

Toby Gifford & Andrew R. Brown

Queensland University of Technology

Victoria Park Rd.

Kelvin Grove, 4059,

Australia.

(t.gifford, a.brown)@qut.edu.au

The Ambidrum: Automated Rhythmic Improvisation

Abstract

This paper outlines a system for machine improvisation with a human performer where the focus is limited to the provision of rhythmic complementarity. Complementarity is achieved through the real-time measurement of metrical coherence in currently playing rhythmic material that informs the generation of subsequent material. A robust computational approach building on recent theories of improvisational intelligence and situated cognition is described. This algorithm can be effective across a range of musical styles.

Introduction

This research forms a component of a larger research agenda into the construction of improvisational algorithms for performance collaborations between human musicians and computational agents. The broad aim is to construct a computational musical agent that displays rudimentary improvisational intelligence. In this paper we report on the development of the rhythmic component of this computational agent.

Machine improvisation has been an active area of research for many decades, and includes the work of Dannenberg (1989), Rowe (1993, 2001), Biles (2002), and others. We believe that the work presented here has the potential to be more broadly applicable across styles than previous work given that it is based more firmly in aural cognition and perception theories and relies less on stored data-bases or fixed musical structures than much of the earlier work.

Recent approaches to implementing agent-based improvisational intelligence are described in Bryson (1992), Pachet (2004), Raphael (2003), Suzuki (2002), and Thom (2003). In these approaches a statistical model is estimated by analysing a database of examples of a given musical style, and the estimated model is then used to generate novel musical material in real-time. The focus of these studies is primarily the production of melodic improvised lines given a chord progression, or of chordal accompaniment to a human produced melodic line. These systems model rhythmic and pitch elements jointly, and do not involve any musical ‘knowledge’ other than the database of examples used to train the systems

This paper examines rhythmic improvisation independently of any pitch considerations. We outline a strategy for approaching machine improvisation by starting with the task of rhythmic complementarity. In particular we focus on the problem of maintaining an appropriate level of metrical ambiguity and show how this can be achieved with an algorithmic processes based on statistical theories of expectation and coherence. Finally we discuss how these theories can be applied to real-time interaction with a human performer and discuss various potential mappings for interactions between human and machine in an improvisational setting.

Rhythmic Complementarity

In the performance of an ensemble improvisation an important consideration is complementarity. A central consideration in the act of improvisation is striking a balance between novelty and coherence, as emphasised by Kivy.

“‘good’ music ... must cut a path midway between the expected and the unexpected ... if a work’s musical events are all completely unsurprising .. then the music will fulfil all of the listener’s expectations, never be surprising – in a word, will be boring. On the other hand, if musical events are all surprising ... the musical work will be, in effect , unintelligible” (2002, p.74).

An agent displaying improvisational intelligence should be able to produce output that is complementary to its improvising partner. In this paper we describe an improvisational algorithm that attempts to maintain rhythmic complementarity. We will introduce a system, termed the *Ambidrum*, which will be capable of monitoring the balance of novelty and coherence of existing music and generate complementary material that maintains the appropriate balance between the expected and unexpected.

Expectation and Ambiguity

When designing the Ambidrum we adopt the theory of musical expectations proposed by Leonard Meyer (1956) regarding expectations and affect. In this theory, when listening to music the listener is constantly forming

expectations of what is to come, and that the fulfillment or frustration of these expectations stimulate an affective response in the listener. This theory is not without controversy (Jackendoff, 1992; Kivy, 2002) but is nevertheless widely regarded (Borgo, 2004; Dubnov *et al.*, 2006; Kivy, 2002; Pressing, 1998).

An important aspect of this theory is the role of ambiguity in musical affect. *“Ambiguity is important because it gives rise to particularly strong tensions and powerful expectations. For the human mind, ever searching for the certainty and control which comes with the ability to envisage and predict, avoids and abhors such doubtful and confused states and expects subsequent clarification”* (Meyer, 1956)

Taking Meyer’s theories into account we have, as a first step, developed a measure of metric coherence that has a direct relationship with ambiguity. The metric coherence tracks the degree to which rhythms imply a sense of particular metre.

