How does sequence structure affect the judgment of time? Exploring a weighted sum of segments model

William J. Matthews
University of Essex

William J. Matthews
Dept. of Psychology
University of Essex
Colchester
CO4 3SQ
United Kingdom

Tel.: +44 1206 873818
Fax: +44 1206 873801
E-mail: will@essex.ac.uk

Published as : Matthews, W.J. (2013). How does sequence structure affect the judgment of time? Exploring a weighted sum of segments model. *Cognitive Psychology*, *66*, 259-282.

Abstract

This paper examines the judgment of segmented temporal intervals, using short tone sequences as a convenient test case. In four experiments, we investigate how the relative lengths, arrangement, and pitches of the tones in a sequence affect judgments of sequence duration, and ask whether the data can be described by a simple weighted sum of segments model. The model incorporates three basic assumptions: (i) the judgment of each segment is a negatively accelerated function of its duration, (ii) the judgment of the overall interval is produced by summing the judgments of each segment, and (iii) more recent segments are weighted more heavily. We also assume that higher-pitched tones are judged to last longer. Empirically, sequences with equal-sized segments were consistently judged longer than those with accelerating or decelerating structures. Furthermore, temporal structure interacted with duration, such that accelerating sequences were judged longer than decelerating ones at short durations but the effect reversed at longer durations. These effects were modulated by the number of tones in the sequence, the rate of acceleration/deceleration, and whether the sequence had ascending or descending pitch, and were well-described by the weighted sum model. The data provide strong constraints on theories of temporal judgment, and the weighted sum of segments model offers a useful basis for future theoretical and empirical investigation.

Keywords: Temporal judgment; time perception; internal clock; filled duration illusion

How does sequence structure affect the judgment of time?
Exploring a weighted sum of segments model

People often judge temporal intervals that are divided into subintervals. Actions, speech, travel, and music all involve sequences of events which divide an overall time period into smaller segments. This article examines how sequence structure affects the judgment of duration. We go beyond the historical focus on the difference between intervals which are or are not subdivided, and concentrate instead on the relationships between the elements that make up the stimulus. These relationships are defined by the relative lengths of the segments, their content, and the order in which they occur. We use tone sequences because of the ease with which the duration and content of the sub-intervals can be manipulated, and focus on intervals in the range 0.5-2.0 seconds – long enough to invoke "cognitive" timing, but short enough that counting strategies or external cues are unlikely to come into play. The aims are (i) to provide a substantial body of empirical data which establishes the effects of sequence structure on judgments of duration, and (ii) to see whether one simple model of temporal judgment can account for these effects. In what follows, we first outline existing work on how temporal structure affects temporal judgment. We then consider competing models of time perception. Finally, we outline a simple *weighted sum of segments* model.

Temporal structure

Many studies have focussed on the effects of segmenting a time interval. A typical approach involves comparing an empty interval (defined by two clicks or light flashes) with one which is sub-divided by the presence of additional pulses between the end markers. The usual finding is that segmenting the interval in this way increases its subjective duration, a result which has been known for at least 120 years (Hall & Jastrow, 1889) and replicated many times since (e.g., Adams, 1977; Buffardi, 1971; Israeli, 1930; Nakajima, 1987; Thomas & Brown, 1974). Increasing the subdivision of the interval by adding more markers seems to increase the subjective duration further (Adams, 1977; Buffardi, 1971).

Few authors have examined how the temporal structure of the event sequence – the relative lengths and ordering of the subintervals – affects perceived duration. Israeli (1930) found that the increase in subjective duration caused by interposing a light flash between the end markers diminished as that flash was shifted towards the end of the interval. Similarly, Grimm (1934) found that a sequence of 3 clicks was judged shorter when the second click was near the end rather than near the beginning. More importantly, a sequence of 6 clicks was judged longer when the clicks were evenly-spaced than when they were irregular (see also

Thomas & Brown, 1974). Buffardi (1971) found that clustering the intervening markers near the beginning of the interval led to longer subjective duration than clustering them near the end (see also Adams, 1977). This result held in the auditory, tactile, and visual modalities, although Buffardi only examined one duration. Similarly, Matthews (2011) recently found that shapes moving with constant speed were judged to have been presented for longer than decelerating shapes shown for the same objective duration, which in turn seemed longer than accelerating shapes.

The current article builds on these studies and focuses on three types of temporal structure: constant-rate sequences, where each element has the same duration; accelerating sequences, where each element is shorter than the last; and decelerating sequences, where each element is longer than the last. Unlike the arbitrary clusterings used by Buffardi (1971) and Grimm (1934), the segment lengths of accelerating and decelerating sequences are lawfully related. More importantly, we examine how the differences between these three temporal structures depend on the total length of the interval. We use sequences of contiguous tones rather than using empty intervals segmented by flashes or clicks, so that we can vary the content of each sub-interval as well as the temporal structure, and can examine how these factors interact. The default stimulus is a sequence of tones (not musical notes) whose frequencies uniformly increase on a logarithmic scale (that is, they have uniformly increasing pitch). The tones completely fill up the to-be-timed interval, with no empty gaps between them.

Models of temporal judgment

Many theories of temporal judgment assume that time is represented as the accumulation of counts or pulses from some kind of internal pacemaker (e.g., Creelman, 1962; Killeen & Fetterman, 1988; Taatgen, van Rijn, & Anderson, 2007; Treisman, 1963; Treisman, Faulkner, Naish, & Brogan 1990; see Rammsayer & Ulrich, 2001, for a general case). For example, Scalar Expectancy Theory (SET; Gibbon, Church, & Meck, 1984) posits a pacemaker gated by a switch which controls the flow of pulses into an accumulator, with the accumulated pulses providing a representation of stimulus duration. At stimulus offset, response decisions are based on the proportional discrepancy between the representation of the current duration (the number of pulses in the accumulator) and the representations of other durations stored in long-term memory (e.g., the number of pulses accumulated during intervals associated with a particular outcome). Within this framework, researchers have examined how particular manipulations affect the pacemaker rate (e.g., Matthews, Stewart, & Wearden, 2011; Meck, 1983; Penney, Gibbon, & Meck, 2000; Penton-Voak, Edwards, Percival, & Wearden, 1996), switch latencies (e.g., Matthews, 2011; Wearden, Edwards, Fakhri, & Percival, 1998), comparison process (Wearden & Grindrod, 2003), and the formation of reference memories (e.g., Jones & Wearden, 2003; Ogden, Wearden, & Jones, 2008; Rodríguez-Gironés & Kacelnik, 2001; Taatgen & van Rijn, 2011). Other researchers have added the idea that the flow of pulses depends on the amount of attention directed towards time (Zakay & Block, 1997). Although pacemaker models are widespread, there are alternative accounts. For example, Staddon and Higa (1999) have proposed that temporal information is represented by changes in memory strength; Matell and Meck (2000) suggest an oscillator-based account in which interval timing is based on the detection of coincident neural firing at the criterion duration; and Buonomano and colleagues have argued that temporal information is represented by time-dependent changes in local neural networks (e.g., Buonomano, 2000; Buonomano, Bramen, & Khodadadifar, 2009; Mauk and Buonomano, 2004; see Grondin, 2010, for a recent review of competing accounts).

Despite the success of these models, they are typically applied to cases where the stimulus is homogeneous (a steady light, for example, or a silent interval defined by two clicks). It is not clear what they predict for situations where the to-be-judged interval is divided into segments whose temporal structure and non-temporal features are varied, although some authors have examined the effects of inserting gaps in the stimulus or having participants time multiple intervals (e.g., Buhusi & Meck, 2009; van Rijn & Taatgen, 2008).

A separate class of models deals explicitly with the effects of subdivided time intervals, positing that subjective duration is positively related to the “storage size” of the events defining the interval (Ornstein, 1969), the degree to which the interval is segmented (Poynter, 1983, 1989), or the number of contextual changes taking place during the interval (Block & Reed, 1978). These ideas have received varying support, and the effects of complexity, segmentation, and change depend on moderating factors including the nature of stimulus processing and the presence/absence of a concurrent task (e.g., Predebon, 1996). In addition, these models are most applicable to durations lasting tens of seconds and where judgment is based on memory for the event sequence. For example, Zakay, Tsal, Moses, & Shahar (1994) found that segmenting stimulus sequences by distributing “high priority events” evenly throughout the series only increased temporal judgments in a retrospective (i.e., unexpected) temporal comparison task; when participants knew that they would have to reproduce a time interval, segmentation had no effect, suggesting that prospective absolute judgments are based on an “ongoing register of attended temporal units” (Zakay et al., 1994, p. 35) – as in the pacemaker accounts discussed above. Perhaps most importantly, it is not clear what these kinds of memory-based models predict regarding the effects of temporal structure. In particular, it is unclear whether judged duration will depend only the number of events occurring during an interval, or whether a change in the *rate* of events constitutes additional complexity or contextual change of the type that will lengthen subjective duration.

Jones and Boltz (1989) developed a general model for the effect of temporal structure on duration judgments. These authors were primarily concerned with hierarchically organized time structures, where the time periods of successive layers are related by simple ratios. For such stimuli, the temporal structure allows observers to engage in "future-oriented attending" and to anticipate the end of the sequence. Should the end occur earlier/later than expected, the stimulus sequence is perceived to be relatively short/long. For stimuli which lack this hierarchical, coherent structure, observers engage in "analytic attending" to adjacent elements and focus on low-level relationships by, for example, grouping or counting the elements. Although this dynamic attending model offers a general framework for temporal judgment, it has primarily been applied to musical passages lasting tens of seconds and with high coherence (e.g., Boltz, 1993; Jones, Boltz, & Klein, 1993). Brief, non-musical sequences which lack a coherent, hierarchical structure (like the ones explored in the current work) are likely to elicit analytic attending, so the predictions are unclear.

In short, some models describe in detail the processes by which individual intervals are judged but have little to say about the judgment of sub-divided time, whereas other models explicitly address the judgment of segmented intervals but their applicability to the kinds of stimuli under consideration here and their predictions regarding the effects of temporal and non-temporal structure are unclear.

A weighted sum of segments model

In this paper we explore a very basic weighted sum of segments model of temporal judgment. The model was developed in light of data from experiments examining the effects of temporal structure on temporal judgment, described below.

The core assumption is that the judged duration of a segmented interval is equal to the sum of the judged duration of the individual segments (Thomas and Brown, 1974). For a total time T divided into N segments each labelled t_i, the judged duration of each segment is a function of its physical duration, $j(t_i)$, and the judged duration of the overall time period is:

$$J(T) = \sum_{i=1}^{N} j(t_i) \qquad (1)$$

We further assume that $j(t_i)$ is a negatively-accelerated function – it increases as t_i increases, but the rate of increase diminishes as the interval becomes longer. This was suggested by

Thomas and Brown (1974), who explored it as one possibility within a general framework for temporal encoding and decoding which relates performance on different temporal judgment tasks; more broadly, the idea of a non-linear relationship between judged duration and physical time has been extensively discussed (see e.g., Allan, 1978; Eisler, 1976; Staddon & Higa, 1999; Taatgen et al., 2007; Wackerman & Ehm, 2006; Wearden & Jones, 2007).

Three important predictions follow from Equation 1. First, the judged duration of a given interval will increase as the number of segments is increased: because $j(t_i)$ is negatively-accelerated, $j(t_1 + t_2) < j(t_1) + j(t_2)$. Second, judged duration will be maximal when the segments are equal length. To see why, consider two segments of equal length, $t_1 = t_2$. What happens when t_1 is increased at the expense of t_2? Because $j(t_i)$ is negatively accelerated, the increase in $j(t_1)$ is less than the reduction in $j(t_2)$, so the judged duration of the total interval goes down.

The third prediction of this simple model is that the judgment of the overall interval only depends on the sizes of the segments, not on the order in which they occur. In particular, accelerating and decelerating sequences which are mirror images of one another will be judged equal.

As noted above, the first two predictions are widely supported by existing research. However, the third prediction is not. For example, Buffardi (1971) reported that sequences with markers clustered near the beginning were judged longer than those with markers clustered near the end. Thomas and Brown (1974) offer a plausible ad hoc explanation: perhaps the first few markers set a “pace” which is used to segment the rest of the interval. That is, when the first few segments are short (in a decelerating sequence), the observer continues to mentally divide the interval into small segments, increasing *J(T).*

The current experiments provide further evidence that the ordering of the segments is important, and argue strongly against the basic model outlined above. They are also difficult to reconcile with the pace-setting suggestion offered by Thomas and Brown (1974). In light of these findings (described below), the model incorporates the additional assumption that the contribution of each segment to the overall judgment depends on how recently the segment occurred (e.g., Taatgen & van Rijn, 2011). That is, when summing the $j(t_i)$, more recent segments are weighted more heavily. We also assume that the judged duration of a tone is positively related to its frequency – that is, that higher pitch tones are judged to last longer. Surprisingly little work has examined the effects of tone frequency on apparent duration, but the available evidence tends to suggest a positive relationship (Brigner, 1988; Cohen, Hansel, & Sylvester, 1954; Yu, 2010; for a conflicting result, see Yoblick and Salvendy, 1970). The assumption of a positive effect of pitch was also partly motivated by the differences between ascending and descending tone sequences that were found in the current experiments.

Together these assumptions form the basis for the weighted sum of segments model.

The details of the model

The details of the model are as follows. Judged duration is given by an elaborated version of Equation 1. Several negatively-accelerated functions might be used for $j(t)$. I elected to use a power function, $j(t) = at^b$, because it is convenient and relatively flexible, and because existing work has found that judgments are related to physical durations by a power function with exponent less than one (see e.g., Eisler, 1976; Kowal, 1981; we discuss this choice further below.) To instantiate a positive relationship between the judged duration of a tone and its frequency, the *a* parameter of the power function was set to be linearly related the logarithm of the segment's frequency (that is, linearly related to its pitch): $a = u + v\ln(f_i)$, where f_i is the frequency in Hz.

In order for more recent segments to contribute more to the judgment of the overall interval, the model assumes that when the $j(t_i)$ are summed, each value is weighted according to the time since the end of that segment, with the weighting determined by an exponential decay in physical time (e.g., Wickelgren, 1966). Exponential decay is a convenient, widespread assumption in memory modelling (see e.g., Kahana & Adler, 2012) which "has natural appeal, because many processes in nature (e.g., radioactive decay) manifest the same exponential property" (Anderson & Tweney, 1997, p. 724). Specifically, the weights are given by $w + \exp(-rd_i)$ where d_i is the time, in milliseconds, since the end of the *i*th segment. The weight given to the judgment of a particular segment therefore decays exponentially as the time since the end of that segment increases, from a maximum of *w*+1 to a minimum of *w*. Note that d_i is simply the sum of the durations of the segments that come after the *i*th segment, so

$$d_i = \sum_{j=i+1}^{N} t_j$$

We assume that the weighted summation takes place as soon as the final segment has completed, so d_N is always zero.

Combining these ideas, the judged duration of a sequence can be written as:

$$J(T) = (w+1)(u + v\ln(f_N))t_N^b + \sum_{i=1}^{N-1}(w + e^{-r\sum_{j=i+1}^{N} t_j})(u + v\ln(f_i))t_i^b \qquad (2)$$

The first part of the equation describes the contribution of the last segment, which always receives maximal weighting; the second part (beginning with the sum over *i*) describes the contributions of the earlier segments, which are weighted according to how long in the past they finished.

Finally, the judgment must be mapped onto the response scale. The experiments reported here used a form of category judgment: participants pressed a button to indicate which of (say) 9 possible durations the sequence had. To predict mean responses for a given experiment, the *J*(*T*) values were scaled to lie within the range of observed mean responses. That is, the predicted responses for the conditions with the smallest and largest *J*(*T*) were set equal to the smallest and largest observed mean responses, respectively. The predicted responses for the remaining conditions were then determined according to their distance from the smallest and largest values. (Note that the conditions giving rise to the smallest/largest predicted judgments need not correspond to the conditions actually producing the smallest/largest mean responses.)

