THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2006, 59 (10), 1709–1724

Psychology Press
Taylor & Francis Group

When do auditory/visual differences in duration judgements occur?

J. H. Wearden
Keele University, Keele, UK

N. P. M. Todd and L. A. Jones
Manchester University, UK

Four experiments examined judgements of the duration of auditory and visual stimuli. Two used a bisection method, and two used verbal estimation. Auditory/visual differences were found when durations of auditory and visual stimuli were explicitly compared and when durations from both modalities were mixed in partition bisection. Differences in verbal estimation were also found both when people received a single modality and when they received both. In all cases, the auditory stimuli appeared longer than the visual stimuli, and the effect was greater at longer stimulus durations, consistent with a "pacemaker speed" interpretation of the effect. Results suggested that Penney, Gibbon, and Meck's (2000) "memory mixing" account of auditory/visual differences in duration judgements, while correct in some circumstances, was incomplete, and that in some cases people were basing their judgements on some preexisting temporal standard.

It has been known since the 19th century that the durations of auditory and visual stimuli are usually perceived to be different, even when the stimuli actually last the same length of time. In the words of Goldstone and Lhamon (1974) "sounds are judged longer than lights": In other words, auditory stimuli appear to last longer than visual stimuli. Wearden, Edwards, Fakhri, and Percival (1998) attributed the auditory/visual difference to differences in the operation of an underlying internal clock. The clock was proposed to be of a pacemaker–accumulator type (Gibbon, Church, & Meck, 1984), with stimuli timed in terms of the number of "ticks" from a pacemaker that were accumulated during the stimulus, and the modality effect was interpreted in terms of a faster pacemaker rate for auditory than for visual stimuli. This interpretation essentially regards the auditory/visual difference as a "pacemaker speed" effect, similar to that obtained when trains of clicks are used to apparently increase the speed of the pacemaker of a putative internal clock (Penton-Voak, Edwards, Percival, & Wearden, 1996; see also Burle & Bonnet, 1999; Burle & Casini, 2001; Droit-Volet & Wearden, 2002; Wearden et al., 1998; and Wearden, Philpott, & Win, 1999, for replications of this effect).

More recently, Penney et al. (2000) concurred with the idea that pacemaker speed was faster for auditory stimuli than for visual stimuli, but added an additional qualification. They noted that most studies that had demonstrated

Correspondence should be addressed to J. H. Wearden, School of Psychology, Dorothy Hodgkin Building, Keele University, Keele, Staffordshire ST5 5BG, UK. E-mail: j.h.wearden@psy.keele.ac.uk
We are grateful to Carla Hawkes, Joel White, and students from TP groups for help with data collection.

http://www.psypress.com/qjep
DOI:10.1080/17470210500314729

auditory/visual differences in timing had used a within-group design, with people receiving both auditory and visual stimuli. Penney et al. (2000) compared different experimental conditions using a bisection procedure and found reliable auditory/visual differences only when people were exposed to both auditory and visual stimuli: When auditory stimuli alone, or visual stimuli alone, were used (in their Experiment 2) there was no modality effect. Their explanation was in terms of "memory mixing". Suppose that, as Wearden et al. (1998) proposed, auditory stimuli generate more "clock ticks" per unit time than do visual stimuli. If some standard duration with which subsequent durations are compared is composed of a "mixture" of the number of ticks occurring when this duration is presented in auditory and visual modalities, then on average an auditory stimulus will tend to be judged as longer than this standard and a visual one as shorter, giving rise to the modality effect. If, on the other hand, a person experiences only one modality, then memory mixing cannot occur, and no auditory/visual difference would be expected.

The present article reports four experiments bearing on the issue of when auditory/visual differences in duration judgements are found and when they are absent. The first two experiments use a bisection method, in a procedure similar to, but rather simpler than, that employed by Penney et al. (2000).

Although some earlier work on temporal bisection in humans exists (e.g., Bovet, 1968), most recent studies follow the technique used in an animal experiment by Church and DeLuty (1977), which was developed for use with humans by Allan and Gibbon (1991) and Wearden (1991). These experiments used a discrimination method, where subjects were initially presented with examples of "standard" durations, which were identified as either as a standard "short" duration (*S*) or a standard "long" duration (*L*). After several presentations of these standards, a range of stimuli with varying durations (including *S* and *L*) is presented. In Allan and Gibbon's study, participants had to make a SHORT or LONG response to each stimulus, whereas in Wearden (1991) the SHORT or LONG response indicated whether the participant judged the comparison duration to be more similar to *S* than *L*, or vice versa. Suppose that we define a LONG response as that appropriate to *L*, then all studies with humans find a monotonically increasing proportion of LONG responses as stimulus duration increases (see, for example, Wearden & Ferrara, 1995, 1996).

The bisection method has been used extensively to investigate the conformity of human performance to scalar timing theory (or scalar expectancy theory, SET: Gibbon et al., 1984), which was originally developed as an explanation of the performance of rats and pigeons on temporally constrained reinforcement schedules, but which has enjoyed considerable success as an account of human timing performance (Allan, 1998; Wearden, 2003). In the conventional explanation of bisection according to SET (e.g., Allan & Gibbon, 1991), people are proposed to develop memories of the standards, *S* and *L*, and to compare each presented duration, *t*, with *S* and *L* according to some ratio decision process. Variants of this approach assume slightly different decision processes from those that Allan and Gibbon proposed; see Wearden (2004) for discussion. However, all the research is motivated by potential links between bisection performance and the clock, memory, and decision structure of SET (discussed more fully by Penney et al., 2000), and all is concerned with theoretically relevant features of behaviour on bisection tasks, such as the location of the bisection point or the property of superimposition.

The bisection point is the stimulus duration giving rise to 50% LONG responses, and discussion of where this is located in humans is provided by Allan (2002). The property of superimposition is exhibited when measures of timed behaviour from different conditions (e.g., different stimulus ranges or absolute times) superimpose when plotted on the same relative scale. Superimposition is a requirement of SET and implies an underlying constant-sensitivity timing mechanism as the interval timed varies—a form of conformity to Weber's Law.

EXPERIMENT 1

In Experiment 1, participants received blocks of stimuli consisting of an initial presentation of the standards (*S* and *L*), which differed from block to block, followed by nine comparison durations (*S* and *L*, and durations spaced in 80-ms steps between them). After presentation of each comparison, the participant was required to judge whether its duration was more similar to *S* or to *L* (by making a SHORT or LONG response). In Experiment 1, all participants received four experimental conditions, two of which were unimodal (auditory *S* and *L*, auditory comparisons; visual *S* and *L*, visual comparisons), and two of which were cross-modal (auditory *S* and *L*, visual comparisons; visual *S* and *L*, auditory comparisons). Although all participants received both auditory and visual stimuli (so allowing "memory mixing" as a possibility), the block structure of the experiment segregated the different conditions, so it was always clear to the participant which standard was appropriate to which comparison.

