

CLASSIFICATION OF MUSICAL METRE WITH AUTOCORRELATION AND DISCRIMINANT FUNCTIONS

Petri Toiviainen

Department of Music
University of Jyväskylä
Finland
ptoiviai@campus.jyu.fi

Tuomas Eerola

Department of Music
University of Jyväskylä
Finland
ptee@campus.jyu.fi

ABSTRACT

The performance of autocorrelation-based metre induction was tested with two large collections of folk melodies, consisting of approximately 13,000 melodies in MIDI file format, for which the correct metres were available. The analysis included a number of melodic accents assumed to contribute to metric structure. The performance was measured by the proportion of melodies whose metre was correctly classified by Multiple Discriminant Analysis. Overall, the method predicted notated metre with an accuracy of 75 % for classification into nine categories of metre. The most frequent confusions were made within the groups of duple and triple/compound metres, whereas confusions across these groups were significantly less frequent. In addition to note onset locations and note durations, Thomassen's melodic accent was found to be an important predictor of notated metre.

Keywords: Metre, classification, autocorrelation

1 INTRODUCTION

Most music is organized to contain temporal periodicities that evoke a percept of regularly occurring pulses, or beats. The period of the most salient pulse is typically within the range of 400 to 900 ms [1-3]. The perceived pulses are often hierarchically organized and consist of at least two simultaneous levels, whose periods have an integer ratio. This gives rise to a percept of regularly alternating strong and weak beats, a phenomenon referred to as metre [4,5]. In Western music, the ratio of the pulse lengths is usually limited to 1:2 (duple metre) and 1:3 (triple metre). Metre in which each beat has three subdivisions, such as 6/8 or 9/8, is referred to as compound metre.

A number of computational models have been developed for the extraction of the basic pulse from music. Modelling of metre perception has, however, received less attention. Large and Kolen [6] presented a model of metre perception based on resonating oscillators. Toivi-

ainen [7] presented a model of competing subharmonic oscillators for determining the metre (duple vs. triple) from an acoustical representation of music. Brown [8] proposed a method for determining the metre of musical scores by applying autocorrelation to a temporal function consisting of impulses at each tone onset whose heights are weighted by the respective tone durations. A shortcoming of Brown's study [8] is that it fails to provide any explicit criteria for the determination of metre from the autocorrelation function. Frieler [9] presents a model based on autocorrelation of gaussified onsets for the determination of metre from performed MIDI files. Pikrakis, Antonopoulos, and Theodoridis [10] present a method for the extraction of music metre and tempo from raw polyphonic audio recordings based on self-similarity analysis of mel-frequency cepstral coefficients. When tested with a corpus of 300 recordings, the method achieved a 95 % correct classification rate. Temperley and Sleator [11] present a preference-rule model of metre-finding. An overview of models of metrical structure is provided in [12].

Although there is evidence that the pitch information present in music may affect the perception of pulse and metre [13-15], most models of pulse and metre finding developed to date rely only on note onset times and durations. Dixon and Cambouropoulos [16], however, proposed a multi-agent model for beat tracking that makes use of pitch and amplitude information. They found that including this information when determining the salience of notes significantly improved the performance of their model. Vos, van Dijk, and Schomaker [17] applied autocorrelation to the determination of metre in predominantly isochronous music. They utilized a method similar to that proposed in [8], except for using the melodic intervals between subsequent notes to represent the accent of each note.

In a previous study [18], we applied discriminant function analysis to autocorrelation functions calculated from Brown's [8] impulse functions for classification of folk melodies into duple vs. triple/compound metre. Using two large folk song collections with a total of 12,368 melodies, we obtained a correct classification rate of 92 %. Furthermore, we examined whether the inclusion of different melodic accent types would improve the classification performance. By determining the components of the autocorrelation functions that were significant in the classification, we found that periodicity in note onset locations above the measure level was the most important cue for the determination of metre. Of the melodic accents included, Thomassen's [14] melodic accent provided the most reliable cues for

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2005 Queen Mary, University of London

the determination of metre. The inclusion of five different melodic accents led to a correct classification rate of 96 %.

The present study investigated the capability of the autocorrelation-based metre induction method to carry out a more detailed classification. More specifically, instead of mere classification as duple vs. triple, the dependent variable used in this experiment was the actual notated metre. In the analysis, special attention was paid to the pattern of confusion between metres.

