# MML-based Compressive Models for Musical Melody

Adrian C. Bickerstaffe and David L. Dowe
Monash University, Clayton
Melbourne, VIC 3800, Australia
{adrianb,dld}@csse.monash.edu.au

**Abstract**

Human inference of melodic structure is seemingly innate and intuitive, yet little is known about the cognitive processes that lead to such inference. Subsequently, computer modelling of melodic structure remains a difficult problem. Successful inference of the structure in musical data can provide insight into the process which created the data (e.g. the style of a composer) and result in data compression. Music is structurally rich data, with structure present at several levels, ranging from short term (e.g. note level) to long term (e.g. theme level). To systematically investigate melodic structure, we first begin with short term structure, that is, the structure of individual music notes. This paper provides eight new pitch models and one duration model for musical notes. These models are based upon the Minimum Message Length (MML) principle. Using MML, we discover which models best fit the test melodies and show that the best MML-based models compare favourably to existing compression techniques. We discuss limitations of the proposed methods and, finally, offer possible directions towards improving the MML-based models.

## 1 Introduction

The human mind is capable of intuitively inferring structure in melodies; a person may hear a melody for the first time and yet seemingly know what is to come next as they listen. However, how such models are built in our minds is unknown and thus, efforts to automatically model music have had limited success [5]. Potential applications of accurately inferring melodic structure remain many and varied. For example, a large online database of world folk songs could be compressed efficiently if repeated structure is identified in the songs. Furthermore, melodic animal calls such as bird and whale songs may also be modelled. The inferred structure of animal calls could potentially reveal syntactic and semantic information about animal communication. Clearly, the impact of such a discovery would be significant on a world-scale. The statistical approach used here to model melodies could also be used for other data types such as text, images, and video.

Statistical models for the melodic attributes of pitch interval and duration interval are used as the basis of the research presented here. Modelling melodies using the statistical distribution of various attributes (e.g. note pitch, duration, etc.) has long been used in the study of ethnomusicology [4, 5]. Research has shown that listeners are particularly sensitive to pitch distribution and frequency information within cognitive processes [8]. Hence, we claim that our statistical approach is certainly logical.

Eight distinct pitch interval models were developed alongside a single duration interval model. Using the principle of Occam's Razor in the form of Minimum Message Length (MML) (see Section 2) [9, 10], the "best" models for the test melodies were found. The level of structure captured by the model is compared to GNU compression utilities `gzip` and `bzip`. Better inference of melodic structure will result in a higher data compression rate and, in this way, we determine how much underlying structure is being captured by each model.

Section 2 of this paper details MML, whilst pitch and rhythm models are defined in Section 3. A discussion of results is given is Section 4, while limitations and concluding comments are made in Sections 5 and 6 respectively.

## 2 MML Modelling

Statistical inference concerns data that has been generated by some process which, in turn, can be described by a probability distribution with parameters. Inference involves discovering what the distribution is and finding estimates of the distribution parameters. In other words, given some data, we wish to discover something about how the data came to be [13]. For a melody, our inference involves discovering

the musical structure of the piece. Successful inference of melodic structure can lead to a more concise representation for the melody, that is, a *compressed* version.

Developed by Wallace and Boulton in 1968 [9], MML has proven a powerful Bayesian framework for statistical inference and modelling. MML features a *two part* message, where the first part is a statement of the hypothesis, $H$, and the latter part is an encoding of the data, $D$, given the stated hypothesis (see Figure 1) [9, 12, 10].

