Ratings of speed in real music as a function of both original and manipulated beat tempo

Guy Madison and Johan Paulin
Department of Psychology, Umeå University, SE-901 87 Umeå, Sweden

(Received 19 October 2009; revised 11 August 2010; accepted 6 September 2010)

There is an apparent contradiction between the narrow range of tempi optimal for perceptual judgment and motor synchronization and the wide range of beat tempi found in real music. The relation between listeners' perception of speed and beat tempo was therefore investigated, both for real music excerpts (ME) and metronome sequences. Tempi ranged from 42 to 200 beats per minute (BPM), and some excerpts were further tempo manipulated in four levels from ±5 to ±20%. Regression analyses showed that speed was a shallower function of original tempo for fast (>150 BPM) and slow (<95 BPM) MEs than for MEs with intermediate tempi, describing a non-linear, sigmoid function. Manipulated tempo had twice as large an effect on speed as had original tempo. In contrast, speed was an almost linear function of tempo for metronome sequences. Taken together, these results show that the non-linearity stems from properties of the musical signal, rather than being a subjective perceptual effect. They indicate an inverse relation between tempo and relative event density in real music, and demonstrate that the perception of periodic signals is affected not only by the beat level, but also by faster and slower levels.

© *2010 Acoustical Society of America.* [DOI: 10.1121/1.3493462]

PACS number(s): 43.66.Mk, 43.75.Cd, 43.60.Lq [ADP] Pages: 3032–3040

I. INTRODUCTION

Most music has a periodic temporal structure, which enables predictive timing necessary for ensemble performance of music and dance (Arom, 1991). Such structure creates a sensation of recurrence called beat or pulse, which often induces spontaneous movements such as foot-tapping and head-nodding in synchrony with the perceived pulse (Madison, 2006). People are obviously able to gauge the rate of the beat, as reflected by statements to the effect that one performance of a piece of music may be slow or fast, or faster than another performance (e.g., Levitin and Cook, 1996). It is conceivable that this judgment of speed is determined by several factors, however. The most important of these is beat tempo, the rate of the perceived pulse, which is a psychological construct that is often, but not necessarily, associated with perceptually salient sound events. Another factor is event density, such that the number of sound events per unit time tends to increase the perceived speed regardless of the beat tempo (Behne, 1976). The present study focuses on the relation between beat tempo (from now on referred to as tempo) and the listeners perception of the speed of the pulse (from now on referred to as speed). We specifically address the shape of that relation and how the listener is affected by artificially changing the tempo of real music, defined as commercially available music intended for voluntary listening and not created specifically for experimental purposes.

This study is motivated by the fact that the relation between speed and tempo exhibits apparent contradictions and remains poorly understood. For example, there is substantial evidence for duration-specificity in human perception of temporal recurrence, as in sequences of events that afford temporal predictability. In particular, interonset intervals (IOIs, the times between the onsets of successive events) between 400 and 800 ms are optimal for human temporal processing in several respects, with a distinct peak close to 500 ms IOI for both infants (Baruch and Drake, 1997) and adults (Moelants, 2002). This is demonstrated for human locomotion (MacDougall and Moore, 2005; Styns *et al.*, 2007), sensorimotor synchronization (Bartlett and Bartlett, 1959), serial interval production (Collyer *et al.*, 1994), as well as for temporal judgment and discrimination tasks (Michon, 1967).

It is therefore not surprising that music exhibits a similar peak at the corresponding tempo 120 BPM, both in terms of people's ratings of preference (Moelants, 2002) and of the frequency distribution of tempi found on pop charts and lists of dance music (van Noorden and Moelants, 1999). However, in the light of this strong and ubiquitous preference and perceptual advantage for tempi close to 120 BPM, it may seem surprising that music nevertheless covers a wide range of tempi from about 40 to 250 BPM, as do the music excerpts used in the present study. If these slow and fast tempi are suboptimal for movement and less preferred to listen to, why would they be used at all? How can one or maybe two distinct preferred tempi (100 and 125 BPM) be accommodated with the very wide range of tempi actually found, covering $\frac{1}{2}$ order of magnitude?

Several explanations could be considered. Music can be used for different purposes, and if that purpose is for example mood regulation (Saarikallio and Erkkilä, 2007), preferred tempo or tempi suitable for moving along may not be desirable. Likewise, optimal temporal processing may not be required or appropriate for certain task-specific applications of music, such as slowdance, meditation, and so forth.

A quite different line of explanation is that the perceived rate of recurrence as such differs from tempo. Given that there is a relatively narrow range of preferred tempo one

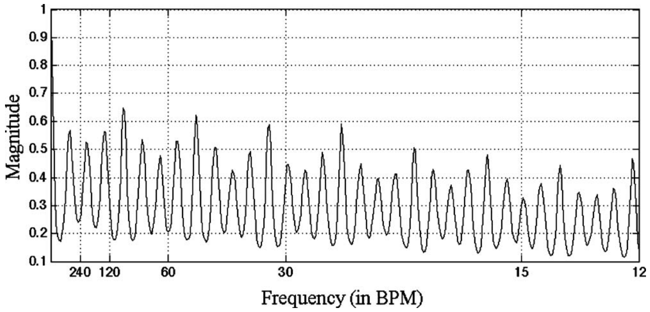

FIG. 1. Rhythm Periodicity Function (RPF) for a samba excerpt.

might expect the relation to be non-linear, such that extremely fast or slow tempi are typically perceived to be less extreme.

The metrical structure characteristic of Western music provides a possible vehicle for such a non-linearity. Multiple metrical levels in principle offer the listener alternative tempi to choose from. This would enable the listener to ignore or downplay metrical levels that they find unsuitable or unpleasant (those that are extremely fast or slow) and focus on that closest to the optimal tempo range (cf. Parncutt, 1994; van Noorden and Moelants, 1999).

