Modelling the Influence of Pitch Duration on the induction of Tonality from Pitch-Use

Niall Griffith
University of Exeter, Dept. of Computer Science.
agr@cs.exeter.ac.uk

Abstract

The patterns of duration of the pitches used in pieces of music are thought to be strongly correlated with profiles of pitch salience identified with keys. However, it is not clear whether this correlation is causal. This report compares different ways of describing pitch duration in classifying pitch use into tonal centres.

1 Introduction

While it is easy to list the characteristics we would like an intelligent musical autonomic to possess, it is much harder to specify and implement them. The structure of Tonality is a case in point. The processes that organise pitch relationships in music into the structures we identify with tonality, such as key, seem to involve complex processes that are concerned with both the intrinsic nature of a sound - pitch and timbre [Terhardt, 1984], as well as how these sounds are used and their context [Krumhansl, 1990a].

[Griiith, 1993]. One important descriptor of pitch use is thought to be the relative duration of pitches. The supposition is that more important pitches are longer in duration and are metrically prominent. While the duration of pitch seems initially a simple quantity to encode, there are in fact a number of ways in which duration information can be derived, and these result in more and less stable descriptions.

2 Duration and Tonality

The summation of pitch duration is the most direct quantification of duration, and this method has been used by Krumhansl in a computational model of key identification [Krumhansl, 1990a]. In this model the duration of pitches predicted the key of pieces of music by correlating the accumulated duration of pitches with profiles of the salience of pitches in different scales. The profile that is closest to the pattern of pitch durations accumulated over a piece identifies the key. The profiles were abstracted from the results of a set of experiments in which subjects were asked to rate the appropriateness of pitches in different contexts. The algorithm was generally successful at keeping a variety of pieces of music. However, it has come under a deal of scrutiny, and criticism [Butler and Brown, 1984; Cook, 1993]. One issue that is not clear is whether or not the correlation between accumulated pitch duration and pitch salience, used within Krumhansl's model, is psychologically causal. It is perhaps just a coincidence arising from the association of longer durations with pitches that are important for other reasons. In fact this problem is quite difficult to resolve through psychological experiment [Butler and Brown, 1984; Brown, 1988; Butler, 1989, Krumhansl, 1990].

Also, if the correlation is causal it is not clear how pitch duration is characterised in the psychological processes we associate with tonality, or whether such characterisations can be learned. However, it is possible to investigate the encoding of the duration of pitches in a computer model. The simulations reported below aim to shed light on what a learning mechanism can induce from patterns of pitch duration in pieces of music. Simulations using the simple frequency of pitch occurrence [Griffith, 1993] established that it is possible to establish stable categories of pitch use that represent tonal centres or keys. However, incorporating information describing the accumulated duration of pitches degraded accuracy.

Several different methods have been used to
generate descriptions of duration of pitches in sets of nursery rhyme melodies. The patterns of values representing pitch duration are classified in an ART2 artificial neural network [Carpenter and Grossberg, 1987]. The descriptors used are accumulated duration, two different measures of average duration, a proportional measure, accumulated duration modified by the time elapsed since the pitch last occurred, duration only accumulated at the start of the bar, the accumulation of duration modified by metrical position, and two compound functions involving frequency, duration, and average duration. Accumulated duration is the simplest descriptor of duration over time and is the description used in [Krumhansl, 1990a]. The purpose of these simulations is to investigate which if any of these encodings of duration describes stable patterns of pitch use that can be classified into consistent categories identified with keys.

3 Experiments

The data set used in the experiments comprised two selections of nursery rhyme melodies, 60 training and 25 test songs. They all use major scales. The representation passed to the ART2 network for classification was a vector of twelve pitch class values, from a to b, which was classified at the end of each nursery rhyme melody. The vector was set to zeros at the start of each melody. The results for all the simulations are shown in Table 1. The first two results are for

<table>
<thead>
<tr>
<th>Type</th>
<th>ρ</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>0.925</td>
<td>98.3</td>
<td>89.5</td>
</tr>
<tr>
<td>Tracked</td>
<td>0.96</td>
<td>107.7</td>
<td>89.4</td>
</tr>
<tr>
<td>Sum</td>
<td>0.9</td>
<td>87.9</td>
<td>91.3</td>
</tr>
<tr>
<td>Mean1</td>
<td>0.925</td>
<td>51.0</td>
<td>62.3</td>
</tr>
<tr>
<td>Mean2</td>
<td>0.91</td>
<td>95.3</td>
<td>92.7</td>
</tr>
<tr>
<td>Fraction</td>
<td>0.905</td>
<td>91.2</td>
<td>93.3</td>
</tr>
<tr>
<td>Elapsed</td>
<td>0.9</td>
<td>73.3</td>
<td>64.0</td>
</tr>
<tr>
<td>Bar1</td>
<td>0.0</td>
<td>61.7</td>
<td>49.3</td>
</tr>
<tr>
<td>Bar2</td>
<td>0.8925</td>
<td>84.7</td>
<td>71.3</td>
</tr>
<tr>
<td>Bar3</td>
<td>0.9</td>
<td>82.6</td>
<td>80.3</td>
</tr>
<tr>
<td>Complex1</td>
<td>0.94</td>
<td>90.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Complex2</td>
<td>0.92</td>
<td>95.3</td>
<td>92.7</td>
</tr>
</tbody>
</table>

Table 1: The percentage of correct classifications of aspects of pitch by an ART2 network. ρ is the response of the ART2 network.

comparative purposes and show figures for two sets of experiments reported in [Griffith, 1993]. These involved a memory that tracked the frequency of occurrence and duration of pitches.