If auditory stimuli are judged as longer than visual stimuli, then the cross-modal conditions should generate a large modality effect. The predictions for the unimodal conditions are less clear. If "memory mixing" occurs here, then a modality effect would be expected, but if the auditory and visual memories are kept separate then no modality effect should manifest itself.

The bisection task used in Experiment 1 is considerably simpler than that employed in most conditions by Penney et al. (2000). Our unimodal conditions resemble those from Penney et al.'s Experiment 2, except that in their study, different participant groups received either auditory or visual stimuli, but never both. Experiments 1 and 3 of Penney et al. employed more complex methods, with participants not only receiving the two different modalities in a single experimental session, but sometimes trials in which the auditory and visual stimuli were simultaneously present, but with asynchronous onset and variable overlap. Penney et al. never actually required explicit comparisons between auditory *S* and *L* and visual comparisons, or vice versa, but rather presented examples of *S* and *L* from both modalities at the start of an experimental session.

Method

Participants

A total of 17 first-year psychology undergraduates from the University of Manchester participated to gain course credits.

Apparatus

Participants were tested individually in a small room, which was insulated from external noise and light. An IBM-compatible computer with a colour monitor controlled the experiment, and the computer keyboard served as the response manipulandum. The experimental program was written in the MEL language (Micro-Experimental Laboratory: Psychology Software Tools Inc.), which assured millisecond accuracy for timing of stimuli and responses.

Procedure

The auditory stimulus employed was a 500-Hz tone produced by the computer speaker, and the visual stimulus was a 14 × 14-cm blue square, which was centred in the middle of the screen. The experimental session, which took around 25 minutes to complete, was divided into a series of blocks. Each block started with four presentations of the standard Short (*S*) and standard Long (*L*) stimuli, followed by the presentation of nine comparison durations. *S* was a random value selected from a uniform distribution running from 150 to 300 ms, and *L* was always 640 ms longer than *S*. The comparison stimuli, which were presented in a random order within a block, included the *S* value for that block and increased in 80-ms increments up to *L*. Thus, for example, if on some block *S* was 225 ms, the comparison stimuli would have been 225, 305, 385, 465, 545, 625, 705, 785, 865 ms. *S* and *L* were presented as either auditory (aud) or visual stimuli (vis), and the comparisons within each block were either auditory or visual stimuli. This gave rise to four conditions: aud/aud (i.e., auditory standard, auditory comparison), vis/vis, aud/vis, and vis/aud. Each condition was repeated three times, so that each participant completed 12 blocks within the experimental session.

Instructions displayed on the screen informed the participants that they would be presented with four examples of *S* and *L*, and that they would then be required to classify the comparison stimuli as being more similar either to *S* or to *L*. Participants began the experiment by pressing the spacebar in response to a prompt displayed on the screen. After the presentation of the standards the participant was again prompted to press the spacebar to begin the next trial. After a random delay of between 1,000 and 3,000 ms a comparison stimulus was presented. The stimulus offset was followed by a display, which asked the subject: "Was that more like the standard SHORT or LONG? Press short (S) or long (L) key." The participant began the next trial by pressing the spacebar in response to a prompt displayed on the screen. After the nine comparison stimuli had been presented, the participant was informed that they were to be presented with new standard stimuli. This procedure was repeated until all 12 blocks had been completed. As there was no "correct" answer for this task, no feedback was given.

Results and discussion

Data were taken from all of the blocks, as there was no learning phase in this experiment. Consider first the mean proportion of LONG responses as a function of stimulus duration and type of condition (e.g., aud/aud, vis/vis, etc.). When all four conditions were treated together, analysis of variance (ANOVA) found effects of condition, $F(3, 48) = 24.45$, $p < .0001$, comparison stimulus duration, $F(8, 128) = 253.39$, $p < .0001$, and a significant condition by stimulus duration interaction, $F(24, 384) = 2.22$, $p < .001$. However, this overall analysis is somewhat uninformative about where the differences between conditions actually lay, whereas considering them pairwise illustrates them. The upper panel of Figure 1 shows data from the within-modality conditions (aud/aud and vis/vis), and the lower panel data from the two cross-modal conditions (aud/vis and vis/aud).

Inspection of the upper panel of Figure 1 shows that the aud/aud and vis/vis conditions produced very similar psychophysical functions. This was confirmed by statistical analysis, which found no effect of condition, $F(1, 16) = 1.12$, $p = .31$, nor condition by stimulus duration interaction, $F(8, 128) = 1.15$, $p = .33$, but there was an obvious effect of stimulus duration, $F(8, 128) = 181.55$, $p < .001$. In the cross-modal comparisons (lower panel of Figure 1), on the other hand, there was a marked difference between conditions. When the standards were visual and the comparisons auditory (vis/aud) the psychophysical function

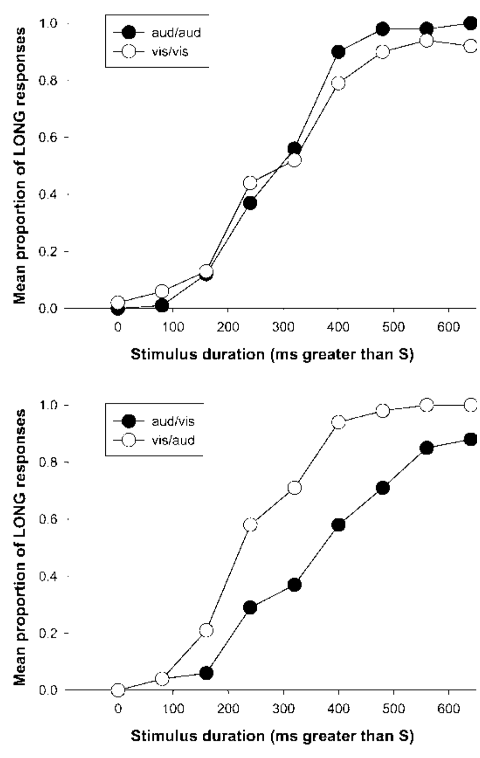

Figure 1. *Psychophysical functions (mean proportion of LONG responses plotted against comparison stimulus duration, given in ms above* S*) from the aud/aud and vis/vis conditions (upper panel), and the aud/vis and vis/aud conditions (lower panel) of Experiment 1.*

was displaced to the left of that obtained when the standards were auditory and the comparisons visual (aud/vis). In this comparison there was a significant effect of condition, $F(1, 16) = 49.50$, $p < .0001$, a significant effect of stimulus duration, $F(8, 128) = 112.06$, $p < .001$, and a significant condition by stimulus duration interaction, $F(8, 128) = 3.80$, $p < .0001$. Such a cross-modal effect is consistent with the notion that for some constant real duration the auditory stimulus was perceived as significantly longer than the visual one.