To exemplify this, Fig. 1 shows a so-called Rhythm Periodicity Function (RPF)[1] for a samba excerpt with a tempo close to 100 BPM. As expected, the tempo corresponds to the strongest periodical level in the audio analysis, and the three cycles with smaller amplitude correspond to the fastest metrical level which is in this case 16th notes, four times faster than the 4th note beat. Figure 1 exemplifies that metrical levels slower than the beat tend to be weaker the slower they are, but also that a significant break in this trend occurs between the whole-note or bar level (~25 BPM) and still lower levels. This is consistent with the rhythmic pattern typical for Samba, as for most popular music in 4/4 time, as it would appear in notation.

In fact, there is at least one study suggesting that listeners may switch between metrical levels when tempi are very fast. Madison (2003) let participants rate the speed of five music excerpts (ME). Each ME occurred in the original tempo and in four tempo-manipulated versions (±10% and ±20%). The speed of the two fastest MEs, with original tempi of 216 and 224 BPM, respectively, was rated more closely to approximately half their tempi as compared to the other MEs. This result was highly consistent across tempo manipulations, in terms of a well-defined slope of speed as a function of tempo for each ME. In addition to this, the slowest ME showed a tendency to be rated faster when its tempo was decreased by 20 percent, from 104 to 83 BPM. Again, this suggests that listeners tended to choose a faster metrical level as their perceived pulse to when the tempo became too slow, only this time as a result of an artificially changed tempo instead of naturally occurring tempo differences among MEs.

An additional component of the meter-based explanation for a possible non-linear relation between speed and tempo is that not only may listeners choose from different available metrical levels, but real music may also afford a different spectrum of metrical levels depending on its tempo. It is conceivable that creators of music will design its manifest metrical structure according to the optimal range of IOIs mentioned above. Arrangers and performers are thus likely to use more metrical levels faster than the beat when the tempo is slow, and more levels slower than the beat when the tempo is fast. Hence, the relative event density may to some extent be tempo-dependent.

As we are now considering large shifts in perceived tempo in terms of subjective halving and doubling, it should be noted that this need not rely on the temporal structure of music, but seems also to be inherent in the perception of pulse. Human temporal processing is in itself metrical in the sense of both detecting and creating multiple temporal levels, as evidenced by rhythmical grouping, a.k.a. subjective rhythmization (Bolton, 1894; Woodrow, 1911). There are thus several avenues for readily changing the speed in relation to tempo. Musicians and non-musicians are highly similar in this respect, but musicians seem to be able to process information over a wider range of hierarchical levels (Drake *et al.*, 1999).

The present study attempts to elucidate the relation between speed and tempo in real music by addressing three issues. First, the apparent contradiction between the narrow range of preferred tempi and the wide range of tempi found in real music could in part be resolved by a non-linear relation between speed and tempo. Specifically, we hypothesize that extremely slow and fast tempi might produce a shallower rating of speed function than tempi in the midrange centered at 100–125 BPM.

Second, this relation was assessed for both naturally occurring tempo differences and manipulated tempo of the same MEs. Tempo manipulation was achieved by means of digital sound processing, so-called time stretch/compression, such that the entire structure of sound events was expanded or contracted in time while the pitch remained the same. Since the durations between sound events changed proportionally to the tempo change, event density in terms of the number of sound events per unit time increased as tempo increased, and decreased as tempo decreased. Thus, it was possible to assess the effect of event density on subjective tempo separately from the effect of original tempo. Given that event density contributes to speed, and that the relative event density varies inversely with natural tempo, as mentioned, we predicted that speed would have a stronger relation to manipulated tempo than to natural tempo.

Third, this combination of natural and manipulated tempi allowed us to address the possibility of inducing discrete shifts in speed as tempo is pushed to the limits for a comfortable speed. Specifically, we predicted that slow natural tempi will tend to be perceived as faster when the tempo is further decreased and that fast natural tempi will tend to be perceived as slower when the tempo is further increased.

Tempo manipulation has previously been effectively used. An early study used an electromechanical device instead of digital sound processing (e.g., Behne, 1972). Quinn and Watt (2006) had participants rate excerpts of Scottish fiddle music occurring in seven (including original) different tempi. Lapidaki and Webster (1991) and Lapidaki (2000) let participants adjust the tempo of stimuli until it sounded right to them. Brennan and Stevens (2006) used time stretch/compression to alter both pitch and tempo.

The choice of real music rather than synthetic stimuli in the present study is motivated by higher ecological validity. This is particularly important since our premises relate to real music; the particular relevant rhythmic properties of which we cannot specify since they are yet to be identified. Speed was the dependent measure, operationally defined as ratings on an 11-point scale of how well the word "fast" applied to each ME. Tempo was the main independent variable, obtained by 'counting' to the perceived beat and measuring the number of counts per unit time. Although these variables would seem to be similar and strongly related, they can in fact be expected to differ according to the preceding discussion. Speed should provide a more direct measure of perceived speed than tempo, for at least two reasons. First, tempo judgments typically involve a selection of the most salient or appropriate metrical level that logically must ignore remaining perceptual influences from non-selected levels. Second, tempo judgments are typically associated with overt motor action, such as physically tapping out the perceived pulse, which again might force participants to focus on just the dominant pulse and thereby disregard other information. Third, motor action and individual differences in motor ability is likely to affect speed judgments in various ways, and speed ratings without motor action is therefore likely to provide a more pure measure of perceptual processes. These considerations are particularly relevant for the present wide range of tempo. Previous studies using perceptual indicators of tempo have provided reliable results, both with direct ratings of speed (Gabrielsson, 1973; Madison, 2006) and in the context of tempo adjustment (Lapidaki and Webster, 1991; Lapidaki, 2000).

Experiment 1 largely confirmed the hypotheses, which raised the question whether the non-linear relation between tempo and speed was due to non-linear speed perception or to systematic differences in the properties of the MEs as a function of tempo. Experiment 2 was therefore a replication with metronome sequences, whose properties were constant across tempi, in which an almost linear speed versus tempo function lends no support for the perceptual explanation.