2 AUTOCORRELATION AND METRE

Below, the method for constructing the autocorrelation function for metre induction is described. For the original description, see [8]. Let the melody consist of N notes with onset times $t_i, i = 1, 2, \dots, N$. Each note is associated with an accent value $a_i, i = 1, 2, \dots, N$; in [8], a_i equals the duration of the respective note. The onset impulse function f is a time series consisting of impulses of height a_i located at each note onset position:

$$f(n) = \sum_{i=1}^N a_i \delta_i(n), n = 0, 1, 2, \dots \quad (1)$$

where

$$\delta_i(n) = \begin{cases} 1, & n = [t_i / dt] \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where dt denotes the sampling interval and $[\]$ denotes rounding to the nearest integer.

Autocorrelation refers to the correlation of two copies of a time series that are temporally shifted with respect to each other. For a given amount of shift (or lag), a high value of autocorrelation suggests that the series contains a periodicity with length equalling the lag. In the present study, the autocorrelation function F was defined as

$$F(m) = \frac{\sum_n f(n)f(n-m)}{\sum_n f(n)^2} \quad (3)$$

where m denotes the lag in units of sampling interval; the denominator normalizes the function to $F(0)=1$ irrespective of the length of the sequence. Often, the lag corresponding to the maximum of the autocorrelation function provides an estimate of the metre. This is the case for the melody depicted in Figure 1.

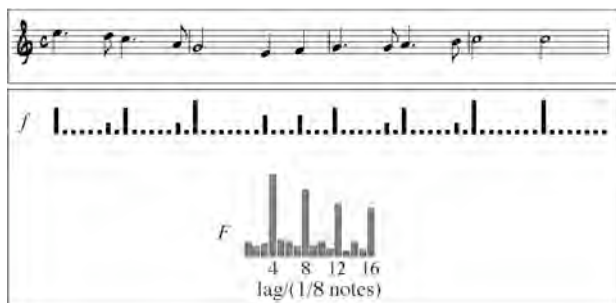


Fig. 1. Excerpt from a melody, its onset impulse function weighted by durational accents, f , and the corresponding autocorrelation function, F . The maximum of the autocorrelation function at the lag of 4/8 indicates duple metre.

Sometimes the temporal structure alone is not sufficient for deducing the metre. This holds, for example, for isochronous and temporally highly aperiodic melodies. In such cases, melodic structure may provide cues for the determination of metre. This is the case, for instance, with the melody depicted in Figure 2. With this isochronous melody, the autocorrelation function obtained from the duration-weighted onset impulse function fails to exhibit any peaks, thus making it impossible to determine the metre. Including information about pitch content in the onset impulse function leads, however, to an autocorrelation function with clearly discernible peaks.

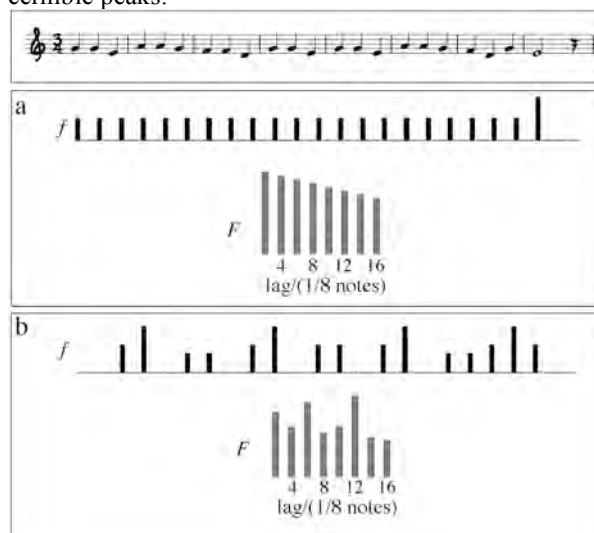


Fig. 2. Excerpt from an isochronous melody; **a)** onset impulse function weighted by durational accents, f , and the corresponding autocorrelation function, F , showing no discernible peaks. **b)** Onset impulse function weighted by interval size, f , and the corresponding autocorrelation function, F . The maximum of the autocorrelation function at the lag of 12/8 indicates triple or compound metre.

