

# EXTRACTING THE PERCEPTUAL TEMPO FROM MUSIC

*Martin F. McKinney*  
Philips Research Laboratories  
Eindhoven, The Netherlands

*Dirk Moelants*  
IPEM-Department of Musicology  
Ghent University, Belgium

## ABSTRACT

The study presented here outlines a procedure for measuring and quantitatively representing the perceptual tempo of a musical excerpt. We also present a method for applying such measures of perceptual tempo to the design of automatic tempo-trackers in order to more accurately represent the perceived beat in music.

Keywords: Tempo, Perception, Beat-tracking

## 1. INTRODUCTION

Tempo is a basic element and useful descriptive parameter of music and has been the focus of many systems for automatic music information retrieval, i.e., automatic tempo trackers [9]. When describing musical tempo, it is often useful to make a distinction between *notated* tempo and *perceptual* tempo. Notated and perceptual tempo can differ in that, for a given excerpt of music, there is only a single notated tempo, while listeners unfamiliar with the score can *perceive* the tempo to exist at different metrical levels [6]. For some pieces of music, the perceptual tempo is quite ambiguous, while for others it is not. It is often desirable to have a representation of perceptual tempo rather than notated tempo, especially in situations where the notated tempo of an audio track is unknown or unavailable.

A common problem with systems for automatic tempo extraction is that they do not distinguish between notated and perceptual tempo and, as a result, cannot reliably represent either form of tempo. A further consequence of this shortcoming is that meaningful performance evaluation of such systems is difficult because the form of the output is poorly defined.

This study provides a method for measuring and characterizing the perceptual tempo of musical excerpts and then applying the characterization to the development and testing of automatic tempo extractors.

Previous studies on the perception of pulse have shown that listeners tend to prefer tempi near a “resonance” of  $\sim 120$  beats per minute [3, 11, 7]. When subjects were

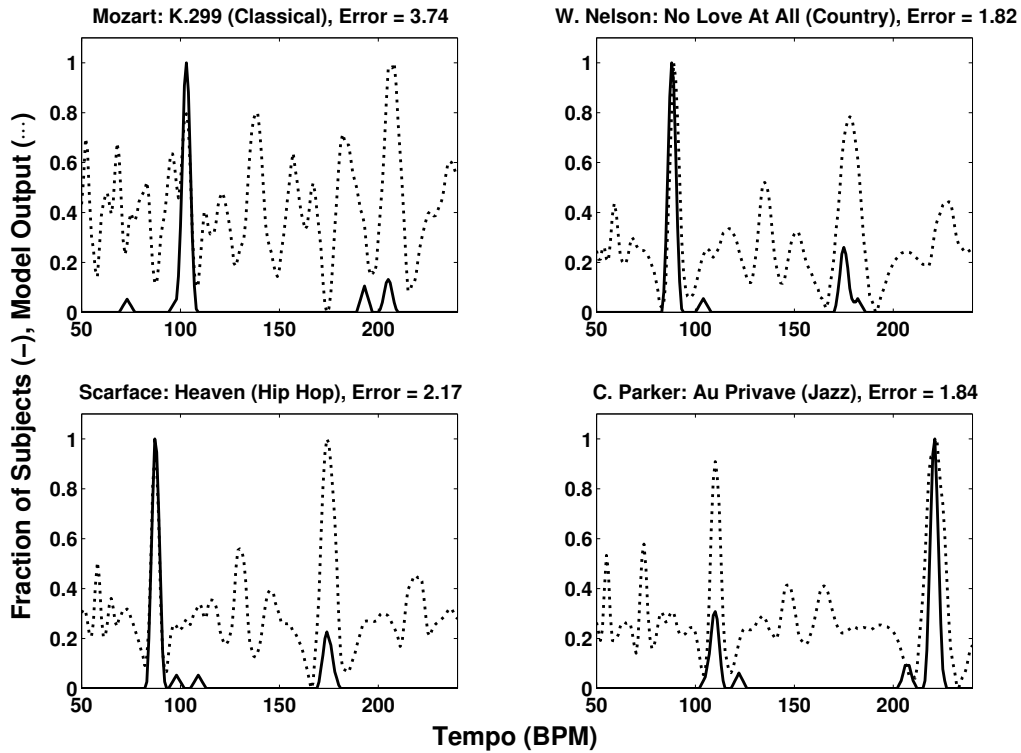
asked to tap to the beat in studies using artificial tone sequences and musical excerpts as stimuli, they would preferentially tap at metrical levels whose tempi were in this resonant range. We have shown in a similar study that, for individual musical excerpts, a resonance model can predict the distribution of subjects’ tapped tempi for some but not all excerpts [6]. Several factors, including various types of rhythmic accents (e.g., dynamic and durational), are likely to cause the distribution of perceptual tempi for some excerpts to deviate from a simple resonance representation [8]. A proper system for tempo extraction should accurately represent the perceptual tempo of music in all cases, including those that are not easily represented by a simple resonant model and those in which the perceptual tempo is ambiguous.

