Review

Judgements of the Duration of Auditory and Visual Stimuli

John H. Wearden^{1,2,*} and Luke A. Jones²

¹School of Psychology, University of Keele, Keele, ST5 5BG, UK

²Division and Neuroscience and Experimental Psychology, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK

Received 27 April 2020; accepted 22 September 2020

Abstract

Studies of judgements of the durations of filled auditory and visual stimuli were reviewed, and some previously unpublished data were analysed. Data supported several conclusions. Firstly, auditory stimuli have longer subjective durations than visual ones, with visual stimuli commonly being judged as having 80–90% of the duration of auditory ones. Secondly, the effect was multiplicative, with the auditory/visual difference increasing as the intervals became longer. Only a small number of exceptions to both these conclusions were found. Thirdly, differences in variability between judgements of auditory and visual stimuli derived from most procedures were small and sometimes not statistically significant, although differences almost always involved visual stimuli producing more variable judgements. Currently, the most viable explanation of the effects appears to be some sort of pacemaker-counter model with higher pacemaker speed for auditory stimuli, although this approach cannot, in its present form, deal quantitatively with all the findings usually obtained.

Keywords

Duration judgements, auditory stimuli, visual stimuli, pacemaker-counter models

1. Introduction

Any beginning student of time perception soon learns, in Goldstone and Lhamon's (1974) words, that 'sounds are judged longer than lights', in other words, auditory

*To whom correspondence should be addressed. E-mail: j.h.wearden@keele.ac.uk.

stimuli have longer subjective durations than visual ones of the same real length. In fact, differences between the perception of durations defined by auditory and visual events have been known since the 19th century, and were remarked on by both Vierordt in 1868 (see Lejeune & Wearden, 2009) and Guyau (1890).

The aim of the present article is to try to characterise differences in duration perception when auditory and visual stimuli are used more precisely than has been done before, so that we have a better idea than previously as to just what facts need to be explained, whatever type of explanation is preferred. We briefly discuss some types of potential explanation later in the article, but our main aim is to address a number of issues which have not previously received detailed attention. To anticipate conclusions to be drawn later, there is overwhelming evidence for Goldstone and Lhamon's statement, but more specific questions might be asked. Are there exceptions and if so when do they occur? Are there situations which systematically make the effect larger or smaller? Given the auditory/visual difference, how should it be more precisely characterised? Is the difference *multiplicative* (that is, grows larger as durations judged increase) or *additive* (constant with changes in duration)? What about variability? Sounds may seem subjectively longer than lights, but do they generate less or more variable temporal representations, or is there no difference between the two?

The present article focusses rather narrowly on potential effects of modality when filled auditory and visual stimuli are judged in fairly simple psychophysical-type paradigms. This is done to keep the discussion manageable, but this decision excludes much interesting work, such as effects of auditory and visual markers when intervals are unfilled (see Grondin, 2003, for a review), or auditory/visual multi-sensory interactions in more complex situations (e.g., Van Wassenhove, 2013).

2. Some General Issues

Suppose that there is a difference between the average subjective duration of a visual stimulus and an auditory one, when their real-time durations are in fact the same. When might this difference be expected to manifest itself? We might define two sorts of comparison that might be made: *intramodal*, where stimuli in the same modality are compared, and *crossmodal*, where stimuli in one modality are contrasted with those in another. In some intramodal experiments, more than one stimulus modality may be presented, but the critical feature is that auditory stimuli are compared with other auditory ones, and visual with visual.

If there are differences in the mean durations perceived when auditory and visual stimuli are presented, then it seems clear that within-subject crossmodal comparisons should be the best way of revealing them, as they represent a 'state change' between one condition and another, which has been an important feature of studies which involve testing the idea that the rate of temporal accumulation is different between conditions. As will be seen later, differential rates of temporal

accumulation (or different ‘clock speeds’) has been a popular explanation of some auditory/visual differences in duration judgements. To illustrate the basic idea, if some sort of standard stimulus is presented in state A, then the state of the organism is changed, for example by drug administration (Meck, 1983) to state B, and if comparison stimuli are presented, then potential effects of the drug on temporal accumulation should be observable. However, if the standard and comparison are both presented in either state A or B, then no effect of differential rates of temporal accumulation will be observed, even if they exist (Meck, 1983). Wearden and Jones (2013) provide a detailed discussion of this argument. In the present case, crossmodal comparisons correspond to state change conditions, whereas intramodal ones do not.

