PINPOINTING THE BEAT: TAPPING TO EXPRESSIVE PERFORMANCES

Simon Dixon and Werner Goebl

Austrian Research Institute for Artificial Intelligence,
Schottengasse 3, A-1010 Vienna, Austria.
{simon,wernerg}@oefai.at

ABSTRACT

In this study we report on an experiment in which listeners were asked to tap in time with expressively performed music, and compare the results to two other experiments using the same stimuli which investigated beat and tempo perception through other modalities. Many computational models of beat tracking assume that beats correspond with the onset of musical notes; we consider the hypothesis that the beat times are rather given by a curve that is "smoother" than the tempo curve of the note onset times, which nevertheless can be derived from the onset times. The tapping results show a tendency to underestimate the tempo changes, which supports the smoothing hypothesis, and agrees with listening experiments and other tapping studies.

1. INTRODUCTION

Tempo and beat are well-defined in the abstract setting of a musical score, but not in the context of analysis of expressive musical performance. That is, the regular pulse, which is the basis of rhythmic notation in common music notation, is anything but regular when the timing of performed notes is measured. These micro-deviations from mechanical timing are an important part of musical expression, although they remain, for the most part, poorly understood. In this study we report on an experiment in which listeners were asked to tap in time with expressively performed music, and compare the results to two other experiments using the same stimuli which investigated beat and tempo perception through other modalities.

In this paper, we define *beat* to be a perceived pulse consisting of a set of *beat times* (or *beats*) which are approximately equally spaced throughout a musical performance. Each pulse corresponds with one of the *metrical levels* of the musical notation, which is usually the quarter note, eighth note, half note or the dotted quarter note level. We refer to the time interval between two successive beats at a particular metrical level as the *inter-beat interval (IBI)*, which is a measure of instantaneous tempo. A more general measure of tempo is given by averaging IBIs over some time period or number of beats. The IBI is expressed in units of time (per beat); the tempo is more often expressed as the reciprocal, beats per time unit (e.g. beats per minute). To distinguish the discussion of the timing of the participants' taps from that of the timing of musical notes by the performer, we use the terms *tapped IBI (t-IBI)* and *performed IBI (p-IBI)*.

1.1. Literature Review

There is a vast literature about finger-tapping, describing experiments requiring participants either to synchronise to an isochronous stimulus (motor synchronisation) or to tap at a constant rate without any stimulus (see Madison, 2001). At average t-IBIs between $300 - 1000$ ms, the reported variability in t-IBI is $3 - 4\%$, increasing disproportionately above and below these boundaries (Collyer, Horowitz, & Hooper, 1997). This variability is slightly greater than the JND for detecting small perturbations in an isochronous sequence of sounds, which is 2.5% at intervals between 240 and 1000 ms (Friberg & Sundberg, 1995). In these tapping tasks, a negative synchronisation error was commonly observed, that is, participants tend to tap earlier than the stimulus. This error is typically between -20 and -60 ms (Wohlschläger & Koch, 2000), and therefore above the temporal order threshold for the perception of asynchronies, which is of the order of 20 ms (see Hirsh, 1959). In more recent research, even subliminal perturbations in a stationary stimulus (below the perceptual threshold) are corrected for by tappers (Thaut, Tian, & Sadjadi, 1998; Repp, 2000).

However, there are very few attempts to investigate tapping along with music (either deadpan or expressively performed). One part of scientific effort is directed to investigate at what metrical level and at what phase listeners tend to synchronise with the music and what cues in the musical structure influence these decisions (e.g. Parncutt, 1994; Drake, Penel, & Bigand, 2000; Snyder & Krumhansl, 2001). They did not analyse the timing deviations of the taps at all. Another approach is to systematically evaluate the deviations between taps and the music. In studies by Repp (1999b), participants, tapping in synchrony with a metronomic performance of the first bars of a Chopin study, showed systematic variation related to music structure. They slowed down at phrase boundaries although the stimulus lacked any timing perturbations. In another study by Repp (1999a), pianists tapped to different expressive performances (including their own). It was found that they could synchronise well with these performances, but they tended to underestimate long IBIs, compensating for the error on the following tap.