Metre and Rhythm

Metre refers to the demarcation of a bar into strong and weak beats (Meyer, 1956:6). Metre is a hierarchical notion where a given metre is potentially composed of sub-metres. In this paper we will consider a slightly more general notion of metre, where any series of beats ranked by strength will be taken to constitute a metre. For example the metre indicated by the 6/8 time signature is described by [a c c b c c] where each letter indicates a beat, the whole sequence spans one bar, and the alphabetic order of the letters indicates the relative strengths of the beats (a being the strongest). This notion of metre captures hierarchical metrical structures as described in Lerhdahl and Jackendoff (1983) and other descriptions of metre such as Yeston (1976).

Rhythm concerns the manner in which accented beats are grouped with unaccented beats. Accenting may be achieved via a number of devices, which we refer to as rhythmic variables. Meyer (1956) identifies three important rhythmic markers

- (i) Stress (dynamic rhythm)
- (ii) Duration (agogic rhythm)
- (iii) Melodic change (tonic rhythm)

We utilise these markers as rhythmic attributes within the Ambidrum system, and correlate their values as a measure of metric coherence.

Meyer’s rhythmic markers are useful attributes for computational processing because metre is a latent quantity; it is not directly observable. Rather, it is a construct in the mind of the listener or performer. Perception of metre is induced by the rhythmic elements of the music (Large & Kolen, 1994). Once established in the mind of the listener, the perceived metre has a tendency to persist despite the subsequent appearance of rhythmic material that suggests a different metre (Epstein, 1995:29). Rhythmic ambiguity can then arise when the different rhythmic markers induce contradictory senses of the metre (Meyer, 1956). It is these perceptual cues that we utilise to enable the Ambidrum to measure the rhythmic ambiguity of musical material.

Coherence and Ambiguity

The Ambidrum ultimately uses any measurement of existing and proposed material in order to generate a new rhythmic pattern. As a step toward rhythmic complementarity the Ambidrum searches for a new rhythm that has a specified degree of coherence with, or similarity to, the currently specified meter. A metre is specified as a series of quanta strengths or emphases, as described in more detail later. The Ambidrum plays, as it’s next pattern, the rhythm that most closely matches a desired degree of coherence. We define coherence as a measure of the correlation between the strength of the rhythmic attributes (markers) at each quanta (subdivision of the beat). At one end of the scale a completely coherent pattern will match the underlying meter exactly, at the other end of the scale an incoherent pattern will have the inverse quanta values to those specified in the meter. As it turns out, rhythms at these two extremes of coherence provide a similar metrical stability and rhythms with a moderate degree of coherence are the least likely to imply a sense of meter. Therefore, we say that the rhythms with moderate coherence values are highly ambiguous with respect to metre and those with either a high or low coherence measure are less ambiguous. This relationship is shown in figure 1.

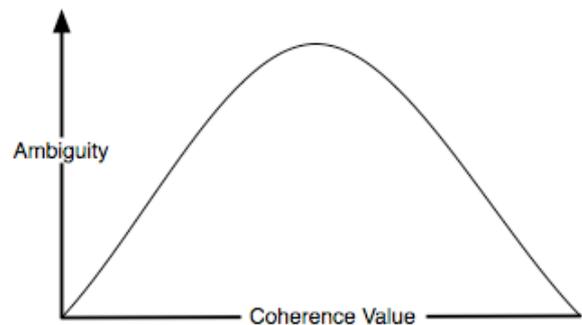


Figure 1. The relationship between metrical ambiguity and coherence.

This relationship presents an interesting musicological or psychoacoustic relationship between statistical correlation and musical ambiguity and, again, reinforces the central insights of Meyer with regard to balance between forces is at play in this computational generation of musical rhythms.

Another way of understanding the relationship between coherence and ambiguity in this context is to imagine that coherence and incoherence are magnetic forces attracting the rhythm into a pattern that moulds itself onto the specified metre. Ambiguity is introduced as these two forces pulling on the rhythm distort it. When the two forces are equally strong the ambiguity is highest because the rhythm bears least resemblance to the metre template. As the rhythm approaches one of the extremes of coherence it becomes less ambiguous by fitting closer to the metre or its inverse.