The cross-modal effect depended on the relation between the standard and comparison durations, rather than on whether the standard was auditory or visual, or the comparisons auditory or visual, although these comparisons are not shown in Figure 1, to save space. For example, significant effects of condition were obtained in the vis/vis versus vis/aud comparisons, $F(1, 16) = 12.49$, $p < .01$, and in the aud/aud versus aud/vis comparisons, $F(1, 16) = 42.89$, $p < .001$.

Table 1 shows bisection points (expressed in terms of ms above *S*) for the different experimental conditions. These were calculated from the psychophysical functions shown in Figure 1 by regressing the proportion of LONG responses against stimulus duration, taking values from the steepest part of the curve, a method introduced by Maricq, Roberts, and Church (1981). The regression line was used to calculate the stimulus duration giving rise to 50% LONG responses—the bisection point. In addition, the regression line was used to calculate the difference limen, half the difference in the duration giving rise to 75% and 25% LONG responses. The difference limen divided by the bisection point is the *Weber ratio*, a measure of temporal sensitivity (essentially a reflection of the steepness of the psychophysical function), where smaller values indicate greater temporal sensitivity.

Table 1. *Bisection points and Weber ratios from Experiments 1 and 2*

Experiment		*BP*	*WR*
1	aud/aud	284	.28
	vis/vis	292	.33
	aud/vis	372	.35
	vis/aud	242	.36
2	Auditory	458	.19
	Visual	561	.16

Note: For Experiment 1 the bisection point is given in ms above *S*. BP = bisection point. WR = Weber ratio.

Inspection of the values in Table 1 shows that the bisection points from the unimodal conditions (aud/aud and vis/vis) were similar and also close to the arithmetic mid-point of the duration range used (320 ms above *S*), whereas the bisection points from the cross-modal conditions deviated markedly from this mid-point and were very different, with the aud/vis bisection point being 130 ms higher than the vis/aud one. The bisection points are, of course, consistent with the picture gained from Figure 1 that the unimodal conditions produced similar psychophysical functions, whereas the cross-modal conditions produced functions that were markedly displaced relative to one another (lower panel of Figure 1). In contrast to the marked effect of condition on bisection point, Weber ratios remained much more similar, ranging from .28 to .36.

The bisection points in Table 1 were used to test the data from Experiment 1 for the property of *superimposition*, the requirement that data from different experimental conditions superimpose when plotted on the same relative scale. The appropriate superimposition test for bisection (Allan & Gibbon, 1991) is to plot the proportion of LONG responses produced for each comparison stimulus against the duration value for that stimulus divided by the bisection point appropriate for the condition being considered. Figure 2 shows the results when this was done. It is clear that the quality of superimposition was good, showing that the main effect of the different conditions was left and right displacement of the psychophysical function.

Our data are very similar in their implications to those of Penney et al. (2000), although they differ slightly in emphasis. It is obviously the case from our results that having auditory and visual durations in the same experimental session does not guarantee an auditory/visual difference, as no

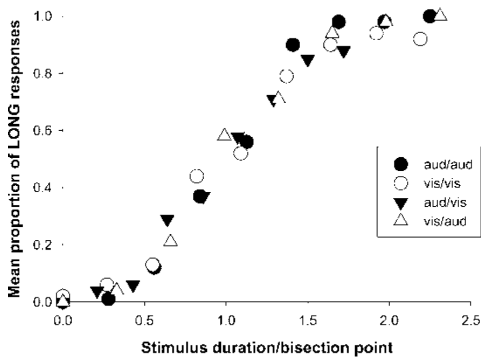

Figure 2. *Psychophysical functions from the four conditions of Experiment 1 (aud/aud, vis/vis, aud/vis, vis/aud) plotted against stimulus duration divided by the bisection point for the condition in force. Bisection point data were taken from Table 1.*

such difference was found in our unimodal conditions. This result implies that the "memory mixing" proposed by Penney et al. (2000) is not a necessary consequence of just experiencing auditory and visual durations. If people in our unimodal conditions had been using some mixture of the auditory and visual standards presented during the experiment, the psychophysical functions from aud/aud and vis/vis would have been displaced relative to one another. Perhaps memory mixing is avoided if the experimental design allows a clear distinction between the standards for different conditions. The design of Penney et al.'s Experiments 1 and 3 intermixed auditory, visual, and simultaneous auditory and visual trials in a test session that followed initial presentation of standards in the two modalities. Feedback was given for responses to comparison durations that actually were *S* and *L* during the test phase, but participants were required to classify each comparison duration as SHORT or LONG, not relative to a standard presented fairly recently (e.g., at the beginning of each block as in our Experiment 1). Their study seems likely to have encouraged memory mixing, as the different conditions were not segregated into blocks as in our study. A slightly different bisection method from the one used above, which might be particularly susceptible to memory mixing, was employed in Experiment 2.

EXPERIMENT 2

Wearden and Ferrara (1995) developed a bisection technique (their *partition* method) in which humans bisected sets of durations without any explicit identification of members of the set as *S* and *L*. We were interested in the question of what performance would look like if members of the to-be-bisected stimulus set were intermixed auditory and visual stimuli. In partition bisection, there are no durations explicitly identified as *S* and *L*, so the participant must presumably construct some kind of standard for themselves. Wearden and Ferrara (1995) suggested that the arithmetic mean of the set of all the presented durations was the basis for judgements in this case. Following the logic of Penney et al. (2000), if auditory durations are subjectively longer than visual durations, and some mixed memory of the auditory and visual durations is used as the basis for decision (e.g., the mean of all the durations, following Wearden & Ferrara, 1995), then visual stimuli will tend to have durations shorter than this mixed standard, and auditory stimuli durations longer, so an auditory/visual duration difference would be expected to manifest itself clearly in this situation. Experiment 2 tested this prediction by performing partition bisection on a set comprising auditory and visual stimuli. Essentially, participants received a set of durations in both modalities and simply had to classify each stimulus duration as SHORT or LONG.

Method

Participants
A total of 15 psychology undergraduates from the University of Manchester participated in the experiment to gain course credits.

