In H. Lappalainen (Ed.), Proceedings of the VII International Symposium on Systematic and Comparative Musicology,
III International Conference on Cognitive Musicology, August, 16–19, 2001. Jyväskylä, Finland, pp. 18–26.

Human Preferences for Tempo Smoothness

Emilios Cambouropoulos, Simon Dixon, Werner Goebl and Gerhard Widmer
Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010, Vienna, Austria
{emilios,simon,wernerg,gerhard}@ai.univie.ac.at

Abstract

In this study we investigate the relationship between beat and musical performance. It is hypothesised that listeners prefer beat sequences that are smoother than beat tracks that are fully aligned with the actual onsets of performed notes. In order to examine this hypothesis, an experiment was designed whereby six different smoothed beat tracks generated are rated by subjects in relation to how well they correspond to a number of performed piano excerpts. It is shown that there is a preference of listeners for beat sequences that are slightly smoother than the onset times of the corresponding musical notes. This outcome was strongly supported by the results obtained from the group of trained musicians whereas it seems to have no bearing for the group of non-musicians.

1 Introduction

Contemporary theories of musical rhythm (Cooper and Meyer 1960; Yeston 1976; Lerdahl and Jackendoff 1983) assume two (partially or fully) independent components: a regular periodic structure of beats and the structure of musical events (primarily in terms of musical accents). The periodic temporal grid is 'fitted' onto the musical structure in a way that the alignment of the two structures is optimal. The relationship between the two is 'dialectic' in the sense that quasi-periodical characteristics of the musical material (patterns of accents, patterns of temporal intervals, pitch patterns etc) induce perceived temporal periodicities while, at the same time, established periodic metrical structures influence the way musical structure is perceived and even performed (see Clarke 1985).

Computational models of beat tracking attempt to determine an appropriate sequence of beats for a given musical piece, in other words, the *best fit* between a regular sequence of beats and a musical structure. Many beat-tracking models attempt to find the beat for a sequence of onsets (Longuet-Higgins and Lee 1982; Povel and Essens 1985; Desain and Honing 1992, Cegmil et al. 2000; Rosenthal 1992; Large et al. 1994, 1999) whereas some more recent attempts take into account elementary aspects of musical salience/accent (Toiviainen and Snyder 2000; Dixon and Cambouropoulos 2000; see also Parncutt 1994). Earlier work took into account only quantised representations of musical scores; modern beat tracking models are usually applied to real *performed* musical data that contain a wide range of expressive timing micro-deviations; in this paper this general case of beat tracking is considered.

An assumption made in the above models is that a preferred beat track should contain as few 'empty' positions as possible, i.e. beats on which no note is played as in cases of syncopation or rests. A related underlying assumption is that musical events may appear only *on* or *off* the beat. In this study we want to introduce a third *just-off-the-beat* option, namely that a musical event may both *correspond* to a beat but at the same time *not coincide* with the beat. This is important as it allows musical events to be said to come *early* or *late* in relation to the beat. Such an event is associated with a specific beat but the two are not fully synchronised.

The proposed hypothesis of *just-off-the-beat* notes affords beat structure a more 'rigid' and independent existence than is usually assumed. A metrical grid is not considered as a flexible abstract structure that can be stretched within large tolerance windows until a best fit to the actual performed music is achieved but a rather more robust psychological construct that is mapped to musical structure whilst maintaining a certain amount of autonomy.

It is herein suggested that the limits of fitting a beat track to a particular performance can be determined in relation to the concept of *tempo smoothness*. Listeners are very sensitive to deviations

that occur in isochronous sequences of sounds (for instance, relative JND constant is 2.5% of tone interonset intervals for sequences with intervals longer than 240ms – Friberg and Sundberg 1995). Despite the fact that this sensitivity decreases for complex real music, it is hypothesised that listeners still prefer smoother sequences of beats and that they are prepared to 'abandon' full alignment of a beat track to the actual event onsets if this results in a smoother beat flow.