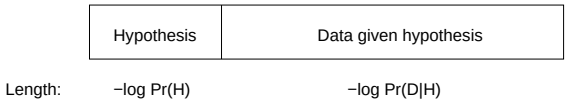


Figure 1: A Two-Part Message

The general MML message length formula for a model with parameters $\vec{\theta}$ and data $x$ is [12]

$$\begin{aligned} msgLen = & - \log\left(\frac{h(\vec{\theta})}{\kappa_n^{\frac{n}{2}}\sqrt{F(\vec{\theta}}}\right) \\ & - \log f(x|\vec{\theta}) + \frac{n}{2} \end{aligned}$$

where $h$ is the prior, $n$ is the number of "free" parameters, $\kappa_n$ is a lattice constant (also known as a dimension constant, $\kappa_1 = \frac{1}{12}$, $\kappa_2 = \frac{5}{36\sqrt{3}}$) [3]), $F$ is the Fisher information, and $f$ is the likelihood function. $F$ is given by the determinant of the expected Fisher information matrix, whose entries $(i, j)$ are

$$\sum_{x\in\mathcal{X}} f(x|\vec{\theta})\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(-\log f(x|\vec{\theta})\right)$$

and where $\mathcal{X}$ is the dataspace. Fisher information indicates the sensitivity of the likelihood function to parameters. Within the MML framework, the Fisher information is used to determine how accurately the model should be stated. Should the second derivatives of the likelihood function be small, the parameters may be stated less precisely. Alternately, large valued second derivatives indicate that the parameters must be stated more accurately.

Given a model of some data, MML can evaluate how well that model explains the data in terms of a message length (i.e. transmission bit-cost). The encoding is assumed to be known to both the transmitting and receiving entities, where transmission occurs over a noiseless transmission channel. In the context of music analysis, MML gives shorter message lengths for more predictive models, and such models can be used to generate new melodies that are more similar to the original melodies. This behaviour relates to Occam's Razor which, loosely paraphrased, is:

> If two theories explain the facts equally well then the simpler theory is to be preferred.

In fact, MML is a quantitative form of Occam's Razor [6]. Wallace's approach gives a trade-off between hypothesis complexity and the goodness of fit to the data. In this way, MML is resistant to overfitting data. MML models the underlying general structure of the data rather than the specifics of any particular data set.

# 3 Pitch and Rhythm Models

Two significant melodic attributes are modelled: pitch interval and duration interval. A pitch interval is the difference in pitch between two successive notes, measured in semitones. Musical rests are modelled as a special case of a musical note; a rest has an associated duration, but has a fixed, reserved pitch of C7 (two octaves above Middle C). Importantly, C7 does not occur naturally in the test data set. A duration interval is the ratio between the duration of a note and the duration of the immediately preceding note. Duration intervals are measured as rational numbers.

Since we are using relative encoding, the pitch and duration of the first note of each melody must be absolutely encoded. The bit-cost of the first pitch, $C_{p_1}(x)$, is calculated using

$$C_{p_1}(x) = \begin{cases} -\log_2(\frac{3}{8}) + L(x), & x > 0 \\ -\log_2(\frac{3}{8}) + L(-x), & x < 0 \\ -\log_2(\frac{1}{4}), & x = 0 \end{cases}$$

where $x$ is the difference between the first pitch and Middle C (C5), measured in semitones, and where $L$ returns the bit-cost of the $\log^*$ integer encoding [7] of $x$. The first duration is transmitted as two integers, $m$ and $n$, where the duration is given by $m \times 2^{-n}$ and where $m$ is odd. The cost of the first note duration is given by $C_{d_1} = \log^*\left(\frac{m+1}{2}\right) + \log^*(n) + 1$. We transmit $\frac{m+1}{2}$ to map $m = 1, 3, 5, ...$ to $1, 2, 3, ....$ This model permits note ties.

As the true distribution of pitch intervals and duration intervals is unknown, a variety of models were

constructed from commonly used statistical distributions. Multinomial, geometric, Poisson, and Gaussian distributions were used to create the models described in Section 3.1. MML message length formulae for these distributions were taken from [12, 11] or hand-derived. The total message length for each model is given by the sum of the message lengths of each component distribution. Eight unique note pitch interval models were constructed, along with a single model for duration intervals.

## 3.1 Pitch Interval Models

To encode a pitch interval, we must first encode whether the pitch interval is negative, positive, or zero. Then, given that a pitch interval is non-zero, we must also encode the amount by which the pitch has changed, conditional on whether the change was negative or positive. We can either encode this amount directly or by encoding it in two parts; as an octave and a semitone within the octave. The manner in which this encoding is achieved is identical for positive and negative intervals; illustration of negative interval encoding is often omitted simply due to diagram space constraints. While models that feature the two-part non-zero interval encoding are more complex than models that use direct encoding, the principle of MML will indicate whether the added complexity is warranted.