Apparatus
The apparatus was the same as that in Experiment 1.

Procedure
The auditory and visual stimuli employed were the same as those in Experiment 1. The experimental

session, which took approximately 20 minutes to complete, was divided into eight blocks. Within each block, nine visual stimuli and nine auditory stimuli were presented in a random sequence. Their durations were 200, 280, 360, 440, 520, 600, 680, 760, and 840 ms. Participants were informed via instructions displayed on the screen that they would be presented with stimuli, which they would have to classify as either SHORT or LONG. The participants were prompted by a display to press the spacebar to begin the experiment. After a random delay of between 2,000 and 3,000 ms, the stimulus was presented, and the participant pressed either the *S* key to classify the stimulus as SHORT or the *L* key to classify the stimulus as LONG. As there was no correct answer for this task, no feedback could, of course, be given. The participant began the next trial by pressing the spacebar, and the experiment continued, until all durations had been delivered.

Results and discussion

In Experiment 2, the participants obviously required some preliminary experience before stable classification of the durations presented as SHORT or LONG could be obtained. We therefore included only data from the last 5, of 8, blocks in calculations. The upper panel of Figure 3 shows the mean proportion of LONG responses plotted against stimulus duration for the auditory and visual stimuli used. Obviously, the psychometric function from the visual stimuli was displaced to the right relative to that for the auditory stimuli, in spite of the fact that Experiment 2 never required explicit cross-modal comparisons.

ANOVA confirmed the difference between the psychometric functions. Auditory stimuli produced more LONG responses overall than did visual stimuli, $F(1, 14) = 35.6$, $p < .0001$, there was the usual effect of stimulus duration, $F(8, 112) = 259.0$, $p < .0001$, and there was also a significant stimulus duration by modality interaction, $F(8, 112) = 11.95$, $p < .0001$.

Bisection points and Weber ratios for the auditory and visual stimuli are shown in Table 1. As in Experiment 1, the auditory bisection point was

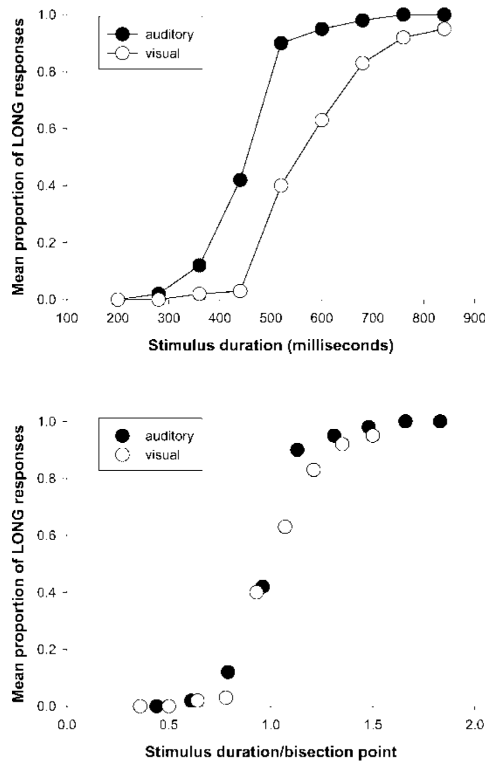

Figure 3. *Upper panel: Psychophysical functions from Experiment 2. Data are shown separately for auditory stimuli (filled circles) and visual stimuli (unfilled circles). Lower panel: Psychophysical functions from Experiment 2, with mean proportion of LONG responses plotted against stimulus duration/bisection point. Bisection point data were taken from Table 1.*

displaced to the left of the arithmetic mean (520 ms), and the visual bisection point was displaced to the right of it. The bisection points were used to test the data in Experiment 2 for superimposition, and the lower panel of Figure 3 shows results when this was done. Obviously, the data from the different modalities superimposed well.

The data obtained in Experiment 2 are completely consistent with Penney et al.'s (2000) idea of memory mixing. Such an explanation fits our data well, particularly if we follow Wearden and Ferrara (1995) and assume that when the partition method is used participants compare the durations presented with the arithmetic mean of all the stimulus lengths. Overall, the subjective mean of the auditory stimuli will be longer than the mean of the visual stimuli, so combining these means by memory mixing will lead to some intermediate value. By comparison to this, auditory stimuli are relatively long, and visual stimuli relatively short, so a cross-modal effect like that observed in the upper panel of Figure 3 would be predicted.

The results of Experiments 1 and 2 generally support the idea that modality effects are obtained readily when (a) there is some sort of comparison between "standards" in one modality and "comparisons" in another (our aud/vis and vis/aud conditions from Experiment 1), and (b) when the method used seems likely to encourage some sort of memory mixing (Experiment 2). Results from the aud/aud and vis/vis conditions of Experiment 1 suggest a modification to Penney et al.'s (2000) conclusions, as it is clear that just experiencing auditory and visual durations in the same experimental session is not sufficient to produce a modality effect when the standards and comparisons are in the same modality.

The results from our two bisection studies using auditory and visual stimuli can be compared with those from another recent study using bisection by Droit-Volet, Tourret, and Wearden (2004). Participants were children of 5 and 8 years, as well as student-age adults, and all received two experimental sessions. In one of these (unimodal), both the short and the long standards (200 and 800 ms) and the comparison durations (200–800 ms in 100-ms steps) were in the same modality, either auditory or visual. In the other session (cross-modal) the standards and comparison durations were in different modalities. At all age groups, participants behaved as if the auditory stimuli were subjectively longer than the visual stimuli in the cross-modal session, as in our conditions requiring explicit (Experiment 1) or implicit (Experiment 2) comparisons between auditory and visual stimuli. Another point of comparison was that, in adults at least, the Weber fractions from the visual/visual comparisons and the auditory/auditory comparisons were very similar, as in our Experiments 1 and 2 (e.g., see Table 1).

Our next two experiments involved between-group comparisons of verbal estimates of the duration of auditory or visual stimuli. In Experiment 3, different participant groups received either visual or auditory stimuli (and no-one received both). This sort of manipulation was not actually used in Wearden et al.'s (1998) study of modality effects, and in their experiment, participants received experimental sessions where auditory and visual durations were intermixed. If in these mixed conditions, people were using some sort of standard resulting from a mixture of auditory and visual durations, then by Penney et al.'s memory-mixing notion, a modality effect would be expected, as in the present Experiment 2, and one was indeed found.