The above hypothesis of beat smoothness has been examined in this study with the design of a preliminary perceptual experiment. For each of three short excerpts of piano music (from Mozart sonatas performed by a professional pianist) six different beat tracks with different degrees of smoothness have been generated and added to the music (according to a simple smoothing function – see next section). Listeners are then asked to rate the 'goodness' of each beat track regarding how well it fits – in a musical sense – with the actual piano performance. The preliminary results show that there is a preference (especially among musicians) for smoothed beat tracks.

The study of tempo smoothing is important as it provides insights into how a better beat tracking system can be developed. It also gives a more elaborate formal definition of beat and tempo that can be useful in other domains of musical research (e.g. in studies of musical expression, additional expressional attributes can be attached to notes in terms of being early or delayed as regards to the local tempo).

2 Tempo Smoothing

Real-time beat prediction implicitly performs some kind of smoothing, especially for ritardandi, as a beat tracker has to commit itself to a solution before 'seeing' any of the forthcoming events - it can't wait indefinitely before making a decision. In the example of Figure 1, an online beat tracker will either predict early beats for the fourth onset in both onset sequences or predict on-the-onset beats for the fourth onset in both sequences – the beat tracking solution given in the example is not possible unless *a posteriori* beat correction is enabled. It is herein suggested that a certain amount of beat correction that depends on the forthcoming musical context is important for a more sophisticated alignment of a beat track to the actual musical structure.

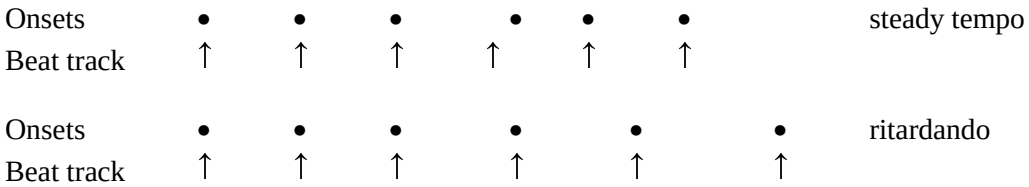


Figure 1 In the first onset sequence the fourth onset is *just-off-the-beat* (delayed) whereas in the second sequence it is *on* the beat. The two sequences are exactly the same up to the fourth note onset; the difference in the positioning of the two beats on the fourth note is possible only if *a posteriori* judgements of beat tracking are allowed.

Some might object to the above suggestion by stating that human beat tracking is always a real-time process. This is in some sense true, however, it should be mentioned that previous knowledge of a musical style or piece or even a specific performance of a piece allows better time synchronisation and beat prediction. In a sense tapping along to a certain piece for a second or third time enables a listener to use previously acquired knowledge about the piece and the performance for making more accurate beat predictions.

The aim of the current study is to determine the best fit between a beat sequence and given musical performance. As there were no real-time restrictions, a two-sided smoothing function (i.e. taking into account previous and forthcoming beat times) was applied to the performance data in order to derive a number of 'smoothed' beat tracks.

Starting with the beat positions that coincide with the performed onsets of events in the musical segments (beat track version *s0*), the simple smoothing function (below) is used for generating a number of smoothed beat track versions (see section 3.1.1). In the case of chords, the onset time was taken to be that of the highest pitch note.

Smoothing is performed by averaging each inter-beat interval (IBI) with adjacent inter-beat intervals. For each beat onset a new smoothed onset is calculated by taking the average of the IBIs within a window centred on this onset. The window widths used in the experiment below are for 1, 3 and 5 IBIs on either side of the window centre.

If the initial sequence of beat onsets is $t_1, t_2, \ldots t_n$ then the IBI sequence is: $d_i = t_{i+1} - t_i$ (i = 1, ..., n-1) and the sequence of smoothed inter-beat intervals is:

$d'_i = \frac{1}{2w+1} \cdot \sum_{j=-w}^{w} d_{i+j}$ (i = 1, ..., n-1) where w is the smoothing width.