Pitch Model 1 features geometric distributions to directly encode negative and positive intervals (see Figure 2).

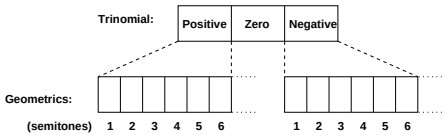


Figure 2: Pitch Interval Model 1

Numbered segments of figures indicate the pitch interval value measured in semitones. Pitch Model 2 replaces the geometric distributions with Poisson distributions.

Pitch Model 3 (see Figure 3) uses a geometric distribution to model the octave in which the pitch interval falls (noting that there are twelve semitones to an octave). For each octave range there is a $12^{th}$ order multinomial to encode the individual intervals of that octave.

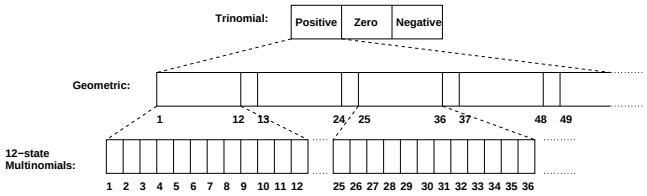


Figure 3: Pitch Interval Model 3

Model 4 is identical to Model 3 except that it features a *single* $12^{th}$ order multinomial to model the individual intervals. Here, intervals are recorded mod12; since the octave is already specified, the exact pitch interval value can be decoded. Model 3 is more general than Model 4.

Model 5 (see Figure 4) models non-zero intervals using Poisson distributions on a per-octave basis, unlike Model 3, where a geometric distribution is used. For each octave range modelled by the Poisson distribution, a $12^{th}$ order multinomial encodes the specific interval value within the octave. Pitch Model 6 is identical to Model 5 except that it features a *single* multinomial distribution to model the intervals mod12. Model 5 is more general than Model 6.

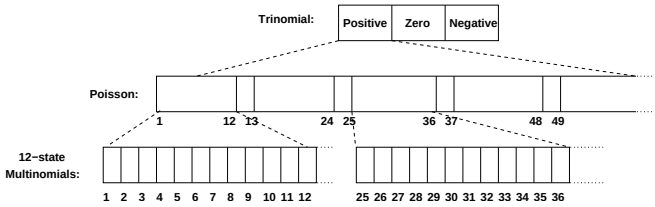


Figure 4: Pitch Interval Model 5

Model 7 (see Figure 5) utilizes a $12^{th}$ order multinomial to encode intervals mod12. That is, the first event of this multinomial accounts for intervals 1, 13, 25, etc., and the second event accounts for intervals 2, 14, 26, etc. Geometric distributions then model the multiples of the multinomial events.