If different groups receive only either auditory or visual stimuli, then no memory mixing can occur, so no modality effect would be expected if this factor alone was responsible for auditory/visual duration differences. Experiment 3 tests this idea simply. Different groups estimated the duration of auditory or visual stimuli in two conditions. In one, the delivery of the duration to be estimated was preceded by a warning signal, and in another case the warning signal was absent, but in both cases, a between-group comparison of estimation of the duration of auditory or visual stimuli was possible. The experiment was actually intended to test the effect of the warning signal, but as will be seen later, this produced no significant effect, so the data are used here to address the question of possible between-group effects on auditory/visual duration judgements.

EXPERIMENT 3

Method

Participants

A total of 30 Manchester University psychology undergraduates participated for course credit. They were arbitrarily allocated to two equal-sized groups, one receiving auditory stimuli and the other visual stimuli.

Apparatus

The apparatus was the same as that in Experiments 1 and 2.

Procedure

All participants received a single experimental session lasting about 30 minutes. An individual made judgements of the duration of either auditory stimuli (500-Hz tones produced by the computer speaker), or visual stimuli (14 × 14-cm light-blue squares presented in the centre of the monitor screen), with no participant receiving both. Consider first the procedure for the auditory group. The 10 stimulus durations used were 77, 203, 348, 461, 582, 707, 834, 958, 1,063, and 1,183 ms, and a block of trials consisted of the presentation of each duration twice, once preceded by a warning signal and once without. To produce each stimulus, the participant pressed the spacebar in response to a "Press spacebar for next trial". On no-warning-signal trials, this response was followed by a random delay between 3,000 and 7,000 ms and the presentation of the duration to be estimated. On warning-signal trials, the delay before the warning signal was random, between 1,000 and 5,000 ms, and the warning signal (a 50-ms 2,000-Hz tone) preceded the presentation of the stimulus by 2 s. When the stimulus terminated, the participant typed in the estimated duration of the stimulus in milliseconds using the numeric keypad on the computer keyboard. This was followed by the "Press spacebar . . ." prompt and the next trial. No feedback as to accuracy of estimates was given.

The procedure for the visual group was identical except for the stimuli presented. Before any trials were delivered, participants in both groups were instructed to estimate the duration in milliseconds, using a scale where 1,000 = 1 s, and were also told that all durations presented were between 50 and 1,500 ms. They were further informed that on half the trials a brief beep would precede the presentation of the stimulus by exactly 2 s.

Participants in both groups received five blocks of 20 trials (10 durations with and without warning signal), with the trials being randomly ordered within the block.

Results and discussion

Figure 4 shows mean verbal estimates of the duration of the auditory and visual stimuli presented with data taken from the last four of the five blocks of the experiment. Verbal estimates were filtered to remove all values outside the 50–1,500-ms range. The upper panel shows data from trials without a warning signal, and the lower panel data from the warning-signal trials. Inspection of both panels immediately suggests that a modality effect was present in data obtained, as mean verbal estimates of the duration of auditory stimuli were always higher than those for visual stimuli of the same real duration. Furthermore, the difference between the mean estimates increased with increasing duration, suggesting that the estimation functions differed in slope.

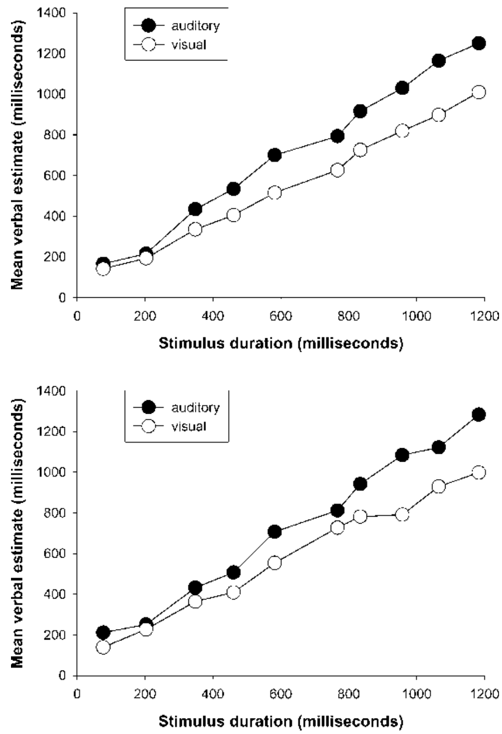

Figure 4. *Mean verbal estimates from Experiment 3, plotted against stimulus duration in ms. Upper panel: no-warning-signal trials; lower panel: warning-signal trials. Data are shown separately for auditory (filled circles) and visual (unfilled circles) stimuli.*

An overall ANOVA found no significant effect of presence or absence of warning signal, although there was a trend towards an effect, $F(1, 298) = 3.76$, $p < .07$, but there were significant effects of stimulus modality, $F(1, 28) = 6.24, p < .02$, and stimulus duration, $F(9, 252) = 257.70, p < .001$, and a significant Modality (auditory or visual) × Duration interaction, $F(9, 252) = 3.97, p < .001$. The first of these significant effects shows that auditory stimuli produced longer duration estimates than did visual stimuli, even though no participant experienced more than one modality, the second shows the obvious effect of stimulus duration on mean estimate, and the third shows that the effect of modality increased with increasing stimulus duration. No other interactions were significant, or approached significance.

When the no-signal and warning-signal conditions were analysed separately, the results were identical in form: There were significant effects of modality: no signal, $F(1, 28) = 6.65, p < .05$; signal, $F(1, 26) = 5.52, p < .05$, and stimulus duration: no signal, $F(9, 252) = 184.35, p < .001$; signal, $F(9, 252) = 193.82, p < .001$, and a significant Modality × Stimulus duration interaction: no signal, $F(9, 252) = 2.83, p < .01$; signal, $F(9, 252) = 3.59, p < .01$.

The results of Experiment 3 demonstrate a very clear modality effect in duration judgements, even when different groups experienced only one stimulus modality. The modality effect was not only very clear in the between-group comparisons, but also seemed to have the same form as the modality effect reported from within-group verbal estimation in Wearden et al. (1998)—that is, with the difference between the mean estimates from auditory and visual stimuli increasing with increasing duration. As discussed in Wearden et al. (1998), such a result is exactly what would be predicted if the pacemaker of the internal clock ran faster for auditory than for visual stimuli. Wearden et al. (1998, p. 104) show that according to the standard mathematical description of a pacemaker–accumulator clock, pacemaker speed multiplies real-time duration, so if two conditions differ in pacemaker speed, the effect should be more marked at longer durations than short durations (see also Droit-Volet & Wearden, 2002). Alternative accounts of auditory/visual differences, such as the idea that switch processes of the internal clock operated with different latencies for auditory and visual stimuli, would, as Wearden et al. showed, generate a constant difference between auditory and visual duration judgements, rather than the multiplicative effect shown in Figure 4.