To clarify: we first normalize the *J*(*T*) value for each condition to lie between 0 and 1:

$$J_k = \frac{J_k - J_{\min}}{J_{\max} - J_{\min}} \qquad (3)$$

where J_k is the *J*(*T*) for the *k*th condition of the experiment and $J_{\min}$ and $J_{\max}$ are the smallest and largest *J*(*T*) values for the experiment. We then produce the predicted mean response for the *k*th condition:

$$R_k = O_{\min} + (O_{\max} - O_{\min})J_k \qquad (4)$$

where $O_{\min}$ and $O_{\max}$ are the smallest and largest observed mean responses in the experiment. This kind of normalization is common in category judgment and provides a straightforward way to examine the relative judgments of the various experimental conditions.

The model as outlined here is deterministic: there is no random variation in the predicted judgments. Such variation could be added, but for now we focus on fitting the model to mean judgments. This is partly for simplicity, partly because the primary empirical focus concerns perceived durations rather than variability across trials (Grondin, 2010), and partly because, for category judgment tasks like the ones used here, it may be hard to disentangle noise in the subjective representation of time from noise that arises from mapping this representation onto a numeric response scale. Note also that although the model makes broadly plausible assumptions, the mathematical formulation is not grounded in specific hypotheses about the mechanisms at work. The power function for the judgment of each subinterval, the linear function for pitch, and the exponential decay for recency-weighting are conveniences which are not based on detailed theorizing about the underlying processes.

Overview of the current research

Below we report four experiments that investigate judgments of tone sequences with accelerating, decelerating, or constant-rate temporal structures. Experiment 1 compares accelerating, decelerating, and constant-rate sequences of various durations. Experiment 2 examines the effects of changing the rate of acceleration and deceleration. Experiment 3 manipulates the number of tones. Experiments 4 compares sequences with ascending and descending pitch. The data provide a thorough investigation of the effects of sequence structure on the judged duration of one type of stimulus.

Alongside the empirical investigation, we explore the adequacy of the weighted sum of segments model. We focus on whether the model can capture the qualitative effects of the various manipulations. The model has five free parameters. Four of these were fixed for all experiments: u = 0.5, v = 0.2, w = 3, and r = 0.0075. The b parameter (the exponent of the power function) was varied slightly from experiment to experiment. Better fits could be obtained by formal model fitting with separate parameter adjustment for each study, but the emphasis here is on whether the simple weighted sum model can capture the key qualitative patterns in this data set, and the parameter values were chosen to give a reasonably good fit to the data.

General Method

The stimuli were sequences of contiguous tones. All tones had equal amplitude and included 2.5 ms cosine ramps at onset and offset to avoid click artefacts. The stimuli were prepared at a sampling frequency of 44100 Hz and recorded as 16 bit wav files. They were presented binaurally over Sennheiser HD580 Precision headphones. Stimulus presentation was controlled by DMDX (Forster & Forster, 2003).

Participants completed the experiment in sound-attenuating booths. Each trial began with a blank interval for 1 second followed by presentation of the stimulus, after which participants indicated their judgment of the duration of the sequence by pressing a response button. Each button corresponded to a particular duration: in Experiments 1 and 2, 8 buttons were labelled from 0.4 to 1.8 seconds in 0.2 second increments; in the other experiments, 9 buttons were labelled from 0.4 to 2.0 seconds in 0.2 second increments (the larger response scale was motivated by the use of longer durations in these experiments). This response format was used because it provides a straightforward way for participants to register their judgments on a familiar, socially-agreed scale which avoids some of the problems associated with reproduction tasks (e.g., Droit-Volet, 2010; Matthews, 2011c). Matthews (2011b) found that the effects of stimulus dynamics were the same when participants made judgments in milliseconds as when they completed a category rating task anchored at “very short” and “very long”.)

The participants were staff and students from the University of Essex. They took part for course credit or a small payment. Unless otherwise noted, fresh participants were recruited for each experiment.

Experiment 1

Experiment 1 simply compared accelerating, decelerating, and constant-rate tone sequences with durations ranging from 0.6-1.2 seconds.

Method

Thirty three participants took part (20 female, ages 18-38 years, *M* = 24.1 years, *SD* = 4.6 years). The stimuli were sequences of 5 consecutive tones of increasing pitch. The first tone had a frequency of 400 Hz and each successive tone had a frequency 1.25 times that of the previous one. The total duration of each sequence was 0.6, 0.8, 1.0, or 1.2 seconds.

In the constant-rate condition, each tone in the sequence had equal duration (120 ms for the 0.6 s sequence, 160 ms for the 0.8 s sequence, and so on). In the decelerating condition, each tone in the sequence was longer than the previous one by a factor of 1.5. Similarly, in the accelerating condition each tone was shorter than the last by a factor of 1/1.5 = 0.66. Thus, the temporal structures of the accelerating and decelerating conditions were mirror images of each other.

Participants completed 11 blocks of 12 trials. Each block consisted of one of each duration-temporal structure combination, in random order. The first block was treated as practice and not analysed.

Results

The mean judgments are plotted in the left panel of Figure 1. (Raw data from all experiments are available from the author.) A 3x4 within-subjects ANOVA showed that judgments were larger for longer stimuli, $F(1.84, 58.93) = 293.23$, $p < .001$, $\eta_p^2 = .90$, with both linear and quadratic trends [$F(1,32) = 401.12$, $p <. 001$, $\eta_p^2 = .3$ and $F(1,32) = 31.46$, $p < .001$, $\eta_p^2 = .50$, respectively]. (Here and at various points below, a Huynh-Feldt correction has been used because of violations of sphericity.) Judgments differed between the accelerating, decelerating, and constant-rate conditions, $F(1.17, 37.42) = 5.54$, $p = .022$, $\eta_p^2 = .14$. However, the effects of temporal structure were moderated by duration, $F(6, 192) = 3.36$, $p = .004$, $\eta_p^2 = .10$. At the shortest duration, the accelerating stimuli were judged longer than the decelerating ones, but this pattern reversed as the physical duration of the stimuli increased.

Discussion

Experiment 1 produced three results. First, the judgment function was slightly curvilinear. Second, judgments were longest for constant-rate sequences. Third, temporal structure interacted with physical duration: at short durations, accelerating sequences were judged longer than decelerating ones but at longer durations this difference reversed. These three results are repeated in all of the current experiments, and the pattern depicted in Figure 1 provides a basic constraint on models of time perception.

The right panel of Figure 1 shows that the weighted sum of segments model captures the results fairly well (*b* was set to 0.35). Indeed, this pattern partly motivated the model. The prediction that constant-rate sequences will be judged longest follows from the negatively-accelerated relationship between the length of a segment and its internal representation; so does the curvilinear form of the judgment function. The cross-over pattern for the accelerating and decelerating structures is more surprising, and motivates the assumption that recent segments are weighted more heavily.

Consider first the shortest (600 ms) stimuli. Recall that the weighting of each segment depends on the time between the end of that element and the end of the whole sequence. For the accelerating structure, the ends of the segments are clustered near to the end of the sequence, so all segments are subject to little decay and contribute substantially to the sum (specifically, segments 1-5 receive weightings of 3.06, 3.20, 3.43, 3.71, and 4.00; note that with $w = 3$, the weights decay from a maximum of 4 to a minimum of 3). For the decelerating structure, the last segment makes a large contribution but the weight given to the earlier segments is relatively small: the length of the final segment means that the representations of the earlier segments have decayed considerably (the weights for segments 1-5 are 3.02, 3.03, 3.06, 3.18, and 4.00). Thus, the weighted sum is larger for the accelerating sequence, which is correspondingly judged to be longer. Now consider the longer-duration (1200 ms) stimuli obtained by doubling the length of each segment of the 600 ms stimuli. This doubling increases the decay time for each non-terminal segment, diminishing their contribution to the sum so that the final segments dominate the judgment for the accelerating structure as well as for the decelerating one (for the accelerating sequence the weights are 3.00, 3.04, 3.18, 3.51, and 4.00; for the decelerating sequence they are 3.00, 3.00, 3.00, 3.03, 4.00). Because the highly-weighted, final segments are longer in the decelerating case, this type of sequence has the greater subjective duration. In addition, the sequences in this experiment had ascending pitch, so the last tones also had the highest frequency. In the model, pitch multiplies with duration to determine the judgment of each segment, which serves to amplify the advantage for decelerating stimuli at longer durations.

In short, for brief stimuli the accelerating structure was judged longer because many of the segments occurred recently and were weighted heavily. At longer durations, the decelerating structure seemed longer because the preferential weighting of recent segments only extends to the final items in the sequence, which were longer (and higher-pitched) in the decelerating sequence.

The model's predictions regarding this cross-over pattern depend on the parameter values, as illustrated in Figure 2. The top two panels show the effects of changing the exponential decay parameter (with all other parameters as for Figure 1). With slow decay (r = 0.005, left panel), many of the segments in the accelerating sequence continue to receive high weighting even at quite long overall durations, so the cross-over point is shifted to the right. With more rapid decay (r = 0.01, right panel), the last segments dominate even at short overall durations, so decelerating sequences are judged longer than accelerating ones even for brief stimuli, and the cross-over point is shifted to the left. Similarly, the bottom panels of Figure 2 show the effects of changing the b parameter (with all other parameters as for Figure 1). When b is 0.25 (left panel), the relationship between the subjective and objective duration of each segment is highly non-linear and sub-dividing a long interval into multiple short segments produces a big increase in judged duration. Correspondingly, accelerating sequences (where the final, heavily-weighted part of the sequence comprises many small segments) are judged longer than decelerating ones even at long overall durations (that is, even when decay means that the preferential weighting of the last few segments is weak). When b is increased to 0.45 (right panel), the relationship between the subjective and objective duration of each segment is more linear, reducing the advantage that comes from giving high weighting to lots of small segments and shifting the cross-over point to the left. (In the limit where b = 1, decelerating structures are consistently judged longer than both accelerating and constant-rate sequences.) Thus, the model's ability to capture the empirical pattern in Figure 1 rests on the interplay between the non-linear growth of the judged duration of individual segments and the diminishing weight given to segments which occurred further in the past.

Experiment 2

Experiment 2 manipulated the rate of acceleration/deceleration.

Method

Thirty three participants took part (22 female, ages 18-35 years, M = 23.5 years, SD = 4.6 years). As in Experiment 1, the stimuli were sequences of 5 contiguous tones of increasing pitch where the first tone had a frequency of 400 Hz and each successive tone had a frequency 1.25 times that of the previous one. The total duration of each sequence was either 800 ms or 1400 ms. In the constant-rate condition, each tone in the sequence had equal duration (160 ms for the 800 ms sequence and 280ms for the 1400 ms sequence). In the decelerating conditions, each tone in the sequence was longer than the previous one, by a factor of 1.25 in the gradual deceleration condition and by a factor of 1.5 in the rapid deceleration condition. Similarly, in the gradual acceleration condition each tone was shorter than the last by a factor of 1/1.25=0.8 and in the rapid acceleration condition each tone was shorter than the last by a factor of 1/1.5 = 0.66. (Thus, the temporal structures of the accelerating and decelerating sequences were mirror images of each other.)

Participants completed 11 blocks of 12 trials. Each block consisted of two accelerating (one rapid acceleration, one gradual acceleration), two decelerating (one rapid deceleration, one gradual deceleration) and two constant-rate stimuli at each of the two durations, in random order. The first block was treated as practice and not analysed.

Results

Mean judgments are plotted in the top left panel of Figure 3. To clarify the findings, the second panel shows the results for the accelerating and decelerating sequences after collapsing over gradual and rapid change. The third panel shows the interaction between acceleration/deceleration and rate of change. The bottom panel shows the results for each temporal structure after collapsing over duration.

As in Experiment 1, the constant-rate stimuli were judged longer than all others at both durations. At short durations, accelerating sequences stimuli were judged longer than decelerating ones, and rapid change stimuli were judged shorter than gradual change sequences for both accelerating and decelerating structures. The differences between conditions diminished as the duration of the sequence increased to 1400 ms, mimicking the interaction between temporal structure and duration seen in Experiment 1.

These impressions were supported by analysis of variance. A 2x5 within-subjects ANOVA with duration (800 ms, 1400 ms) as one factor and temporal structure (constant rate, gradual deceleration, gradual acceleration, rapid deceleration, rapid acceleration) as the other found that judgments were larger for longer stimuli, $F(1,23) = 134.27$, $p<.001$, $\eta_p^2 = .85$. Judgments were also affected by temporal structure, $F(1.71, 39.43) = 6.22$, $p = .006$, $\eta_p^2 = .21$, and this effect depended on the physical duration, $F(4,92) = 3.56$, $p = .010$, $\eta_p^2 = .13$.

To clarify the effects of the rate of acceleration and deceleration, an additional 2x2x2 ANOVA was conducted with duration, temporal structure (accelerating vs. decelerating) and rate (gradual vs. rapid) as within-subject factors. Again, longer stimuli were judged longer, $F(1,23) = 128.1$, $p < .001$, $\eta_p^2 = .85$, and the difference between the accelerating and decelerating conditions diminished as duration increased, $F(1,23) = 4.91$, $p = .037$, $\eta_p^2 = .18$ (see second panel of Figure 3). More interestingly, there was a main effect of rate of change: gradual-change stimuli were judged longer than rapid-change stimuli, $F(1,23) = 14.26$, $p = .001$, $\eta_p^2 = .38$, and this effect was more pronounced for the decelerating sequences than for the accelerating ones, $F(1,23) = 5.41$, $p = .029$, $\eta_p^2 = .19$; these effects can be seen in the third panel of Figure 3. No other effects were significant (*F*s<2.4, *p*s >.13).

Discussion

Experiment 2 replicates and extends the results of Experiment 1. Constant-rate sequences were judged longest and there is an interaction between temporal structure and duration such that accelerating stimuli were judged longest at short durations but this advantage disappears as the duration lengthens. Furthermore, the gradually accelerating/decelerating stimuli were judged longer than the rapidly changing ones. This is what one would expect from the weighted sum of segments model; the segments of the gradually accelerating and decelerating sequences are closer to equal size, and therefore closer to the arrangement that maximizes judged duration.

The right-hand panels of Figure 3 show that the weighted sum of segment model captures these qualitative trends well. The only difference from Experiment 1 is that the power function exponent *b* has been reduced slightly to 0.22 to accommodate the empirically-observed change in the cross-over point between accelerating and decelerating sequences. (Recall that in Experiment 1stimuli identical to the rapidly decelerating sequences here were judged longer than rapidly accelerating sequences at 1.2 seconds, whereas in this experiment the lines have not quite crossed at 1.4 seconds.) The bottom right panel shows that the model slightly underestimates the judged duration of the rapidly accelerating items as compared to the gradually decelerating ones, but in general the weighted sum of segments model provides a good description of data.

Experiment 3

Experiment 3 varied the number of tones making up each sequence.

Method

Thirty two participants took part (16 female, ages 18-50 years, M = 25.1 years, SD = 6.5 years). As before, the stimuli were sequences of contiguous tones of increasing pitch. The sequences consisted of 4 tones in the 4-tone condition and of 6 tones in the 6-tone condition. As in the preceding experiments, the first tone had a frequency of 400 Hz and each successive tone had a frequency 1.25 times that of the previous one.

The total duration of each sequence was 0.6, 1.2, or 1.8 seconds. In the constant-rate condition, all tones in the sequence had equal duration. In the decelerating condition, each tone in the sequence was longer than the previous one by a factor of 1.5. In the accelerating condition each tone was shorter than the last by a factor of 1/1.5 = 0.66. (Thus, the accelerating and decelerating conditions were like those of Experiment 1 and like the rapid change conditions of Experiment 2.) Participants completed 9 blocks of 18 trials. Each block comprised one of each duration x temporal structure x number of tones combination, in random order. It was clear after the first two experiments that participants had no trouble understanding the task, so practice blocks were not included in this and subsequent studies.