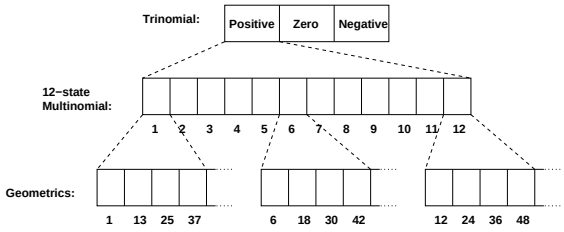


Figure 5: Pitch Interval Model 7

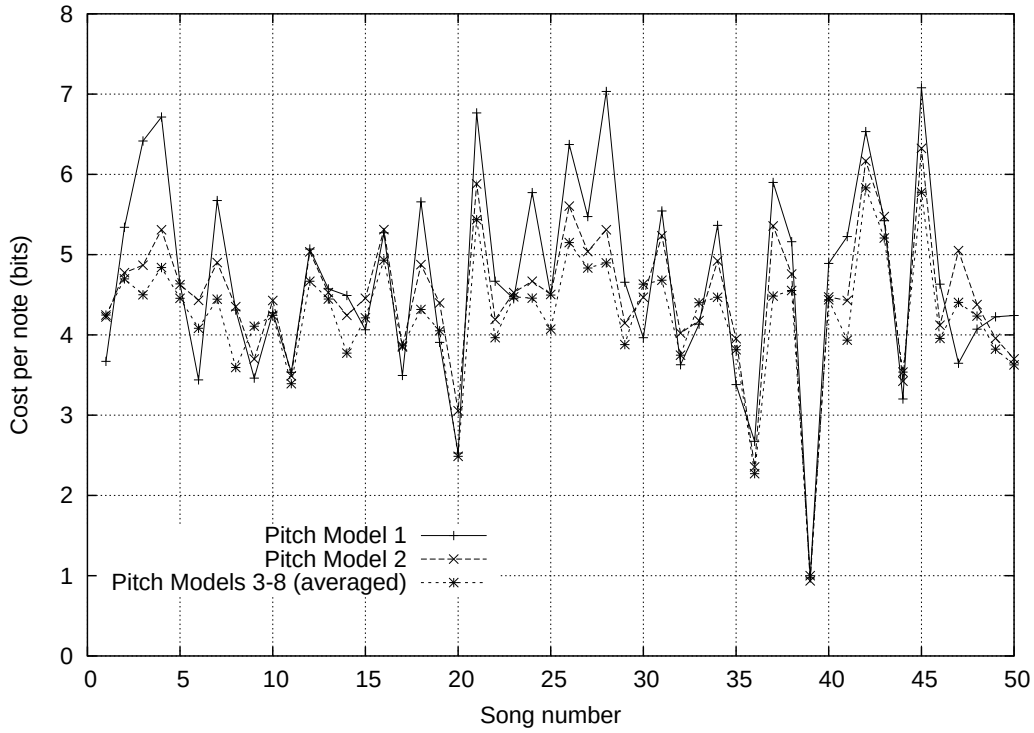


Figure 6: Comparison of MML-based Models

Model 8 is identical to Model 7 except that the geometric distributions are replaced by Poisson distributions.

### 3.2 Duration Interval Model

Each duration ratio, $d_r$, is recorded as $\log_2(\frac{d_i}{d_{i-1}})$ where $d_i$ is the current note duration and $d_{i-1}$ is the previous note duration. This encoding applies to all notes and rests following the first note/rest of a melody.

The logarithm of the ratio is taken since standard music note durations are logarithmic in nature (i.e. whole note, half note, quarter note, etc.), and the resulting values are modelled using a Gaussian distribution. The MML message length formula for Gaussian distributions can be found in [11].

## 4 Results and Discussion

The models of Section 3 were implemented in C++ using Donncha O'Maidin's "Common Practice Notation View" (CPNView) library. CPNView is obtainable via private correspondence (email: Donncha.OMaidin@ul.ie). Test data originates from an online archive of music copyright infringement maintained by Columbia University's Law Library [1]. This archive details legal cases regarding music plagiarism, with pieces ranging from pre-1900s to the present time. A total of 50 randomly selected melodies were hand-transcribed by the first author and used for model comparison. All melodies are monophonic.

Figure 6 gives a direct comparison of the MML-based pitch models described in Section 3, each coupled with the Gaussian-based note duration model. Pitch Models 3-8 ranked very closely to one another in terms of transmission efficiency, showing negligible differences in transmission costs per note. Subsequently, the bit-costs per note of these pitch models have been plotted as the average of values to improve the readability of Figure 6. Pitch Model 2 ranks closely to the average of Models 3-8. On average, Model 2 is within 1 bit per note of Models 3-8. Generally, Pitch Model 1 produces the highest bit-cost per note, with the exception of test melodies 1, 6, 9, 15, 17, 30, 32, 35, 44 and 47 (20% of test cases). However, for these exceptional cases, the transmission cost using Model 1 is within approximately 0.8 bits per note of the next best model. Interestingly, there are three obvious local minima on the graph; all pitch models provide very similar transmission costs per note for melodies 20, 36, and 39. Upon inspec-