Another situation in which auditory/visual differences, if present, should manifest themselves is one in which auditory and visual stimuli have some common comparison, for example being reproduced by a button press without any stimulus being present. In contrast, some reproduction methods involve stimuli at both the initial presentation and comparison phase — for example, an auditory or visual stimulus is presented and the duration of another auditory or visual stimulus is controlled by the participant to reproduce its duration. Now, if the stimuli are in different modalities in the presentation and reproduction phase, the situation is a crossmodal one, whereas it is intramodal if the same stimulus type is used in both phases (for an example of this procedure see Bratzke & Ulrich, 2019). The crossmodal reproduction case involves state change between presentation and reproduction whereas the intramodal case does not.

Although intramodal comparisons cannot provide a definitive test of the idea of differential rates of temporal accumulation between auditory and visual stimuli, this does not mean that they cannot provide any useful information. As will be seen later, they may provide important clues to the relative variability of judgements of auditory and visual stimuli, although this may not bear directly on the question of whether mean judgements differ. If people only receive one modality, as in a between-group comparison, then they have no basis for direct intermodality comparison. However, as will be seen later, there are what seem to be special cases where auditory/visual differences in mean judgements are found even in between-group comparisons.

3. Studies Finding Auditory/Visual Differences in Mean Judgements

In the present section, we discuss a number of different procedures that have been used to explore auditory-visual differences. We focus specifically in this section on the critical crossmodal comparisons when auditory and visual stimuli were used, although some of the articles report data from other conditions, and some have complex aspects not discussed here.

Table 1 summarises the stimuli used, the method, and the range of durations employed in the articles discussed below.

Table 1.
Summary of studies using auditory and visual stimuli discussed in the text. Table shows the study, the stimuli presented (auditory first), the general method used, and the duration range employed.

<i>Study</i>	<i>Stimuli</i>	<i>Method</i>	<i>Duration range</i>
Behar & Bevan, 1961	White noise, neon lamp	Category scaling	1–5 s
Bobko et al., 1977	1500-Hz tone, black 'X'	Verbal and magnitude scaling	0.25–5 s
Bratzke & Ulrich, 2019	White noise, 7.5 cm blue square	Reproduction	800 or 2400 ms
Brown & Hitchcock, 1965	1000-Hz tone, tachistoscope field	Reproduction	1–17 s
Droit-Volet et al., 2004	500-Hz tone, 4.5 cm filled blue circle	Bisection	200–800 ms
Droit-Volet et al., 2007	500-Hz tone, 4.5 cm filled blue circle	Bisection	3–6 s
Goldfarb & Goldstone, 1964	'Sounds of comfortable intensity', white lights	Category scaling	0.15–1.95 s
Goldstone, 1968	1000-Hz tone, blue-white light	Production and reproduction	1–4 s
Goldstone & Lhamon, 1974	650-Hz tone, 2.5 cm square	Category scaling	0.7–1.3 s
Goldstone et al., 1959	725-Hz tone, white light	Comparison with 1 s	0.1–2 s
Goldstone et al., 1978	1000-Hz tone, red LED	Pair comparison	0.7–1.3 s
Grondin et al., 1998	1000-Hz tone, red LED	Threshold determination	400 or 800 ms
Hirsch et al., 1956	250-Hz tone, 'flash tube'	Reproduction	1–16 s
Indraccolo et al., 2016	White noise, squares	Reproduction	600, 800, or 1200 ms
Klapproth, 2003	500-Hz tones, grey rectangles	Temporal generalization	100–700 ms
Penney et al., 2000	880/440-Hz tones, 5 cm black squares	Bisection	3/6, 2/8, or 4/12 s
Rammesayer et al., 2015	White noise, red LED	Threshold determination	50 and 1000 ms
Szelag et al., 2002	200-Hz tone, green rectangle	Reproduction	1–5.5 s
Ulrich et al., 2006	White noise, green LED	Pair comparison	600–1400 ms
Walker & Scott, 1981	600-Hz tone, neon light	Reproduction	500–1500 ms
Wearden et al., 1998	500-Hz tone, 4 cm blue square	Temporal generalization, verbal estimation	200–800 ms (temporal generalization), 77–1183 ms (verbal estimation)
Wearden et al., 2006	500-Hz tone, 14 cm blue square	Bisection, verbal estimation	150–940 ms (bisection), 77–1183 ms (verbal estimation)
Williams et al., 2019	500-Hz tone, green LED	Threshold determination, verbal estimation	700 ms (threshold), 77–1183 ms (verbal estimation)