3 MATERIAL

The material consisted of monophonic folk melodies in MIDI file format taken from two collections: the Essen collection [19], consisting of mainly European folk melodies, and the Digital Archive of Finnish Folk Tunes [20], subsequently referred to as the Finnish collection. From each collection, melodies that consisted of a single notated metre were included. Moreover, for each collection only metres that contained more than 30 exemplars were included. Consequently, a total of 5,592 melodies in the Essen collection were used, representing nine different notated metres (2/4, 3/2, 3/4, 3/8, 4/1, 4/2, 4/4, 6/4, 6/8). From the Finnish collection, 7,351 melodies were used, representing nine different notated metres (2/4, 3/2, 3/4, 3/8, 4/4, 5/2, 5/4, 6/4, 6/8). For each collection, the number of melodies representing each notated metre is shown in Table 1.

4 METHOD

For each of the melodies in the two collections, we constructed a set of onset impulse functions weighted by various accent types (Eqs. 1 and 2). In each case the sampling interval was set to 1/16 note. The accents consisted of (1) durational accent (a_i equals tone duration); (2) Thomassen's melodic accent [14]; (3) interval size in semitones between previous and current tone (e.g. [17]); (4) pivotal accent ($a_i = 1$ if melody changes direction, $a_i = 0$ otherwise); and (5) gross contour accent ($a_i = 1$ for ascending interval, $a_i = -1$ for descending interval, $a_i = 0$ otherwise). Since the note onset times alone, without regard to any accent structure, provide information about metrical structure, we further included (6) constant accent ($a_i = 1$). The analysis was carried out using the MIDI Toolbox for Matlab [21].

For each melody, each of the onset impulse functions was subjected to autocorrelation. The components of the obtained autocorrelation functions corresponding to lags of 1, 2, ..., 16 eighth notes were included in the subsequent analyses. Figure 3 depicts the onset impulse functions and the respective autocorrelation functions constructed from a melodic excerpt using each of the accent types described above.

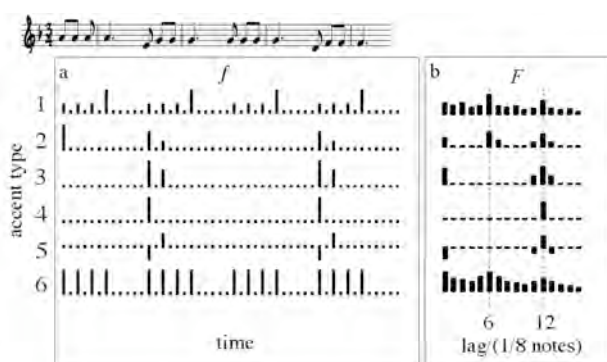


Fig. 3. a) Onset impulse functions constructed from a melodic excerpt using the six accent types described in the text; b) the respective autocorrelation functions. As can be seen, the melodic accents frequently fail to co-occur either with each other or with the durational accents. All the autocorrelation functions, however, have maxima at lags of either 6/8 or 12/8, indicating triple or compound metre.

The classification of metres was performed with Multiple Discriminant Analysis (MDA) [22], a simple yet efficient classification method widely used in various application areas. With n groups, the MDA produces $n-1$ discrimination functions, each of which is a linear combination of the independent variables. In the current classification task, the independent variables comprised the autocorrelation functions obtained using all the accent types and the dependent variable was the notated metre. In testing the classification performance, the leave one out cross-validation scheme [23] (i.e. k -fold cross-validation with $k=N$) was utilized. The perform-

ance was assessed by means of a confusion matrix. Furthermore, for both collections the precision and recall values as well as the F-score were calculated for each metre [24]. For a given metre, precision is defined as the number of melodies having the metre and being correctly classified, divided by the total number of melodies being classified as representing the metre. Similarly, for each metre, recall is defined as the number of melodies being notated in the metre and being correctly classified, divided by the total number of melodies being notated in the metre. The F-score is defined as the harmonic mean of precision and recall and is regarded as an overall measure of classification performance.

