the direction that was selected). The distribution takes the form of a half Gaussian 'bell' curve. The steepness of the curve depends on the mean interval of previous pitches. If the mean interval is large, the curve will be steeper, forcing the selection of a pitch that lies near the previous pitch. If the mean interval is small, bigger intervals are allowed, so the curve will be less steep. An example is given in figure 3.

The choice of melodic direction is also based on a probability distribution. The chances of changing melodic direction increase as a function of the previous interval. Larger intervals increase the probability of directional change. This is in accordance with Narmour's principles of registral implication [6]. An alternative to taking only the previous interval into account, would be to add up all intervals since the last change of direction. This prevents long decreasing or increasing sequences of small interval steps.

4 The motif process

The motif process provides an alternative way of generating durations and pitch types. This process uses preexistent melodic fragments. One possibility is to store a repertory of motifs that are used for every improvisation. This would be in accordance with the assertion in [8], that musicians often have a personal repertoire of 'licks' which they use in their solo improvisations.

Another possibility, that is encountered in human jazz improvisations as well, is to draw motifs from the from the original theme of the song. This is the approach that was chosen in JIG, although the two possibilities are certainly not mutually exclusive.

The extraction of motifs from the original melody requires a way to identify groups of notes as motifs. Conveniently, SaxEx already maintains an elaborate representation of the songs, including a Narmour analysis of the melody. The Narmour analysis comprises the structure of the melody, in terms of basic structures (also called 'Narmour structures'). Such structures normally consist of three notes, but the structures can appear in forms containing fewer or more notes. The Narmour analysis of a melody specifies for each note of the melody, to which basic structure it belongs. In this way, it is possible to retrieve the structure to which an arbitrary note belongs, thus retrieving a group of notes that form a

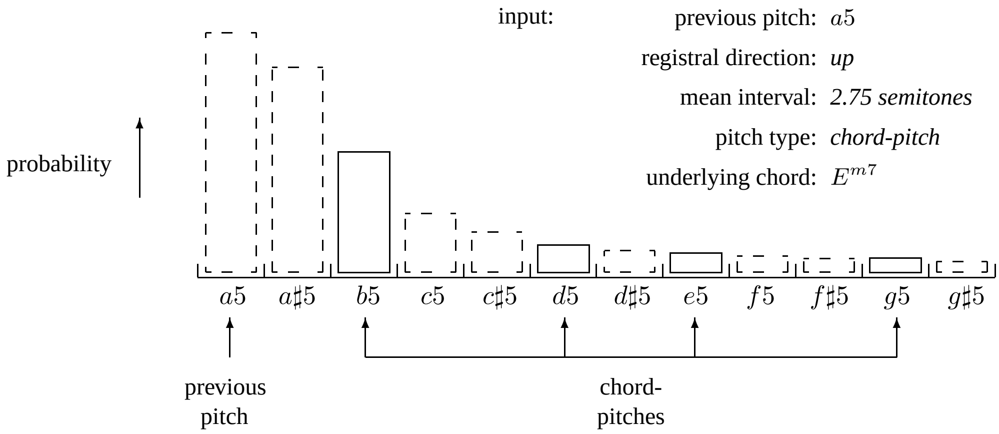

Figure 3: Probability distribution for pitch-selection

perceptual whole. It is difficult to say whether Narmour structures always co-incide with motifs, because there is no clear-cut description of what precisely is a motif, beyond a group of notes that form a whole. However, the boundaries of a motif are usually perceived by certain 'closural' phenomena, such as a rest, a long note or a harmonic resolution. The Narmour structures bear 'closural' or 'non-closural' as an attribute, so if a good Narmour analysis is available, selecting Narmour structures should be a good way of retrieving motifs.

Once a motif is retrieved, JIG stores the durations of the notes in the motif and compares the pitches of the motif with the underlying chord(s) and the scale that was selected for that chord, thus inferring the pitch type of each note in the motif. Additionally, it stores the direction of the intervals between the notes. With this information, an abstraction is made from the literal motif. From the pitch type and direction attributes of the abstract motif, the appropriate pitches for the current improvisation can then be inferred, using the same mechanism as described in section 3.3. The durations can be adopted directly.

Before actually incorporating the motif in the improvisation, it can be transformed. Some obvious transformations are: registral reversion (i.e. mirroring the pitches horizontally), temporal reversion (mirroring the pitches vertically) and temporal contraction/expansion (respectively shortening and lengthening of durations). In [3], Dean surveys a variety of ways to generate variations on existing motifs.

5 Results

Recalling the three desired properties initially stated, the improvisations generated by JIG can be evaluated. I will evaluate one such improvisation here. The score of the improvisation is given in figure 4. The first desired property is *tonality*. The requirement of tonality is properly met by the improvisation. Nearly all notes were drawn from the $e\ minor$ scale, which is harmonic to all but one chords and is therefore a natural choice. The $e\ minor$ scale is not harmonic to the B^7 chord, because that chord contains a $d\sharp$ pitch. The most obvious solution would be to augment the seventh degree of the $e\ minor$ scale (pitch d), to become $d\sharp$. The resulting scale is $e\ harmonic\ minor$. As can be read from the score, JIG indeed chose for this solution, playing a $d\sharp$ on the B^7 chord in measure 7.