Coherence Level

The Ambidrum is a real-time system that produces a rhythm one note at a time by analysing the coherence of its previous output and taking action to maintain the coherence of its output at a given target level. To this end it constructs a measure of rhythmic coherence, which we refer to as the coherence level.

The inputs to the Ambidrum process are a tempo, a metre, and a matrix of target coherence levels. The metre is defined as being a series of stress levels of quanta in a bar. For example 4/4 time could be represented by the series [a c b c] where each quanta is a quarter-note and a represents the strongest value and c the weakest value. At a higher quantisation level the same time signature could be represented with more quanta by [a d d d c d d d b d d d c d d d] providing sixteenth-note resolution. The Ambidrum takes the quantisation as being effectively determined by the quanta-length of the metre series relative to the time signature.

Following the above discussion, the Ambidrum considers three rhythmic variables: velocity, timbre and duration. When the process is running it generates MIDI messages which are sent to a drum machine. These variables are mapped to the velocity, pitch and duration parameters for a MIDI note-on/note-off pair. In the context of a drum machine the pitch parameter of the MIDI message is not directly related to frequency but rather to timbre, determining which drum sound is triggered.

At every quanta a value is set for these rhythmic variables. The variables take on discrete values selected from the range determined by the metre. So, for example, for the metre defined by [a d d d c d d d b d d d c d d d] the rhythmic variables may take the values a, b, c and d. Where a represents a strong rhythmic event through to d which represents a weak rhythmic event.

For the variable of velocity, a strong value is mapped to a high velocity. For duration, a long duration is taken to be stronger than a short duration. For timbre (which really amounts to choice of drum) it is not always clear which timbres are stronger or weaker. In the case of a classic drum machine kit with kick-drum, snare-drum, high-hat and tom, probably the most obvious assignment would be;

Timbre	Value
kick-drum	a
tom	b
snare-drum	c
high-hat	d

The generated rhythm is described by a series of values for each of the rhythmic variables. The Ambidrum selects values for these variables that attempt to create a rhythm that is suitably coherent, as determined by the input target coherence matrix. Following the above discussion, the process considers the rhythmic ambiguities created by latent metrical dissonances induced by disparate metric suggestions of the different rhythmic variables.

A metrically unambiguous (eg., completely coherent) rhythm would have all of the rhythmic variables matching the metre, as shown in figure 2.

metre	[a d d d c d d d b d d d c d d d]
velocity	[a d d d c d d d b d d d c d d d]
timbre	[a d d d c d d d b d d d c d d d]
duration	[a d d d c d d d b d d d c d d d]

Figure 2. A metrically unambiguous rhythm matrix.

However, let us consider a more ambiguous (less coherent) rhythm, shown in figure 3, where the rhythmic variables are not perfectly aligned to the metre, nor to each other.

metre	[a d d d c d d d b d d d c d d d]
velocity	[a c a c d b b d a c d d c d a d]
timbre	[b d d c b d d c b d a d c b d b]
duration	[c c d d c c c d a a b d c d d a]

Figure 3. A metrically ambiguous rhythm matrix.

The Ambidrum uses a measure of how closely aligned these sequences are to each other as a proxy for the coherence of the rhythm. The particular measure employed is a correlation statistic for each pair of these sequences. To calculate the correlation we assign each of the possible variable values a numeric value centred around zero. In the above example this would translate to mapping

- a ⇒ 2
- b ⇒ 1
- c ⇒ -1
- d ⇒ -2

Then considering each series of variable values as a vector we calculate the correlations via the formula

$$corr(x,y) = \frac{x^T y}{\sqrt{(x^T x)(y^T y)}}$$

for each pair of variables.

The correlation value lies between 1 and -1. If the variables have identical values then their correlation will be equal to 1. When a pair of variables are inverse to each other, their correlation will be -1. When two variables are unrelated to each other (or orthogonal) their correlation will be zero.

The collection of pairwise correlations of the rhythmic variables to themselves and to the metre forms a correlation matrix. For example, the preceding values for the variables yield the correlation matrix shown in figure 4.