Results

The mean judgments are plotted in the left hand panels of Figure 4. The 6-tone sequences show a pronounced effect of temporal structure that depends on sequence duration, whereas the 4-tone sequences do not.

An initial ANOVA established that judgments were larger for longer stimuli, $F(1.22, 37.86) = 445.87$, $p<.001$, $\eta_p^2 = .94$, with both linear and quadratic trends [$F(1,31) = 491.35$, $p < .001$, $\eta_p^2 = .94$ and $F(1,31) = 5.33$, $p = .028$, $\eta_p^2 = .15$, respectively]. Overall, 6-tone sequences were judged longer than 4-tones sequences, $F(1, 31) = 10.02$, $p = .003$, $\eta_p^2 = .24$. There was also a three-way interaction between duration, temporal structure, and number of tones, $F(4, 124) = 5.76$, $p < .001$, $\eta_p^2 = .16$, so the data for the 4-tone and 6-tone sequences were analysed with separate 3 (duration) x 3 (temporal structure) ANOVAs. For the 4-tone sequences, judgments were larger for longer stimuli, $F(1.28, 39.64) = 408.61$, $p < .001$, $\eta_p^2 = .93$, but the effect of temporal structure was not significant, $F(1.58, 48.90) = 3.33$, $p = .055$, $\eta_p^2 = .10$, and there was no interaction between structure and duration, $F(4,124) = .89$, $p = .471$, $\eta_p^2 = .03$. For the 6-tone sequences, judgments were larger for longer stimuli, $F(1.50, 45.22) = 352.12$, $p < .001$, $\eta_p^2 = .92$, and there was an effect of structure, $F(2,62) = 19.42$, $p < .001$, $\eta_p^2 = .39$ which was modulated by stimulus duration, $F(4,124) = 8.19$, $p < .001$, $\eta_p^2 = .21$. As before, accelerating sequences were judged longer than decelerating sequences at short durations, but decelerating sequences were judged longer at long durations.

Discussion

The finding that 6-tone sequences were judged longer than 4-tone sequences replicates previous work showing that empty intervals that contain more markers are judged to last longer (e.g., Buffardi, 1971). The data for the 6-tone sequences replicate the cross-over interaction between temporal structure and duration found above.

The right hand panels of Figure 4 show the predictions of the weighted sum of segments model. (The power exponent *b* has again been set to 0.35.) The model captures the qualitative trends reasonably well. As noted above, the negatively accelerated function relating the judgment of each segment to its physical duration entails the prediction that 6-tone sequences will be judged longer than 4-tone sequences (see Thomas & Brown, 1974). However, the model over-estimates the effect, a shortcoming we return to in the General Discussion. The model also captures the greater overall effect of temporal structure for the 6-tone sequences than for the 4-tone sequences. (Collapsing across duration, the predicted difference between the constant-rate stimuli and the accelerating stimuli is 25 ms for the 4-tone sequences and 58 ms for the 6-tone sequences. In the empirical data the corresponding values are 29 ms and 112 ms.) However, although the model successfully predicts the cross-over pattern for the 6-tone stimuli, it rather underestimates the magnitude both of this effect and of the difference in the size of this interaction for the 6-tone and 4-tone sequences (where, empirically, the interaction is not significant).

In short, the model captures the qualitative effects fairly well, but the quantitative predictions sometimes slightly off. We return to this in the General Discussion, but it is worth noting in passing that some of the mis-prediction may reflect noise and instability in the data rather than a problem with the model. For example, other studies which use stimuli identical to the 6-tone sequences of this experiment sometimes produce weaker interactions between duration and temporal structure which are more in line with the model predictions (see Supplementary Materials). Similarly, there is reason for caution about the lack of significant interaction in the 4-tone case: Experiments 1 and 2 both produced significant interactions between duration and temporal structure with similar sequences of 5 tones.

Experiment 4

Up to this point, all sequences have had ascending frequencies. Experiment 4 examined the effects of pitch direction.

Method

Forty participants took part (29 female, ages 18-35 years, M = 25.0 years, SD = 5.3 years)[1]. The stimuli had the same temporal structure as the 6-tone stimuli from Experiment 3. The sequences were 0.6, 1.2, or 1.8 seconds long with accelerating, decelerating, or constant-rate structures, as before. In the ascending condition, the first tone had a frequency of 400 Hz and each successive tone had a frequency 1.25 times the previous one (up to 1220.7 Hz for the last tone). In the descending condition, the first tone had a frequency of 1220.7 Hz and each successive tone had a frequency 0.8 times the previous one. Thus, the ascending sequences were identical to the 6-tone stimuli of the previous experiment, whilst the descending stimuli were similar except that the pitch relations between successive elements of the sequence was reversed.

Participants completed 15 blocks of 18 trials. Each block contained one occurrence of each duration x temporal structure x pitch change direction combination in random order, with the opportunity to take a break between blocks.

Results

The mean judgments are plotted in the left panels of Figure 5. The first and second panels show the results for the ascending and descending sequences, respectively. Inspection suggests that pitch direction moderates the effects of temporal structure, and this was

[1] Sixteen of these had participated in one of the earlier experiments in this series (including those in the Supplementary Materials). The time since this past experience was several months, and preliminary analysis showed that prior experience did not moderate any of the effects reported here.

confirmed by a three-way interaction between duration, temporal structure, and pitch direction, $F(4,156) = 2.73$, $p = .031$, $\eta_p^2 = .07$. For the ascending sequences, the pattern is the same as in previous experiments: constant-rate stimuli were judged longest, and there is an interaction between structure and duration such that the decelerating sequences come to be judged longer than the accelerating ones as the total duration increases. Correspondingly, a 3x3 within-subjects ANOVA showed a positive effect of sequence duration, $F(1.23, 48.02) = 236.93$, $p<.001$, $\eta_p^2 = .86$ [linear trend $F(1,39) = 259.63$, $p <.001$, $\eta_p^2 = .87$; quadratic trend $F(1,39) = 33.73$, $p<.001$, $\eta_p^2 = .46$], a main effect of temporal structure, $F(2,78) = 9.84$, $p <.001$, $\eta_p^2 = .20$, and a structure x duration interaction $F(4,156) = 7.44$, $p <.001$, $\eta_p^2 = .16$).

The pattern is different for the descending sequences; constant-rate stimuli were still judged longest, but there is little indication of the cross-over between accelerating and decelerating stimuli that characterizes the ascending sequences. Correspondingly, a 3x3 ANOVA show a main effect of duration, $F(1.14, 44.35) = 202.49$, $p<.001$, $\eta_p^2 = .84$ [linear trend $F(1,39) = 217.72$, $p<.001$, $\eta_p^2 = .85$; quadratic trend $F(1,39) = 42.09$, $p<.001$, $\eta_p^2 = .52$], a main effect of temporal structure, $F(2,78) = 12.89$, $p<.001$, $\eta_p^2 = .25$, but no interaction between structure and duration, $F(4,156) = 1.86$, $p=.120$, $\eta_p^2 = .05$.

Discussion

The right hand panels of Figure 5 show the predictions of the weighted sum of segments model (with *b* set to 0.45). The model captures the patterns in the data, including the finding that the interaction between temporal structure and duration is more pronounced for ascending sequences than descending ones. To understand why, recall the model's explanation for why duration modulates the influence of temporal structure. At short durations there is little decay in the representation of the last few segments, so accelerating sequences (which have a large number of short recent segments) are judged longer than decelerating ones (which have a single long recent segment). As duration increases, the weighted sum is increasingly influenced by the last segment of the sequence. This is larger in decelerating sequences, so these come to be judged progressively longer than accelerating stimuli. For ascending sequences this *weighting-based* effect is amplified by a *pitch-based* effect. Pitch multiplies with (transformed) duration to produce the judgment of each segment (see equation 2). For ascending sequences, this increases the judgment of decelerating stimuli relative to accelerating ones, because in the decelerating case the longest segments also have the highest pitch. For descending sequences, the reverse obtains: judgments tend to be larger for accelerating sequences. This two-way interaction between temporal structure and pitch direction is visible in the third row of Figure 5. Furthermore, the multiplication of pitch and duration in the model produces a three-way interaction in which decelerating stimuli are judged progressively longer than accelerating ones for ascending sequences but the effect is weakened for descending sequences (because the weighting-based effect and the pitch-based effect act in opposition, the former producing an increasing advantage for decelerating stimuli but the latter producing an increasing advantage for accelerating stimuli). Thus, the interaction between temporal structure and duration is more pronounced for ascending sequences than descending ones, as shown in the top panels of Figure 5.

General Discussion

A summary of the empirical findings is given in Table 1. In what follows, we first consider the effects of temporal structure and the adequacy of the weighted sum of segments model as an account of these effects. Next we consider alternative theoretical perspectives. Finally, we outline challenges for the model and important directions for future research.

Temporal Structure

The experiments show that the temporal structure of a sequence (the relative lengths and arrangement of sub-intervals) has a substantial effect on temporal judgment. Sequences with elements of equal size were consistently judged longer than accelerating or decelerating structures. This mirrors the result reported by Matthews (2011) for continuously-moving visual shapes, and findings from studies of empty intervals filled with clicks reported by Buffardi (1971), Grimm (1934), and Thomas and Brown (1974). In addition, temporal structure interacted with physical duration: at longer durations (c. 1.8 seconds) the decelerating items were judged substantially longer than accelerating ones, but at shorter durations (c. 0.6 s) the effect was weaker or reversed. This novel pattern replicated throughout the experimental series. Furthermore, increasing the rate of acceleration/deceleration shortened subjective duration, and decreasing the number of tones both shortened subjective duration and lessened the effects of temporal structure.

The weighted sum of segments model captures these effects reasonably well. Indeed, it was designed to. Perhaps the most conspicuous shortcoming is that the model overestimates the effect of increasing the number of tones in the sequence (Figure 4). This discrepancy can be reduced by adjusting the model parameters, but there are also at least two reasonable post hoc explanations for the problem. One is that the 4-tone and 6-tone sequences may be sufficiently perceptually different as to be treated as two different categories of stimulus. In category judgment tasks like the ones used here, items are typically judged relative to one another (e.g., Brown, McCormack, Smith, & Stewart, 2005; Parducci, 1965). That is, each item is compared with the other items in the set and the final response represents the relative magnitude of the item. People can simultaneously maintain two comparison groups in parallel (e.g., Petzold & Haubensak, 2004) and may therefore judge each 4-tone sequence relative to the other 4-tone stimuli, and similarly for the 6-tone sequences, with the result that the category judgments will be similar for both types. (Taatgen and van Rijn, 2011, have likewise argued that people may maintain two different categories of duration whose representations nonetheless influence one another.) A second possibility is that judgments are actually a compromise between the weighted sum of segments and a separate representation of the overall interval. In other words, as well as timing each sub-interval and adding them up, participants form a separate representation of the overall sequence length, and their final judgment is a combination of the two. This would again have the effect of reducing the difference between 4-tone and 6-tone sequences, because their overall physical duration is the same.

The weighted sum of segments model attributes the larger judgments for constant-rate stimuli to the summation of negatively-accelerated representations of each segment. To the extent that the empirical effect was found in every experiment here, the assumption seems reasonable. However, there is considerable debate about the function relating physical time to subjective time (see Allan, 1978, and Wearden & Jones, 2007, for an overview). Pacemaker-based mechanisms typically generate linear timing (see e.g., Gibbon, Church, & Meck, 1984). However, other timing mechanisms assume a non-linear, negatively-accelerated relationship. Wackerman and Ehm's (2006) "dual klepsydra" model of temporal reproduction, for example, incorporates a negatively-accelerated exponential relationship, and Staddon and Higa's (1999) memory-based model of animal timing posits a logarithmic function arising from a simple habituation process. Brown, Neath, and Chater's (2007) SIMPLE model of memory similarly assumes that psychological time is a logarithmic transformation of physical time. Elsewhere, studies of magnitude estimation typically report that judgments are related to physical durations by a power function with exponent less than one (e.g., Eisler, 1976; Kowal, 1981). Similarly, Friedman and Kemp (1998) found that estimates of the temporal distance to past events are a power function of event age, with an exponent of 0.2. In the current work it was assumed that the judged duration of each segment was a power function of its physical extent, but it may well be possible to replace this with an alternative such as a logarithmic or exponential function. One additional point to bear in mind is that responses in discrimination, reproduction, or category judgment tasks are not pure indicators of the subjective experience of time, and the function mapping physical time to temporal judgment need not be the same across tasks. For example, the weighted sum of segments account does not necessarily imply that participants sum the subjective durations of the sub-intervals; they may sum the *judgments* of those sub-intervals, which could be a power-law transformation of physical time without there necessarily being a power law "time code" (see, for example, Laming, 1997).

The other key component of the weighted sum of segments model is the assumption that more recent segments are weighted more heavily. This arose from the observed interaction between structure and duration. The particular form of the recency weighting used here was based on simple exponential decay, with the pragmatic adjustment that the representation does not fade to zero. The idea is that more recent elements of a sequence are more vivid, but that there is not an inexorable loss of duration information over time: events in the distant past do not seem to have infinitesimal duration. The idea that representations of basic perceptual properties like pitch and loudness decay over time is widespread (e.g., Clément, Demany, & Segal, 1999), and using an exponential function to model this process is a common strategy (e.g., Wickelgren, 1966). Of course, one could choose other weighting functions and other mechanisms for the weighting. For example, Taatgen and van Rijn (2011) used the declarative memory component of the ACT-R framework (Anderson et al., 2004) to model the effects of lag on the accessibility of reference durations in a temporal reproduction task, and one might try to apply the same approach here; similarly, accounts based on interference or item distinctiveness could also adapted to provide the segment weights. One intriguing possibility is to assume a step function, such that duration representations formed in (say) the last 100 ms receive one weighting and representations formed earlier receive another, smaller weighting. Preliminary exploration suggests that this approach captures some of the current data fairly well (although not as well as the exponential decay used here). The possibility that there is a relatively small "window" of heavily weighted sensation might be worth pursuing in future.

In sum, the weighted sum of segments model offers a reasonable account of the effects of temporal structure on judgments of duration. There are potential problems. The model is relatively flexible, so substantial changes in parameters produce large changes in the predictions. Moreover, although the parameter values were restricted (fixing four of the five parameters for all participants/experiments), it was necessary to adjust slightly the power exponent b between studies. (This seems reasonable because the same stimuli were judged differently in different experiments.) Finally, the model's success at capturing the different effects of temporal structure for accelerating and decelerating structures is partly due to the assumption that high-pitched tones have longer subjective duration (which amplifies the effect of having the longest segments at the end of the sequence in decelerating stimuli). As discussed below, the relationship between tone frequency and judged duration is rather vexed, and this aspect of the model may need modification in future.

Alternative models

We have explored a simple weighted sum of segments model in some detail. It is worth briefly considering how other models might cope with the data.

Nakajima(1987). Nakajima (1987) proposed an explanation for the illusion of divided time (the finding that intervals filled with more clicks are judged longer). As in the weighted sum of segments model, overall subjective duration is the sum of transformations of each sub-interval. However, the transformation is linear, with a constant intercept attributable to the time taken to process each marker. A linear transform cannot account for the difference between accelerating, decelerating, and constant rate stimuli or the effects of physical duration that were found throughout the current experiments.

Wackermann and Ehm (2006). Wackermann and Ehm (2006) have developed a model of temporal judgment based on the dynamics of a water clock (klepsydra). During stimulus timing, inflow is constant and outflow occurs at a rate proportional to the current accumulation (that is, the water clock leaks). Leakage continues after stimulus offset. The rate of change in accumulation is therefore $dy = (i - \kappa y)dt$, where *i* is the (constant) inflow rate. Solving this equation gives $y_t = y_0 e^{-\kappa t} + \bar{y}(1 - e^{-\kappa t})$, where y_0 is the state at *t* = 0 and $\bar{y} = i / \kappa$, the steady state equilibrium.