II. EXPERIMENT 1

A. Methods

1. Participants

Sixty participants (30 females, 30 males) were recruited for the experiment. They were between 23 and 68 years of age ($M=30.08$, $SD=8.09$) and received the equivalent of 8 euro for attending. They were treated in accordance with relevant institutional and national regulations and with the Helsinki Declaration of the World Medical Association.

2. Stimuli

A large sample of authentic popular music ($N>200$) was selected and grouped into three ranges based on tempo. The tempo of each music excerpt (ME) was measured by two staff members by individually tapping the perceived beat on the key of an electronic metronome, which showed the tapped tempo on a display. In no case did the tempo estimates differ by more than 1 BPM. The tempo ranges were based on equal distances from 122.5 BPM, which is the average of 120 (Moelants, 2002) and 125 BPM (van Noorden and Moelants, 1999): a slow range up to 95 BPM, medium from 95–150 BPM, and fast above 150 BPM. MEs in the slow and fast ranges likely to be familiar to the listeners were discarded. Since these songs were to be time-stretched/compressed, and high familiarity with a song has been shown to instill an accurate long-term memory for its tempo (Brennan and Stevens, 2006; Levitin and Cook, 1996), we wanted to avoid reactions related to the song being found manipulated or unnatural. This unfamiliarity demand was relaxed for the medium range that was not to be tempo manipulated. A total of 50 MEs were finally selected, 10 for the medium range (mean tempo=116.6 BPM) and 20 for each of the extreme ranges (mean tempo=67.4 and 169.2 BPM, respectively). A 32 s excerpt was cut out from each track using Audacity, an open source software (http://audacity.sourceforge.net). The edits begun on a beat, if there was no singing, and on a syllable if there was singing. To assume some modicum of representativity for music in general, MEs were chosen from diverse genres such as metal, punk, pop, jazz, klezmer, and world music.

The tempo of each slow ME was decreased by −5%, −10%, −15%, and −20%, and the tempo of each fast ME was increased with +5, +10, +15, and +20%. The medium range ME occurred only in their original tempi. Since 40 of the total 50 original MEs occurred in 5 tempi (one original and 4 manipulated), the total number of MEs was $210(10+40\times 5)$. All musical samples were in 16 bit PCM stereo format with a sample rate of 44100 Hz, and were adjusted to equal peak loudness levels (between −14 and −16 dB relative to clipping limit).

Tempo change was performed in Audacity with the "Change tempo" effect, which employs the tdstretch function from the SoundTouch library written by Olli Parvainen (http://www.surina.net/soundtouch/). SoundTouch employs a so-called Synchronous-OverLap-Add (SOLA) technique that operates in the time domain, using 32-bit floating-point arithmetic. The sound waveform is cut into segments of 10–100 ms, which are then joined closer (for tempo increase) or farther away in time (for tempo increase). Both the segment length and the overlap are optimized with respect to the periodicity of the waveform to reduce artifacts, such as humming, clicking, drifting, or reverberation. Since all music in the present study has a well-defined periodicity, we did not specify tempo but let the program compute this automatically. This is done by downsampling the waveform to <500 Hz, calculating the autocorrelation function for segments of the enveloped signal, and finding the highest peak of the autocorrelation function. Another important aspect of the SOLA technique is that the exact joining points between segments are determined by the best waveform fit within a seek window, a small proportion of the nominal overlap. This is achieved by computing the cross-correlation between the end of the first segment and the beginning of the next for all possible points within the seek window, and choosing the point with highest cross-correlation. The segments are joined by a gradual amplitude transition. A time domain method for time stretch was chosen because it generally preserves rise time, intensity, and spectrum of sounds better than do frequency domain methods, which is particularly critical for percussive sounds. Since the periodicity is clearly defined in the present music examples, the automatic determination of local tempo ensures good optimization of the time stretch process, reducing the magnitude of adjustment of joining points between segments. The maximum adjustment allowed is proportional to the tempo, and is no larger than 28 ms at 100 BPM (600 ms IOI), which corresponds to 4.6 percent of the IOI. Since this is below the detection threshold in a musical signal, it should have no consequences for the present study (cf. Ellis, 1991; Geringer and Madsen, 1984; Miller and Eargle, 1990).

3. Apparatus and design

The MEs were presented through a pair of headphones by means of a specially written computer program running on a PC. This software also recorded listeners' ratings and administered the experiment. The signal was sent from the built-in soundcard of the PC to the headphones.

Participants were randomly assigned to one of five presentation conditions, labeled A-E. Each of these conditions featured a different subset containing 50 out of the total 210 MEs, designed such that no ME appeared more than once for each participant. This was to avoid that the participant heard the same ME in different tempi, which might affect the second rating by the association that the first version of that ME was the "correct" one. Furthermore, the within-ME tempo changes were rotated so that each condition A-E had different levels of tempo change for each fast and slow ME. This was to avoid that any participants heard all ME in a specific level of tempo change, which would likely bias those participants' subjective scaling of tempo. Participants in all five conditions heard the same 10 medium tempo MEs in their original tempi. In summary, each participant heard 20 different music excerpts in each of the fast and slow ranges, four of which occurred in each of the five levels of tempo change, plus the 10 MEs in the medium range. MEs were presented in a different random order for each participant to avoid possible order effects.

4. Rating scales

Participants gave their subjective ratings relative to visual scales presented on a computer screen. Six scales were used, assessing how familiar, preferable, pleasurable, movement inducing, impressive, and how fast the music was perceived. Only ratings of fast were considered in the present study. The scales were represented by horizontal lines with 11 vertical lines numbered from 0 to 10; the anchor points were labeled 'not at all' for 0 and 'very much' for 10. The scale was defined in the instructions as: *The music feels fast ('Not at all' thus meaning very slow).*

Participants used the mouse to move a slider on the computer screen to the desired position along the scale for each adjective. Once satisfied with their rating the participants clicked a button to proceed to the next ME. It was not possible to go back to change ratings of previous MEs, nor was it possible to hear any ME more than once. There were no time constraints on the experiment.