To correct for missing values at the ends, y was extended so that $d_{1-k} = d_{1+k}$ (k = 1, ..., w) and $d_{n-1+k} = d_{n-1-k}$ (k = 1, ..., w).

The smoothed onset times t'_i are given by: $t'_i = t_1 + \sum_{j=1}^{i-1} d'_j$

3 Experiment

In this experiment six different smoothed beat tracks generated according to the smoothing function above are rated by subjects in relation to how well they correspond to the performed musical excerpts. The main hypothesis to be tested is whether listeners show a preference towards smoothed beat tracks in relation to the beat track that corresponds to the performed onsets.

3.1 Methods

3.1.1 Materials

Three excerpts from professional performances of Mozart piano sonatas K281 (3rd movt, bars 8-17), K284 (3rd movt, bars 35-42) and K331 (1st movt. bars 1-8) were used in this experiment (duration of excerpts 15-25 seconds). The main criterion for choosing these excerpts was the existence of rather large local tempo deviations in the specific performances (the standard deviation of inter-beat intervals was 31, 47 and 74ms respectively – see Figures 2, 3 and 4). In the excerpt from sonata K281 the deviations relate to the existence of triplets, in sonata K284 to the performance of grace notes, and in the opening of sonata K331 to the fact that the beat was tracked at the 'unnatural' eighth-note level (the 2:1 rhythm distorts the note onset sequence at this level as the shorter notes are lengthened – see Gabrielsson, 1987)

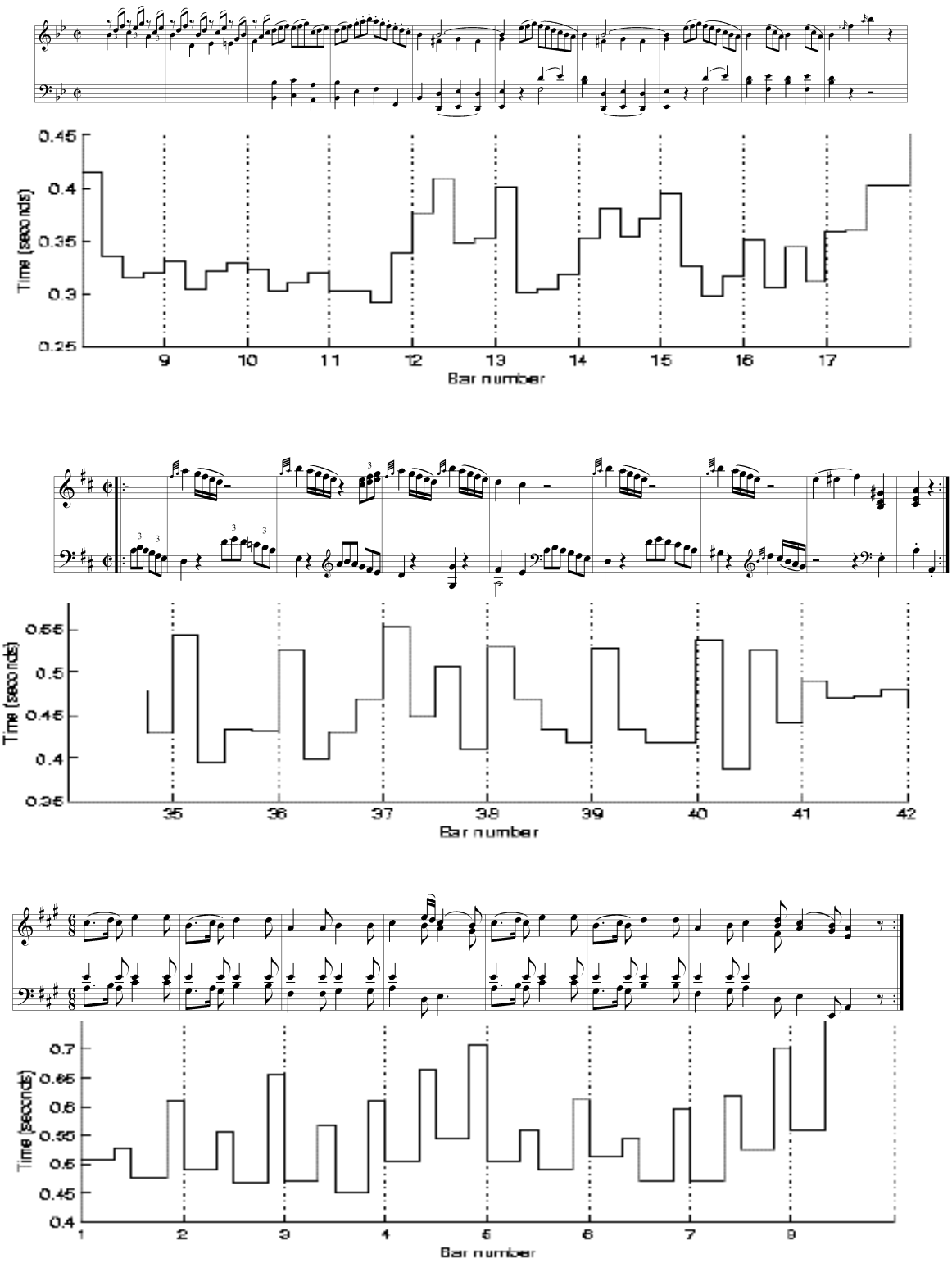
For each of these excerpts 6 beat tracks were generated as explained in section 2:

s0: beat track positions coincide with event onsets
s1: the *s0* beat track is smoothed by taking into account the previous and next beat (w=1)
s3: the *s0* beat track is smoothed by taking into account 3 previous and 3 next beats (w=3)
s5: the *s0* beat track is smoothed by taking into account 5 previous and 5 next beats (w=5)
anti: the smoothing effect of *s1* is reversed resulting in an anti-smoothed beat track
rand: random noise uniformly distributed in the range $-30\text{ms} \le r \le 30\text{ms}$ was added to *s1* beat track

For the excerpt from sonata K284 that contained grace notes two different *s0* beat track versions were constructed: in the first the onset of the first grace note was chosen whereas in the second the onset of the main note following the grace notes was selected. It is clear that the performer plays the grace notes as accented grace notes on the down-beat; for this reason the second version was disregarded from the final analysis as will be discussed in section 3.3. The beat track was realised as a sequence of woodblock clicks and was mixed with the recorded stereo piano performance at an appropriate level.

3.1.2 Participants

A group of 25 listeners (average age 30) were asked to rate the 'goodness' of fit of the various beat tracks for each musical excerpt. In the analysis below, the 25 listeners were split into two sub-categories: 15 musicians (average number of years of musical training and practice is 19.5 years) and 10 non-musicians (average number of years of training and practice is 2.2 years).



Figures 2, 3, 4 The three excerpts K281, K284 and K331 accompanied by the corresponding inter-beat interval curves.

3.1.3 Procedure

The material presented to the subjects comprises of 5 musical excerpts (i.e. K281 twice, K284 twice and K331). Excerpt K281 is presented two times for control reasons, namely so as to exclude subjects that are not consistent in their responses (if required). Excerpt K284 is presented twice once for each of the two different onset selections (see previous paragraph). For each musical excerpt a *group* of 6 different versions is created according to the 6 beat smoothing conditions described above.

Subjects were asked to rate each beat track for each different group, i.e. overall 30 different ratings. They were asked to rate 'how well the timing of the woodblock corresponds to the piano performance (in a musical sense)'. They were advised to listen to the tracks of a complete group in any order and as many times as they like before choosing their ratings. The given rating scale ranged from 1 (best) to 5 (worst).