Goldstone et al. (1959) conducted three studies using auditory and visual (and sometimes bimodal) stimuli, in ascending or descending series of duration, where the task was to decide whether each stimulus had a duration greater than or less than 1 s, although no example of 1 s was presented. Results when more than one modality was presented on the trial were complex, but for present purposes the effects when only one stimulus was present on each trial are most pertinent. Here, a clear auditory-visual difference emerged, with the ‘auditory second’ being judged as much shorter than the visual one. In other words, less physical time in the auditory modality was needed to equal the subjective 1 s than in the visual modality, so auditory stimuli are judged as relatively longer than visual ones. This effect was obtained in groups who received either auditory or visual stimuli, as well as in groups that received both.

Behar and Bevan’s (1961) article was mostly concerned with the influence of ‘anchor’ stimuli on judgements of other stimuli, rather than auditory/visual differences *per se*, but their Experiment 3 presented people with an intermixed series of auditory and visual stimuli, ranging in duration from 1 to 5 s. Each stimulus had to be classified on an 11-point scale ranging from *very very very short* to *very very very long*. The auditory stimuli received higher category judgements at all duration values. Goldfarb and Goldstone (1964) obtained the same sort of auditory/visual difference with a very similar method.

Goldstone (1968) asked participants to produce or reproduce time intervals defined by auditory or visual stimuli. People attempted to produce 40 durations of 1, 2, 3, and 4 s by controlling the duration of auditory or visual stimuli. Two groups were used, which differed only in the order in which the stimuli were presented: visual first then auditory, or the reverse. In both cases, shorter auditory stimuli were produced than visual ones, for all the times produced, but the effect was most marked when auditory stimuli were employed first.

A second experiment used reproduction, where visual or auditory stimuli 1, 2, 3, and 4 s long served as standards, and these were reproduced by controlling either visual or auditory stimuli. Two intramodal cases (auditory/auditory and visual/visual) were used, as well as two crossmodal cases. Once again, different groups received the conditions in different orders. In the crossmodal cases, reproductions of all durations were much shorter in the visual-auditory case than the auditory-visual one, once again illustrating that the auditory stimuli appeared to have longer subjective durations than the visual ones. Likewise, Walker and Scott (1981) used reproduction of durations of 500, 1000, and 1500 ms, where the reproduction was effected by holding down a button (the stimulus to be reproduced was not presented during the reproduction part of the trial). Auditory stimuli resulted in significantly longer reproductions than visual ones in their Experiment 1. For a similar result from a more recent reproduction experiment see Bratzke and Ulrich (2019).

Goldstone and Lhamon (1974) used a category judgement procedure, similar to that of Behar and Bevan (1961). Two stimuli were presented on each trial, either an intramodal or crossmodal pair. In their Experiment 1, one of these stimuli always had a duration of 1 s, the other a variable duration, and the task was to compare the two, using a category scale running from 1 (shorter) to 5 (longer) with 3 as equal. In the crossmodal case higher categories were used when the stimulus to be judged was auditory and the standard visual than the other way around, indicating that the auditory stimuli had longer subjective durations.

Other experiments varied stimulus type, with constant and warbling tones, and solid and patterned or moving visual stimuli being used. Goldstone and Lhamon's summary in their Fig. 8 of some of these conditions shows that the auditory-visual difference was generally found despite changes in stimulus type, although results were complex and the reader is referred to the original article for details.