This account has some similarity to the weighted sum of segments model. We can imagine that each sub-interval is timed by a separate klepsydra. During each subinterval, the accumulation process gives negatively-accelerated growth in subjective time; at the end of the subinterval the activity decays until the end of the stimulus sequence, at which point the residual accumulation in all klepsydrae is summed to give the overall representation of duration. However, closer inspection shows that this idea will not work because the number and arrangement of sub-intervals will make no difference to the total accumulation. For example, consider a single interval of length $t_1 + t_2$. At the end of stimulus presentation, $y_{t_1+t_2} = {}^{i}\!/\!_{\kappa}\left(1 - e^{-\kappa(t_1+t_2)}\right)$. Now consider two consecutive intervals, t_1 and t_2. At the end of the second interval the associated klepsydra has accumulated $i/_{\kappa}\left(1-e^{-\kappa t_2}\right)$ and the klepsydra recording the first duration contains $i/_{\kappa}\left(1-e^{-\kappa t_1}\right)e^{-\kappa t_2}$, giving a total of $i/_{\kappa}\left(1-e^{-\kappa(t_1+t_2)}\right)$, as before.

Temporal Expectancy/Memory/Change-based accounts. As noted in the Introduction, most memory/expectancy-based theories do not make clear predictions regarding the kinds of stimuli and manipulations used here. However, it is worth mentioning an account from Block (2003; Block & Zakay, 2008) in which prospective timing involves comparing the ages of the start-of-duration and end-of-duration events in memory. As in the weighted sum of segments model, the perceived duration of a sequence is the sum of power transformations of the individual elements, although Block's (2003) account emphasizes segmentation based on spontaneous (attention-dependent) sampling of memory age whereas the current work emphasizes external demarcations causes by changes in tone pitch. As it stands, the memory age model cannot account for the interaction between temporal structure and duration, or for the effects of the non-temporal properties of the sequence (e.g., direction of pitch change). Nonetheless, it may provide a useful mechanistic basis for the more abstract weighted sum of segments model. Moreover, conceptualizing the weighted sum of segments model in this way ties it to the notion of attending to time in a manner that would make testable predictions about the interactions between temporal structure and the division of attention.

Pacemaker models. Many models assume that timing is based on the accumulation of "pulses" from a pacemaker or some kind of neural counting, typically producing linear growth in subjective time (e.g., Gibbon, Church, & Meck, 1984). Without additional assumptions, such models are silent about the judgment of sequences. One possibility is that the rate of the pacemaker is linked to the rate of change in the environment. Matthews (2011) suggested this as a possible explanation for the effects of changing speed on judgments of the duration of moving shapes. However, Beckmann and Young (2009) have modelled the effects of (constant) speed on temporal perception and argue that changes to pacemaker rate provide a poor description of the data. Rather, they suggest that stimulus change combines additively with physical duration to give perceived duration, which provides the basis for temporal judgments. In this view, physical duration is a proxy for the amount of background environmental change, and the moving stimulus is just another type of change which contributes to subjective time. As it stands, this idea will not work for the current data because the differences between accelerating, decelerating, and constant rate stimuli occur despite the fact that the total amount of change (the pitch-distance travelled and number of subintervals per unit of physical time) is the same across these conditions. One might introduce the assumption that second order dynamics (acceleration and deceleration) constitute more environmental change, in which case these stimuli should be judged longer than the constant-rate sequences – the opposite of what was found.

An alternative is to assume a pacemaker with a non-linear growth in subjective time. Taatgen et al. (2007) have developed such a model, embedding a pacemaker-based timing module within the ACT-R framework (Anderson et al., 2004), thereby connecting time perception to a broad-ranging account of cognition. Their model assumes that the interval between pulses increases over time, creating a negatively accelerated growth in subjective time similar to that posited by the weighted sum of segments model. This model has also been extended to incorporate the idea that stored durations decay over time. Taatgen and van Rijn (2011) had participants produce two intervals on alternating trials, with feedback, and the duration of one target interval was gradually shifted over the course of the experiment. These adjustments influenced productions of the other duration, indicating that times are represented by a pool of experiences with confusion between items. The results were captured by assuming memory traces whose accessibility declines logarithmically over time (and which also depend on the relevance of the encoded representation to the current goal). The negatively accelerated growth of subjective time and the decay of representations are similar to the assumptions of the weighted sum of segments model, although specific mathematical formulations are different.

Van Rijn and Taatgen (2008) have also applied Taatgen et al.'s (2007) pacemaker model to a timing task in which participants attempted to time two overlapping intervals. The start of each interval was signalled by a coloured light and the participant's task was to make a response when the interval was a certain length (2 or 3 seconds). The intervals overlapped such that the onset of the second interval occurred before the end of the first. van Rijn and Taatgen modelled performance by assuming a single pacemaker whose rate decreases over time and whose output is stored by a single accumulator. The first response is produced when the accumulated pulses match the number associated with the first target duration; the second response is produced when the pulses equal the number associated the first target duration, plus the number associated with the target-onset asynchrony, plus or minus the difference between the number associated with the two target durations (depending on whether the second interval is longer or shorter than the first). This approach has some commonality with the weighted sum of segments model, in that individual segments of an interval are timed and combined to produce temporal judgments. It differs, however, because the weighted sum model assumes that each sub-interval is represented as a separate negatively-accelerated function of its duration, whereas van Rijn and Taatgen's modelling posits that the durations of sub-intervals are abstracted from a single pacemaker-accumulator with a negatively-accelerated growth in accumulated pulses. The latter approach does not predict the finding that constant-rate stimuli will be judged longer than other temporal structures (indeed, it does not predict any effect of temporal structure) because the total number of pulses will be the same in all conditions. However, one could adapt van Rijn and Taatgen's approach by assuming that, for contiguous tone sequences like the ones used here, the accumulator resets at the start of each new segment with the outputs being stored and combined to obtain the overall estimate of duration.

In short, the pacemaker model of Taatgen and colleagues may well be able to capture the effects reported here, although one would need to adapt the assumptions about how subintervals are timed and combined and to assume that the pacemaker runs faster for higher-pitched tones. Examining the qualitative and quantitative performance of such a model will be a useful direction for the future.

Future challenges

Experiment 4 examined the effect of pitch direction and found that whether a sequence was ascending or descending modulated the effects of temporal structure and the interaction between structure and duration. In future experiments, it will be important to explore how other aspects of pitch structure interact with temporal structure. One obvious issue concerns how the size of the changes in pitch between successive tones affects the perception of time. Much of the previous work on pitch-jump effects has focussed on the auditory kappa effect, where increasing the frequency difference between a pair of tones increases the apparent duration of the empty interval between them (e.g., Cohen et al., 1954; Crowder and Neath, 1994; Henry & McAuley, 2009; Shigeno, 1986; Yoblick and Salvendy, 1970). Such studies focus on the judgment of individual temporal intervals between sounds rather than on judgments of the overall duration of a sequence of contiguous tones, which is the topic of the current work and is much less researched. In one relevant study, Sivyer and Finlay (1982) manipulated the average frequency change from one tone to the next and found no effect on the judged duration of 4-tone sequences, but their study had a very small sample size (10 participants). Tipps (1980) did find evidence that the size of pitch changes affects judgments of sequence duration, but the direction of the effect was "inconsistent" (p. 697), with larger pitch jumps sometimes lengthening and sometimes shortening apparent duration. More recently, Henry and McAuley (2011) presented four 100 ms tones with 400 ms between each one and had participants judge the overall sequence. Increasing the pitch jumps between successive tones (from 1 semitone to 7 semitones) had no effect on judgements of duration. However, this may be because of the large silent gaps between the tones: in a separate study of pitch velocity judgments, the effect of total frequency change was much greater for continuous pitch glides than for sequences of discrete, separated tones. Thus, an effect of frequency change on apparent duration may emerge if the tones are contiguous. Kowal (1981) used longer sequences (15-65 seconds, with tones and inter-tone intervals both lasting 300 ms) and a magnitude estimation procedure. She found a larger power-law exponent for repetitive sequences (the same note played over and over again) than for random sequences and familiar melodies (which involve pitch changes). This superficially implies a more rapid growth of subjective time for repetitive sequences, but intercept differences meant that the judgment functions crossed (e.g., at short durations repetitive stimuli were underestimated relative to random/melodic sequences).

In short, it is unclear how the pitch jumps between consecutive tones influences the apparent duration of a sequence. In a series of preliminary experiments, I have attempted to explore this issue by creating three conditions (see Supplementary Materials). In the large-jump condition, the lowest tone was 400 Hz and each tone's frequency differed from the previous one by a factor of 1.25, as in the experiments reported above. In the small jump_low condition, successive tone frequencies differed by a factor of only 1.05 and the lowest tone in the sequence matched the lowest tone from the large-jump condition. In the small jump_high condition, tones again differed by a factor of 1.05 and the highest tone matched the highest of tone from the large-jump condition. Despite the simplicity of this manipulation, the experiments found that the effects of pitch jump are highly unstable and vary markedly from one experiment to the next, even when identical stimuli are presented for judgment. This instability makes it difficult to establish and model the effects of pitch change on temporal judgment, and may reflect a context dependency in which perceptions of a given item depend on the other stimuli presented during the experimental session. (For examples of such context effects involving tone frequencies see Matthews & Stewart, 2008; 2009; for examples involving judgments of duration see Bobko, Schiffman, Castino, & Chiappetta, 1977, and Matthews, 2011a. Attempts to model the effect of context on duration judgments are provided by Brown et al., 2005, Taatgen & van Rijn, 2011, and Wearden & Ferrara, 1995). By contrast, the differences between ascending and descending sequences seem to be robust and have been replicated across experiments (see Supplementary Materials).

A related issue concerns the ordering of the tone frequencies that make up the stimulus. The experiments reported here used orderly sequences of ascending or descending tones. In a preliminary study I found that these orderly sequences were judged longer than random ones (see Supplementary Materials). This result is rather surprising because some theoretical accounts assert that more complex or unpredictable stimuli will have longer subjective duration (e.g., Ornstein, 1969; Pariyadath & Eagleman, 2007), and researchers have found evidence for this [e.g., Aubry et al., 2008; Block, 1978; Schiffman & Bobko, 1974; but see Kowal (1981) for a summary of conflicting results]. Moreover, the difference between orderly and random sequences is potentially problematic for the weighted sum of segments model, which predicts little effect of randomness (because the average random sequence is the same as the average of the ascending and descending sequences). Given the instability of the effects of frequency spacing on temporal judgment outlined above, it will be important to establish the robustness and generality of the effects of sequence randomness before exploring the implications for formal models of timing.

There are several other issues to address in future. The first concerns more complex temporal structures. We have focussed upon accelerating, decelerating, and constant-rate structures because they are simple and correspond to naturally-occurring patterns, but it will be important to examine (for example) the judgment of sequences which involve acceleration followed by deceleration and vice-versa. Similarly, we should ask whether the results generalize to other modalities. Buffardi (1971) found that temporal structure had comparable effects across vision, hearing, and touch when looking at empty intervals filled with clicks, but it would be helpful to have a more thorough cross-modal investigation of the types of duration, structure, and content manipulations examined here. Relatedly, the current experiments focussed on brief stimuli, but it will be important to see whether the results generalize to longer intervals (for example, pieces of music or journey times), particularly given the observed interaction between temporal structure and duration. Finally, it will be helpful to see how the effects of temporal structure are affected by the amount of attention devoted to temporal aspects of the stimuli. Manipulations of cognitive load and of the priority that should be given to temporal judgment exert a marked effect on many types of temporal judgment and can shape theorizing about the processes underlying the formation of temporal representations (e.g., Block, Hancock, & Zakay, 2010; Brown, 1997; Casini & Macar, 1997; Taatgen et al., 2007).

Conclusions

These experiments have produced a large and complex set of results which constrain any account of temporal judgment. We have focussed on relatively brief tone sequences because (a) they provide a straightforward way to manipulate the temporal structure of the interval and the non-temporal content of each subinterval, (b) tones permit precise, arbitrary durations (unlike visual stimuli, where durations are usually constrained to be whole numbers of screen refreshes), and (c) using relatively short durations reduces the likelihood of chronometric counting. The experiments have found robust effects of temporal structure which are modulated by duration, the number of tones, the rate of acceleration/deceleration, and the direction of pitch change. The weighted sum of segments model offers one simple account of these effects that is grounded in previous empirical and theoretical work. In future, it will be important to establish how the effects of temporal structure are influenced by the non-temporal structure of the sequence, although preliminary studies suggests that the effects of pitch relations on sequence judgments are volatile. Future work will also need to account for noise in the representations of sequence durations. This will be particularly important for temporal discrimination tasks, where the emphasis is on the precision of representations rather than perceived duration. It is hoped that the data and model described here will provide a useful building block for these theoretical developments.

Acknowledgements

This research was supported by the University of Essex Research Promotion Fund and the British Academy. I am grateful to Tim Rakow and Geoff Ward for helpful discussion, to Silke Paulmann for translating Grimm (1934), and to Eric-Jan Wagenmakers, Hedderik van Rijn, and two anonymous reviewers for their constructive comments on a previous version.

References

Adams, R.D. Intervening stimulus effects on category judgments of duration. *Perception & Psychophysics, 21*, 527-534.

Allan, L.G. (1978). Comments on current ratio-setting models for time perception. *Perception & Psychophysics, 24*, 444-450.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036-1060.

Anderson, R.B., & Tweney, R.D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, *25*, 724-730.

Aubry, F., Guillaume, N., Mogicato, G., Bergeret, L., & Celsis, P. (2008). Stimulus complexity and prospective timing: Clues for a parallel process model of time perception. *Acta Psychologica*, *128*, 63-74.

Beckmann, J. S., & Young, M. E. (2009). Stimulus dynamics and temporal discrimination: Implications for pacemakers. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 525-537.

Block, R.A. (1978). Remembered duration: Effects of event and sequence complexity. *Memory & Cognition*, *6*, 320-326.

Block, R.A. (2003). Psychological timing without a timer: The roles of attention and memory. In H.Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 41-59). Göttingen: Hogrefe & Huber.

Block, R.A., Hancock, P.A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*, 330-343.

Block, R.A., & Reed, M.A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 656-665.

Block, R.A., & Zakay, D. (2008). Timing and remembering the past, the present, and the future. In S.Grondin (Ed.), *Psychology of Time* (pp. 367-394). Bingley: Emerald.

Bobko, D.J., Schiffman, H.R., Castino, R.J., & Chiappetta, W. (1977). Contextual effects in duration experience. *American Journal of Psychology*, *90*, 577-586.

Boltz, M.G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, *53*, 585-600.

Boltz, M.G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics, 60,* 1357-1373.

Brigner, W.L. (1988). Perceived duration as a function of pitch. *Perceptual and Motor Skills, 67,* 301-302.

Brown, G. D. A., McCormack, T., Smith, M., & Stewart, N. (2005). Identification and bisection of temporal durations and tone frequencies: Common models for temporal and nontemporal stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 919-938.

Brown, G.D.A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539-576.

Brown, S.W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59,* 1118-1140.

Buffardi, L. (1971). Factors affecting the filled-duration illusion in the auditory, tactual, and visual modalities. *Perception & Psychophysics, 10,* 292-294.

Buhusi, C.V., & Meck, W.H. (2009). Relativity theory and time perception: Single or multiple clocks? *PLoS ONE*, *4*, e6268.

Buonomano, D.V. (2000). Decoding temporal information: A model based on short-term synaptic plasticity. *Journal of Neuroscience*, *20,* 1129-1141.

Buonomano, D.V., Bramen, J., & Khodadadifar, M. (2009). Influence of the interstimulus interval on temporal processing and learning: testing the state-dependent network model. *Philosophical Transactions of the Royal Society B, 364,* 1865-1873.