2. METHOD

In this experiment, the participants were asked to tap the beat in time to a set of musical excerpts. The excerpts were taken from Mozart piano sonatas, expressively performed by a professional pianist on a Bösendorfer SE290 computer-monitored grand piano, for which audio recordings and precise measurements of note onset times (within 1.25ms) were available.

2.1. Participants

The experiment was performed by 25 musically trained participants (17 male, 8 female; average age 29 years). The participants have played their instruments for an average of 19 years; 19 participants studied their instrument at university level (average length of study 8.6 years); 14 participants play piano as their main instrument.

Label	Sonata	Movt	Bars	Dur.	p-IBI	ML
K284:1	K284	1st	1–9	14s	416ms	1/4
K331:1	K331	1st	1–8	25s	539ms	1/8
K281:3	K281	3rd	8–17	13s	336ms	1/4
K284:3	K284	3rd	35–42	15s	463ms	1/4

Table 1: Stimuli used in tapping experiment. The p-IBIs shown are averages over the excerpt, at the given metrical level (ML).

2.2. Stimuli

Four excerpts from professional performances of Mozart piano sonatas were used in the experiment, summarised in table 1. Each excerpt was repeated 10 times with random duration gaps ($2-5$ seconds) between the repetitions, and recorded on a compact disk (total duration 13 minutes 45 seconds).

2.3. Equipment

Participants heard the stimuli through AKG K270 headphones, and tapped with their finger or hand on the end of an audio cable. The use of the audio cable as tapping device eliminated the delay associated with a button, between the contact time of the finger on the button and the electronic contact of the button itself. The stimuli and taps were recorded to disk on separate channels of a stereo audio file, through an SB128 sound card on a Linux PC. The participants also received audio feedback of their taps in the form of the buzz produced.

2.4. Procedure

The participants were instructed to tap in time with the beat of the music, as precisely as possible, and were allowed to practise tapping to one or two excerpts, in order to familiarise themselves with the equipment and clarify any ambiguities in instructions. The tapping was then performed, and results were processed using software developed for this experiment. The tap times were automatically extracted with reference to the starting time of the musical excerpts, using a simple thresholding function.

In order to match the tap times to the corresponding musical beats, the *played beat times* had to be generated from the Bösendorfer piano performance data. First, a suitable metrical level for each excerpt was chosen (given in table 1). These metrical levels corresponded to the tapping rates of the majority of participants. Then the onset times of the notes occurring "on the beat" (i.e. according to the score, at the chosen metrical level) were extracted, with the onset of the melody note being taken where more than one note was on the beat according to the score. In the case of grace notes, the main note was taken, except in excerpt K284:3, where the grace notes were played on the beat (Cambouropoulos *et al.*, 2001), so the first grace note in each group was taken to define the beat. Beats with no corresponding played notes were interpolated linearly.

The matching algorithm then matched each tap to the nearest played beat time, deleting taps which were more than 40% of the average p-IBI from the beat time or which matched a to beat which already had a nearer tap matched to it. The metrical level was then calculated by a process of elimination: metrical levels which were contradicted by at least 3 taps were deleted, which always left a single metrical level and phase if the tapping was performed consistently. The initial synchronisation time was defined to be the first of three successive beats which matched the calculated metrical level and phase. Taps occurring before the initial synchronisation were deleted. If no such 3 beats existed, we say that the tapper failed to synchronise with the music.

	Metrical level (phase)					
Excerpt	1	2 (in)	2 (out)	3 (in)	3 (out)	Fail
K284:1	250	0	0	0	0	0
K331:1	164	0	0	86	0	0
K281:3	220	16	11	0	0	3
K284:3	153	89	8	0	0	0

Table 2: Number of excerpts tapped at each metrical level and phase (in/out), where the metrical levels are expressed as multiples of the default level given in table 1.

Excerpt	Av. sync. time
K284:1	3.29
K331:1	3.46
K281:3	3.88
K284:3	3.82

Table 3: Average synchronisation time (i.e. the number of beats until the tapper synchronised with the music).