Overall, 83.2 % of the melodies from the Essen collection and 68.0 % of those from the Finnish collection were correctly classified. The notably low correct classification rate for the Finnish collection can be mainly attributed to the fact that a large proportion (43.4 %) of melodies representing 4/4 metre were classified as being 2/4 (see below).

To obtain a more detailed view of the classification performance, we calculated the confusion matrices for the both collections. Table 1 shows the precision, recall, and F-values for each metre as well as the most common confusions between metres.

Table 1. Classification performance for each collection and metre. R = recall; P = precision; F = F-score; the Errors column displays the two most common confusion and their prevalence.

| Metre (N) | R | P | F | Errors |
|------------------------------------|------|------|------|----------------------|
| <i>Essen Collection (N=5592)</i> | | | | |
| 2/4 (1285) | 0.88 | 0.86 | 0.87 | 4/4 (10%), 3/4 (2%) |
| 3/2 (100) | 0.65 | 0.92 | 0.76 | 4/4 (16%), 3/4 (11%) |
| 3/4 (1215) | 0.77 | 0.90 | 0.83 | 6/4 (9%), 4/4 (7%) |
| 3/8 (291) | 0.58 | 0.53 | 0.55 | 6/8 (31%), 2/4 (8%) |
| 4/1 (39) | 0.92 | 0.75 | 0.83 | 4/2 (5%), 3/2 (2%) |
| 4/2 (173) | 0.86 | 0.87 | 0.86 | 4/4 (6%), 4/1 (6%) |
| 4/4 (1598) | 0.91 | 0.85 | 0.88 | 2/4 (6%), 3/4 (2%) |
| 6/4 (110) | 0.73 | 0.43 | 0.54 | 3/4 (19%), 4/4 (8%) |
| 6/8 (781) | 0.82 | 0.87 | 0.84 | 3/8 (14%), 3/4 (2%) |
| <i>Finnish Collection (N=7351)</i> | | | | |
| 2/4 (3293) | 0.74 | 0.69 | 0.71 | 4/4 (22%), 3/4 (2%) |
| 3/2 (74) | 0.61 | 0.44 | 0.51 | 4/4 (20%), 2/4 (7%) |
| 3/4 (902) | 0.77 | 0.86 | 0.81 | 2/4 (11%), 6/4 (7%) |
| 3/8 (129) | 0.36 | 0.48 | 0.42 | 6/8 (43%), 3/4 (13%) |
| 4/4 (2205) | 0.55 | 0.60 | 0.57 | 2/4 (43%), 5/2 (1%) |
| 5/2 (39) | 0.67 | 0.38 | 0.49 | 4/4 (28%), 5/4 (3%) |
| 5/4 (413) | 0.91 | 0.95 | 0.93 | 2/4 (8%), 3/2 (1%) |
| 6/4 (78) | 0.49 | 0.32 | 0.39 | 4/4 (33%), 3/4 (12%) |
| 6/8 (218) | 0.61 | 0.68 | 0.65 | 3/8 (22%), 3/4 (7%) |

Table 1 reveals that, in terms of the F-score, the most accurately classified metres were 4/4 and 2/4 for the Essen collection and 5/4 and 3/4 for the Finnish collection. Similarly, the least accurately classified metres were 6/4 and 3/8 for both collections. For both collections, metres 2/4 and 4/4 displayed the highest mutual

confusion rate, followed by metres 3/4 and 6/4. A large proportion of these misclassifications can probably be attributed to the effect of tempo on the choice of notated metre (cf. [25]). Take, for instance, a melody that is played in a fast tempo (e.g., $MM > 160$) and notated in 6/8 metre. If the same melody is played in a much slower tempo (e.g., $MM < 70$), it could be notated in 3/8 metre. As tempo information was not available for either of the collections, the effect of tempo could not be assessed. Table 1 suggests that the most frequent confusions were made within the groups of duple and triple/compound metres, whereas confusions across these groups were less frequent. To investigate this, we calculated the proportions of confusions within and across these groups for both collections and both metre groups. These are shown in Table 2. As can be seen, the proportion of melodies misclassified across the metre groups is for both collections and both metre groups smaller than the proportion of melodies misclassified within the metre group.