	Metre	Velocity	Timbre	Duration
Metre	1	0.44	0.43	0.24
Velocity	0.44	1	-0.32	0
Timbre	0.43	-0.32	1	0.34
Duration	0.24	0	0.34	1

Figure 4. A calculated correlation matrix.

The Ambidrum considers its output each quanta based on a sliding window of its own historical output - generally a fixed number of bars. So, for example, using

a metre of [a c b c] the process might find itself in the following situation depicted in figure 5.

```
metre      [a c b c] [ a ...
velocity   [b c c b] [ ? ...
timbre     [a c b b] [ ? ...
duration   [c c b a] [ ? ...
```

Figure 5. A calculated correlation matrix.

The question marks signify that the Ambidrum must choose a value for each of these variables for the next quanta. The choice is made so as to have the resulting sequences as close as possible to the target coherence levels, determined by a target correlation matrix, which is an input to the generative process.

	Metre	Velocity	Timbre	Duration
Metre	1	1	1	1
Velocity	1	1	1	1
Timbre	1	1	1	1
Duration	1	1	1	1

Figure 6. A coherent target correlation matrix

For example using the target correlation matrix shown in figure 6 the Ambidrum would choose the velocity, timbre and duration [v t d] of the next note so as to make the series

```
metre      . [c b c a]
velocity   . [c c b v]
timbre     . [c b b t]
duration   . [c b a d]
```

have intercorrelations as close to 1 as possible. In this case the choice would be v = 1, t = 1, d = 1.

The target correlation matrix in figure 6 is the completely coherent (totally unambiguous) target matrix. Any other choice of target matrix is possible and would result in different choices for the next note generated.

A useful metaphor for the coherence level is a VU metre, that constantly monitors the level of some property of an audio stream in real-time.



Figure 7. A coherence level metre.

The Ambidrum monitors the coherence of its generated rhythm and attempts to maintain it at a target level. This target level is externally controlled, and may

be changed during the course of performance. In fact, the Ambidrum essentially monitors a coherence level bridge – comprising of a coherence level metre for each of the pairwise correlations of the rhythmic variables and the metre. The target correlations may be set independently, and comprise external control parameters that will affect the operation of the Ambidrum in real-time. The mute button on the picture in figure 7 alludes to the option of turning off tracking for any of the variable pairs.

Example Results

Rhythms generated by the system quickly locate a pattern that closely matches the target coherence value and then falls into a stable cycle which results in repeating that pattern indefinitely.

As an example we show the resulting patterns produced for a few target coherences using the metre [a d d d c d d d b d d d c d d d]. The completely coherent target matrix reproduces the metre exactly

Stable cycle for all target correlations = 1

```
metre      [a d d d c d d d b d d d c d d d]
velocity   [a d d d c d d d b d d d c d d d]
timbre     [a d d d c d d d b d d d c d d d]
duration   [a d d d c d d d b d d d c d d d]
```

However when we allow the rhythmic variables to be independent by setting the target correlations to zero we obtain a rhythm that is more ambiguous

Stable cycle for all target correlations = 0

```
metre      [a d d d c d d d b d d d c d d d]
velocity   [d a a c d d d d a d d d b a d]
timbre     [b d d d d d d d b d a a a a a a]
duration   [b d d d c d d d b d a a a a a a]
```

Setting the target correlations to -1 results in

Stable cycle for all target correlations = -1

```
metre      [a d d d c d d d b d d d c d d d]
velocity   [d b a a c d d d d d d b c d b d]
timbre     [b d c c c a a a d a a b c a b a]
duration   [b d c c c a a a d a a b c a b a]
```

Target Automation

To create variation, and interest, in the generated rhythm pattern the target coherence values can be continuously adjusted. A simple way to do this is to modulate the target values by some simple function, for example a low frequency sine wave or selection of a random value.

Automating the target value by small degrees produces subtle and interesting variations that can sound almost evolutionary in nature, frequent large variations tend to produce unstable rhythmic behavior, while infrequent shifts from one value to another introduce sudden changes followed by periods of rhythmic stability.

The automation of the target cohesion value is an effective method for controlling the rate of change and the general interest of the generated rhythm patterns. However, the modulating functions usually become

tiresome after some extended listening due to their lack of large-scale direction. It is more effective, and closer to the intention of this research, to have the target coherence level controlled by a human performer.