The typical structure of a system for tempo extraction can be divided into two stages: 1) a stage that generates a representation of temporal dynamics either by extracting it from audio (e.g., taking the derivative of the signal energy in a number of frequency bands [9]) or deriving it from a symbolic representation such as MIDI; and 2) a secondary stage that tabulates periodic regularities in the driving signal produced in the first stage, for example, through the use of resonator filter banks [9, 5], multi-agent methods [4], or probabilistic models [1]. It is common that a single tempo value or a list of candidate values is generated from these tabulations of periodicities to represent the tempo of a given piece of music. However, a more intermediate representation, such as beat histograms (see [10]) can be a more valuable description of the tempo when trying to relate it to the actual perceived tempo. Here we show how such representations can be used in conjunction with perceptual data to tune systems for tempo extraction so that they more accurately represent perceptual tempo.

## 2. METHOD

We performed an experiment in which listeners were asked to tap to the beat of 24 10-second musical excerpts covering a wide range of musical styles (see Appendix). We derived a measure of perceived tempo from the tapping times using linear regression and generated histograms of all subjects’ perceived tempi for each excerpt. These histograms of perceived tempo served as the “group response” for each excerpt and were taken to represent the overall perceived tempo for a particular excerpt.

Analogues of the perceived-tempo histograms were automatically generated from the audio waveforms of the ex-



**Figure 1.** Perceived-tempo histograms and histograms calculated using the RFB method. Each plot shows, for a single musical excerpt, the perceived-tempo histogram calculated from subjects’ tapping responses (solid line) and a tempo histogram calculated from the audio signal using the RFB method (dotted line). The error between the histograms (Eq. 1) is shown for each plot.

cerpts using three forms of common beat extractors [9, 2]: (1) a multi-band resonator filterbank after Scheirer [9]; (2) an envelope autocorrelation method; and (3) an inter-onset interval (IOI) histogram method. For each method, an internal representation of “periodicity energy” as a function of tempo was taken to be an analogous representation of perceived tempo and quantitatively compared with the perceived-tempo histograms.

### 2.1. Resonator filterbank method

The resonator filterbank (RFB) method of beat extraction was taken from [9] and consists of two stages: 1) a set of driving signals are generated from the derivative of the temporal envelope in six frequency bands of the audio signal; 2) each driving signal is filtered by a bank of resonators and then the output of like-frequency resonators are summed across frequency bands (See [9] for details).

The output of the extractor’s resonator filterbanks were summed and then plotted against resonant frequency to obtain a histogram in the same format as the perceived-tempo histogram. Direct comparisons of the extractor-generated histograms to the perceived-tempo histograms were made in order to evaluate the performance of the extractor (see Results).

### 2.2. Autocorrelation method

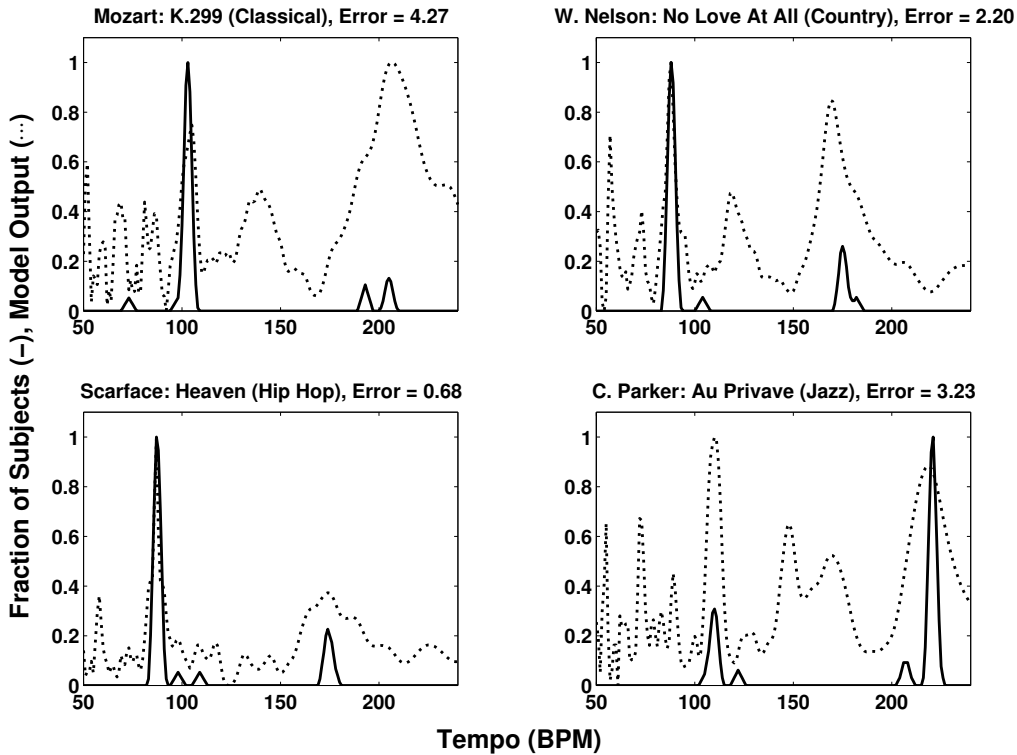
For the autocorrelation (AC) method of beat extraction we used the same driving signal as in the RFB method (the derivative of the temporal envelope in six frequency bands). However, instead driving a bank of resonators, we simply took the autocorrelation of the driving signal in each band and then summed across bands. The autocorrelation was normalized by the inverse of the lag in order to eliminate edge effects due a relatively short signal (10 seconds). The lag was translated to beats-per-minute to allow comparison to the perceived-tempo histograms.