Table 2. Proportion of melodies misclassified within and across the groups of duple and triple/compound metres.

| <i>Essen Collection (N = 5592)</i> | | | |
|------------------------------------|-------|------------------------|--------|
| | | <i>Predicted metre</i> | |
| <i>Notated metre</i> | | Duple | Triple |
| | Duple | | 0.083 |
| Triple | | 0.086 | 0.159 |

| <i>Finnish Collection (N = 6899)</i> | | | |
|--------------------------------------|-------|------------------------|--------|
| | | <i>Predicted metre</i> | |
| <i>Notated metre</i> | | Duple | Triple |
| | Duple | | 0.309 |
| Triple | | 0.133 | 0.178 |

Certain confusions imply more severe misattributions by the algorithm. For instance, 11.7 % of the melodies in the Essen collection notated in 3/4 metre were misclassified as representing binary metre (4/4 or 2/4), the corresponding figure for the Finnish collection being 12.6 %. In general, duple metres were less frequently misclassified as representing triple/compound metre as vice versa. This asymmetry may be due to the fact that the MDA attempts to maximize the total correct classification rate, as a result of which the most common metres receive the best classification rates. To investigate this, we performed for both collections a MDA with an equal number of melodies representing the most common duple and triple metres. For the Essen collection we used all the 1215 melodies notated in 3/4 metre and an equal number of melodies notated in 4/4 metre, randomly chosen. The leave-one-out classification yielded correct classification rates of 96.7% and 96.5% for the 3/4 and 4/4 metres, respectively. Similarly, for the Finnish collection we used all the 902 melodies notated in 3/4 metre and an equal number of melodies notated in

2/4 metre, again randomly chosen. This yielded correct classification rates of 95.5% and 95.1% for the 3/4 and 2/4 metres, respectively. There were thus no significant differences in the classification rates between the metres, which suggests that the asymmetry in classification rates can be attributed the differences in group sizes and the characteristics of the classification algorithm used.

To assess the relative importance of features (i.e., types of accent and lags) that contribute to the discrimination between metres, we examined the magnitudes of the standardised beta coefficients of the variables for each discrimination function. In particular, we took the mean of the absolute values of the beta weight across the discriminant functions to represent the relative importance of each feature. The first 48 most important features, ordered according to the respective maximal beta values, are shown in the Appendix. According to this result, the components of the autocorrelation function derived from the durational and constant accents were the most significant predictors of metre for both collections. The next most important predictor for both collections was Thomassen's melodic accent [14], followed by the interval size accent.

To further inspect the relationships between metres, we performed a hierarchical cluster analysis separately for both collections. To this end, we calculated the distance between each metre from the confusion matrix according to the formula

$$d_{ij} = 1 - \left(\frac{c_{ij} + c_{ji}}{c_{ii} + c_{jj}} \right), \quad (4)$$

where d_{ij} denotes the distance between metres i and j , and c_{ij} the number of cases where a melody in metre i has been classified as being in metre j . By definition, the larger the proportion of melodies confused between metres, $c_{ij} + c_{ji}$, to the number of melodies correctly classified for both metres, $c_{ii} + c_{jj}$, the smaller the distance d_{ij} between the metres.

Figure 4 displays the dendrograms obtained from the clustering algorithms. In the dendrograms, the stage at which given metres cluster together reflects the algorithm's rate of confusion between the metres. For both collections, the metres to first cluster together are 3/8 and 6/8. For the Essen collection, this is followed by the clustering of the metres 3/4 and 6/4 as well as 2/4 and 4/4, in this order. Also for the Finnish collection these pairs of metres cluster next, albeit in reverse order, that is, the clustering of 2/4 and 4/4 precedes that of 3/4 and 6/4. A further similar feature between the two dendrograms is that the last clustering occurs between the cluster formed by the metres 3/8 and 6/8 and the cluster formed by all the other metres. This suggests that, in terms of the autocorrelation functions, metres 3/8 and 6/8 are most distinct from the other metres.

One peculiar feature of the dendrogram for the Essen collection is the relatively late clustering of metres 4/1 and 4/2 with metres 2/4 and 4/4. In particular, the former two metres cluster with metre 3/2 before clustering with the latter two metres. A potential explanation for this is the difference in the average note durations be-

tween the metres. More specifically, the average note durations for metres 4/1, 4/2, and 3/2 exceed those of metres 2/4 and 4/4 by a factor of two.