Verbal estimation of duration, where people assign verbal labels in conventional time units like seconds and milliseconds to stimuli usually yields very clear auditory/visual differences. This technique might be particularly useful for distinguishing additive and multiplicative effects, as a wide range of durations (e.g., a 10-fold or more ratio between the shortest and longest stimulus) can be presented without the intervention of chronometric counting at the longest intervals. Using this method, Wearden et al. (1998) presented participants with intermixed series of auditory and visual stimuli ranging from 77 to 1183 ms in length. Auditory stimuli were consistently judged as longer. Wearden et al. (2006) replicated this result in their Experiment 4, and also explored between-group intramodal conditions, where people received either auditory or visual stimuli, but not both. The usual auditory/visual difference was obtained, and was of about the same magnitude whether or not auditory/visual comparisons were within-group or between-group. In the literature this is a very unusual result, but was similar to that obtained by Goldstone et al. (1959). Wearden et al. (2006) attribute the effect to the use of conventional time units, and suggested that people might be using some extra-experimental standard (e.g., what one second 'feels like'). Support for this idea comes from the finding that between-group effects were obtained by Goldstone et al. (1959) when stimuli were judged relative to an imagined one second standard.

Williams et al. (2019) also used verbal estimation of the same durations as in Wearden et al. (1998), with the same result. Auditory stimuli were estimated, on average, as longer than visual ones.

There are several different ways of conducting the procedure of *temporal bisection*. The commonest one (Allan & Gibbon, 1991; Wearden, 1991) initially presents participants with two stimuli with standard durations, one a *Short* standard (e.g., 200 ms) the other a *Long* one (e.g., 800 ms). Usually, these standards are easily distinguishable. Following experience with the standards (which can be a few presentations with student-age adults, or more extensive training with children) comparison stimuli are presented, and these usually include the standard

durations with other durations in between, for example, stimuli in 100-ms steps between 200 and 800 ms. After each stimulus, the participant's task is usually to judge whether each comparison stimulus is more similar in duration to the previously presented *Short* or *Long* standards. Feedback is sometimes given as to response correctness when comparisons which have the same duration as the *Short* and *Long* standards are presented, but cannot be given after the other comparisons as there is no correct response on this task.

A psychophysical function, usually the proportion of '*Long*' responses (that is, judgements that a comparison is more similar to the *Long* than the *Short* standard) plotted against stimulus duration, is constructed, and from this various measures of performance can be calculated. The most important are bisection point (BP; the comparison duration giving rise to 50% '*Long*' responses), and the difference limen (DL), and Weber fraction or Weber ratio. The latter two are measures of sensitivity, usually considered to reflect underlying timing or memory variability, and will be discussed in more detail in a later section.

The simplest case for auditory/visual comparisons would involve conditions where the standards were presented in one modality and the comparisons in another one, perhaps contrasted with the situation where both were in the same modality. In the crossmodal comparisons, if auditory stimuli seem to last longer than visual ones, the BP derived from visual comparisons with auditory standards would be higher than if the standards were visual and the comparisons auditory.

In fact, most experiments have used slightly more complicated procedures, but the one that perhaps comes closest to the simple case outlined above comes from Wearden et al. (2006). Each block began with four presentations each of a *Short* and *Long* standard, either an auditory stimulus or a visual one. The *Short* standard duration varied between blocks but was constant within a block, and was randomly chosen from a uniform distribution between 150 and 300 ms. The *Long* standard was 640 ms longer than the *Short* one, whatever that was. Following standard presentations, seven comparison stimuli including the standards but with other stimuli spaced between them in 80 ms steps were presented, and the task was to judge whether each was more similar to the *Short* and *Long* standards presented at the start of the block. Bisection points (expressed in ms above the average of the *Short* standards) were 372 ms for the auditory/visual case and 242 ms for the visual/auditory one. In contrast, when both standard and comparison stimuli were in the same modality, BPs were very similar (284 ms for auditory/auditory, 292 ms for visual/visual).

In their Experiment 1, Penney et al. (2000) initially presented standards 3 and 6 s long, half of which were auditory, and half visual. Following that, comparison stimuli with durations equal to the standards or spaced between them were presented. Comparisons were auditory or visual stimuli presented singly, or simultaneous, but overlapping, auditory/visual stimuli. When the comparisons were

presented singly, there was a large modality difference in BP (4.08 s for auditory and 4.80 s for visual). Penney et al. (2000) also found a similar difference for what they described as ‘auditory/simultaneous’ and ‘visual/simultaneous’; presumably this refers to which stimulus type came first when simultaneous auditory/visual stimuli were used.

Their Experiment 2 used only intramodal conditions, with different groups receiving either auditory standards and comparisons or visual ones. *Short/Long* stimulus pairs of 3–6 s, 2–8 s, and 4–12 s were used. Bisection points varied with standard duration, but not with modality.