The between-group modality effect poses problems for the memory mixing interpretation of auditory/visual differences in duration judgements, as no memory mixing could have occurred in Experiment 3, yet a clear effect was found. It seems as if memory mixing is not necessary for auditory/visual differences to occur, although it may be sufficient (e.g., Experiment 2).

EXPERIMENT 4

The final experiment was a more elaborate version of Experiment 3, involving three participant groups. For two of these, verbal estimation was performed on either auditory or visual stimuli (as Experiment 3), and for the third group participants received both auditory and visual stimuli, as in the verbal estimation experiments of Wearden et al. (1998).

Method

Participants

A total of 60 Manchester University psychology undergraduates served.

Apparatus

The apparatus was the same as that in Experiment 3.

Procedure

The participants were arbitrarily allocated to three equal-sized groups. For two of these (aud and vis), the stimuli whose duration had to be estimated were either 500-Hz tones (aud), or blue squares (vis), as Experiment 3. The 10 stimulus durations were the same as those in Experiment 3 and ranged from 77 to 1,183 ms. Participants received five blocks of trials, where each block consisted of presentation of the 10 stimulus durations in a random order. For the third group (aud/vis), participants received both auditory and visual stimuli, with the 10 stimulus durations being presented once in each modality in a block of 20 trials. Participants received five blocks of trials, in each of two identical experimental sessions. All other details were as for the no-warning-signal trials of Experiment 3 except that when the spacebar was pressed to present the stimulus, the delay before stimulus presentation was a random value between 2,000 and 3,000 ms, rather than the longer values used in Experiment 3.

Results and discussion

Data were taken from Blocks 2 to 5 of the experiment and were filtered to remove all estimates outside the specified range (50–1,500 ms). The filtering process resulted in 1 participant in the aud/vis group having no data for some stimulus durations, so all data from this participant were discarded, leaving 19 participants in the aud/vis group, and data for 1 participant in the aud group were lost as a result of a data storage error, leaving 19 participants in this group. Figure 5 shows mean verbal estimates plotted against stimulus from the aud and vis groups (i.e., the participants receiving only one modality: upper panel), and the aud/vis group (participants receiving both modalities: lower panel). Data are shown separately for the two experimental sessions, and for auditory and visual stimuli.

Consider first data from the between-groups comparison (aud and vis: upper panel of Figure 5). Inspection of the data suggests that there was (a) little effect of experimental session, (b) a large effect of stimulus duration on mean verbal estimate, and (c) a marked modality effect, which took the form of the estimates for auditory stimuli being longer than those for visual stimuli. In addition, it appeared as if the auditory/visual difference increased with increasing stimulus duration.

All these suggestions were confirmed by statistical analysis. ANOVA found no effect of session on mean verbal estimates, $F(1, 37) = 2.96$, $p = .10$, but there were significant effects of stimulus duration, $F(9, 333) = 426.90$, $p < .001$, and stimulus modality, $F(1, 37) = 8.90$, $p < .01$, and a significant Stimulus Modality × Stimulus Duration interaction, $F(9, 333) = 8.05$, $p < .001$. No other interactions (Session × Modality, Session × Stimulus Duration, or Session × Duration × Modality) approached significance.

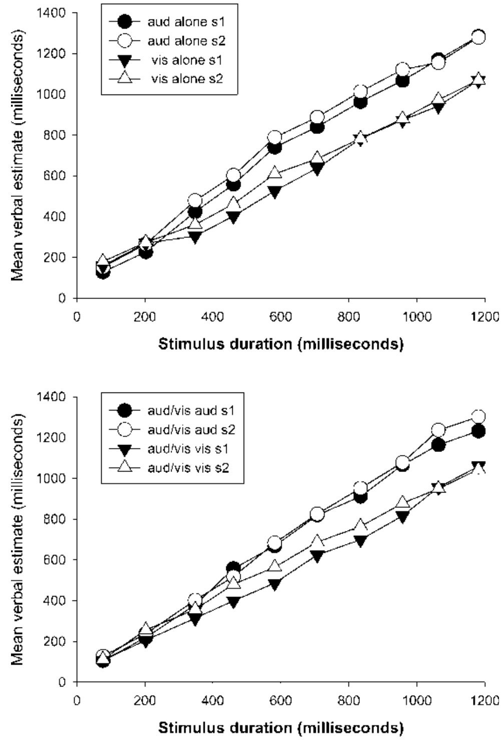

Figure 5. *Upper panel: Mean verbal estimates from the between-group comparisons of Experiment 4 (groups aud and group vis) plotted against stimulus duration in ms. Data are shown separately for auditory and visual stimuli, and for the two experimental sessions (s1 and s2), see key for details. Lower panel: Mean verbal estimates from the aud/vis group of Experiment 4. See key for details.*

The first of the significant effects shows that the auditory/visual difference was replicable over sessions, the second shows the obvious effect of stimulus duration on mean estimates, the third confirms that there were significantly higher estimates when auditory stimuli were used than when visual stimuli were used, and the final result, the interaction, shows that the auditory/visual difference increased with increasing stimulus duration (i.e., was multiplicative with duration).

The lower panel of Figure 5 shows mean verbal estimates from auditory and visual stimuli, from both experimental sessions, from the aud/vis group who received auditory and visual stimuli intermixed within blocks. Once again, inspection suggests that effects of session were small, but that stimulus duration and stimulus modality had large effects, and that these two variables interacted so that the auditory/visual difference was greater at longer durations. The pattern of results, in fact, appears identical to that shown in the upper panel of Figure 5, where data come from between-group comparisons.

These suggestions were confirmed by statistical analysis. There was no effect of session, $F(1, 18) = 1.93$, $p = .18$, but there were significant effects of stimulus duration, $F(9, 162) = 544.59$, $p < .001$, and of stimulus modality, $F(1, 18) = 69.88$, $p < .001$, as well as a significant Modality × Stimulus Duration interaction, $F(9, 162) = 12.32$, $p < .001$. No other effects approached significance.

It seems, therefore, that not only could auditory/visual differences in duration judgements be obtained from between-group comparisons (as in Experiment 3), but the effects of within-group comparisons, where people had received both auditory and visual stimuli, and those of between-group comparisons were highly similar. Another way of examining whether the between-group modality effect and the within-group one are the same is to plot the data in the form used in Figure 6.