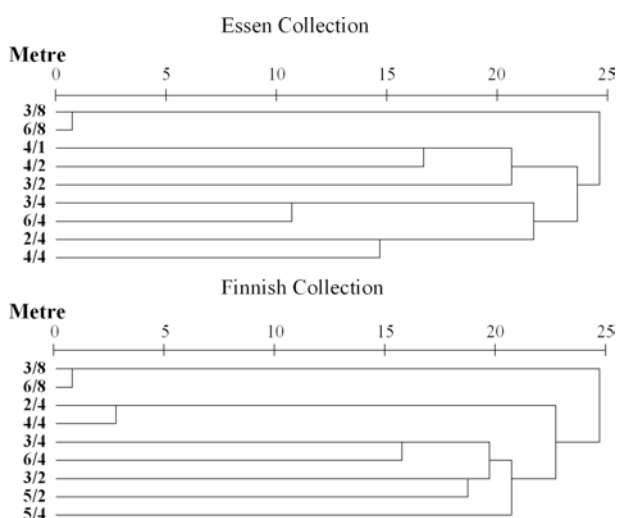


Fig. 4. Dendrograms obtained from the confusion matrix using the similarity measure of Eq. 4. The leftmost column displays the average note durations in quarter notes for the melodies representing each metre.

5 CONCLUSIONS

We studied the classification performance of the autocorrelation-based metre induction model, originally introduced in [8]. Using Multiple Discriminant Analysis, we provided an explicit method for the classification. Furthermore, we included a set of melodic accents that in a previous study [18] were found to improve the classification performance.

The overall correct classification rate was approximately 75%. While this rate appears to be relatively low compared to what has been obtained in some other similar classification studies [e.g., 10], it must be noted that the material used in the present study consists of monophonic melodies, which by their nature provide fewer cues for metre than polyphonic material. We would expect that human subjects, when presented with the material used in this study, would not significantly exceed the correct classification rate achieved by the model. This hypothesis should, however, be verified with listening experiments.

The most frequent confusions were made within the groups of duple and triple/compound metres, whereas confusions across these groups were significantly less frequent. For both collections, metres 2/4 and 4/4 displayed the highest mutual confusion rate, followed by metres 3/4 and 6/4. A large proportion of these misclassifications can probably be attributed to inherent disambiguity between certain pairs of metre as well as the effect of tempo on the choice of notated metre.

A finding that calls for further study was the significant difference between the correct classification rates for melodies in duple and triple/compound metre. More

specifically, melodies in duple metre were more often correctly classified than melodies in triple/compound metre. When the classification was performed with an equal number of melodies representing duple and triple/compound metres, this asymmetry was however absent, suggesting that it was originally due to the weighting of the classification by the frequency of occurrence of metres.

Investigation of the standardised beta coefficients of the discriminant functions revealed that the components of the autocorrelation functions derived using durational and constant accents were the most significant predictors of metre. This suggests that, in conformance with the general view, the most important features in the prediction of metre were based on note onset locations and note durations. Of the melodic accents included in the study, Thomassen's accent was found to be the next most important predictor, followed by the interval size accent. This result conforms to findings in a previous study by the present authors [18].

An apparent limitation of the method presented in this paper is its inability to deal with melodies that contain changes of metre. For a melody that, say, starts in 2/4 metre and changes to 3/4 metre, the algorithm gives unpredictable results. This is due to the fact that the algorithm considers the melody as a whole. The limitation may be overcome by applying a windowed analysis.

The present study utilized melodies that were represented in symbolic, temporally quantized form. The choice of stimuli was mainly based on the availability of correct (notated) metres for the melodies in the collections. In principle the method could, however, be applied to performed music in acoustical form as well, at least with a monophonic input. This would require algorithms for onset detection [26], pitch estimation [27, 28], beat tracking [6, 29-31], and quantization [32].

Acknowledgement

This work was supported by the Academy of Finland (grant No. 102253).

REFERENCES

- [1] Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D. (Ed.), *Psychology of music* (pp. 149-180). New York: Academic Press.
- [2] Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11, 409-464.
- [3] van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28, 43-66.
- [4] Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
- [5] Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D. (Ed.), *Psychology of music* (pp. 149-180). New York: Academic Press.