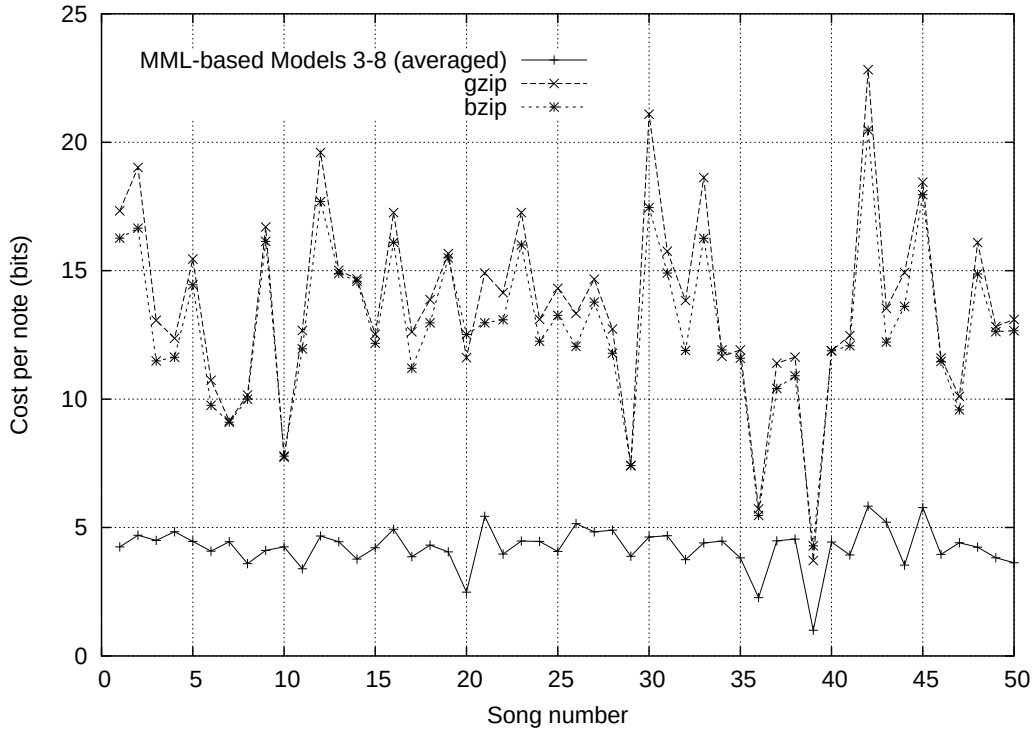


Figure 7: MML-based Model vs. Standard Compression Tools

tion, it was discovered that these melodies contain very few unique pitches and durations, and that they all contain highly repetitive bars of notes.

Pitch Models 1 and 2 feature higher transmission costs than Models 3-8 for the majority of test melodies. Hence, the extra complexity of stating pitch intervals in two parts (octave, interval within octave), as seen in Pitch Models 3-8, is justified. We claim that Pitch Models 3-8 are preferable to Pitch Models 1 and 2, and that the differences between Models 3-8 are negligible.

Figure 7 shows the average transmission costs of Pitch Models 3-8 (the preferable pitch models, as discussed above) versus standard GNU compression tools `gzip` and `bzip`, run with default parameter settings. As before, each pitch model was coupled with the Gaussian-based duration interval model and used to model each test melody. Lempel-Ziv (LZ) [14] based `gzip` performs slightly worse than `bzip`, which is based on the Burrows-Wheeler block sorting text compression algorithm [2] and Huffman coding. `gzip` gives an average transmission cost of 13.65 bits per note, which equates to an average compression rate of 2.98 : 1. `bzip` provides an average transmission cost of 12.80 bits per note and yields an average compression rate of 3.10 : 1. These transmission costs are reasonable given that a pitch-duration *pair* must be transmitted to specify each note.

Clearly though, the MML-based models are far more efficient than both GNU compression tools. With an average cost of only 4.29 bits per note, the MML models achieve an excellent average compression ratio of 9.31 : 1. Hence, the MML models are capturing more of the underlying structure of the melodies than the GNU compression tools. Whilst both `gzip` and `bzip` must learn that there is a restricted alphabet of symbols, the MML models were constructed with this knowledge a priori.