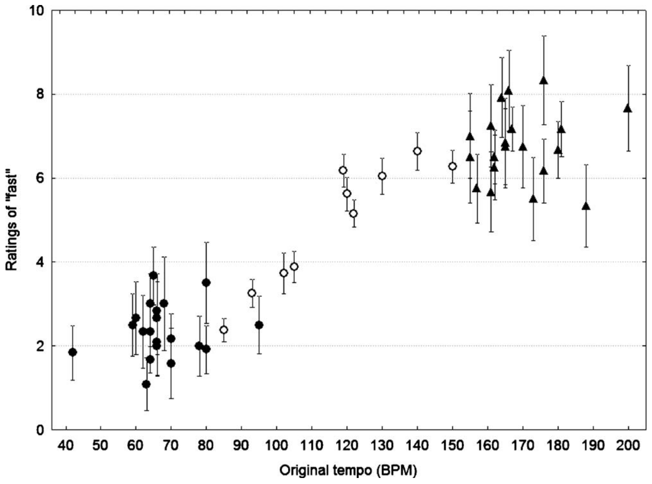

FIG. 2. Mean ratings of Fast for each ME as a function of its original tempo (N=50). Error bars depict 0.95 confidence intervals. Black circle: slow MEs; White circle: medium MEs; Black triangle: fast MEs.

5. Procedure

At arrival the participants were given oral and written instructions, and completed a training block under the supervision of the experimenter. The training block consisted of eight MEs in the range 77–200 BPM that were not used in the experiment proper and had not been subjected to time stretch/compression. The experimenter supervised the first 3–5 training trials, until it was clear that the participant fully understood the task and the procedure. The experiment took about 50 min in total, including instructions and training. Participants performed the main part of the experiment individually without the supervision of the experimenter. After completion a short interview and debriefing took place.

B. Results and discussion

Interviews indicated that the participants felt comfortable with the task, and experienced no particular difficulties. Figure 2 provides an overview of the relation between tempo and speed, in terms of ratings of Fast for all ME in their original tempo. It is notable that some of the MEs differ in how consistently they were rated, as indicated by the confidence intervals. The different number of ratings (60 vs. 12) accounts for the smaller confidence intervals for MEs in the medium range. The relation describes a sigmoid shape with shallower slopes for the extreme tempo ranges compared to the middle range.

As seen in Fig. 3, separate linear regressions fitted to each tempo range confirmed that slopes for the slow and fast ranges were clearly shallower than for the medium range, expressed in the following as $\text{intercept}+\text{slope}\pm 0.95$ confidence interval. Ratings corresponded to $1.738\,56+\text{tempo}\times 0.009\,32\pm 0.016\,38$ for the slow range and $5.530\,17+\text{tempo}\times 0.007\,283\pm 0.018\,08$ for the fast range, and this difference in slope is not significant according to the confi-

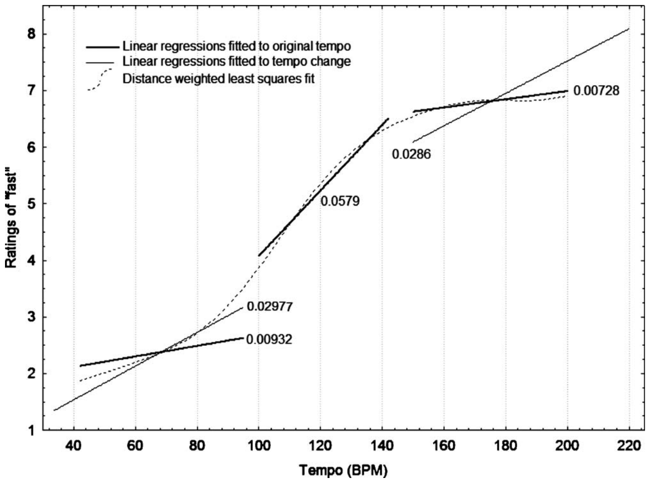

FIG. 3. Linear regressions fitted to ratings of Fast for each tempo range, both as a function of original tempo (thick lines) and manipulated tempo (thin lines).

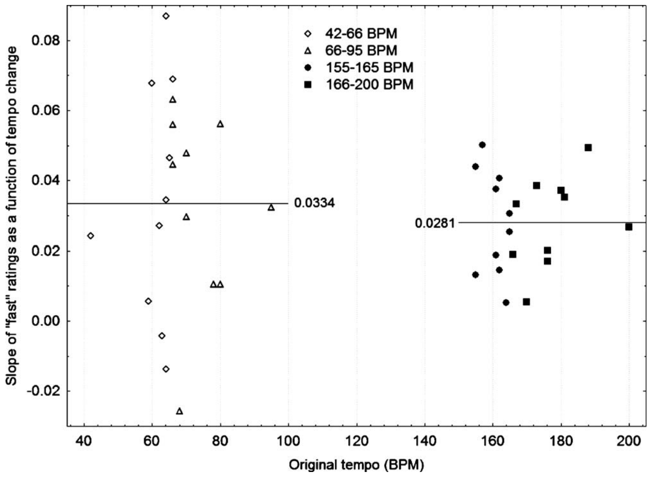

FIG. 4. The mean slopes of ratings of Fast as a function of tempo for individual tempo-manipulated MEs. Horizontal lines indicate mean slope across slow and fast MEs respectively.

dence intervals. Both these slopes were however significantly different from the $\text{slope}-1.710\,11+\text{tempo}\times 0.0579\pm 0.011\,86$ obtained for the medium range.