The order of the tracks for each group was randomly determined and 3 different CDs were created with different orderings within the groups – each CD was given to 1/3 of the participants. This provision along with the advice to listen to the tracks in any order was taken in order to eliminate any possible effects of ordering of the materials.

3.3 Results and Discussion

All 25 subjects were very consistent in their ratings of the tracks of the repeated excerpt (K281) – even though the ratings were overall slightly lower (i.e., better) for the second listening of this group (see K281a,b in Figure 5).

As mentioned above, in the performance of the excerpt K284 it is clear that the grace notes are accented and appear on the beat. The second version of this excerpt with the beats appearing not on the first grace note but on the main note following the grace notes was 'unnatural'. This is very clear in the results of Figure 5 (smoothing condition: *s0* for K284b): listeners considered this track much worse than any of the corresponding tracks for the other excerpts. For this reason we decided to discard all the results that relate to the second version of excerpt K284b in the rest of our analysis. It is still very interesting to notice that simply by applying some smoothing to the 'awkward' *s0* beat track it is transformed into 'good' rating beat tracks *s1*, *s3* and *s5* (Figure 5). This observation is very important as it may contribute to determining the onsets themselves of musical events that consist of more than one note, such as in cases of 'significantly' asynchronous chords, arpeggiated chords, grace notes etc. If the onset of a musical event is not unambiguously obtainable from its constituent tones, then a smoothed beat track may indicate a tentative perceptual onset for that event.

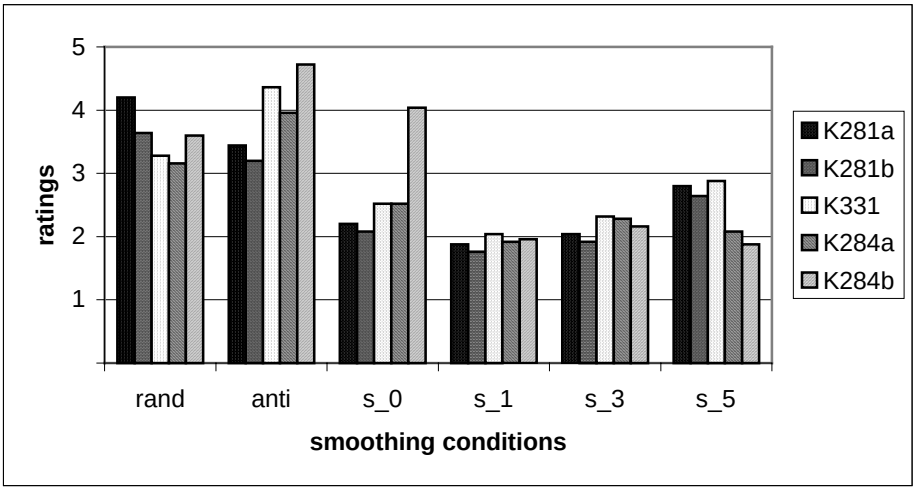


Figure 5 Average ratings of the 25 listeners for the 5 groups of tracks.

As the number of rating values available is quite small subjects tended to use the full range of values. An analysis of variance using an unrelated one-way ANOVA showed that there is a significant effect

of the independent beat smoothing variable on the dependent 'goodness' ratings of subjects (F = 53.45; df = 5, 594; p = 0.000).