- [6] Large, E. W. & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, 6(1), 177-208.
- [7] Toiviainen, P. (1997). Modelling the perception of metre with competing subharmonic oscillators. In A. Gabrielsson (Ed.), *Proceedings of the Third Triennial ESCOM Conference*. Uppsala: Uppsala University, 511-516.
- [8] Brown, J. C. (1993). Determination of meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94, 1953-1957.
- [9] Frieler, K. (2004). Beat extraction using gaussified onsets. In *Proceedings of the 5th International Conference on Music Information Retrieval - ISMIR 2004*.
- [10] Pikrakis, A., Antonopoulos, I., & Theodoridis, S. (2004). Music meter and tempo tracking from raw polyphonic audio. In *Proceedings of 5th International Conference on Music Information Retrieval - ISMIR 2004*.
- [11] Temperley, D. & Sleator, D. (1999). Modeling Meter and harmony: a preference rule approach. *Computer Music Journal*, 15(1), 10-27.
- [12] Temperley, D. (2004). An evaluation system for metrical models. *Computer Music Journal*, 28(3), 28-44.
- [13] Dawe, L. A., Platt, J. R., & Racine, R. J. (1993). Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Perception and Psychophysics*, 54, 794-807.
- [14] Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America*, 71, 1596-1605.
- [15] Hannon, E. Snyder, J. Eerola, T. & Krumhansl, C. L. (2004). The Role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 956-974.
- [16] Dixon, S., & Cambouropoulos, E. (2000). Beat tracking with musical knowledge. In *ECAI 2000: Proceedings of the 14th European Conference on Artificial Intelligence* (626-630). IOS Press.
- [17] Vos, P. G., van Dijk, A., & Schomaker, L. (1994). Melodic cues for metre. *Perception*, 23, 965-976.
- [18] Toiviainen, P. & Eerola, T. (2004). The role of accent periodicities in meter induction: a classification study. In *Proceedings of the 8th ICMPC*. Adelaide: Causal Productions, 422-425.
- [19] Schaffrath, H. (1995). *The Essen Folksong Collection in Kern Format*. [computer database]. D. Huron (ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.
- [20] Eerola, T., & Toiviainen, P. (2004). *Digital archive of Finnish folk tunes*. University of Jyväskylä: Jyväskylä, Finland. Available at: <http://www.jyu.fi/musica/sks/>
- [21] Eerola, T. & Toiviainen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyväskylä: Jyväskylä, Finland. Available at: <http://www.jyu.fi/musica/miditoolbox>.
- [22] Huberty, C. J. (1994). *Applied Discriminant Analysis*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics Section. John Wiley & Sons.
- [23] Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics* 10, 1-11.
- [24] Salton, G. & McGill, M. (1983). *Introduction to Modern Information Retrieval*. McGraw Hill, New York.
- [25] London, J. (2002). Cognitive constraints on metric systems: some observations and hypotheses. *Music Perception*, 19, 529-550.
- [26] Klapuri, A. (1999). Sound Onset Detection by Applying Psychoacoustic Knowledge. In *Proc. IEEE Int. Conf. Acoustics Speech and Sig. Proc. (ICASSP)*, pp. 3089-3092, Phoenix AR.
- [27] Brown, J. C., & Puckette. M. S. (1994). A high resolution fundamental frequency determination based on phase changes of the Fourier transform. *Journal of the Acoustical Society of America*, 94, 662-667.
- [28] Klapuri, A. (2003). Multiple fundamental frequency estimation by harmonicity and spectral smoothness. *IEEE Trans. Speech and Audio Processing*, 11, 804-816.
- [29] Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30, 39-58.
- [30] Toiviainen, P. (1998). An interactive MIDI accompanist. *Computer Music Journal*, 22, 63-75.
- [31] Toiviainen, P. (2001). Real-time recognition of improvisations with adaptive oscillators and a recursive Bayesian classifier. *Journal of New Music Research*, 30, 137-1.
- [32] Desain, P., & Honing, H. (1989). Quantization of musical time: a connectionist approach. *Computer Music Journal*, 13(3), 56-66.