Figure 6 shows the same data as those in Figure 5, but this time all the auditory judgements are in the top panel, and all the visual judgements are in the bottom one. So, we can ask whether perceived durations of auditory (or visual) stimuli were the same when people received only one modality as when they received both of them. Inspection of both panels of Figure 6 suggests that the judgements of auditory (or visual) durations were more-or-less identical whether they came from between- or within-group comparisons, as there is little apparent difference in the judgements derived from the different sessions or different conditions (one modality, or both), and no effect apart from the obvious effect of stimulus duration.

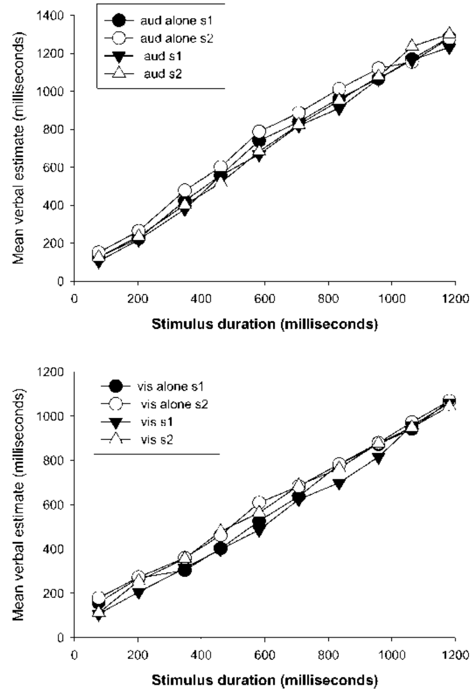

Figure 6. *Mean verbal estimates from Experiment 4 plotted against stimulus duration in ms. Upper panel: Data from judgements of auditory stimuli; aud alone is from the aud group; aud is from auditory judgements by the aud/vis group; s1 and s2 indicate Sessions 1 and 2. Lower panel: Data from judgements of visual stimuli; vis alone is from the vis group; vis is from the vis/aud group.*

These suggestions were confirmed by statistical analysis. ANOVA was used to examine the between-condition effect of one modality or both being experienced by the participant and within-condition effects of stimulus duration and session. Consider first judgements of the duration of auditory stimuli (upper panel of Figure 6). Having one modality or both produced no significant difference in mean verbal estimates, $F(1, 36) = 0.604$, $p = .44$, and there were no significant effects of session, $F(1, 36) = 2.50$, $p = .123$, nor any significant interactions, although there was the obvious effect of stimulus duration, $F(9, 324) = 686.68$, $p < .001$. When judgements of visual stimuli were considered (lower panel of

Figure 6), there was a nearly significant effect of session, $F(1, 37) = 3.96$, $p < .06$, but no other significant effect or significant interaction, apart from that of stimulus duration, $F(9, 333) = 368.56$, $p < .001$. In particular, the effect of having one modality or both did not approach significance, $F(1, 37) = 0.33$, $p = .57$.

The data from Experiment 4 replicated those from the present Experiment 3, as well as those from a within-group auditory/visual comparison in Wearden et al.'s (1998) Experiments 2 and 3. The principal novel contribution of Experiment 4 was the demonstration (shown most clearly in Figure 6) that judgements of auditory and visual durations were not significantly different whether the participant experienced only a single modality or both. So not only were auditory/visual duration differences obtained in within-group and between-group comparisons (Figures 4 and 5), but the behaviour observed in these two different cases was apparently nearly identical and certainly not different statistically (Figure 6). The implications of these results are discussed further below.

GENERAL DISCUSSION

Our results, like almost all previous results, confirm the contention that "sounds are judged longer than lights", but perhaps the most striking finding obtained was that from Experiments 3 and 4. Here, not only could auditory/visual differences in duration judgements be obtained when people received only one stimulus modality, but the effect on verbal estimation of duration seemed identical to that obtained when they received both.

In fact, obtaining auditory/visual differences in duration judgements when people receive only one modality is not novel. Penney et al. (2000, p. 1770) state that Wearden et al. (1998) and Walker and Scott (1981) used both between- and within-subject designs, whereas in fact in all experiments in both articles all participants received both auditory and visual stimuli, but one study that certainly obtained an auditory/visual duration judgement difference from a between-group experiment was Goldstone, Boardman, and Lhamon (1959). In their Experiment 1, four different groups received auditory stimuli alone, visual alone, visual/auditory, or auditory/visual, and in the mixed-modality conditions people received the first-mentioned modality for a block of trials, followed by the second-mentioned modality. The task was to decide whether each stimulus presented was greater or less than 1 s. Durations started at 1 s then either ascended or descended depending on what the response to the 1-s stimulus was. If we consider just data from the auditory- and visual-alone groups, a very large auditory/visual difference was found: The auditory duration judged to be more or less than 1 s 50% of the time was 0.59 s, whereas the visual duration was 1.02 s, a 0.43-s difference. When people received both modalities the difference was 0.31 s for auditory then visual and 0.45 s for visual then auditory. So, just as in our Experiment 4, Goldstone et al. (1959) not only found auditory/visual differences from single-modality conditions, but also found that the difference from these conditions seemed to be of about the same magnitude as when people experienced both modalities.

How can the results obtained from the present experiments, from Penney et al. (2000), and from previous studies be reconciled? Perhaps the starting point is to assert that the "memory mixing" account of Penney et al. is definitely not incorrect (indeed, we provide strong evidence for it in our Experiment 2), but it is certainly incomplete. When memory mixing is likely to occur, auditory/visual differences in duration judgements will be manifested, but they can apparently occur, and be of large magnitude, in other situations too, as in Goldstone et al. (1959) and our Experiments 3 and 4. The critical factor may be the content and nature of the comparison processes used to produce the time judgements. Consider the conditions of our Experiment 1. In the cross-modal conditions, the standards presented at the start of the block and the comparison durations subsequently presented were in different modalities. Fewer "clock ticks" accumulate for the visual than for the auditory stimuli, so auditory comparisons are perceived as relatively long compared to visual standards, and vice versa, and a marked modality effect occurs (see also Wearden et al., 1998, for a similar result from a temporal generalization method). In our unimodal conditions, standards and comparisons are in the same modality, so pacemaker speed does not differ between them, and no modality effect is found. There is no memory mixing in our Experiment 1 because the block structure of the procedure essentially forces participants to use the standards at the start of the block as a basis for judgements of the comparisons. This apparent segregation of the auditory and visual stimuli is perhaps encouraged by the fact that the standard *short* and *long* durations changed between blocks, and participants have been informed that this is the case. There is thus no tendency to compare visual comparison durations with auditory standards, or vice versa, in the unimodal conditions of our Experiment 1.