3. RESULTS

Table 2 shows for each excerpt the total number of repetitions which were tapped by the participants at each metrical level and phase. The only surprising results were that two participants tapped on the 2nd and 4th quarter note beats of the bar (level 2, out of phase) for several repetitions of K281:3 and K284:3. The three failed tapping attempts relate to participants tapping inconsistently — changing phase during the excerpt. Table 3 shows the average beat number of the first beat for which the tapping was synchronised with the music. For each excerpt, tappers were able to synchronise on average by the third or fourth beat of the excerpt, despite differences in tempo and complexity.

The main aim of the study was to investigate the precise timing of taps. In figure 1, the t-IBIs of the mean tap times are plotted against time, with the p-IBIs shown for comparison. (In this and subsequent results, only the successfully matched taps are taken into account.) Two main factors are visible from these graphs: that the t-IBIs describe a smoother curve than the p-IBIs of the played notes, and the following of tempo changes occurs after a small time lag.

In order to test the smoothing hypothesis more rigorously, we calculated the distance of the tap times from the performed beat times and from smoothed versions of the performed beat times. The distance was measured by the RMS difference of the corresponding taps and beats. Four conditions are shown: the unsmoothed beat times; two sets of retrospectively smoothed beats (Double1 and Double3), created by averaging each p-IBI with one (respectively 3) p-IBI(s) on each side of it (Cambouropoulos *et al.*, 2001); and a final set of predictively smoothed beats (Single) created using only the current and past beat times, according to the following equation, where $x[n]$ is the unsmoothed p-IBI sequence, and $y[n]$ is the smoothed sequence:

$$y[n] = \frac{x[n] + y[n-1]}{2}$$

In table 4, the average RMS distance between the smoothed tempo curves and the taps is shown. For each excerpt, at least

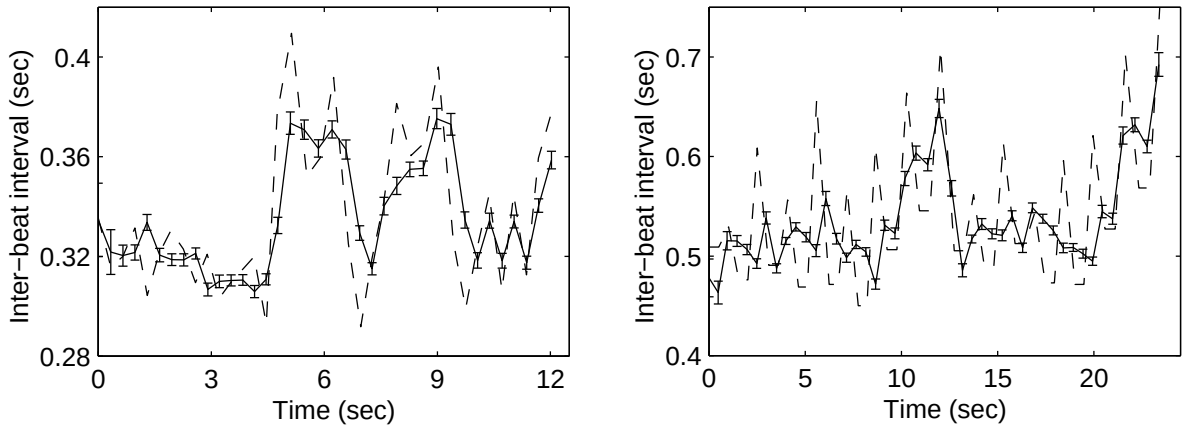

Figure 1: Solid line: t-IBIs calculated from average tap times for all participants, all repeats of K281:3 (left) and K331:1 (right). Error bars show standard error (for 95% confidence). Dotted line: p-IBIs calculated from performed note times.

Smoothing Condition	Excerpt			
	K284:1	K331:1	K281:3	K284:3
Unsmoothed	33	64	41	55
Double1	32	57	39	50
Double3	33	66	40	51
Single	38	93	33	46

Table 4: Average RMS distance between taps and smoothed beats (in ms) for various smoothing conditions.