In the present experiments the first quantification of duration that was calculated was the total duration of each pitch - \( S_1 = \sum_{i=1}^{n} d_{ia} \), where \( d_{ia} \) is the duration of the \( i \)th occurrence of \( a \)th pitch, and \( N \) is the number of notes in the song. This summation is the method used by Krumhansl’s model, however, it resulted in only 87% and 91% of the songs in the training and test sets being attributed to their correct key node. Given the simple diatonic nature of the nursery rhymes and the role of nursery rhymes in children’s musical acculturation this is not overwhelmingly successful. Subsequently, other methods were used. The first of these was the average duration of pitches, which was calculated in two ways. Mean1 was calculated by dividing the total duration of a pitch by the number of its occurrences in the song - \( M_1 = S_1/T \), where \( T = \sum_{i=1}^{n} 1 \) if \( d_{ia} > 0 \). The result for this average duration was very poor, being only 51% and 62% respectively. Mean2 is the result of dividing each pitch’s duration by the number of notes in the tone - \( M_2 = S_1/N \). This creates quite stable patterns, being correct in 95% and 93% cases respectively. Rather than averaging or totalling duration the Fraction method calculates a pitch’s duration as a proportion of the total duration of the pitches in the song - \( F = S_1/D_1 \), where \( D_1 = \sum_{i=1}^{n} d_{ia} \). This was not quite as stable as Mean2, however, it is above 90% in both cases. A variant of the simple accumulation is the Elapsed representation. For this the accumulation of a pitch’s duration is modified by the number of pitches that have occurred since the last occurrence of that pitch - \( E = \sum_{i=1}^{n} d_{ia} + e_i \), where \( e_i \) is the time elapsed since the last occurrence of the \( i \)th pitch. This resulted in a quite low accuracy, 73% and 64% respectively. Another variant of the simple accumulation is the Bar1, Bar2 and Bar3 representations. In these the accumulation of a pitch’s duration is modified by whether the pitch occurs at the start of the bar. Bar ignores pitches that are not the first beat of the bar - \( B_1 = \sum_{i=1}^{n} d_{ia} \) if \( o \) is not 1. This seems to discuss too much information and the accuracy is only 62% and 49% respectively. Bar2 modifies the accumulation of duration by which beat of the bar a pitch lies on - \( B_2 = \sum_{i=1}^{n} d_{ia}/b \) where \( b \) is the best of the bar. This is more accurate but is not as brilliant with 85% and 71% respectively. Bar3

Psychonoacoustics, Perception 36 ICMC Proceedings 1994
also places most emphasis on the first beat of the bar - \( B_1 = \sum_{i=1}^{N} d_i x_i \) where \( x = 1 \) if \( x < 1 \) else 0.5. Like Hark2 this is reasonably accurate with 82% and 89% respectively. Complex1 is the mean value of the normalised values of the frequency of occurrence, total duration (Accumulated) and average duration (as in Mean)) of pitches - \( C_{1} = (F N + SN + \frac{MN}{1})/3 \) where \( FN \), \( SN \) and \( MN \) are the normalised vectors. This was the most successful method with 98% and 100% accuracy respectively. Complex2 is the mean value of the normalised values of a simple frequency of occurrence, and total duration (Accumulated) - \( C_{2} = (FN + SN)/2 \). This was not as stable as Complex1 but gave 90% and 95% accuracy.

The various descriptions derived from the duration of pitch result in a wide variety of classifications of pitch use. While all the representations are classifiable into twelve nodes in the ART2 network that are equivalent to keys, the accuracy of the classifications differs widely, between 51% and 98.3% for the training set. The most successful method seems to be the mixed strategy that takes into account the frequency of occurrence, accumulated duration and average duration. While the results this method gives for the training set are the same as those based upon the frequency of occurrence of pitches, for the test set the figures are improved by 1% from 88% to 100%.

4 Conclusions

The simulations reported here are an investiga- tion into the informational stability of different descriptions of pitch duration. They were undertaken as a preliminary investigation into how to encode duration information in a mechanism that is programmed to induce tonal categories from the statistics of pitch use in melodies. The results show that it is possible to establish stable categories from these representations, and that these conform to the kinds of levels of accu- racy we would expect with these diatonia- cally limited music melodies. However, it is striking that the simple accumulation of duration is not the best method. This may be because melodies in themselves do not contain enough information. Krumpourd's experimenta used multi-part music. However, it does seem surprising that these themes, which are so important for chil- dren learning the ground rules of tonal music, should produce unstable categories when dura- tion is categorised in this way. It suggests that a summation of duration is not the best basis for inducing patterns of pitch duration, and so casts doubt on the causal nature of the correla- tion used in Krumpourd's model of key identifi- cation.

References