### 2.3. IOI Histogram method

In the final method of beat extraction, onsets of auditory events were estimated by thresholding the driving function of the RFB method. Onsets were marked at every point the driving function crossed the threshold in the positive direction. Threshold values were empirically chosen. Intervals between onsets were then calculated and all-order<sup>1</sup> interval histograms were generated. IOI was then translated to beats-per-minute to allow comparison of the IOI histograms to the perceived-tempo histograms.

All histograms were normalized to range from 0 to 1. The error,  $E$ , between a perceived-tempo histogram,  $H_p$ ,

<sup>1</sup> All-order interval histograms include intervals between non-consecutive onsets as well as those between consecutive onsets.



**Figure 2.** Perceived-tempo histograms and histograms calculated using the AC method. Same format as Fig. 1.

and the tempo histogram extracted automatically from the audio,  $H_e$ , was calculated as follows:

$$E = \sqrt{\sum_{BPM=50}^{250} (H_p - H_e)^2 \cdot \max(H_p, H_e)^2} \quad (1)$$

where both  $H_p$  and  $H_e$  are functions of tempo ( $BPM$ ). The error is essentially a weighted mean-square-difference between the histograms over the range of tempi from 50 to 250 BPM. The weighting term,  $\max(H_p, H_e)^2$ , gives more weight to those parts of the histograms with large values, e.g., peaks in the histogram.

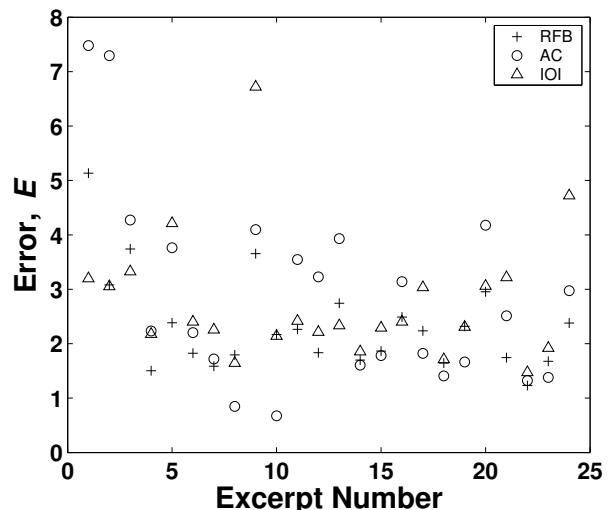
### 3. RESULTS

Results show that a reasonable fit to some of the perceived-tempo histograms can be made. Figure 1 shows four measured histograms along with tempo histograms calculated using the RFB method for the same musical excerpts. Figure 2 shows the same for the Autocorrelation method. Although none of the panels show a perfect match between the two types of histograms, many of them show that they share similar characteristics, for example the place and strength-order of peaks. The peaks of the histograms are important because they reflect tempi that are perceptually salient (in the perceptual-tempo histograms) and tempi at which there is high “periodicity energy” (in the calculated histograms).

The best fit between histograms is shown in the bottom-left panel of Fig. 2, where Autocorrelation method pre-

dicts the perceived-tempo histogram for an excerpt from *Heaven*.

The measure of error ( $E$ ) indicates goodness of fit between the two types of histograms and it can be seen from the plots that an error of around 2 or less indicates that the two primary peaks in both histograms occur at the same place and have the same strength-order (right plots of Fig. 1; upper-right and lower-left plots of Fig. 2).