Next, we compared the effect of original tempo with the effect of the tempo manipulations performed in the extreme tempo ranges. To this end, all MEs were aligned by means of transforming absolute tempi to tempo changes for each ME separately. For example, an ME with an original tempo of 62 BPM received the value 0 in the original tempo level, −3.1 for 5% decrease and −6.2, −9.3, and −12.4 BPM for 10%, 15%, and 20% decrease, respectively. Similarly, a fast ME with 176 BPM received positive values of 8.8, 17.6, 26.4, and 35.2 for the four levels of tempo increase. Thus, this procedure made original tempo and tempo change equivalent with respect to the linear regression on ratings of fast.

Figure 3 shows that the tempo manipulations yielded steeper slopes than the original tempi. The statistical significance of all differences between slopes was assessed with confidence intervals. The slopes were $2.3681+\text{tempo}\times .029\,766\pm .015\,83$ for the slow range and $6.666\,85+\text{tempo}\times .0286\,09\pm .007\,639$ for the fast range. Both these slopes differ significantly from the slope for the medium range, but not from each other. The differences between original and manipulated tempo were statistically significant for both the slow and fast range.

Third, we addressed the hypothesis that slow natural tempi will tend to be perceived as faster when the tempo is further decreased and that fast natural tempi will tend to be perceived as slower when the tempo is further increased. To this end we considered mean ratings for each level of tempo change for each ME separately, which indicated that a few of the MEs in the slow range indeed exhibited a tendency to be rated as faster when their tempi were decreased. No such shift was statistically significant according to 95 confidence intervals, however. We also adopted a different approach to the same question by computing regression slopes as a function of tempo change for each ME separately. These are plotted in Fig. 4, and show that there was a considerable variability among the slopes, and that this variability was furthermore much larger for slow MEs than for fast MEs. Specifically, the steepest slopes in the fast range were similar to the mean slope for original tempi in the medium range (0.0579), and slopes for slow MEs ranged from approximately 0.07 down to below zero, the latter indicating shifts to higher perceived tempi when tempo was decreased. None of these slopes were significantly different from neighboring ones with positive slopes. This could be attributed both to the relatively small amount of data and to large individual differences, as suggested by the large confidence intervals seen in Fig. 2.

III. EXPERIMENT 2

That the slopes of the speed versus tempo relation were less steep for the extreme tempo ranges, both for original and tempo-manipulated MEs, invites two different explanations. One is that perceived speed differs across the range of tempi, such that it is more compressed for very slow and very fast tempi than for medium tempi. As mentioned in the introduction, this may be due to such a subjective relation in itself or to a biased choice among alternative metrical levels in the MEs. The alternative explanation is that other properties differ systematically among MEs as a function of tempo, and that this, in turn, affects perceived speed. To discriminate between these alternatives, Experiment 2 replicated Experiment 1 for metronome sequences instead of real music. Metronome sequences contain only one metrical level and their properties are in all respects constant across tempi, which should prevent both phenomena pertaining to the first explanation. That the effects on speed were considerably larger for manipulated tempo than for original tempo in Experiment 1 indicates that stimulus properties provide the source of the non-linearity. It was therefore predicted that this result would not be replicated in Experiment 2. In order to maintain similar conditions as in Experiment 1, and not to introduce context effects, participants listened to a mix of real music and metronome samples, both with matching BPM-rates.

A. Methods

The apparatus, design, scales and procedure were identical to Experiment 1, with the following exceptions.

1. Participants

Ten men and 14 women between 19 and 38 years of age (M=24.54, SD=4.87) were recruited and paid the equivalent of 5 euro. None of them had participated in Experiment 1.

2. Stimuli

The music excerpts ranged from 40–200 BPM in steps of 10 BPM and were close to 15 s in duration. In addition to the short samples of real music, samples of metronome-clicks with the same tempi were included. None of the MEs were time-stretched/compressed or otherwise altered except for equalizing the sound levels across samples.

3. Design

The experiment consisted of two identical blocks with 34 conditions (2 type × 17 tempo), in which the 17 tempi were represented by isochronous metronome samples and different MEs.

This was replicated once by presenting two such blocks in succession. The 34 stimuli within each block were played in a different random order for each participant. These conditions were identical for all participants, as opposed to the fragmented design across participants employed in Experiment 1. Only the metronome stimuli were considered in the analysis.

4. Rating scales

Four of the six scales from Experiment 1 were used, namely interesting, preferable, movement inducing, and how fast the music was perceived. Only speed was subjected to analysis.

5. Procedure

Each participant performed a training trial before the experiment proper, consisting of 10 samples not appearing in the experiment; 5 MEs and 5 metronome samples with 44, 78, 132, 161, and 200 BPM. The experimenter supervised the first 3–5 trials until it was clear that the participant fully understood the task and the procedure. All in all the instructions, training, the experiment proper, and a brief interview took about 35 min.

B. Results and discussion

Again, interviews indicated that the participants felt comfortable with the task, although they found it tedious to listen to the metronome excerpts in the long run. Figure 5 depicts mean ratings of speed for the metronome samples as a function of tempo. Similarly to Experiment 1, participants used the whole scale, resulting in a mean range of 1–8. There is a marked difference in the slope of the function, however, which is almost perfectly linear. Confidence intervals (0.95) indicate no substantial departure from a linear regression fitted to the raw data, except for the slowest tempo 40 BPM. If anything, the slope is steeper for the slowest tempi, contrary to Experiment 1, while there is a tendency for a shallower slope for the highest tempi. Nevertheless, even in this range the slope is substantially steeper (0.42) than in Experiment 1 (0.29), in which it was instead steeper in the medium range (0.58). We suggest that these absolute differences among slopes are a trivial consequence of participants' ability to adjust their ratings to the range of the scale.