The second desired property is *continuity*. To a certain degree, the melody is fluent, but there are some rather large intervallic jumps. The jump from pitch b to c at the end of measure 2 sounds rather unnatural. The same jump is made between the first two notes of measure 7. Here, however it sounds less dramatic. A possible explanation could be that a large melodic interval is more acceptable if the first note has a longer duration, i.e. if there is more distance between the onset positions of the notes. This could also explain why rather great intervals like those between the first and second note of measure 6 sound more natural than the first jump that was mentioned.

The structure in the improvisation is not present very explicitly. The program trace shows that the motif process generated the notes from the beginning of the improvisation up to and including the first note of measure 3; then the melodying process took over, up to and including the second note of measure 6; after which the motif process generated the notes to the end. The first motif part is clearly recognizable as the motifs from the original melody (in the case of Autumn Leaves, the motifs that constitute the melody are all very similar). The last motif part is less recognizable, due to transformations. The last four notes, however show a melodic contour that is similar to the motifs from the original melody.

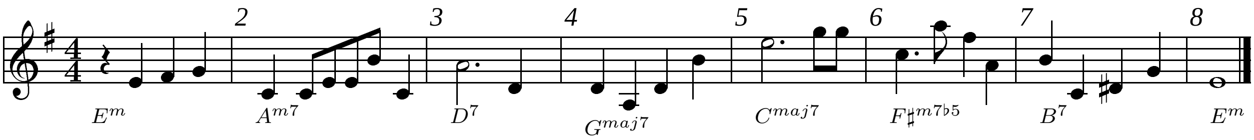

Figure 4: Improvisation on Autumn Leaves

6 Conclusions

Tonality is generally the easiest requirement to satisfy. Sometimes however, basing the choice of scales only on the chords underlying one theme resulted in queer choices of scales. This problem was solved when the chord sequence underlying other song themes were taken into account as well.

In some cases, jumps in the improvisation sounded rather unnatural. This might be solved by either increasing the steepness of the (discrete) bell curve shown in figure 3 (avoiding large jumps altogether), or by adding a constraint that specifies the conditions under which a jump can occur.

At this point, the structure in the improvisation consists of the occurences of motifs from the original melody, at irregular intervals. In principle, this is the desired kind of structure. However, further improvements are necessary to make the structure really apparent, preferably by some kind of cross references between different motif parts within the improvisation.

All in all, the general approach taken in JIG seems promising. The way of structuring the improvisation is modeled after analysis of realistic improvisations [8], [10]. Also, the constraint based approach in combination with probabilistic selection methods, yields a lot of control without determining the improvisations in advance.

7 Future work

A considerable improvement might be achieved if the transition probabilities for the transition model that generates the rhythm, were somehow computed from the themes of the song. In this way, the parts of the improvisation that were generated by the melodying process might better capture the character of the theme.

Further progression could be made by extending the sources of motifs: in addition to extracting motifs from the theme, a permanent body of motifs (representing the 'personal repertoire') could be kept. Possibly, this body might be mutated as new motifs are encountered in newly presented songs. Furthermore, in [10], motifs are not necessarily literal sequences of notes, but often a procedural rule, e.g. 'play play the uneven degrees of scale X when a transition from chord Y to chord Z is encountered'. The use of such rules, as alternative motifs could also be considered.

References

[1] Josep Lluís Arcos, Ramon López de Mántaras, and Xavier Serra. Saxex: A case-based reasoning system for generating expressive musical performances. *Journal of New Music Research*, 27(3):194–210, 1998.

[2] Josep Lluís Arcos and Ramon López de Mántaras. An interactive case-based reasoning approach for generating expressive music. *Applied Intelligence*, 14(1):115–129, 2001.

[3] Roger T. Dean. *Creative improvisation: jazz, contemporary music and beyond*. Open University Press, Milton Keynes, 1989.

[4] Maarten Grachten. JIG: an approach to computational jazz improvisation. Master's thesis, University of Groningen, The Netherlands, 2001.

[5] P. N. Johnson-Laird. Jazz improvisation: a theory at the computational level. In Peter Howell et al., editors, *Representing musical structure*, chapter 9. Academic press ltd, 1991.

[6] Eugene Narmour. *The analysis and cognition of basic melodic structures: the implication realization model*. University of Chicago Press, 1990.

[7] Jeff Pressing. Cognitive processes in improvisation. In W.R. Crozier and A.J. Chapman, editors, *Cognitive processes in the perception of art*, pages 345–63. North Holland, Amsterdam, 1984a.

[8] S. Sadie, editor. *New Grove dictionary of music*. Macmillan, London, 1980.

[9] John Sloboda. *The musical mind: the cognitive psychology of music*, chapter Composition and improvisation. Oxford universitiy press, 1985.

[10] David Sudnow. *Ways of the hand*. Routledge & Kegan Paul Ltd, 1978.