Their Experiment 3 used a rather complex design involving presentation of auditory/visual simultaneous comparisons for some groups, and single stimuli for others, with different *Short/Long* pairs. The reader is directed to the original article for details, but in summary evidence strongly supported the idea that auditory comparisons had longer subjective durations than visual ones (e.g., see the BPs in their Table 4, p. 1780).

Overall, then, data from Penney et al. (2000) supported the idea that auditory stimuli seem longer than visual ones of the same duration, and that the effect depended on a contrast between a standard in one modality and a comparison in another one, as the effect disappeared in their intramodal Experiment 2.

Wearden et al. (2006) also used a different bisection method, derived from the *partition bisection* procedure introduced by Wearden and Ferrara (1995). Nine auditory and nine visual stimuli, ranging in duration from 200 to 840 ms, were presented intermixed in a random order. The task was to classify each stimulus as *Long* or *Short*, although no standards were presented. The BP for auditory stimuli was 458 ms and for visual ones 561 ms.

Wearden et al. (1998) used a temporal generalisation technique in crossmodal comparisons. The procedure involved blocks where standards were presented at the start of a block, followed by comparisons that had to be judged as equal in duration (or not) to the standard which started the block. The standard for a block was randomly chosen from a uniform distribution between 400 and 600 ms, and was repeated four times. Following that, seven comparison durations ranging from 300 ms shorter than the standard, up until 300 ms longer, in 100-ms steps, were presented. Standards were either auditory or visual, and comparisons were either auditory or visual giving rise to two intramodal comparisons and two crossmodal ones.

Temporal generalisation gradients were plotted in the form of the proportion of responses where the stimuli were judged equal in duration to the standard, plotted against the difference between the comparison and the standard. In the crossmodal cases, this gradient was skewed markedly to the left when the standard was visual and the comparisons auditory, and to the right in the opposite case. Data from the intramodal conditions are discussed later.

Although Wearden et al.’s results were clearly consistent with the idea that auditory stimuli seemed to last longer than visual ones, a later similar experiment

by Klapproth (2003) found different results. Now, generalisation gradients were skewed to the right (indicating that longer stimuli were identified as the 400 ms standard) for both the situation where the standard was auditory and the comparisons visual, but also for the reverse condition, a result which has no obvious interpretation.

Ulrich et al. (2006) used a discrimination technique in which two stimuli were presented on each trial, one a standard, presented first, and the other a comparison. The task was to decide whether the comparison was shorter or longer than the standard. The stimuli could be intramodal (both auditory or visual) or cross-modal. Standards were either 100 or 1000 ms. The duration of the comparison was adjusted to estimate the values at which the response ‘comparison longer’ was given with probabilities of 0.25 or 0.75, and from these a DL was computed. Another measure used was the constant error (CE) the difference between the standard duration and the comparison value that resulted in ‘comparison longer’ probability of 0.5. The CE values indicated that people judged the auditory stimuli as lasting longer than the visual ones. The DL results are discussed later.

4. Potential Exceptions

In considering potential exceptions, care needs to be taken over the type of comparisons made. For example, Indraccolo et al. (2016) reported that visual stimuli were reproduced as longer than auditory ones (their Fig. 3c), which at first sight seems contrary to the usual direction of auditory/visual effects. However, their reproduction procedure was intramodal: that is, auditory samples of 600, 800, or 1000 ms in length were reproduced by turning off another auditory stimulus, and likewise for visual stimuli, so the result does not in fact contradict the general trend, as the authors themselves conclude. A similar study by Szelag et al. (2002), which used three groups of children, found no auditory/visual difference in reproduction but, likewise, employed an intramodal procedure where auditory and visual samples were reproduced using the same stimuli, with the children receiving the different modalities in different blocks.

Penney et al. (2000) identified Bobko et al. (1977) as an exception to the usual auditory/visual differences, but not only were their comparisons intramodal, they were also between-group. They used two procedures. The first they called verbal estimation. People received a standard that they were told was 2.5 s long, and then had to judge other stimuli with respect to this standard. Durations ranged from 0.5 to 5 s. The second procedure (magnitude estimation) was similar except that the standard stimulus (2.5 s) was identified with the number ‘100’ and other numbers had to be assigned on this basis. Their first experiment used auditory stimuli, and the second visual ones. Both methods produced power function relations between estimates and real magnitudes, and there was no modality effect although there were ‘clear trends for auditory intervals to be judged longer than physically equal

intervals defined by a visual stimulus' (pp. 707–708). However, this is another case where comparisons were intramodal, so no effects would perhaps be expected.