Casini, L., & Macar, F. (1997). Effects of attention manipulation on judgments of duration and of intensity in the visual modality. *Memory & Cognition*, *25*, 812-818.

Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America*, *106*, 2805-2811.

Cohen, J., Hansel, C.E.M., & Sylvester, J.D. (1954). Interdependence of temporal and auditory judgments. *Nature*, *174*, 642-644.

Creelman, C.D. (1962). Human discrimination of auditory duration, *Journal of the Acoustical Society of America*, *34*, 582-593.

Crowder, R.G., & Neath, I. (1994). The influence of pitch on time perception in short melodies. *Music Perception*, *12*, 379-386.

Droit-Volet, S. (2010). Stop using time reproduction tasks in a comparative perspective without further analyses of the role of the motor response: The example of children. *European Journal of Cognitive Psychology*, *22*, 130-148.

Eagleman, D.M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Philosophical Transactions of the Royal Society B*, *364*, 1841-1851.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, *83,* 1154-1171.

Forster, K.I., & Forster, J.C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116-124.

Friedman, W.J., & Kemp, S. (1998). The effect of elapsed time and retrieval of young children's judgments of the temporal distances of past events. *Cognitive Development, 13*, 335-367.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences, Volume 423: Timing and time perception.* New York: New York Academy of Sciences.

Grimm, K. (1934). Der Einfluß der Zeitform auf die Wahrnehmung der Zeitdauer. *Zeitschrift für Psychologie*, *132*, 104-132.

Grondin, S. (2010). Timing and time perception: A review of recent behavioural and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics, 72*, 561-582.

Hall, G.S., & Jastrow, J. (1889). Studies of rhythm. *Mind*, *11*, 55-62.

Henry, M.J., & McAuley, J.D. (2009). Evaluation of an imputed pitch velocity model of the auditory kappa effect. *Journal of Experimental Psychology*: *Human Perception and Performance*, *35,* 551-564.

Henry, M.J., & McAuley, J.D. (2011). Velocity perception for sounds moving in frequency space. *Attention, Perception, & Psychophysics, 73*, 172-188.

Israeli, N. (1930). Illusions in the perception of short time intervals. *Archives of Psychology*, *19*, No. 113.

Jones, M.R., & Boltz, M. (1989). Dynamic attending and responding to time. *Psychological Review*, *96*, 459-491.

Jones, M.R., Boltz, M.G., & Klein, J.M. (1993). Expected endings and judged duration. *Memory & Cognition*, *21*, 646-665.

Jones, L.A., & Wearden, J.H. (2003). More is not necessarily better: Examining the nature of the temporal reference memory component in timing. *Quarterly Journal of Experimental Psychology*, *56B*, 321-343.

Kahana, M.J., & Adler, M. (2012). Note on the power law of forgetting. *Manuscript submitted for publication.*

Killeen, P.R., & Fetterman, J.G. (1988). A behavioral theory of timing. *Psychological Review*, *95*, 274-295.

Kowal, K.H. (1981). Growth of apparent duration: Effect of melodic and non-melodic tonal variation. *Perceptual and Motor Skills*, *52*, 803-817.

Laming, D. *The measurement of sensation.* Oxford: Oxford University Press.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Matell, M.S., & Meck, W.H. (2000). Neuropsychological mechanisms of interval timing behavior. *BioEssays*, *22*, 94-103.

Matthews, W.J. (2011a). Can we use verbal estimation to dissect the internal clock? Differentiating the effects of pacemaker rate, switch latencies, and judgment processes. *Behavioural Processes*, *86*, 68-74.

Matthews, W.J. (2011b). How do changes in speed affect the perception of duration? *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1617-1627.

Matthews, W.J. (2011c). Stimulus repetition and the perception of time. *PLoS ONE,* e19815.

Matthews, W.J., & Stewart, N. (2008). The effect of stimulus range on two-interval frequency discrimination. *Journal of the Acoustical Society of America*, *123*, EL45-EL51.

Matthews, W. J., & Stewart, N. (2009). The effect of interstimulus interval on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology, 62,* 2014-2029.

Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 303-313.

Mauk, M.D., & Buonomano, D.V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience*, *27,* 307-340.

Meck, W.H. (1983). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behaivor Processes*, *9,* 171-201.

Nakajima, Y. (1987). A model of empty duration perception. *Perception*, *16*, 485-520.

Ogden, R.S., Wearden, J.H., & Jones, L.A. (2008). The remembrance of times past: Interference in temporal reference memory. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1524-1544.

Ornstein, R. E. (1969). *On the experience of time*. Harmondsworth, Middlesex: Penguin.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407-418.

Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE*, *11*, e1264.

Parker, S., Murphy, D.R., & Schneider, B.A. (2002). Top-down gain control in the auditory system: Evidence from identification and discrimination experiments. *Perception & Psychophysics*, *64,* 598-615.

Penney, T.B., Gibbon, J., & Meck, W.H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1770-1787.

Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 307-320.

Petzold, P., & Haubensak, G. (2004). The influence of category membership of stimuli on sequential effects in magnitude judgment. *Perception & Psychophysics, 66*, 665-678.

Poynter, W.D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition*, *11*, 77-82.

Poynter, W. D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 305-321). Amsterdam: Elsevier.

Predebon, J. (1996). The relationship between the number of presented stimuli and prospective duration estimates: The effect of concurrent task activity. *Psychonomic Bulletin & Review*, *3*, 376-379.

Rammsayer, T., & Ulrich, R. (2001). Counting models of temporal discrimination. *Psychonomic Bulletin & Review, 8,* 270-277.

Rodríguez-Gironés, M.A., & Kacelnik, A. (2001). Relative importance of perceptual and mnemonic variance in human temporal bisection. *Quarterly Journal of Experimental Psychology, 54A,* 527-546.

Schiffman, H.R., & Bobko, D.J. (1974). Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology, 103,* 156-159.

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics*, *40,* 9-19.

Sivyer, M., & Finlay, D. (1982). Perceived duration of auditory sequences. *Journal of General Psychology*, *107*, 209-217.

Staddon, J.E.R., & Higa, J.J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, *71*, 215-251.

Taatgen, N., & van Rijn, H. (2011). Traces of times past: Representations of temporal intervals in memory. *Memory & Cognition*, *39*, 1546-1560.

Taatgen, N.A, van Rijn, H., & Anderson, J. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review*, *114*, 577-598.

Thomas, E.A.C, & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception & Psychophysics, 16,* 449-458.

Tipps, R.S. (1980). Time estimation of auditory patterns of different pitch intervals. *Perceptual and Motor Skills, 51*, 695-698.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs, 77,* 1-31.

Treisman, M., Faulkner, A., Naish, P.L.N., & Brogan, D. (1990). The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*, 705-743.

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics, 66,* 1171-1189.

van Rijn, H., & Taatgen, N.A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica*, *129,* 365-375.

Wackerman, J., & Ehm, W. (2006). The dual klepsydra model of internal time representation and time reproduction. *Journal of Theoretical Biology*, *239,* 482-493.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "Sounds are judged longer then lights": Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology, 51*, 97-120.

Wearden, J.H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection in humans. *Quarterly Journal of Experimental Psychology*, *48B*, 289-310.

Wearden, J.H., & Grindrod, R. (2003). Manipulating decision processes in the human scalar timing system. *Behavioural Processes, 61,* 47-56.

Wearden, J.H., & Jones, L.A. (2007). Is the growth of subjective time in humans a linear or nonlinear function of real time? *Quarterly Journal of Experimental Psychology*, *60*, 1289-1302.

Wickelgren, W.A. (1966). Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, *72*, 250-259.

Yoblick, D.A., & Salvendy, G. (1970). Influence of frequency on the estimation of time for auditory, visual, and tactile modalities: The kappa effect. *Journal of Experimental Psychology*, *86,* 157-164.

Yu, A.C.L. (2010). Tonal effects on perceived vowel duration. *Laboratory Phonology*, *10*, 151-168.

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science, 6*, 12-16.

Zakay, D., Tsal, Y., Moses, M., & Shahar, I. (1994). The role of segmentation in prospective and retrospective time estimation processes. *Memory & Cognition*, *22,* 344-351.

Table 1.

Effect	**Evidence**	**Captured by model?**
Judged duration is an increasing, negatively-accelerated function of overall duration	All experiments[1]	Yes
Constant rate stimuli judged longest	All experiments	Yes
Temporal structure interacts with duration	All experiments	Yes
Sequences with gradual acceleration/deceleration judged longer than those with rapid change; this effect more pronounced for decelerating sequences	Experiment 2	Yes
Sequences of more tones judged longer	Experiment 3	Yes (although the model rather overestimates the effect)
Larger effects of temporal structure for sequences comprising more tones	Experiment 3	Yes (although the model rather underestimates the effect)
Interaction between temporal structure and pitch direction (descending longer than ascending for decelerating sequences but not for accelerating ones)	Experiment 4	Yes
Interaction between temporal structure and duration more pronounced for ascending sequences than descending ones	Experiment 4	Yes

[1]Experiment 1 only used two durations, so the quadratic trend in the judgment function could not be tested.

Figure Captions

Figure 1. Results of Experiment 1.The left panel shows the mean judgments for the decelerating (Dec), accelerating (Acc), and constant-rate (Con) sequences at each of the four durations. The right panel shows the predictions of the weighted sum of segments model.

Figure 2. Effects of changing parameters on the predictions of the weighted sum of segments model. The top panels show the effects of decreasing and increasing the decay parameter *r*. The bottom panels show the effects of changing the *b* parameter in the power function that relates the judgment of an individual segment to its objective duration. All other parameters have the same values as for the model predictions shown in Figure 1.

Figure 3. Results of Experiment 2. The top left panel shows the mean responses for all temporal structures (rapid deceleration, gradual deceleration, constant rate, gradual acceleration, and rapid acceleration) at each of the two durations. To clarify the findings, the second panel down shows the results after averaging over rate of change, and the third panel shows the two-way interaction between type of change (accelerating vs decelerating) and rate of change (gradual vs rapid). The bottom panel shows the differences between the five temporal structures, averaged over the two durations. The right-hand plots show the predictions of the weighted sum of segments model.

Figure 4. Results of Experiment 3. The top left panel shows the mean judgments for each temporal structure at each duration when the sequence consisted of four tones. Neither the main effect of temporal structure nor the interaction between structure and duration is significant. The bottom left panel shows the results when the sequences consisted of six tones; overall, the judgments are larger than for the 4-tone case, and both the main effect of temporal structure and the structure x duration interaction are significant. The right hand panels show the predictions of the weighted sum of segments model.

Figure 5. Results of Experiment 4. The top left panel shows the mean judgments for each temporal structure at each duration for ascending sequences; the panel below shows the results for descending sequences. The third panel down shows the two-way interaction between pitch direction (Descending vs Ascending) and temporal structure; the bottom panel shows the lack of a two-way interaction between pitch direction and duration. The right-hand panels show the predictions of the weighted sum of segments model.

Figure 1.

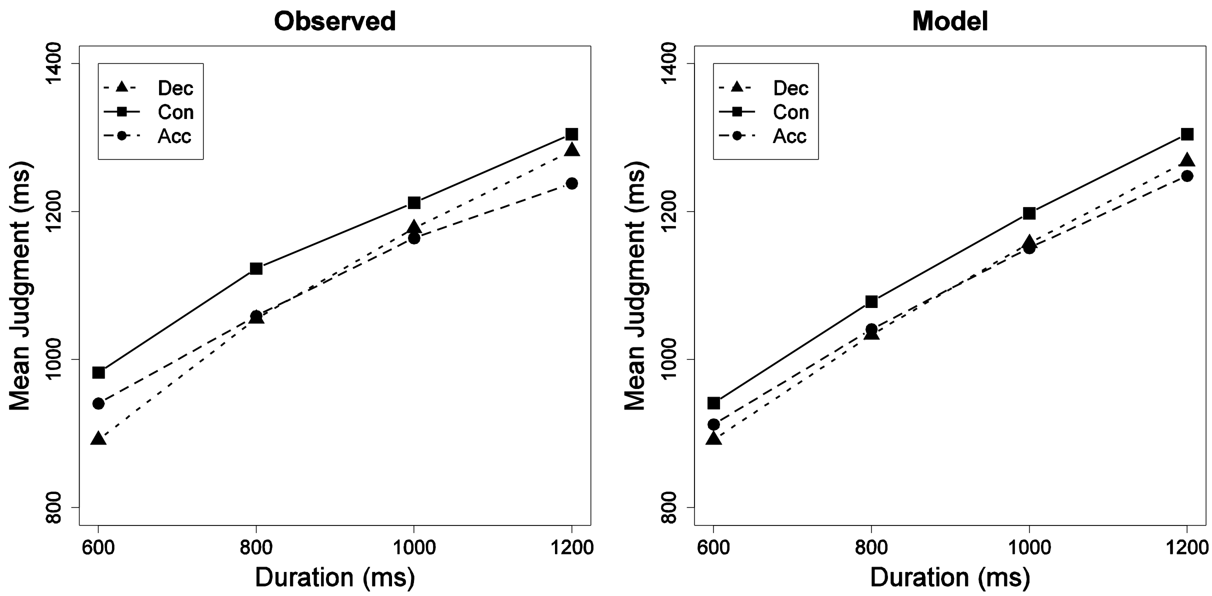

Observed
Model
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
800
1000
1200
1400
600
800
1000
1200

Figure 2.