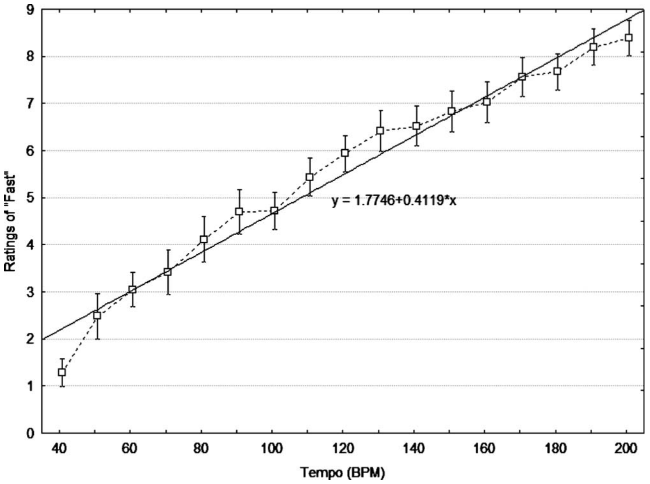

FIG. 5. Mean ratings of Fast, across participants and replications, for metronome samples as a function of tempo. Error bars depict 0.95 confidence intervals. A least-squares linear regression line is fitted to the raw data.

The hypothesis was supported, which leads to the conclusions that people can accurately gauge tempo by means of ratings of speed, and that the non-linearity found in Experiment 1 was caused by systematic differences in stimulus properties across the range of tempo.

IV. GENERAL DISCUSSION

In this first study of the relationship between tempo and speed across a wide range of tempi in real music we asked three main questions. First, we asked whether the contradiction between the narrow range of preferred tempi and the wide range of tempi found in real music could be resolved by a non-linear relation between speed and tempo. Second, we assessed whether the effect of tempo on speed would be stronger for manipulated tempo than for naturally occurring tempo differences, according to the premise that the latter are counteracted by an inverse change in relative event density. Third, it was considered if pushing tempo in the extreme direction might induce a shift in perceived speed in the opposite direction. In order to discriminate between a perceptual and physical source for the sigmoid function found, in accord with the two first hypotheses, an additional experiment with isochronous sound sequences was performed. It gave further support for the physical explanation; that slow music tends to have a larger number of manifest faster metrical levels than has fast music. Indeed, manipulated tempo had an almost three times larger effect than natural tempo on speed in both the slow and fast tempo ranges, which indicates that structural properties other than tempo play a major role. No evidence was found for mental subdivision or doubling of the speed because of pushing tempi in the extreme direction (cf. Madison, 2003). A tendency for three MEs to exhibit such switching indicates that it might nevertheless occur, but that it would be better revealed with stimuli in which tempo and event density are independent, unlike both the metronome sequences, the manipulated MEs, and the natural tempo and ME combinations. It is possible that these three MEs had more pronounced metrical levels that were faster than the dominant pulse. Specifically, one obvious confound is that the overall manipulation of tempo changed all temporal properties in the MEs, thus changing the event density on all metrical levels from what the composers had initially intended, and making the slow MEs more sparse and the fast MEs more busy.

Another aspect of the results is that speed of individual MEs does not always follow the general tempo trend, although it is nevertheless rated quite consistently. Significant differences among MEs shown in Fig. 2, in spite of the small number of participants, suggest that structural properties other than tempo affect ratings in a systematic fashion. For example, some MEs with almost the same tempo close to 80, 120, and 175 BPM are rated very differently. That the results exhibit such systematic effects make it worthwhile to explore the underlying structural properties in the physical audio signal. However, it would have been premature to apply such an analysis within the present study since, for example, the relation between the frequency and magnitude of rhythmic events corresponding to the respective metrical levels and their perceptual salience is highly unclear. With improved empirical mapping of perceived and physical properties, this will probably provide a useful approach for future research.

Another tendency in the data is that ratings for the slow and fast MEs were more diverse than in the medium range. MEs with similar tempo were rated with great variation, sometimes approaching distances that correspond to the double or half tempo. In addition to the statistically significant differences in this respect just mentioned, confidence intervals vary more among MEs in both the fast and slow ranges. This indicates that there was some disagreement between participants on how to rate the fast and the slow music; possibly influenced by properties of the MEs. In contrast, the results from Experiment 2 demonstrate that people are able to rate less ambiguous signals very accurately. This in turn suggests that there is one or several properties found in real music that causes the different results in ratings observed in Fig. 2.

Another noteworthy result is that the fast range did overall seem to elicit less ambiguous ratings than did the slow range. However, the selected MEs were not evenly distributed along the tempo dimension, which raises the concern that the unequal sampling of tempo affected the participants' expectations, thereby creating a locally compressed scale, and that this limits the possibility to draw conclusions regarding differences between the tempo ranges. Future studies with more evenly distributed tempo and larger samples of MEs are needed to examine these issues in more detail.