Source Following

While it would be easy to have a performer directly control the coherence level via a dial or slider, we can utilise the existing coherence measuring techniques to follow a human rhythmic performance in real-time. This approach enables improvisation by the machine in direct response to the performance of the human, and is elegant in that the same rhythm coherence technique is used for both the performance tracking and the algorithmic generation.

Given that the metre and tempo are specified in advance, sections of the human performance can be captured and their coherence value calculated. These values can be used to adjust the machine's coherence value and thus the generated rhythms. The mapping between human and machine coherence values is a matter of choice depending upon the desired musical outcome. Two obvious mappings include a) that the machine use the same coherence values as the performer which results in the reinforcement of the coherence or ambiguity dictated by the performer, or b) that the machine use an inverse coherence mapping such that as the performer played less metrically obvious rhythms the machine would tighten-up and play quite 'straight' or conversely as the human played regular metrical patterns the computer would provide greater rhythmical interest and freedom. This latter scenario shows how our objective to achieve rhythmic complementarity has finally been realised, albeit in a simplistic way. Further scaling and offsetting of the coherence mappings could increase the range of interactions and the adjustment of the mappings over time would provide even greater interest and variety.

Conclusion

We have outlined a method to enable unsupervised complementary rhythmic improvisation between a human performer and a computational agent. This method has been implemented as the Ambidrum system in the Impromptu environment (Sorensen 2005). At the current stage of this research a number of assumptions need to be maintained about the improvisation, in particular the metre and tempo are assumed to be constant, but within these constraints the Ambidrum is a robust interactive rhythmic improvising system. In future research on the Ambidrum system we plan to utilise beat induction techniques to remove the need for the tempo and metre assumptions and will also examine control structures for larger scale organisation of musical structure so that the evolution of the improvisation is not solely controlled by the human performer.

References

Biles, J. (2002). *Genjam: Evolutionary computation gets a gig*. Paper presented at the 3rd

- Conference on Information Technology Curriculum, Rochester.
- Borgo, D. (2004, April 2004). *Sync or swarm: Group dynamics in musical free improvisation*. Paper presented at the Conference of Interdisciplinary Musicology, Graz, Austria.
- Bryson, J. (1992). *The subsumption development strategy of a music modelling system*. University of Edinburgh.
- Dubnov, S., McAdams, S., & Reynolds, R. (2006). Structural and affective aspects of music from statistical audio signal analysis. *to appear in Journal of the American Society for Information Science and Technology, Special Issue on Style*.
- Epstein, D. (1995). *Shaping time: Music, the brain, and performance*. New York: Schirmer Books.
- Jackendoff, R. (1992). *Languages of the mind*. Cambridge, MASS: MIT Press.
- Kivy, P. (2002). *Introduction to a philosophy of music*. Oxford: Oxford University Press.
- Large, E., & Kolen, J. (1994). Resonance and the perception of musical meter. *Connection Science, 6*(1).
- Lerdahl, F. a. J., Ray. (1983). *A generative theory of tonal music*. Cambridge, Massachusetts: MIT Press.
- Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- Pachet, F. (Ed.). (2004). *On the design of a musical flow machine*: IOS.
- Pressing, J. (1998). Psychological constraints on improvisational expertise and communication. In B. Nettl (Ed.), *In the course of performance*. Chicago: University of Chicago Press.
- Raphael, C. (2003). Orchestra in a box: A system for real-time musical accompaniment. *IJCAI*.
- Sorensen, A. (2005). Impromptu: An interactive programming environment for composition and performance. In A. R. Brown and T. Opie (eds.) *Australasian Computer Music Conference 2005*. Brisbane: ACMA pp. 149-153.
- Suzuki, K. (2002). *Machine listening for autonomous musical performance systems*. Paper presented at the International Computer Music Conference, Gothenburg.
- Thom, B. (2003). Interactive improvisational music companionship: A user-modelling approach. *The User Modelling and User-Adapted Interaction Journal*.
- Yeston, M. (1976). *The stratification of musical rhythm*. New Haven: Yale University Press.