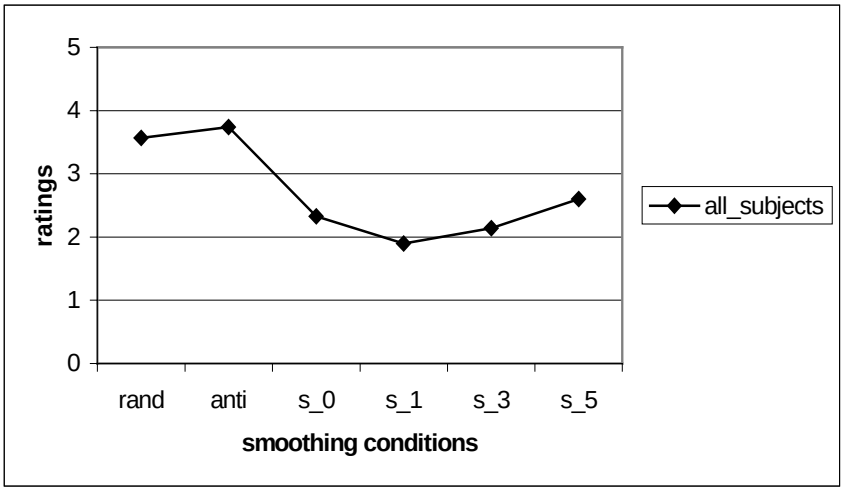


Figure 6 Overall average ratings of the 25 listeners for the six different tempo smoothing conditions (excluding excerpt K284b)

The post-hoc Scheffe test was used to compare pairs of group means in order to assess where the differences lie (Table 1). The mean difference significance values (p = 0.000) for the anti-smoothing and the random conditions indicate that these are significantly different ('disliked' by listeners) from the means of the *s0, s1, s3* and *s5* smoothed conditions. Regarding *s0, s1, s3* and *s5* smoothing conditions, *s1* has the lowest mean (i.e. most preferred condition – see Figure 6) and the mean difference between *s1* and *s0* is significant (p = 0.043). Overall, the smoothed beat track *s1* is the most preferred track and is significantly better than the beat track *s0* that coincides with the note onsets.

anti	0.862				
s0	0.000	0.000			
s1	0.000	0.000	0.043		
s3	0.000	0.000	0.795	0.586	
s5	0.000	0.000	0.452	0.000	0.024
	rand	*anti*	*s0*	*s1*	*s3*

Table 1 Significance values of the mean differences for all pairs of smoothing conditions (post hoc Scheffe test).

Further analysis was performed for the two main sub-categories of *musicians* and *non-musicians* (see Figure 7). Musicians seem to be much more acute in their perception of the differences between the *s0, s1, s3* and *s5* smoothing conditions - showing a clear preference for condition *s1* - than are non-musicians (following further analysis of variance tests, there is no significant difference among these conditions for non-musicians). This result seems to suggest that trained listeners are better equipped to perceive the refined micro-timing deviations that relate to beat timing and expressive performance. Of course these are only preliminary results; further studies would be necessary to substantiate such a claim.

4. Conclusions

In this study we investigated the relationship between musical performance and beat. It has been shown that there is a preference of listeners for beat sequences that are slightly smoother than the onset times of the corresponding musical notes. This result was strongly supported by the results obtained from the group of trained musicians whereas it seems to have no bearing for the group of non-musicians.

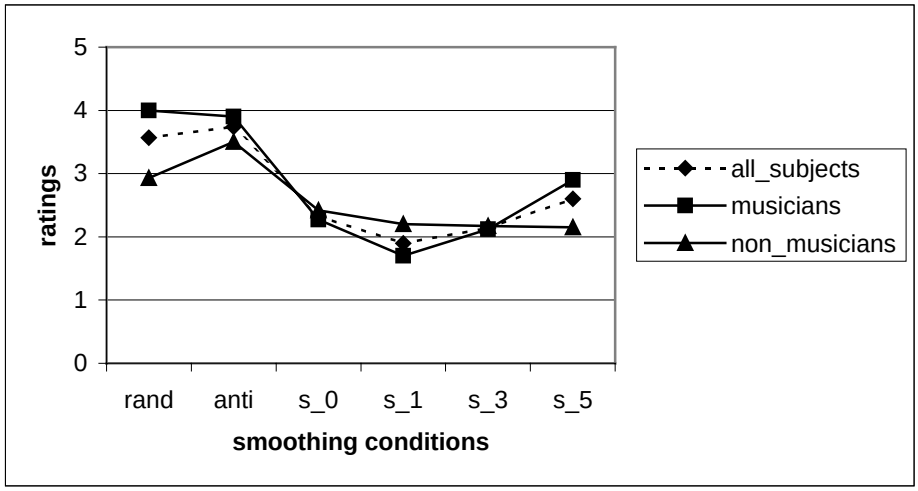


Figure 7 Overall average ratings of a) all 25 listeners, b) 15 musicians and c) 10 non-musicians for the six different tempo smoothing conditions (excluding excerpt K284b)