In summary, the results can be interpreted such that the temporal structure of real music differs systematically as a function of its tempo. A high tempo seems to be counteracted by a lower relative event density and vice versa. Furthermore, event density alone was shown to contribute substantially to speed, since the manipulation of tempo changed event density while keeping all other structural properties intact, and this had a larger effect on speed than natural tempo. In other words, the perception of periodicity is affected not only by the beat level, but also by faster and slower levels. This may seem trivial and quite obvious to musically experienced people, but it is nevertheless understated in the literature. The strong focus on the beat level, and the almost ubiquitous use of metronome-like sound sequences, may give an impression that other levels are of no consequence. Recent years have brought an increasing number of studies considering multiple temporal levels (Keller and Repp, 2005; Madison, 2009; Patel *et al.*, 2005; Repp, 2003a, 2003b; Repp, 2004, 2007). These kinds of studies, including the present one, find that multiple levels are actually used regardless of whether they correspond to the beat level or not (Patel *et al.*, 2005; Repp, 2003a, 2003b, 2004, 2007). For example, it has been shown that faster metrical levels increase the predictability of very long intervals, in terms of precise synchronization with events separated by up to several seconds (Madison, 2009). On the other hand, sounds not related to any metrical level do also affect synchronization behavior (Wohlschläger and Koch, 2000). One outstanding question is therefore to what extent and within which boundaries musical meter corresponds to an internal perceptual model. Just like our auditory system is pre-wired to handle complex (i.e., natural) tones, with their physically determined spectral properties, time intervals with small-integer relationships may reinforce each other in the perceptual processes. On the other hand, it may be that musical meter is just one way to exploit a more general "temporally hierarchical" organization, the precise properties of which are yet unknown. While musical meter is mainly based on powers of two, it is not clear if multiple or subdivision factors of 3, 4, or 5 would entail significant differences in human performance. Musical rhythmic structure is likewise based on canonical time values, in principle without deviation from isochrony. Yet, there is always some amount of variability in human performances of music, including systematic patterns of timing deviations. How are these accommodated by the perceptual system? Are there for example differences as a function of the between magnitude of deviation from strict isochrony (Madison and Merker, 2002)? A characterization of these processes based on limit boundaries would be very valuable, inasmuch as it would facilitate the formulation of distinct and testable hypotheses for both behavioral and neural activity studies regarding the perception of periodicity in auditory signals.

APPENDIX

The 50 music excerpts used as stimuli in the listening experiment.

Tempo	Artist	Album	Title/track
Slow			
42	David Gilmore	On an island	On an island
59	Marie Bergman	Fruit	Let me be the first
60	Elexir	Feel real	Utsikt

62	Ayreon	The universal migrator pt1	2084
63	Nils Landgren Funk unit	Live in Stockholm	Cheyenne
64	Blackfield	Blackfield II	1000 people
64	Chris Cornell	Euphoria morning	Preaching the end of the world
64	Steve Vai	Passion and warfare	Sisters
65	Cranberries	No need to argue	The icicle melts
66	IQ	Ever	Came down
66	Ayreon	The universal migrator pt1	And the druids turned to stone
66	Landberk	Indian summer	Why do I still sleep
66	Crippled black phoenix	A love of shared disasters	When you're gone
68	Marie Bergman	Fruit	Just one of those things
70	Nils Landgren Funk Unit	Funk da music	Anytime anywhere
70	Leonard Cohen	Ten new songs	That don't make it Junk
78	Massive Attack	Mezzanine	Teardrop
80	Led Zeppelin	Physical graffiti	Kashmere
80	The streets	A grand don't come for free	Dry your eyes
95	Elvis Presley	Elvis' Christmas album	Blue Christmas
Medium			
85	Duffy	Rockferry	Warwick avenue
93	Tananas wide	Unamunacua	Funky Bumpkins
102	Anathema	A natural disaster	Are you there?
105	Pussycat dolls	PCD	Buttons
119	Wolf	Evil star	Wolfs blood
120	Robyn	Robyn	With every heartbeat
122	Nine Inch Nails	The slip	Discipline
130	September	September	Cry for you
140	Cascada	Truly, madly, deeply (CD single)	Truly madly deeply
150	Radio tarifa	Temporal	El mandil de Carolina
Fast			
155	Coryell, Coster, Smith	Cause and effect	These are odd times
155	The black eyed peas	Monkey business	Pump it
157	Nils Landgren Funk Unit	Funk da music	From Stockholm to Beijing
161	Gogol Bordello	Voi-la intruder	Greencard husband
161	Porcupine Tree	Deadwing	Deadwing
162	Faith No More	King for a day. fool for a lifetime	Digging the grave
162	Rancid	… and out comes the wolves	Roots radical
164	Jeff Loomis	Zero order phase	Race against disaster
165	Stratovarius	Stratovarius	Fight
165	Candlemass	Candlemass	Black dwarf
166	Rage	Carved in stone	Gentle murder
167	Den flygande bokrullen	Den flygande bokrullen	Ternovka
170	Den flygande bokrullen	20th century klezmer	Klaunowicz
173	Takida	Bury the lies	The dread
176	Pitchshifter	Deviant	Hidden agenda
176	Gogol Bordello	Gypsy punks: Underdog world strike	I would never wanna be young
180	Grand Magus	Wolf's return	Blood oath
181	Den Flygande Bokrullen	Den flygande bokrullen	Odessa bulgar
188	Belá Fleck and the Flecktones	Flight of the cosmic hippo	Michelle
200	Exodus	The atrocity exhibition: Exhibit A	Funeral hymn

[1] The Rhythm Periodicity Function (RPF) measures the amount of self-similarity as a function of time lag. First, the audio data are pre-processed into a representation of lower dimensionality that highlights energy changes (cf. Klapuri *et al.*, 2006), upon which the RPF is computed as the autocorrelation function (ACF), r(τ), of the energy time series [denoted by x(n)] as follows: $r(\tau)=\sum_{n=0}^{N-\tau-1}x(n)x(n+\tau), \forall \tau\in\{0\cdots U\}$ where N is the number of samples of the signal and U is the upper limit for the autocorrelation lag τ. This function yields an estimate of the relative magnitude of different metrical levels in the signal that has some level of correspondence to that perceived by humans. In Fig. 1 the function is normalized so that r(0) = 1 and the maximum lag (upper limit U) is set to 5 s (i.e., a frequency of 0.2 Hz, or 12 Beats Per Minute). Further details on the implementation are provided by Gouyon (2005).

Arom, S. (**1991**). *African Polyphony and Polyrhythm: Musical Structure and Methodology* (Cambridge University Press, Cambridge, UK), pp. 179–200.

Bartlett, N. R., and Bartlett, S. C. (**1959**). "Synchronization of a motor response with an anticipated sensory event," Psychol. Rev. **66**, 203–218.

Baruch, C., and Drake, C. (**1997**). "Tempo discrimination in infants," Infant Behav. Dev. **20**, 573–577.