**Figure 3.** Error between perceived-tempo and calculated tempo histograms. The error (calculated with Eq. 1) is shown for each tempo calculation method (RFB: Resonant Filter Bank; AC: Autocorrelation; IOI: Inter-onset Interval) as a function of excerpt number.

Individual histograms generated by the IOI method are not shown, however a summary of the errors for all tempo extraction methods is shown in Fig. 3. On average, the RFB method provides the lowest error (2.33) followed by the IOI method (2.75) and the AC method (2.88). There is, however, quite a bit of variation and for some excerpts (8 and 10), the AC method performs the best.

Analysis of excerpts whose tempo distributions are not well fit by the tempo extractor can provide useful cues to the aspects of the signal relevant to the perceived tempo. We are currently analyzing those excerpts for which tempo is not easily predicted in an effort to find characteristics that may be useful for next-generation tempo trackers.

#### 4. CONCLUSIONS

Distinguishing between notated and perceptual tempo has advantages not only for cognitive models of tempo perception, but also for the development of systems for automatic tempo extraction. The histogram representation of perceived tempo serves well as a basis for the tuning and quantitative assessment of automatic tempo extractors. Using the techniques presented here, we have shown that we can generate accurate representations of perceived tempo, which is an important component of systems for musical information retrieval.

#### 5. APPENDIX

Excerpt Number	Composer/Performer	Title	Genre
1	Fibich	Poeme	Classical
2	Mendelssohn	Spring Song	Classical
3	Mozart	K.299	Classical
4	Cumberland Highlanders	Cumberland Mtn. Home	Country
5	The Soggy Bottom Boys	I Am a Man of Const. Sorrow	Country
6	W. Nelson & W. Jennings	No Love At All	Country
7	Blue Sonix	Come On	Dance
8	The Youngsters	Spanish Harlem	Dance
9	Large Prof.	Kool	Hip Hop
10	Scarface	Heaven	Hip Hop
11	U God	Bizarre	Hip Hop
12	C. Parker	Au Privave	Jazz
13	F. Morgan	Mood Indigo	Jazz
14	Miles Davis	So What	Jazz
15	F. Villalona	Musica Latina	Latin
16	M. Bethania & Gal Costa	Sonho Meu	Latin
17	Bob Marley	Could You Be Loved	Reggae
18	Bob Marley	Jammin'	Reggae
19	Alannah Miles	Just One Kiss	Rock
20	Jimi Hendrix	Hey Joe	Rock
21	Lenny Kravitz	Come On And Love Me	Rock
22	Bill Withers	Grandma's Hands	Soul
23	C. Mayfield	Think	Soul
24	D. Hathaway	What's Going On?	Soul

**Table 1.** Sources of musical excerpts used in this study.

#### 6. REFERENCES

- [1] A.T. Cemgil, B. Kappen, P. Desain, and H. Honing. On tempo tracking: Tempogram representation and kalman filtering. *Journal of New Music Research*, 29(4):259–274, 2000.
- [2] S. Dixon, E. Pampalk, and G. Widmer. Classification of dance music by periodicity patterns. In *Proceedings of the 4th International Conference on Music Information Retrieval*, Baltimore, MD, October 2003. Johns Hopkins University.
- [3] P. Fraisse. Rhythm and tempo. In D. Deutsch, editor, *The Psychology of Music*, pages 149–180. Academic Press, New York, 1982.
- [4] M. Goto and Y. Muraoka. Music understanding at the beat level: Real-time beat tracking for audio signals. In D.F. Rosenthal and H.G.Okuno, editors, *Computational Auditory Scene Analysis*, pages 157–176. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1998.
- [5] Edward W. Large and Caroline Palmer. Perceiving temporal regularity in music. *Cognitive Science*, 26:1–37, 2002.
- [6] Martin F. McKinney and Dirk Moelants. Deviations from the resonance theory of tempo induction. In R. Parncutt, A. Kessler, and F. Zimmer, editors, *Conference on Interdisciplinary Musicology*, Graz, Austria, 2004. Dept. of Musicology, University of Graz.
- [7] Dirk Moelants. Preferred tempo reconsidered. In C. Stevens, D. Burnham, G. MacPherson, E. Schubert, and J. Renwick, editors, *Proceedings of the International Conference on Music Perception and Cognition*, Adelaide, 2002. Causal Productions.
- [8] Richard Parncutt. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11(4):409–464, 1994.
- [9] E. D. Scheirer. Tempo and beat analysis of acoustical musical signals. *Journal of the Acoustical Society of America*, 103:588–601, 1998.
- [10] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002.
- [11] Leon van Noorden and Dirk Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66, 1999.