What seems at first to be an early definite exception was Hirsch et al. (1956), who used the reproduction of auditory or visual stimuli by a subsequent button press. No effect of modality *per se* was found, although there was an effect of background stimulation. This involved either quiet or loud noise, or a dark or lighter visual field. Presenting the target stimulus against a quiet background and performing the reproduction with a loud one resulted in longer reproductions for both auditory and visual stimuli than the other way around. However, this was an intramodal procedure, as the button press 'turned on the same stimulus' (p. 563), presumably the stimulus whose duration had to be reproduced, so may not argue against modality differences in average duration perception between auditory and visual stimuli.

Brown and Hitchcock (1965), another exception identified by Penney et al. (2000), used crossmodal as well as intramodal comparisons, and a reproduction technique. Durations to be reproduced ranged from 1 to 17 s. No effect of modality was found even in crossmodal comparisons. This was an apparently very thorough and substantial study with 80 participants, so its failure to obtain the usual auditory/visual effect is very striking. One possibility is that, given the length of the durations used, participants used chronometric counting (see Wearden, 2016, for a discussion) to judge the duration of the standards and the subsequent reproductions. This would, of course, remove any auditory/visual differences if people counted at the same rate for both types of stimuli. Nothing in the reported procedure discouraged counting, or even mentioned it, so this potential explanation for this unusual result, while possible, remains speculative. This single clear exception used reproduction of multi-second intervals, whereas reproduction of shorter intervals (as in Bratzke & Ulrich, 2019) produced the usual auditory/visual difference.

Williams et al. (2019) in their auditory/visual verbal estimation study, obtained the usual result on average but, in addition, they looked at individual differences in the effect. Using their raw data, we explored this in a slightly different way from in the original paper. Firstly, inspection of their data showed that not all participants judged that auditory stimuli lasted longer than visual ones. We computed the average estimate for the auditory and visual stimuli over the 10 durations used (from 77 to 1183 ms). Fourteen participants had longer average estimates for visual stimuli than auditory ones; the other 38 had longer estimates for the auditory stimuli. Figure 1 shows the mean estimates for the two stimulus types, plotted against stimulus duration, for the participants who showed the usual auditory/visual effect (upper panel), and those for whom the visual stimuli were judged to last longer (lower panel). For both groups there was a significant effect of modality and stimulus duration, but the interaction (indicating a slope difference) was significant only for the people who judged the auditory stimuli as longer. For the

Figure 1. Reanalysis of verbal estimation data from Williams et al. (2019). Upper panel: mean estimate plotted against stimulus duration for auditory and visual stimuli for participants whose average estimate of auditory stimuli was longer than that for visual stimuli. Lower panel: The same measure for participants whose average estimate of visual stimuli was higher.

others the estimates from the two modalities appeared approximately parallel. Although too much should not be read into data from the ‘anomalous’ findings, which were derived only from about a quarter of the participants, the results suggest that whatever is causing the auditory/visual difference in people who judged

the auditory stimuli as longer is not exactly the same as for people who judge the durations the other way around: one is multiplicative, and other appears additive.

5. Interim Conclusions

Overall, then, although a few exceptions can be noted, many studies have found that auditory stimuli have longer subjective durations than visual ones. This conclusion holds over a wide range of different procedures: production, reproduction, judgement relative to one second, verbal estimation, category assignment, bisection, temporal generalisation, and discrimination procedures. It also holds over a range of different sorts of stimuli. Pre-computer-controlled studies often used taped tones, or white noise, delivered through speakers or headphones, and computer-controlled studies also sometimes used headphones, but often the computer speaker. A wide variety of visual stimuli have also been used in studies finding auditory/visual differences: tachistoscopic fields, black Xs, illumination of lights or diodes, and more recently a variety of stimuli displayed on computer screens, as outlined in Table 1.