Behne, K.-E. (**1972**). *Der Einfluß des Tempos auf die Beurteilung von Musik (The Influence of Tempo on the Judgment of Music)* (Arno Volk Verlag, Köln, Germany), p. 183.

Behne, K.-E. (**1976**). "'Zeitmaße'-zur psychologie des musikalischen tempoempfindens ('Tempos'-concerning the psychology of the perception of musical tempos), " Die Musikforschung **29**, 155–164.

Bolton, T. L. (**1894**). "Rhythm," Am. J. Psychol. **6**, 145–238.

Brennan, D., and Stevens, C. (**2006**). "The effect of pitch, tempo and proportional pitch and tempo manipulation on memory of familiar musical excerpts," in Proceedings of the 9th International Conference on Music Perception and Cognition, Alma Mater Studiorum, University of Bologna, Bologna, Italy, edited by M. Baroni, A. R. Addessi, R. Caterina, and M. Costa, pp. 1771–1778.

Collyer, C. E., Broadbent, H. A., and Church, R. M. (**1994**). "Preferred rates of repetitive tapping and categorical time production," Percept. Psychophys. **55**, 443–453.

Drake, C., Gros, L., and Penel, A. (**1999**). "How fast is that music? The relation between physical and perceived tempo," in *Music, Mind, and Science*, edited by S. W. Yi (Seoul National University Press, Seoul), pp. 190–203.

Ellis, M. (**1991**). "Thresholds for detecting tempo change," Psychol. Music **19**, 164–169.

Gabrielsson, A. (**1973**). "Adjective ratings and dimension analyses of auditory rhythm patterns," Scand. J. Psychol. **14**, 244–260.

Geringer, J. M., and Madsen, C. K. (**1984**). "Pitch and tempo discrimination in recorded orchestral music among musicians and non-musicians," J. Res. Music Educ. **32**, 195–203.

Gouyon, F. (**2005**). "A computational approach to rhythm description," Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, Spain, p. 188.

Keller, P. E., and Repp, B. H. (**2005**). "Staying offbeat: Sensorimotor syncopation with structured and unstructured auditory sequences," Psychol. Res. **69**, 292–309.

Klapuri, A., Eronen, A. J., and Astola, J. T. (**2006**). "Analysis of the meter of acoustic musical signals," IEEE Trans. Audio, Speech, Lang. Process. **14**, 342–355.

Lapidaki, E. (**2000**). "Stability of tempo perception in music listening," Music Educ. Res. **2**, 25–44.

Lapidaki, E., and Webster, P. R. (**1991**). "Consistency of tempo judgments when listening to music of different styles," Psychomusicology **10**, 19–30.

Levitin, D. J., and Cook, P. R. (**1996**). "Memory for musical tempo: Additional evidence that auditory memory is absolute," Percept. Psychophys. **58**, 927–935.

MacDougall, H. G., and Moore, S. T. (**2005**). "Marching to the beat of the same drummer: The spontaneous tempo of human locomotion," J. Appl. Physiol. **99**, 1164–1173.

Madison, G. (**2003**). "Perception of jazz and other groove-based music as a function of tempo," in Proceedings of the 5th Triennial ESCOM Conference, Hannover, Germany, edited by R. Kopiez, A. C. Lehmann, I. Wolther, and C. Wolf, pp. 365–367.

Madison, G. (**2006**). "Experiencing groove induced by music: Consistency and phenomenology," Music Percept. **24**, 201–208.

Madison, G. (**2009**). "An auditory illusion of infinite tempo change based on multiple temporal levels," PLoS ONE **4**, e8151.

Madison, G., and Merker, B. (**2002**). "On the limits of anisynchrony in pulse attribution," Psychol. Res. **66**, 201–207.

Michon, J. A. (**1967**). "Magnitude scaling of short durations with closely spaced stimuli," Psychonomic Sci. **9**, 359–360.

Miller, L. K., and Eargle, A. (**1990**). "The contribution of development versus music training to simple tempo discriminations," J. Res. Music Educ. **38**, 294–301.

Moelants, D. (**2002**). "Preferred tempo reconsidered," in Proceedings of the 7th International Conference on Music Perception and Cognition, Causal Productions, Adelaide, C. Stevens, D. Burnham, G. McPherson, E. Schubert, and J. Renwick, pp. 580–583.

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

Patel, A. D., Iversen, J. R., Chen, Y., and Repp, B. H. (**2005**). "The influence of metricality and modality on synchronization with a beat," Exp. Brain Res. **163**, 226–238.

Quinn, S., and Watt, R. J. (**2006**). "The perception of tempo in music," Perception **35**, 267–280.

Repp, B. H. (**2003a**). "Phase attraction in sensorimotor synchronization with auditory sequences: Effects of single and periodic distractors on synchronization accuracy," J. Exp. Psychol. Hum. Percept. Perform. **29**, 290–309.

Repp, B. H. (**2003b**). "Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision," J. Motor Behav. **35**, 355–370.

Repp, B. H. (**2004**). "On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences," Hum. Mov. Sci. **23**, 389–413.

Repp, B. H. (**2007**). "Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization," Cognition **102**, 434–454.

Saarikallio, S., and Erkkilä, J. (**2007**). "The role of music in adolescent's mood regulation," Psychol. Music **35**, 88–109.

Styns, F., van Noorden, L., Moelants, D., and Leman, M. (**2007**). "Walking on music," Hum. Mov. Sci. **26**, 769–785.

van Noorden, L., and Moelants, D. (**1999**). "Resonance in the perception of musical pulse," J. New Music Res. **28**, 43–66.

Wohlschläger, A., and Koch, R. (**2000**). "Synchronization error: An error in time perception," in *Rhythm Perception and Production*, edited by P. Desain and W. L. Windsor (Swets and Zeitlinger, Lisse, The Netherlands), pp. 115–127.

Woodrow, H. (**1911**). "The role of pitch in rhythm," Psychol. Rev. **18**, 54–77.