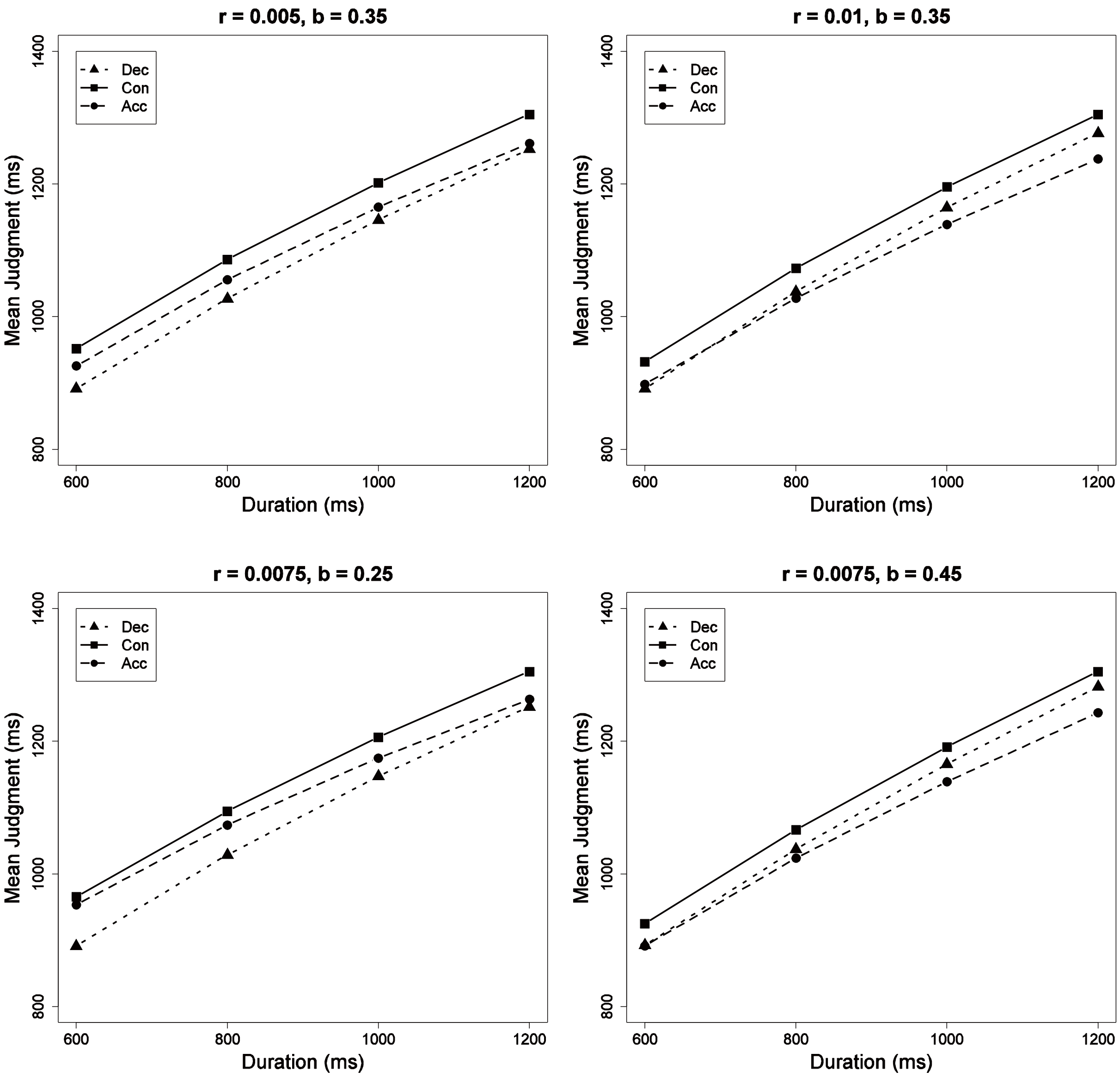
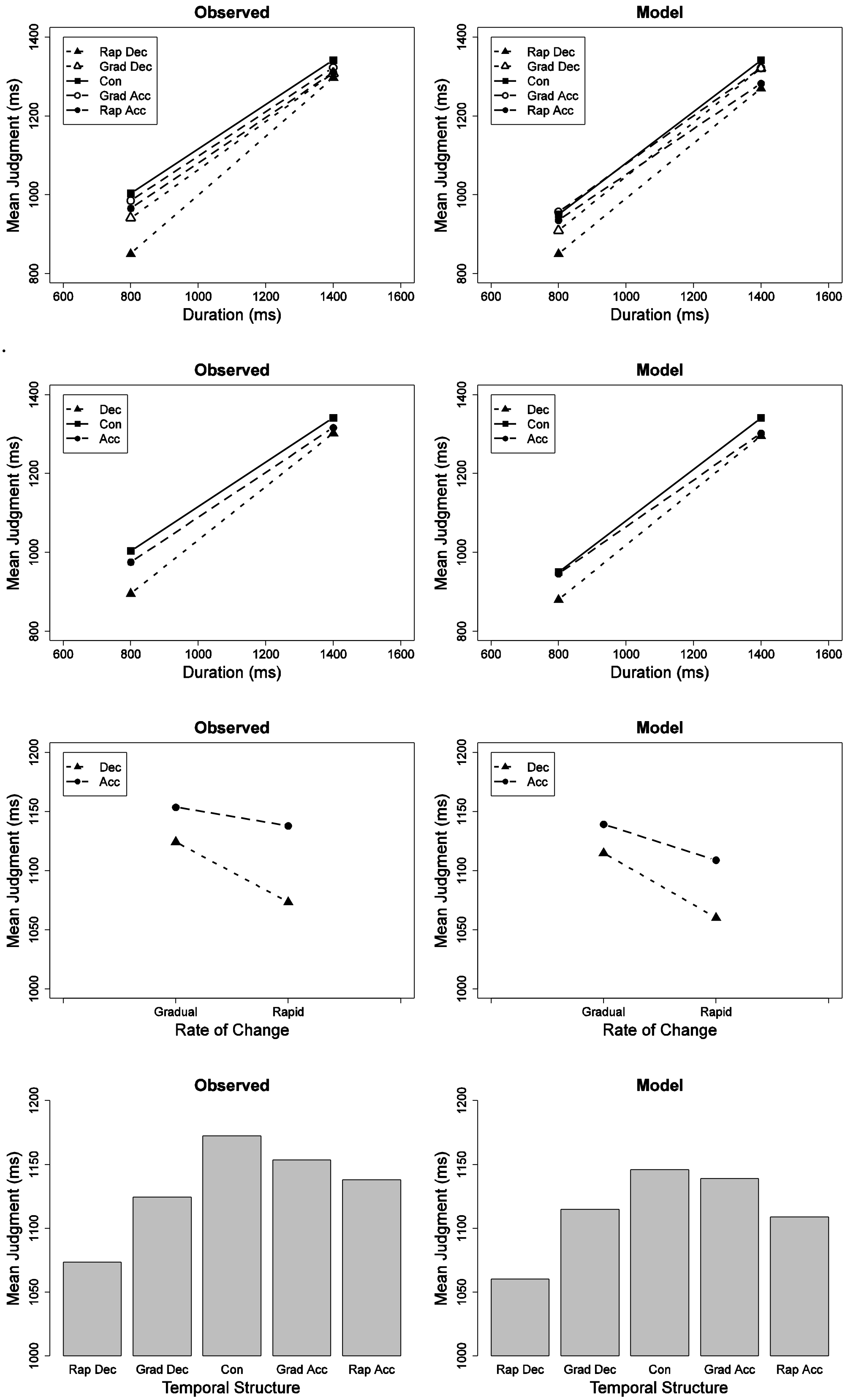

r = 0.005, b = 0.35
r = 0.01, b = 0.35
r = 0.0075, b = 0.25
r = 0.0075, b = 0.45
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
800
1000
1200
1400
600
800
1000
1200

Figure 3



Observed
Model
Rap Dec
Grad Dec
Con
Grad Acc
Rap Acc
Mean Judgment (ms)
800
1000
1200
1400
600
1600
Duration (ms)
Dec
Acc
1050
1100
1150
Gradual
Rapid
Rate of Change
Temporal Structure

Figure 4

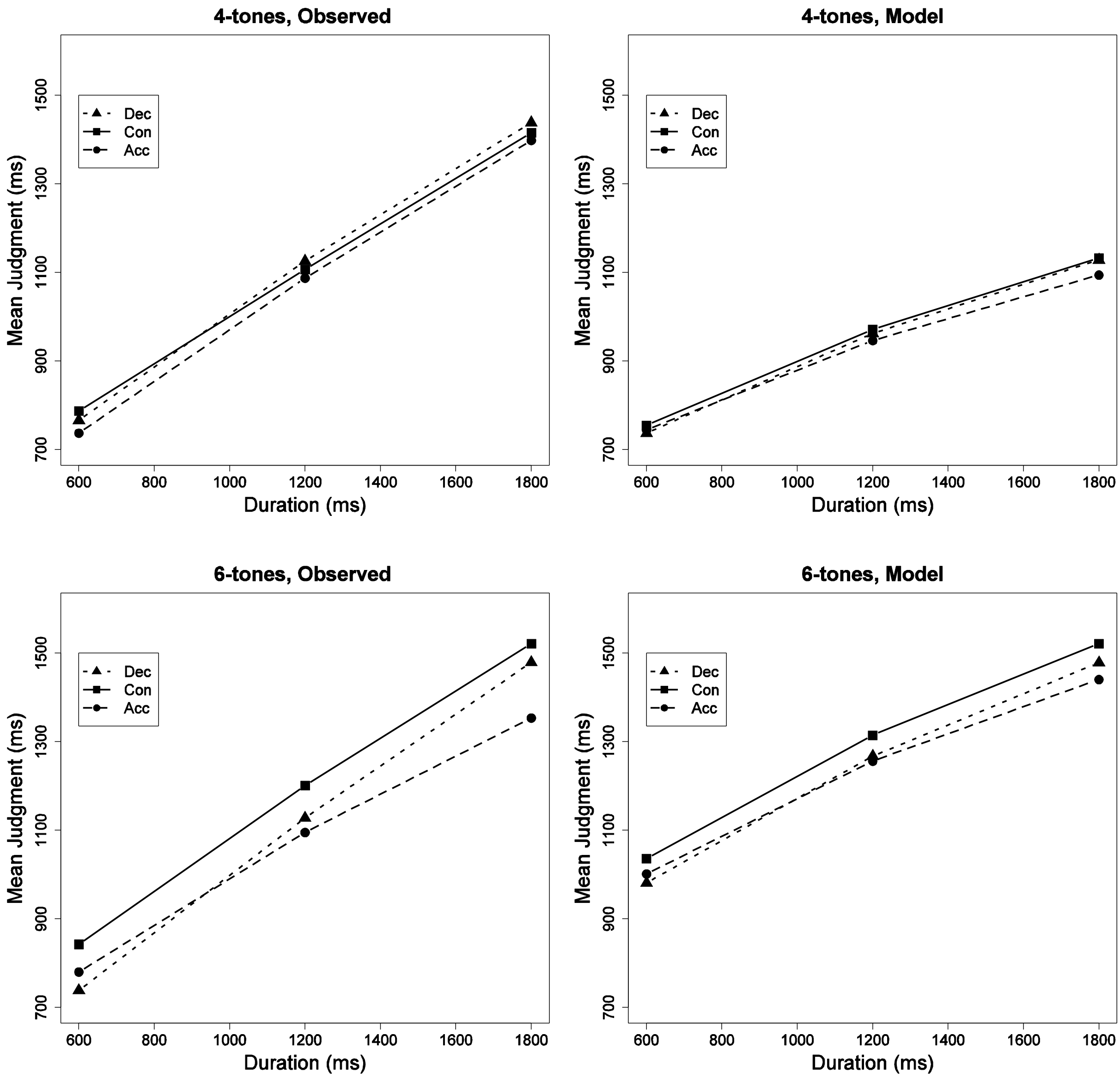
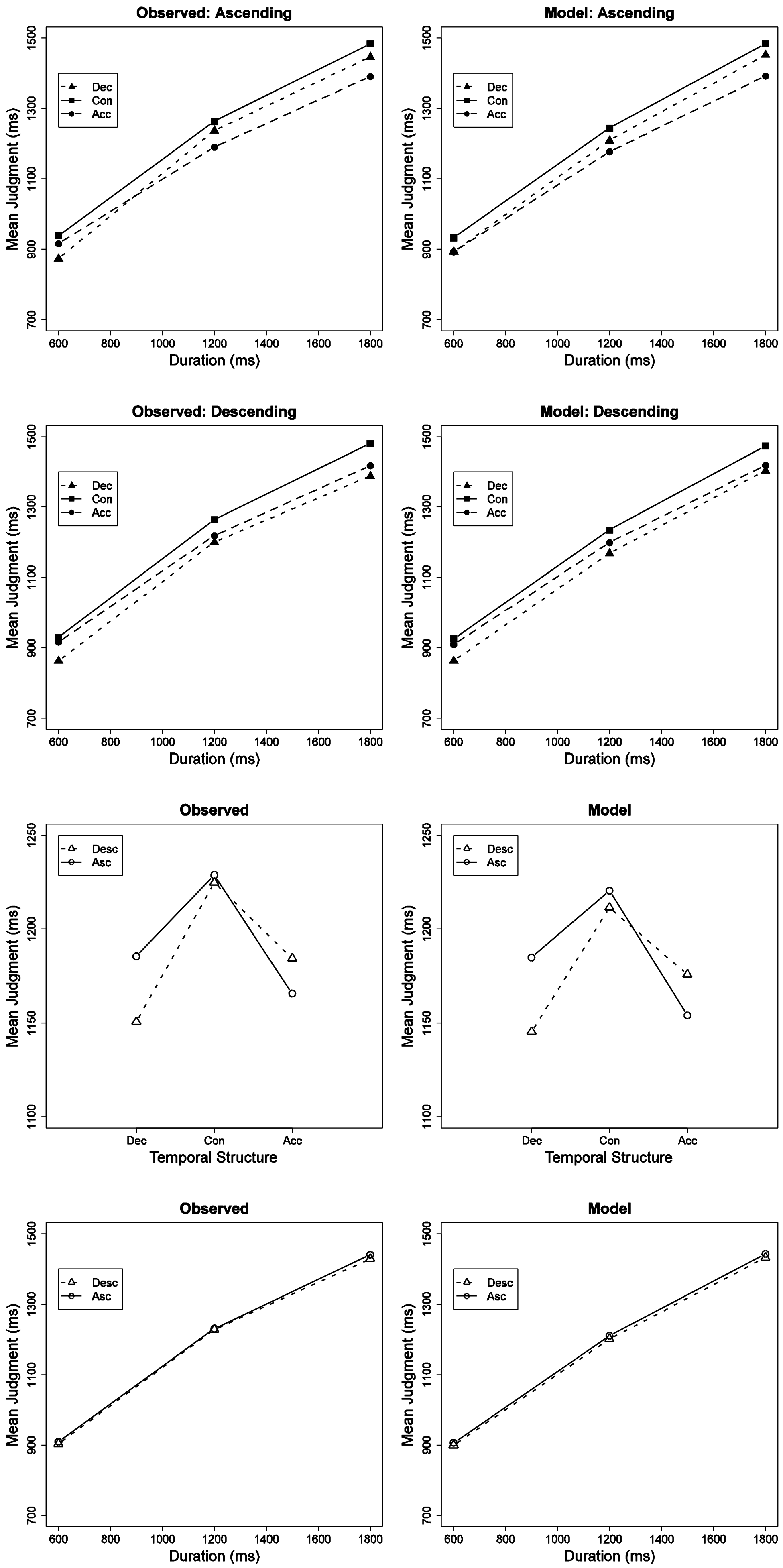

4-tones, Observed
4-tones, Model
6-tones, Observed
6-tones, Model
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
700
900
1100
1300
1500
600
800
1000
1200
1400
1600
1800

Figure 5


Observed: Ascending
Model: Ascending
Observed: Descending
Model: Descending
Observed
Model
Observed
Model
Mean Judgment (ms)
Duration (ms)
Temporal Structure
Dec
Con
Acc
Desc
Asc

Supplementary Materials

These materials describe four additional experiments examining the effects of temporal structure on judgments of duration, conducted as part of the series described in the main text. These experiments manipulated the pitch relations between the tones making up each sequence. In particular, they examined the effect of changing the “jump” in pitch between successive tones in the sequence. The overall finding is that, although the differences between accelerating, decelerating, and constant-rate stimuli described in the main text are robust, the effects of changing the pitch jumps between tones are labile, with identical stimuli producing different results in different studies. The experiments below also examine the effect of pitch direction (ascending vs. descending) and produce results similar to those in the main text. A final experiment provides preliminary evidence that random sequences are judged to be briefer than ordered ones. The experiments are described here as a potential starting point for further investigations of the complex effects of pitch structure on judged duration.

Experiment S1

Method

Thirty six participants took part (21 female; ages 19-45 years, M = 25.3, SD = 6.3). The experiment employed a 3 x 3 x 2 x 2 mixed design with duration (0.6, 1.2, 1.8 s), temporal structure (accelerating, decelerating, constant rate) and size of pitch change (small jump, large jump) manipulated within subjects, and pitch of the small-jump stimuli (low, high) manipulated between subjects. In this and all subsequent studies, the temporal structures of the accelerating, constant, and decelerating sequences were the same as for the 6-tone stimuli of Experiments 3 and 4 (main text).

Stimuli in the large-jump condition were identical to the 6-tone stimuli of Experiment 3 in the main text: the fist tone in the sequence had a frequency of 400 Hz and the frequency of each subsequent tone was 1.25 times that of the previous one. Stimuli in the small-jump conditions were the same except that the change in frequency was reduced: each successive tone was only 5% higher in frequency than the previous one. Two groups of participants were tested. For participants in the low condition (N = 18), the first tone of the small-jump sequences had a frequency of 400 Hz (the same starting frequency as in the large-jump case). For participants in the high condition (N = 18), the first tone of the small-jump sequences was 956.5 Hz, meaning that the frequency of the final tone was 1220.7 Hz (the same final frequency as in the large-jump case).

Each block of 18 trials comprised one occurrence of each duration x temporal structure x size of pitch jump combination in random order. Participants completed 10 blocks.