Acknowledgements

This research is part of the project Y99-INF, sponsored by the Austrian Federal Ministry of Education, Science, and Culture in the form of a START Research Prize and support to the Austrian Research Institute for Artificial Intelligence. We would like to thank all the participants in the experiment.

References

Cemgil A.T., Kappen B., Desain P. and Honing H. (2000) On Tempo Tracking: Tempogram Representation and Kalman Filtering. In *Proceedings of ICMC2000 (International Computer Music Conference),* 28 Aug – 1 Sep 2000, Berlin.

Clarke, E.F. (1985) Structure and Expression in Rhythmic Performance. In *Musical Structure and Cognition,* P. Howell et al. (eds), Academic Press, London.

Cooper, G.W. and Meyer, L.B. (1960) *The Rhythmic Structure of Music.* The University of Chicago Press, Chicago.

Desain, P. and Honing H. (1992) *Music, Mind and Machine*. Thesis Publishers, Amsterdam.

Dixon S. and Cambouropoulos E. (2000) Beat Tracking with Musical Knowledge. In *Proceedings of ECAI 2000 (14th European Conference on Artificial Intelligence),* W.Horn (ed.), IOS Press, Amsterdam.

Friberg A. and Sundberg, J. (1995) Time Discrimination in a Monotonic, Isochronous Sequence. *Journal of the Acoustical Society of America* 98(5): 2524-2531.

Gabrielsson, A. (1987) The Theme from Mozart's Piano Sonata in A Major (K331). In A. Gabrielsson (Ed.*) Action and Perception in Rhythm and Music,* Vol. 55, pp.81-103. Publications issued by the Royal Swedish Academy of Music, Stockholm.

Large, E.W. and Kolen, J.F. (1994) Resonance and the perception of Musical Meter. *Connection Science*, 6(2-3), 177-208.

Large, E.W. and Jones M.R. (1999) The Dynamics of Attending: How people Track Time-Varying Events. *Psychological Review,* 106(1): 119-159.

Lerdahl, F. and Jackendoff, R. (1983) *A generative Theory of Tonal Music,* The MIT Press, Cambridge (Ma).

Longuet-Higgins, H. C. and Lee, C. S. (1984) The Rhythmic Interpretation of Monophonic Music. *Music Perception,* 1:424-441.

Longuet-Higgins, H. C. and Lee, C. S. (1982) The Perception of Musical Rhythms. *Perception,* 11:115-128.

Parncutt, R. (1994) Template-Matching Models of Musical Pitch and Rhythm Perception. *Journal of New Music Research,* 23:145-167.

Parncutt, R. (1994a) A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms. *Music Perception*, 11(4):409-464.

Povel, D. J. and Essens, P. (1985) Perception of Temporal Patterns. *Music Perception,* 2:411-440.

Rosenthal, D. (1992) Emulation of Human Rhythm Perception. *Computer Music Journal,* 16(10):64-76.

Steedman, M. J. (1977) The Perception of Musical Rhythm and Metre. *Perception,* 6:555-569.

Toiviainen, P. and Snyder, J. (2000) The Time-Course of Pulse Sensation: Dynamics of Beat Induction. In *Proceedings of ICMPC 2000 (International Conference on Music Perception and Cognition),* 5-10 Aug. 2000, Keele, U.K.

Yeston, M. (1976) *The Stratification of Musical Rhythm.* Yale University Press, New Haven.