Smoothing Condition	Excerpt			
	K284:1	K331:1	K281:3	K284:3
Unsmoothed	26	74	31	57
Double1	24	45	26	45
Double3	24	47	27	45
Single	23	48	24	43

Table 5: Average RMS distance between t-IBIs and smoothed p-IBIs (in ms) for various smoothing conditions.

one of the smoothed tempo curves is closer to the tap times than the original beat times. For excerpt K331:1, only the Double1 smoothing produces a tempo curve closer to the taps. The reason for this can be understood from figure 1 (right): the tempo curve is highly irregular due to relatively long pauses, which are used to emphasise the phrase structure, and if these pauses are spread across the surrounding or following beats, the result contradicts musical expectations.

On analysing these results, it was found that part of the reason that smoothed tempo curves model the tapped beats better is that the smoothing function creates a time lag similar to the response time lag found in the tapping. To remove this effect, we computed a second set of distances using p-IBIs and t-IBIs instead of onset times and tap times. The results, shown in table 5, confirm that even when synchronisation is factored out, the tap sequences are closer to the smoothed tempo curves than the performance data.

We also checked for a learning effect, to see whether tapping moved from an initially smooth sequence of taps to a sequence fitting closer to the unsmoothed data as participants learnt the tempo changes. It was found that the distances decreased with repetition, but the ranked order of distance by condition remained as shown in tables 4 and 5.

Finally, to find the time lag between tempo changes and changes in tapping rate, the p-IBI and t-IBI sequences were cross-correlated, and the lags corresponding to the highest correlation were found for each repetition. Table 6 shows for each lag how often this lag gave the best correlation. The results show that the lag of 1 tap is most common, that is, participants respond to a tempo change on the tap after it occurs. It was expected that with

Excerpt	Lag			
	0	1	2	3
K284:1	10.0	64.4	9.2	5.2
K331:1	45.1	43.3	2.4	0.6
K281:3	31.4	57.7	7.3	2.7
K284:3	58.2	13.7	3.9	10.5

Table 6: Analysis of time lags of responses to tempo changes, measured by correlation of t-IBIs and p-IBIs, shown as percentages of repetitions for which each lag had the highest correlation.

repetition, the lag would decrease, as the participants would remember and predict the tempo changes in their tapping. Table 7 shows this effect for excerpts K331:1 and K281:3, where with increasing repetitions, the 0 lag has the best correlation more frequently. For the other two excerpts, no learning trend is seen; K284:3 has a high correlation at lag 0 even from the initial repetitions, and K284:1 has much smaller tempo deviations, to which, it appears, the participants are able to respond but not to learn. It may be the case that such learning requires conscious recognition of timing fluctuations, or a greater number of repetitions. At least some participants were aware of the learning effect; one commented after the experiment: “it was like a chamber music rehearsal — you get it right after the third time”.

4. DISCUSSION

The main hypothesis, that the perceived beat is smoother than the played notes would indicate, is well supported by the results. This agrees with the findings of Repp (1999a) that tappers un-

Excerpt	Rpt 1–3		Rpt 5–7		Rpt 8–10	
	0	1	0	1	0	1
K284:1	9.3	61.3	10.7	66.7	8.0	66.7
K331:1	26.0	60.0	52.9	35.3	54.2	33.3
K281:3	13.4	65.7	33.3	56.1	47.0	50.0
K284:3	55.1	10.2	62.2	13.3	56.8	18.2

Table 7: Analysis of time lags of responses to tempo changes, showing the effects of learning on the lag 0 and lag 1 percentages.

derestimate timing changes. It is still unclear as to the nature and extent of the smoothing that occurs. We observed that applying rather arbitrarily chosen smoothing functions to the note onset times gave a closer match to the tapping times than the onset times themselves gave. But different functions perform better for different excerpts, and there is clearly a dependence on musical context which is not modelled by a simple smoothing function. It remains to be shown whether more accurate models can be found.

We now briefly compare the results with two other experiments using the same stimuli (without K284:1) — a listener preference test, and an offline beat marking task. In the first experiment, listeners were asked to rate how well the timing of sequences of clicks corresponded to the performed musical excerpts, presented simultaneously (Cambouropoulos *et al.*, 2001). The sequences of clicks corresponded to the Unsmoothed, Double1 and Double3 conditions in tables 4 and 5, plus 3 further conditions. Musically trained listeners showed greatest preference for the click sequence corresponding to the Double1 condition, which agrees with the tapping results reported here.