APPENDIX. Most important features in the classification and their mean standardized canonical discriminant function coefficients (β). Abbreviations: *dur* = durational accent (accent 1); *mel* = Thomassen's melodic accent (accent 2); *int* = interval size accent (accent 3); *piv* = pivotal accent (accent 4); *con* = gross contour accent (accent 5); *non* = constant accent (accent 6). Numbers in the feature columns refer to lag in units of one eighth note.

| <i>Essen collection</i> | | | <i>Finnish collection</i> | | |
|-------------------------|---------|---------|---------------------------|---------|---------|
| Rank | Feature | β | Rank | Feature | β |
| 1 | dur11 | 0.793 | 1 | dur5 | 1.597 |
| 2 | dur7 | 0.692 | 2 | dur9 | 1.186 |
| 3 | non4 | 0.664 | 3 | dur3 | 1.016 |
| 4 | non7 | 0.652 | 4 | dur1 | 0.883 |
| 5 | non5 | 0.588 | 5 | dur11 | 0.868 |
| 6 | dur3 | 0.568 | 6 | non7 | 0.814 |
| 7 | dur5 | 0.553 | 7 | dur7 | 0.802 |
| 8 | non11 | 0.501 | 8 | dur10 | 0.801 |
| 9 | dur15 | 0.500 | 9 | dur15 | 0.800 |
| 10 | non2 | 0.483 | 10 | non5 | 0.775 |
| 11 | dur12 | 0.476 | 11 | non13 | 0.688 |
| 12 | dur10 | 0.473 | 12 | non10 | 0.688 |
| 13 | non15 | 0.471 | 13 | non9 | 0.684 |
| 14 | dur6 | 0.468 | 14 | dur13 | 0.664 |
| 15 | non8 | 0.464 | 15 | non3 | 0.663 |
| 16 | non3 | 0.441 | 16 | non11 | 0.628 |
| 17 | non16 | 0.394 | 17 | non1 | 0.625 |
| 18 | dur1 | 0.391 | 18 | non6 | 0.549 |
| 19 | non13 | 0.371 | 19 | non15 | 0.538 |
| 20 | dur16 | 0.371 | 20 | dur2 | 0.531 |
| 21 | dur9 | 0.364 | 21 | non12 | 0.509 |
| 22 | dur14 | 0.353 | 22 | non14 | 0.470 |
| 23 | non6 | 0.347 | 23 | non16 | 0.425 |
| 24 | dur13 | 0.342 | 24 | dur14 | 0.410 |
| 25 | non12 | 0.334 | 25 | dur4 | 0.376 |
| 26 | non14 | 0.332 | 26 | dur6 | 0.368 |
| 27 | non1 | 0.331 | 27 | dur16 | 0.360 |
| 28 | non10 | 0.319 | 28 | non2 | 0.352 |
| 29 | dur4 | 0.303 | 29 | dur12 | 0.315 |
| 30 | non9 | 0.293 | 30 | non8 | 0.301 |
| 31 | dur2 | 0.273 | 31 | dur8 | 0.242 |
| 32 | dur8 | 0.255 | 32 | non4 | 0.225 |
| 33 | mel8 | 0.175 | 33 | mel11 | 0.196 |
| 34 | mel4 | 0.129 | 34 | mel15 | 0.182 |
| 35 | mel9 | 0.116 | 35 | mel3 | 0.152 |
| 36 | mel12 | 0.111 | 36 | mel6 | 0.135 |
| 37 | mel7 | 0.107 | 37 | int4 | 0.124 |
| 38 | mel6 | 0.106 | 38 | mel10 | 0.123 |
| 39 | mel3 | 0.097 | 39 | int5 | 0.122 |
| 40 | mel15 | 0.086 | 40 | int12 | 0.122 |
| 41 | mel16 | 0.083 | 41 | int1 | 0.114 |
| 42 | mel11 | 0.075 | 42 | mel5 | 0.113 |
| 43 | con2 | 0.072 | 43 | int6 | 0.106 |
| 44 | mel5 | 0.069 | 44 | mel2 | 0.105 |
| 45 | piv12 | 0.069 | 45 | con12 | 0.099 |
| 46 | mel2 | 0.066 | 46 | piv1 | 0.095 |
| 47 | piv3 | 0.065 | 47 | int10 | 0.095 |
| 48 | int3 | 0.064 | 48 | mel13 | 0.092 |