Concerning the size of the auditory/visual difference in mean judgements usually found, it is obviously problematic to draw general conclusions from studies which have used diverse procedures and different stimuli, but some tentative comparisons can be made from such measures as BP or verbal estimation ratios, or values derived from reproduction and production experiments. There is a, perhaps surprising, degree of consensus between some studies using different duration ranges and different procedures. A common result is that visual stimuli appear to have about 80–90% of the subjective duration of auditory ones of the same real length. For example, Penney et al. (2000) found BP ratios of between 0.83 and 0.91 in different conditions, and used visual/auditory duration ratios of between 0.85 and 0.94 in their modelling. Wearden et al.'s (1998) verbal estimation study found visual/auditory slope ratios of 0.81, 0.85, and 0.86 in three different conditions, and inspection of data in figures suggests similar values from Wearden et al. (2006). Goldstone (1968) obtained ratios of 0.73–0.89 (average 0.82) from their production study and ratios of 0.7–0.89 (average 0.82) from reproduction (values estimated from Goldstone's Figs 1 and 2).

It would, of course, be surprising if stimulus features such as intensity played no role at all, and some evidence suggests that they do. For example, Walker and Scott's (1981) Experiment 1 found significant auditory/visual differences in the normal direction. However, when the intensity of the auditory stimulus was reduced from 75 dB to 50 dB, visual stimuli were reproduced as longer at 500 ms, but not at 1000 and 1500 ms. Overall, however, the overall significant modality effect found with the more intense auditory stimulus was abolished.

Indracoello et al.'s (2016) study provided further information about effects of intensity, and some rather complex results. As mentioned earlier, the task used

Figure 2. Standard deviations of verbal estimates from Williams et al. (2019), calculated from unpublished data. Upper panel: Standard deviation of estimates for auditory and visual stimuli plotted against stimulus duration. Lower panel: Standard deviations for the two modalities plotted against mean estimates.

reproduction, and samples could be high- or low-intensity auditory or visual stimuli, and comparisons could also be auditory or visual with different intensities. Effects of modality were highest when the sample stimuli were low intensity, and negligible when they were high. In contrast, effects of modality were highest when the comparisons were of high intensity and negligible when they were low. However, this study was intramodal so we cannot know whether the intensity manipulations they used would have affected the expected auditory/visual difference had a crossmodal condition been used.

Another intramodal study with an intensity manipulation was that of Goldstone et al. (1978). Here, a category judgement procedure was used, and higher intensity for both auditory and visual stimuli increased perceived duration, but no intersensory comparisons were made.

One factor which has been suggested to play a role in determining auditory/visual differences is age. Droit-Volet et al. (2004) tested children of five and eight years, and adults, on a bisection task with standards of 200 and 800 ms. When the standards were visual, only the five-year-olds showed a crossmodal difference, but when the standards were auditory all age groups did. Droit-Volet et al. (2007) found a similar result. The standards were 3 and 6 s, and during training the modalities of the standards were intermixed. Then groups of five-year-olds, eight-year-olds, and adults received auditory and visual comparisons, and in this case all age groups showed the usual crossmodal effect, although this was most pronounced in the five-year-olds.

6. Auditory/Visual Differences in Duration Perception: Multiplicative or Additive?

Although there may be a few exceptions, studies with diverse procedures confirm that auditory stimuli seem to last longer than visual ones, but exactly how should the difference be characterised? There seem to be two clear possibilities. One is that the difference is *additive*, that is, regardless of the duration, the auditory stimuli seem to be some constant amount longer than the visual ones. If, for example, verbal estimation of a range of intervals were used, the auditory and visual estimates would have the same slopes. If bisection were used, the BPs would differ by a constant amount, regardless of the durations used. In contrast, the difference could be *multiplicative*, that is, get larger as durations grow, so verbal estimation would yield different slopes for the different modalities, and the BP differences would increase with the intervals judged.

In Goldstone's (1968) production study, auditory-visual differences increased as the durations to be produced increased over values of 1–4 s, but the effect was clearest (according to his Fig. 1, p. 758) when the auditory productions preceded visual ones. The effect was much smaller in the reverse case. Similar results were found in the reproduction study, where the difference in the crossmodal cases

increased as the intervals reproduced lengthened (see his Fig. 2, p. 759). Walker and Scott (1981) in their Experiment 1 also found a significant modality $\times$ duration interaction, indicating that the auditory-visual difference was multiplicative. Bratzke and Ulrich (2019) likewise found larger auditory-visual differences in crossmodal reproduction when the stimulus to be reproduced was 2400 ms compared with 800 ms (see their Fig. 3, p. 1222).