# 5 Limitations and Future Work

Perhaps the most obvious limitation of the research presented here is that only low-level sequential properties of melodies are modelled. Furthermore, many of the low-level musical attributes that may occur in melodies are not included in the models presented here. Attributes such as note accents, slurs, staccato, volume, etc., are to be included in future models. The discrete memoryless models detailed in Section 3 are zero-order Markov models; experiments with increased Markov order are also planned.

High-level structure such as the repetition of music bars and repetition of sequences of bars is not

currently modelled. Modelling higher-level structure is likely to increase the compressibility of melodies which feature recurring themes. Accounting for higher-level structure is not only likely to increase compressibility but will also assist in the generation of "musical" sounding melodies using inferred models.

# 6 Conclusion

This paper has described a statistical method for modelling melodies using the MML principle. New pitch and duration models have been developed and examined for goodness of fit to musical data. Pitch Models 3-8 performed better than Pitch Models 1 and 2 for the majority of test melodies. Using MML, we have shown that the added complexity of stating the octave range in which the interval falls is justified when modelling melodies.

All MML-based models consistently produced lower transmission bit-costs than standard GNU compression tools. This indicates that MML more effectively captured underlying melodic structure than traditional compression approaches. The inference of structure in melodies has far-reaching applications including, but not limited to, the inference of structure in bird and whale songs, amongst other animal calls. Importantly, the MML approach could also be applied to infer structure and style in text, and to compress text. Clearly, other data types as images and video may also benefit from MML analysis.

# Acknowledgements

The authors thank Dr. Peter Tischer for his insightful comments on drafts of this paper.

# References

[1] Columbia Law Library, Music Copyright Infringement Online Archive, http://library.law.columbia.edu/music_plagiarism/index2.html, Columbia University, USA, 2002.

[2] Burrows M., Wheeler D. J. A Block-sorting Lossless Data Compression Algorithm. Technical Report 124, Digital Equipment Corporation, Palo Alto, California, 1994.

[3] Conway J. H., Sloane N. J. *Sphere Packings, Lattices, and Groups.* Springer-Verlag, New York, U.S.A., 1993.

[4] Freeman L. C., Merriam A. P. Statistical Classification in Anthropology: An Application to Ethnomusicology. 58:464–472, 1956.

[5] Lomax A. *Folk Song Style and Structure (with the Cantometrics staff).* American Association for the Advancement of Science, Washington D.C., Transaction Books, New Brunswick, N. J., 1968.

[6] Needham S., Dowe D. L. Message Length as an Effective Ockham's Razor in Decision Tree Induction. In *Proc. 8th International Workshop on Artificial Intelligence and Statistics*, pages 253–260, 2001.

[7] Rissanen J. Modeling by Shortest Data Description. *Automatica*, 14:465–471, 1978.

[8] Saffran J. R., Johnson E. K., Aslin R. N., Newport E. L. Statistical Learning of Tone Sequences by Human Infants and Adults. *Cognition*, 70:27–52, 1999.

[9] Wallace C. S., Boulton D. M. An Information Measure for Classification. *The Computer Journal*, 11(2):185–194, 1968.

[10] Wallace C. S., Dowe D. L. Minimum Message Length and Kolmogorov Complexity. *The Computer Journal*, 42(4):270–283, 1999.

[11] Wallace C. S., Dowe D. L. MML Clustering of Multi-State, Poisson, von Mises Circular and Gaussian Distributions. *Statistics and Computing*, 10:73–83, January 2000.

[12] Wallace C. S., Freeman P. R. Estimation and Inference by Compact Coding. *Journal of the Royal Statistical Society (Series B)*, 49(3), 1987.

[13] Wallace C. S., Georgeff M. P. A General Objective for Inductive Inference. Technical Report 32, Department of Computer Science, Monash University, Clayton, Australia, March 1983.

[14] Ziv J., Lempel A. A Universal Algorithm for Sequential Data Compression. *IEEE Transactions On Information Theory*, IT-23(3):337–343, 1977.