The second experiment involved the use of an interactive multimedia computer program to mark the times of beats on a display of the performances, followed by iterative correction of the beat times using audio and visual feedback until the participants were satisfied with the results (Dixon, Goebl, & Cambouropoulos, 2001). Once again, the sequences of beats chosen were smoother than the performed IBIs, but this effect was greatly reduced for most participants when they could see the onset times on the display and align the beats visually. Further work is required to ascertain whether the offline nature of the task influenced the results as compared to an online task such as tapping.

5. CONCLUSION

Although the experiments are not broad enough to suggest a complete model of beat perception, the evidence from all of the experiments supports the hypothesis that perceived beat sequences are smoother than the timing of the performed notes. This implies that timing fluctuations are not necessarily perceived as tempo changes. Beat perception shows a resistance to change and to random fluctuations; it is only when timing changes persist that one perceives an intended tempo change.

Possible explanations for the smoothing effect are that nominally on-beat notes are perceived as anticipating or following the beat, rather than defining the beat, or that p-IBIs are perceived categorically, so that by classifying intervals in units of beats, the perceptual system minimises the deviations from strictly metrical time. Several aspects of this study require further analysis and discussion which we defer to future work, including the analysis of the results with respect to the relationship between tempo and timing, and an analysis of the effect of modality on the results from the three experiments.

6. ACKNOWLEDGEMENTS

This research is part of the project Y99-INF, sponsored by the Austrian Federal Ministry of Education, Science and Culture (BMBWK) in the form of a START Research Prize. The BMBWK also provides financial support to the Austrian Research Institute for Artificial Intelligence. We thank Roland Batik for permission to use the performance data, and the L. Bösendorfer Company, Vienna, for providing the data. Thanks also to Emilios Cambouropoulos and Guy Madison for comments on a draft of this paper.

7. REFERENCES

Cambouropoulos, E.; Dixon, S.; Goebl, W.; and Widmer, G. 2001. Computational models of tempo: Comparison of human and computer beat-tracking. In *Proceedings of VII International Symposium on Systematic and Comparative Musicology and III International Conference on Cognitive Musicology*, 18–26.

Collyer, C.; Horowitz, S. B.; and Hooper, S. 1997. A motor timing experiment implemented using a musical instrument digital interface (midi) approach. *Behavior Research Methods, Instruments and Computers* 29(3):346–352.

Dixon, S.; Goebl, W.; and Cambouropoulos, E. 2001. Beat extraction from expressive musical performances. Technical Report 2001–22, Austrian Research Institute for Artificial Intelligence. Presented at 2001 Meeting of the Society for Music Perception and Cognition (SMPC2001), Kingston, Ontario.

Drake, C.; Penel, A.; and Bigand, E. 2000. Tapping in time with mechanically and expressively performed music. *Music Perception* 18(1):1–23.

Friberg, A., and Sundberg, J. 1995. Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America* 98(5):2524–2531.

Hirsh, I. 1959. Auditory perception of temporal order. *Journal of the Acoustical Society of America* 31:759–767.

Madison, G. 2001. *Functional Modelling of the Human Timing Mechanism*. Uppsala University.

Parncutt, R. 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception* 11(4):409–464.

Repp, B. 1999a. Control of expressive and metronomic timing in pianists. *Journal of Motor Behaviour* 31(2):145–164.

Repp, B. 1999b. Detecting deviations from metronomic timing in music: effects of perceptual structure on the mental timekeeper. *Perception and Psychophysics* 61(3):529–548.

Repp, B. 2000. Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research* 63(2):106–128.

Snyder, J., and Krumhansl, C. 2001. Tapping to ragtime: Cues to pulse-finding. *Music Perception* 18(4):455–489.

Thaut, M.; Tian, B.; and Sadjadi, M. A. 1998. Rhythmic finger tapping to cosine-wave modulated metronome sequences: Evidence of subliminal entrainment. *Human Movement Science* 17(6):839–863.

Wohlschläger, A., and Koch, R. 2000. Synchronisation error: an error in time perception. In Desain, P., and Windsor, W. L., eds., *Rhythm perception and production*. Lisse: Swets and Zeitlinger. 115–127.