Results and Discussion

The data are plotted in Figure S1 and were analysed with a 3 x 3 x 2 x 2 mixed ANOVA, with duration (0.6, 1.2, 1.8 s), temporal structure (accelerating, decelerating, constant rate), and size of pitch change (small jump, large jump) as within-subjects factors, and pitch group (low, high) as a between-subjects factor. As usual, judgments were larger for longer stimuli, $F(1.19, 40.37) = 296.02$, $p < .001$, $\eta_p^2 = .90$ with both linear and quadratic trends [$F(1,34) = 314.04$, $p <.001$, $\eta_p^2 = .90$ and $F(1,34) = 39.45$, $p <.001$, $\eta_p^2 = .54$]. Judgments were also affected by whether the sequences were accelerating, decelerating, or constant-rate, $F(1.82, 61.76) = 12.64$, $p< .001$, $\eta_p^2 = .27$, and this effect was modulated by duration, $F(3.69, 125.49) = 3.77$, $p = .008$, $\eta_p^2 = .10$. This interaction is plotted in the left panel of Figure S1, and mirrors the pattern from previous experiments. There was no main effect of the size of the pitch jump, $F(1, 34) = 2.61$, $p = .116$, $\eta_p^2 = .07$, but this factor interacted with whether the small-jump sequences were low or high pitch , $F(1,34) = 24.07$, $p < .001$, $\eta_p^2 = .41$. This interaction is plotted in the right panel of Figure S1. For the low group, large-jump sequences were judged longer than small-jump sequences; for the High group, this pattern was reversed. (Repeating the foregoing ANOVA separately for the low and high groups revealed significant but opposite effects of jump size: $F(1, 17) = 5.37$, $p = .033$, $\eta_p^2 = .24$ for the low group and $F(1,17) = 21.44$, $p <.001$, $\eta_p^2 = .56$ for the high group).

Note that judgments of the small-jump sequences are similar for the low and high groups whilst judgments of the large-jump sequences are very different. This is surprising, given that the large-jump stimuli were physically identical in the low and high conditions. The implication is that judgment of a given stimulus is highly sensitive to the context established by the other items in the experimental session. None of the other interactions were significant (all *F*s < 1).

Experiment S2

This experiment was like Experiment S1 but used a fully within-subjects design.

Method

Thirty participants took part (25 female; ages 18-35 years, M = 20.2, SD = 3.9). The stimuli were the same as in Experiment S1, but all participants judged all of the stimuli. The experiment used a 3 (duration: 0.6, 1.2, 1.8 seconds) x 3 (temporal structure: decelerating, constant, accelerating) x 3 (pitch spacing: small-jump_low, large jump, small-jump_high) fully within subjects design. Each block of 27 trials contained one presentation of each stimulus in random order; participants completed 10 blocks.

Results and Discussion

The mean judgments for each pitch spacing are plotted in the Figure S2. A 3 x 3 x 3 ANOVA was conducted with duration (0.6, 1.2, 1.8 s), temporal structure (accelerating, decelerating, constant rate), and pitch spacing (small-jump_low, large jump, small-jump_high) as within-subjects factors. Like the previous experiments, judgments were larger for longer stimuli, $F(1.18, 34.20) = 279.23$, $p < .001$, $\eta_p^2 = .91$ [linear trend: $F(1,29) = 304.57$, $p<.001$, $\eta_p^2 = .91$; quadratic trend $F(1,29) = 45.20$, $p <.001$, $\eta_p^2 = .61$], judgments were affected by temporal structure, $F(2, 58) = 23.20$, $p < .001$, $\eta_p^2 = .44$, and this effect was modulated by stimulus duration, $F(3.37, 97.62) = 3.49$, $p = .015$, $\eta_p^2 = .11$. The pattern is similar to other experiments: the decelerating items are judged longer than the accelerating ones at long durations but not at short durations, with the constant-rate stimuli consistently judged longest.

There was no main effect of pitch spacing, $F(2,58) = .90$, $p = .413$, $\eta_p^2 = .03$, but spacing did interact with duration, $F(4,116) = 2.59$, $p = .040$, $\eta_p^2 = .08$. This two-way interaction is visible in the bottom panel of Figure S2 (although the effect is small). At short durations the small-jump_high stimuli were judged briefer than the other spacing conditions, but at long durations they were judged longer than the other conditions. Neither the interaction between temporal structure and spacing nor the three-way interaction were significant [$F(4,116) = 2.42$, $p = .053$, $\eta_p^2 = .08$ and $F(8,232) = 1.91$, $p = .060$, $\eta_p^2 = .06$, respectively].

Experiment S1 found that the large-jump sequences were judged differently depending on whether they were presented in a session with low- or high-pitched small-jump sequences. The within-subjects design of the current experiment meant that the pattern found in Experiment S1 could not occur here. Instead, the data show a rather puzzling interaction between pitch spacing and duration.

Experiment S3

Experiment S3 was similar to Experiment S2 but also manipulated the direction of pitch change (ascending vs. descending).

Method

Forty participants took part (22 female; ages 20-42 years, *M* = 26.4, *SD* = 5.4; one had previously participant in Experiment 2 of the main text). The experiment used a 3 (duration: 0.6, 1.2, 1.8 seconds) x 3 (temporal structure: decelerating, constant, accelerating) x 3 (pitch spacing: small-jump_low, large jump, small-jump_high) x 2 (pitch change direction: descending, ascending) fully within subjects design. The ascending stimuli were identical to the tone sequences from Experiment S2; the descending stimuli were the same but with the order of the tone frequencies reversed. Each block of 54 trials contained one presentation of each stimulus in random order; participants completed 10 blocks and were given the opportunity to take a break after every 18 trials.

Results and Discussion

The mean judgments for each temporal structure at each duration are shown separately for the ascending and descending sequences in the top two panels of Figure S3. A 3 x 3 x 3 x 2 ANOVA was conducted with duration (0.6, 1.2, 1.8 s), temporal structure (accelerating, decelerating, constant rate), pitch spacing (small-jump_low, large jump, small-jump_high) and direction (descending, ascending) as within-subjects factors. Judgments were larger for longer stimuli, $F(1.06, 41.45) = 250.66$, $p < .001$, $\eta_p^2 = .87$ with both linear and quadratic trends [$F(1,39) = 255.77$, $p<.001$, $\eta_p^2 = .87$ and $F(1,39) = 92.39$, $p<.001$, $\eta_p^2 = .70$]. As in previous experiments, judgments were affected by temporal structure, $F(2, 78) = 12.03$, $p < .001$, $\eta_p^2 = .24$, and this effect depended on stimulus duration, $F(2.98, 116.34) = 4.41$, $p = .006$, $\eta_p^2 = .10$.

More importantly, the effect of temporal structure also depended on whether the sequence was ascending or descending, $F(2, 78) = 10.09$, $p <.001$, .21 (see the bottom left panel of Figure S3), and was modulated by duration in a three-way interaction, $F(4, 156) = 3.36$, $p = .011$, $\eta_p^2 = .08$. The pattern is visible in the top two panels of Figure S3 and replicates that from Experiment 4 of the main text: for ascending sequences, the relationship between accelerating and decelerating sequences changes markedly as the stimulus duration increases, but this effect is much less conspicuous for the descending sequences.

The bottom right panel of Figure S3 shows judgments as a function of the direction and spacing of the pitch changes between successive tones. Ascending sequences were on average judged longer than descending ones, $F(1, 39) = 15.36$, $p < .001$, $\eta_p^2 = .28$. Judgments were also affected by the frequency spacing, $F(2, 78) = 14.59$, $p < .001$, $\eta_p^2 = .27$, and this effect was more pronounced for the descending stimuli than the ascending ones, $F(1.69, 65.71) = 3.55$, $p = .042$, $\eta_p^2 = .08$. As can be seen in the figure, the two small-jump conditions produced similar judgments but the large-jump sequences were judged to be shorter. No other effects were significant (all *F*s < 1.56, *p*s > .13).

The effects of pitch spacing in this experiment differed from those in Experiments S1 and S2, despite the fact that (for the ascending sequences) the stimuli are identical. It seems that the effects of pitch spacing are volatile and vary from study to study for no obvious reason. Indeed, combining the data for the ascending sequences from this experiment and the previous one in a single ANOVA revealed a significant interaction between frequency spacing and experiment, $F(2,136) = 5.24$, $p = .006$, $\eta_p^2 = .07$. This confirms that the effects of frequency spacing were very different despite the stimuli being identical. The participants were similar in both studies, so it seems likely that it is the context defined by the other items in the experiment that changes the effects of frequency spacing on temporal judgment.

Experiment S4

This experiment investigated the effects of changing the order of the tone frequencies that make up the sequence. It compared ascending, descending, and randomly-ordered sequences. For the random ordering, two types of sequence were included. In one, each random sequence was unique; in the other, the same random sequence was presented repeatedly throughout the experiment. Comparing these two conditions helps establish whether differences between the ordered (i.e., ascending and descending) and unordered sequences are due to the lack of predictable structure within the item, or to the increased familiarity with ascending and descending items that arises over the course of the experimental session.

Method

Thirty six participants took part (27 female; ages 18-54 years, M = 21.3, SD = 6.4). The experiment used a 3 (duration: 0.6, 1.2, 1.8 seconds) x 3 (temporal structure: decelerating, constant rate, accelerating) x 4 (pitch order: ascending, descending, random_unique, random_repeated) fully within-subjects design. The ascending and descending sequences were identical to those used in the preceding experiment and in Experiment 4 of the main text. It is convenient to label the frequencies of the tones used in these sequences from 1 (lowest) to 6 (highest). In the random_unique condition, the sequence on each trial was constructed by randomly permuting frequencies 1 to 6, subject to the constraints that (a) the sequence contained no runs of more than two adjacent frequencies (e.g., 1,2,5,3,6,4 was allowed but 3,2,1,5,4,6 was not), and (b) the sequence had not already been used for that participant. In the random_repeated condition, one sequence meeting the criterion for the random_unique condition was used repeatedly throughout the experiment. Each block of 36 trials contained one presentation of each type of stimulus in random order; participants completed 10 blocks.

Results and Discussion

For each pitch order, the mean judgments for accelerating, decelerating, and constant-rate sequences are shown in the top four panels of Figure S4. A 3 x 3 x 4 ANOVA was conducted with duration (0.6, 1.2, 1.8 s), temporal structure (accelerating, decelerating, constant rate), and frequency structure (descending, ascending, random_unique, random_repeated) as within-subjects factors.

Judgments were larger for longer stimuli, $F(1.20, 42.11) = 519.99$, $p < .001$, $\eta_p^2 = .94$ [linear trend: $F(1,35) = 566.03$, $p<.001$, $\eta_p^2 = .94$; quadratic trend: $F(1,35) = 57.23$, $p<.001$, $\eta_p^2 = .62$] and were affected by stimulus type, $F(1.66, 58.22) = 25.01$, $p < .001$, $\eta_p^2 = .42$ in a way which depended on the stimulus duration, $F(4, 140) = 3.64$, $p = .008$, $\eta_p^2 = .09$. The pattern is the same as in previous experiments, and was not modulated by pitch order, $F(12, 420) = .78$, $p = .672$, $\eta_p^2 = .02$.

There was, however, a main effect of pitch order, $F(2.47, 86.38) = 17.33$, $p < .001$, $\eta_p^2 = .33$, and this effect was modulated by both duration and temporal structure, [$F(6, 210) = 4.66$, $p <.001$, $\eta_p^2 = .12$ and $F(6,210) = 2.76$, $p = .013$, $\eta_p^2 = .07$, respectively]. These two-way interactions can be seen in the bottom two panels of Figure S4.

To help understand the interactions, two separate ANOVAs were conducted, one using the data from the ascending and descending sequences and the other using data from the two random conditions. For the ascending and descending sequences, judgments were affected by duration and temporal structure as in the overall analysis, although the interaction was no longer significant, $F(4,140) = 1.64$, $p = .167$, $\eta_p^2 = .05$. More importantly, there was no main effect of whether the sequence is ascending or descending, $F(1,35) = .37$, $p = .548$, $\eta_p^2 = .01$, and no interaction between duration and pitch order, $F(2,70) = .21$, $p = .811$, $\eta_p^2 = .01$. However, pitch order did interact with temporal structure, $F(2, 70) = 5.93$, $p = .004$, $\eta_p^2 = .15$: ascending sequences were judged longer than descending ones in the decelerating condition but not in the accelerating and constant-rate conditions. This is visible in the bottom left panel of Figure S4. The threeway interaction was not significant, $F(3.44, 120.36) = 1.06$, $p = .374$, $\eta_p^2 = .03$.

Focusing on the data from the random_unique and random_repeated conditions, the effects of duration, temporal structure, and the temporal structure x duration interaction were all significant, mirroring the overall analysis. There was no overall difference between the random_repeated and random_unique conditions, $F(1, 35) = .81$, $p = .376$, $\eta_p^2 = .02$, and no interaction between this factor and temporal structure, $F(2,70) = 1.81$, $p = .171$, $\eta_p^2 = .05$. However, there was an interaction between stimulus repetition and duration, $F(2,70) = 14.03$, $p < .001$, $\eta_p^2 = .29$, with the repeated stimuli judged fractionally longer at the shortest duration and the unique items judged longer at the longest duration. This pattern is visible in the bottom right panel of Figure S4. The threeway interaction was not significant, $F(4,140) = .87$, $p = .482$, $\eta_p^2 = .02$.

The ascending and descending stimuli used in this experiment are identical to those from the large-jump condition of Experiment S3 and from Experiment 4 of the main text. Combining the data for these stimuli in a single ANOVA with experiment included as a between subject-factor showed that none of the terms involving the experiment factor were significant (all *F*s < 3.04, *p*s > .05; three participants in Experiment 8 who had participated in one of the two previous experiments were excluded from this ANOVA to avoid non-independence). Thus, the judgments were similar and the effects of ascending vs. descending pitch structure seem to be reasonably stable. The combined data from Experiments S3 and S4 are plotted in Figure S5, and are similar to those from Experiment 4 in the main text.