As mentioned above, the methods used by Penney et al. (2000) in their Experiments 1 and 3 did not segregate auditory and visual stimuli into blocks, and all had to be compared with standards learned at the start of the experimental session. In addition, the instructions did not emphasize comparing auditory comparisons with auditory standards, and visual comparisons with visual standards, so memory mixing seems more likely to have occurred than with our procedure. Penney et al.'s Experiment 2, on the other hand, used auditory *S* and *L* with auditory comparisons, and visual *S* and *L* with visual comparisons, and did not obtain a between-group modality effect. However, the critical feature in avoiding memory mixing in this experiment may not be the fact that people only experienced one modality, but that by doing so it was necessarily clear which comparison durations should be compared with which standards. Our Experiment 1 shows that people can experience both auditory and visual durations and still not show memory mixing, if experimental trials are appropriately arranged.

The between-group effect obtained in our Experiments 3 and 4, and by Goldstone et al. (1959), can also be explained in terms of pacemaker speed differences between auditory and visual stimuli, albeit more speculatively. Penney et al. (2000, p. 1770) argue that studies that have obtained between-group modality effects have "required participants to compare the [auditory or visual] signal to an internal representation held prior to the experiment (e.g., a clock second)", and this is certainly true of the study of Goldstone et al. (1959), although this work is not cited by Penney et al. We might also extend this idea to verbal estimation of duration. Suppose that people perform verbal estimation by comparing each duration presented to some preexisting internal standard—for example, some representation of 1 s that gives rise to a response of "1,000". Any comparison of auditory and visual stimuli with a common standard will then give rise to a modality effect, if the pacemaker speed interpretation of auditory/visual differences is correct, as for any real-time value an auditory duration will produce more "ticks" than would a visual one and will thus appear longer by reference to some common standard. There is no requirement that the common standard be accurate (i.e., what a person believes to be 1 s may be above or below the real time value of 1 s), although the mean verbal estimates produced in our experiments suggest that the standard cannot be wildly inaccurate; it is only necessary that the same internal standard be used for both auditory and visual comparisons. If the internal standard is conceived of in terms of "ticks" of the internal clock, there is also no need for the rate of the pacemaker operative during creation of this internal standard to have any special relation to the rate during auditory or visual stimuli. That is, if the putative standard is based on x "clock ticks" per s, x can be between the rate obtaining for auditory stimuli and the, lower, rate for visual stimuli, can be higher than both, or can be lower than both. In all cases, auditory/visual modality effects would be obtained if a common standard were used.

This interpretation implies that between-group modality effects may be commonly obtained, and it also implies that even in within-group auditory/visual comparisons (like those in Experiment 4, and in Wearden et al., 1998), the participants may not actually be comparing the auditory and visual durations directly, nor producing any "mixture" of the two, but merely comparing each stimulus independently with some common standard. Indeed, the fact that between-group and within-group modality effects were identical in our Experiment 4 suggests that the processes involved might have been the same in both cases.

Although our Experiments 3 and 4 have demonstrated large and reliable between-group differences in the estimation of the duration of auditory and visual stimuli, and they have attributed this difference to the existence of some common "standard", which probably exists prior to the experimental procedure (as Penney et al., 2000, themselves assume), we have no definite suggestions at present as to what this putative standard might be, how it is generated, or how it is used in different situations. Many questions remain: For example, is the behaviour that a particular person exhibits when asked to produce or estimate "1 second" useful in predicting how this individual will respond when the duration to be produced or estimated is different? Do temporal "standards" remain constant within individuals between tasks, or constant over time with the same task? What characteristics of the individual (such as age or measured intellectual ability) predict the value of temporal "standards" that the individual uses on timing tasks? Although it is a cliché of psychology to suggest that "further research is needed", in the present case the search for putative extraexperimental temporal "standards" may be central to our understanding of how individuals perform on some commonly used timing tasks.

To return to the question posed in the title of our article, differences in the judgements of the duration of auditory and visual stimuli will certainly occur when a stimulus in one modality is compared with one in another modality, but may also occur without exposure to both stimulus types in some conditions, probably as the result of use of some common standard for judgement of the duration of both sorts of stimulus.

Original manuscript received 13 February 2003
Accepted revision received 5 August 2005
First published online 30 January 2006

REFERENCES

Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, *44*, 101–117.

Allan, L. G. (2002). The location and interpretation of the bisection point. *Quarterly Journal of Experimental Psychology*, *55B*, 43–60.

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, *22*, 39–58.

Bovet, P. (1968). Echelles subjectives de durée obtenues par une méthode de bissection. *L'Année Psychologique, 68*, 23–36.

Burle, B., & Bonnet, M. (1999). What's an internal clock for? From temporal information processing to the temporal processing of information. *Behavioural Processes*, *45*, 59–72.

Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time estimation: Implications for clock models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 195–205.

Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes, 3*, 216–228.

Droit-Volet, S., Tourret, S., & Wearden, J. (2004). Perception of the duration of auditory and visual stimuli in children and adults. *Quarterly Journal of Experimental Psychology, 57A*, 797–818.

Droit-Volet, S., & Wearden, J. H. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *Quarterly Journal of Experimental Psychology, 55B*, 193–211.

Gibbon, J., Church, R. M., & Meck, W. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences: Vol. 423: Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.

Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1959). Intersensory comparisons of temporal judgments. *Journal of Experimental Psychology*, *57*, 243–248.

Goldstone, S., & Lhamon, W. T. (1974). Studies of auditory–visual differences in human time judgment: 1. Sounds are judged longer than lights. *Perceptual and Motor Skills*, *39*, 63–82.

Maricq, A. V., Roberts, S., & Church, R. M. (1981). Metamphetamine and time estimation. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 18–30.

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.

Walker, J. T., & Scott, K. J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1327–1339.

Wearden, J. H. (1991). Human performance on an analogue of an interval bisection task. *Quarterly Journal of Experimental Psychology*, *43B*, 59–81.

Wearden, J. H. (2003). Applying the scalar timing model to human time psychology: Progress and challenges. In H. Helfrich (Ed.), *Time and mind II*: *Information-processing perspectives* (pp. 21–39). Gottingen, Germany: Hogrefe & Huber.

Wearden, J. H. (2004). Decision processes in models of timing. *Acta Neurobiologiae Experimentalis, 64*, 303–317.

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

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

Wearden, J. H., & Ferrara, A. (1996). Stimulus range effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology, 49B*, 24–44.

Wearden, J. H., Philpott, K., & Win, T. (1999). Speeding up and (. . . relatively . . .) slowing down and internal clock in humans. *Behavioural Processes, 46*, 63–73.