Figure S1

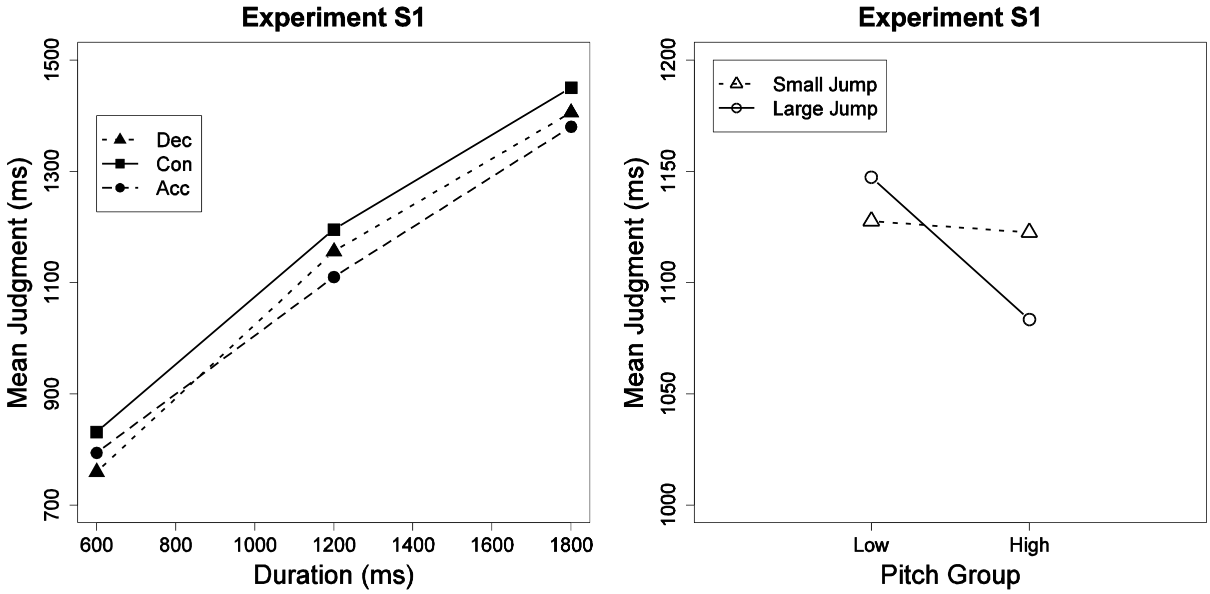
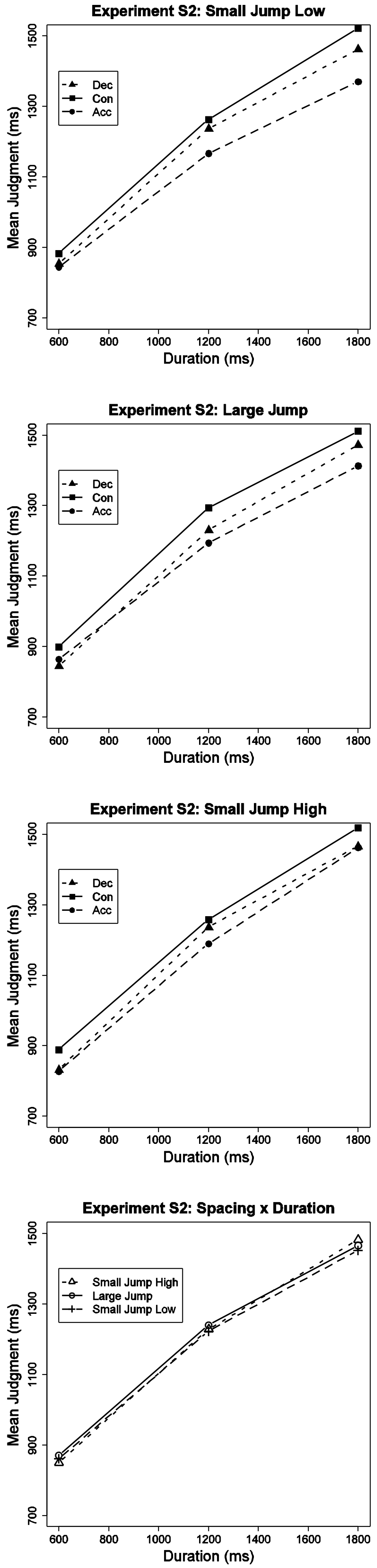
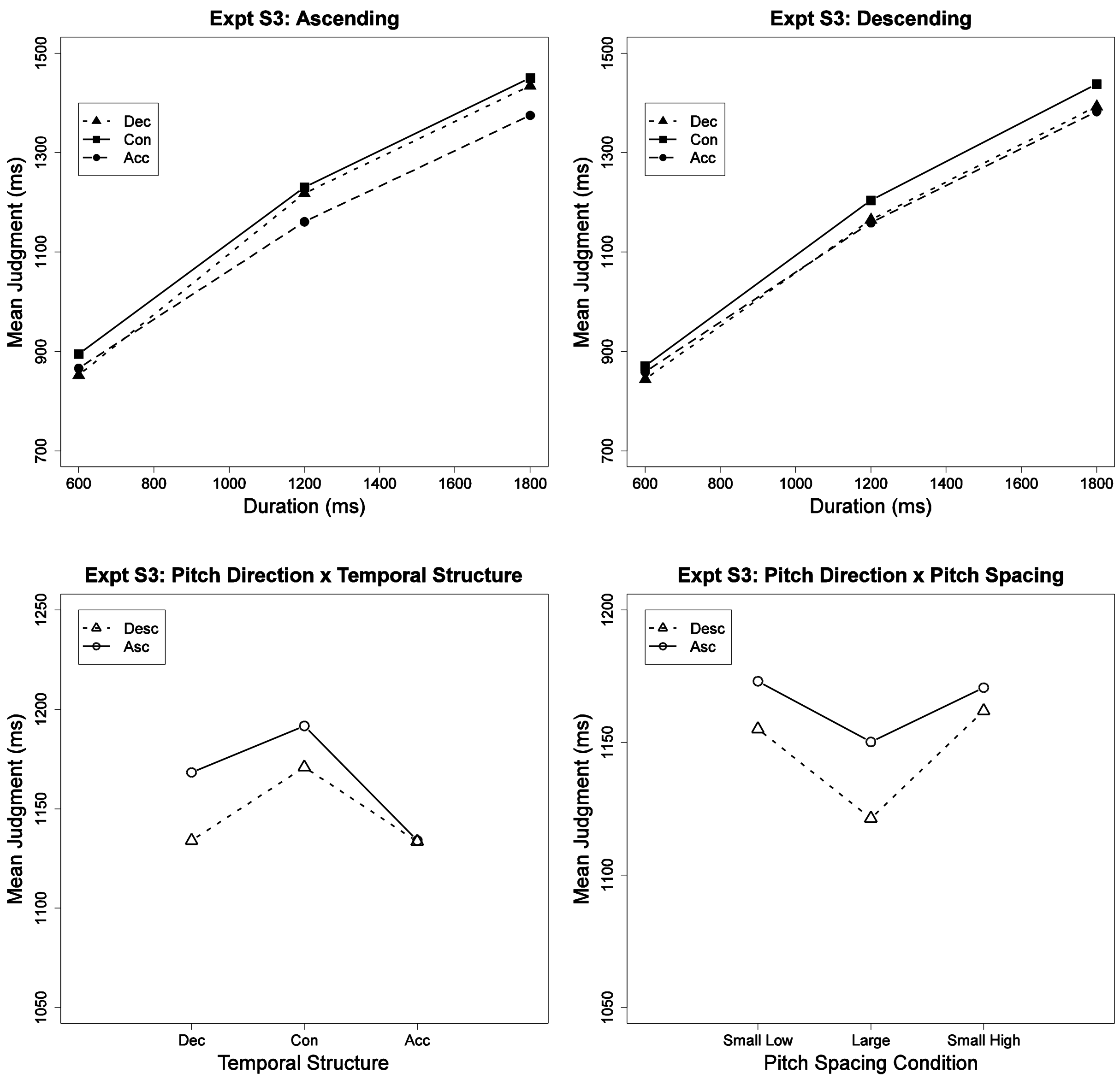

Experiment S1
Mean Judgment (ms)
700
900
1100
1300
1500
Dec
Con
Acc
600
800
1000
1200
1400
1600
1800
Duration (ms)
Experiment S1
Mean Judgment (ms)
1000
1050
1100
1150
1200
Small Jump
Large Jump
Low
High
Pitch Group

Figure S2



Experiment S2: Small Jump Low
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
Experiment S2: Large Jump
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
Experiment S2: Small Jump High
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
Experiment S2: Spacing x Duration
Small Jump High
Large Jump
Small Jump Low
Mean Judgment (ms)
Duration (ms)

Figure S3



Expt S3: Ascending
Mean Judgment (ms)
Duration (ms)
Dec
Con
Acc
Expt S3: Descending
Mean Judgment (ms)
Duration (ms)
Dec
Con
Acc
Expt S3: Pitch Direction x Temporal Structure
Mean Judgment (ms)
Temporal Structure
Desc
Asc
Dec
Con
Acc
Expt S3: Pitch Direction x Pitch Spacing
Mean Judgment (ms)
Pitch Spacing Condition
Desc
Asc
Small Low
Large
Small High

Figure S4

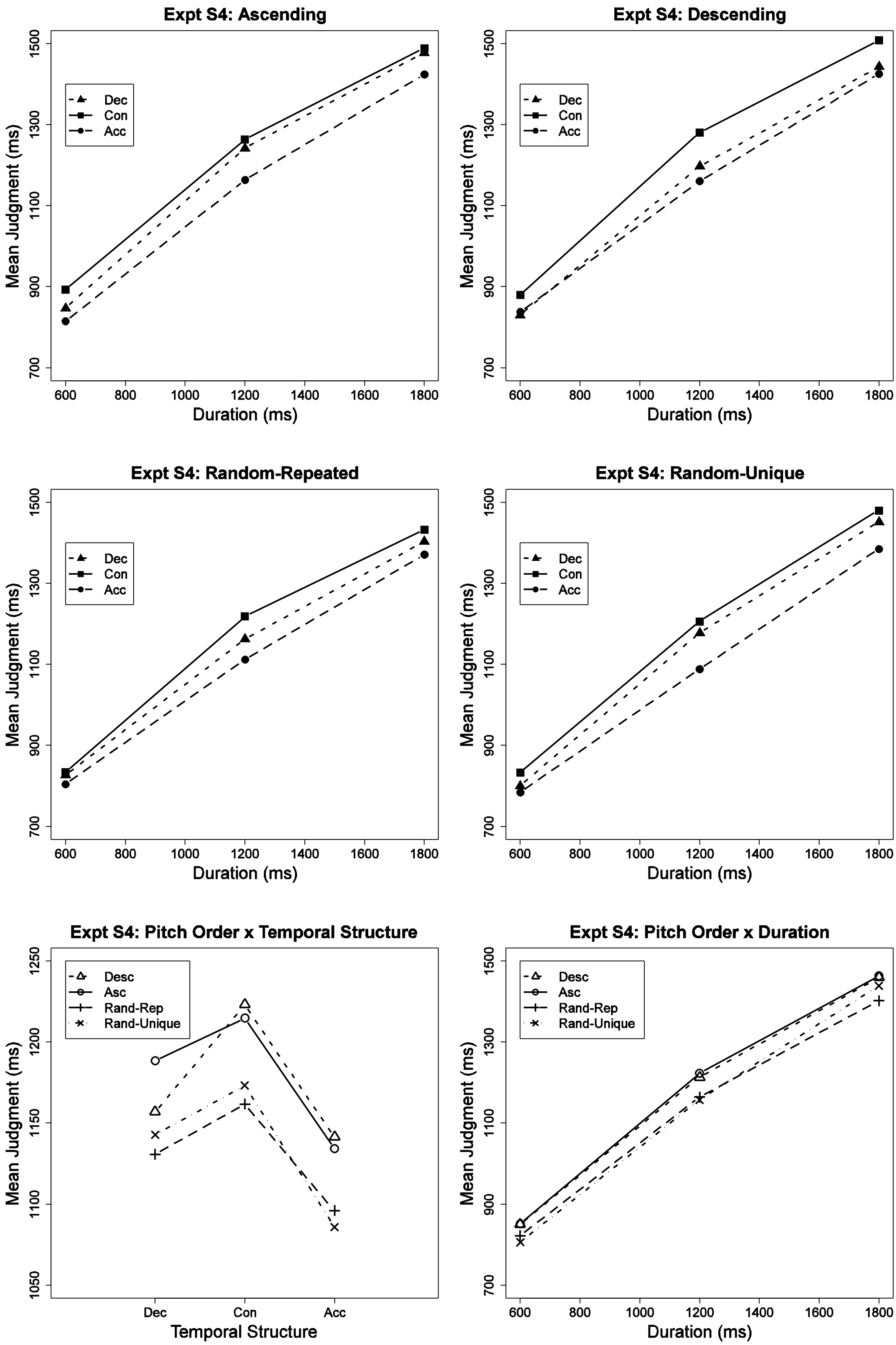
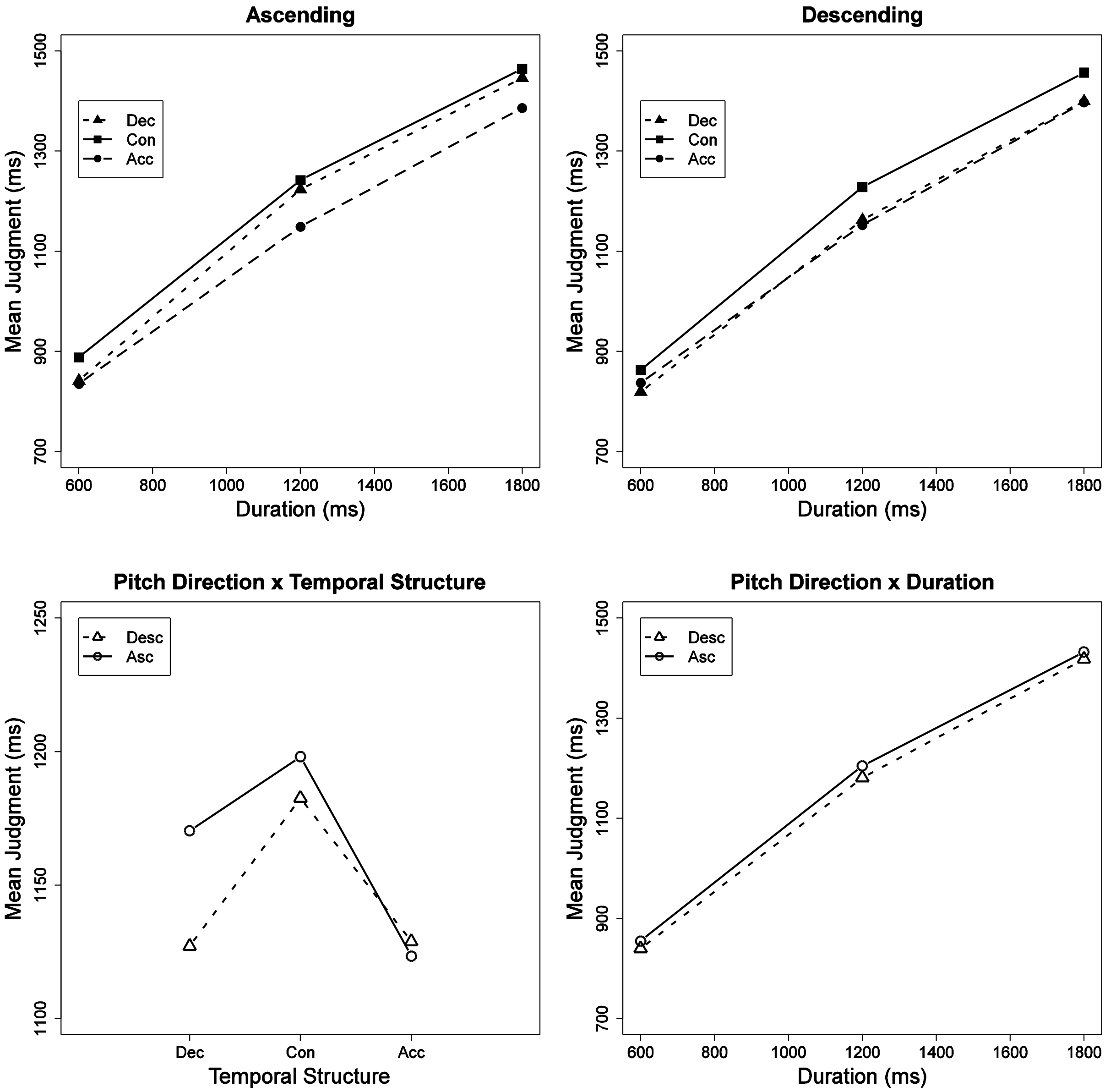

Expt S4: Ascending
Expt S4: Descending
Expt S4: Random-Repeated
Expt S4: Random-Unique
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
700
900
1100
1300
1500
600
800
1000
1200
1400
1600
1800
Expt S4: Pitch Order x Temporal Structure
Expt S4: Pitch Order x Duration
Desc
Asc
Rand-Rep
Rand-Unique
1050
1100
1150
1200
1250
Temporal Structure

Figure S5


Ascending
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
Descending
Dec
Con
Acc
Mean Judgment (ms)
Duration (ms)
Pitch Direction x Temporal Structure
Desc
Asc
Mean Judgment (ms)
Temporal Structure
Pitch Direction x Duration
Desc
Asc
Mean Judgment (ms)
Duration (ms)