In bisection, relevant data come from Penney et al.'s (2000) rather complex Experiment 3, which used crossmodal comparisons with two ranges of durations (3–6 s and 4–12 s). In their sequential-same conditions, the BP differences were 0.35 s for the former case and 1.46 s for the latter; in the simultaneous-same conditions, the differences were 0.29 and 0.66 s for the same duration values.

In verbal estimation, both Wearden et al. (1998), in their Experiment 2, and Wearden et al. (2006), in their Experiments 3 and 4, found slope differences between the estimates of auditory and visual stimuli, with auditory stimuli having higher slopes. Results from their Experiment 4 are particularly noteworthy, as slope differences were found when participants received both auditory and visual stimuli, and when they received only one of them. Williams et al. (2019) found a similar result overall (see also the upper panel of Fig. 1, above).

An exception to multiplicative effects comes from Behar and Bevans's (1961) Experiment 3, where auditory and visual stimuli from 1 to 5 s in length were used, and these had to be rated on a scale (discussed above). Here, although the auditory stimuli were judged as consistently longer than the visual ones, there was no significant interaction between duration and modality, and inspection of their Fig. 3 (p. 23), suggests that the category rating vs. duration functions for the two modalities were completely parallel.

Overall, then, the limited data available, where the intervals judged have been varied between conditions, mostly support the idea that the auditory/visual difference gets bigger as the intervals judged become longer, Behar and Bevan's study is the exception.

7. Modality Differences in Variability

Do auditory stimuli give rise to more or less variable temporal representations than visual ones? Questions about variability are more difficult than the other questions to answer for two reasons. One is that older articles often provide no data than might be used as a measure of variability. The second is that when more modern articles do provide a measure it is often a relative rather than absolute one. For example, coefficients of variation (standard deviation/mean) are often presented from verbal estimation studies or Weber fractions (DL divided by BP) from bisection. If auditory and visual stimuli systematically generate different mean estimates or different BPs, as data reviewed in a previous section suggests they generally do, then do the relative measures of variability differ only because

mean estimates or BPs do? Posing this question is not mere pedantry: in Wearden et al.'s (2007) study of verbal estimation of the duration of filled and unfilled intervals, means differed very markedly in favour of higher means for filled intervals, and coefficients of variation for filled intervals were lower. However, this difference appeared to depend solely on the mean difference: standard deviations of estimates were not significantly different between estimates of filled and unfilled stimuli. So, do filled interval produce more precise temporal estimates than unfilled ones? In relative terms they do, but not in absolute terms.

Goldstone and associates performed many studies of auditory/visual duration judgements up until the early 1970s but, unfortunately, few provide data that might bear on potential variability differences. However, two of his articles might be useful in this respect. One is Goldfarb and Goldstone (1964) who used a classification technique. Two stimuli, either auditory or visual, were presented on each trial. The first always had a 1 s duration, the second ranged from 0.6 to 1.4 s. The judgement of the second stimulus relative to the first used a scale running from 1 (very much shorter) through 5 (equal) to 9 (very much longer). The interval between the stimuli was varied over values of 1, 2, and 4 s. No variability data are provided but their Fig. 1 (p. 484), plots the mean judgement category against the duration of the second stimulus. Inspection suggests that the slope of the category change was slightly shallower, suggestive of higher variability, for the intramodal visual judgements compared with the auditory ones, at least when the interstimulus intervals were 2, and 4 s, but no statistical analysis was provided. Goldstone and Lhamon (1974) reported a similar study as their Experiment 1 and, once again, inspection of data (in their Fig. 2, p. 68) suggested that the slope for intramodal visual judgements was shallower than for auditory ones, perhaps more clearly than in the earlier article.

When temporal bisection is used, the usual measure of variability is derived from the slope of the psychophysical function (the proportion of *Long* responses plotted against stimulus duration), with calculation of the DL (half the duration difference between stimuli giving rise to 75 and 25% *Long* responses), or the Weber fraction of ratio (WF: the DL divided by BP, the stimulus duration giving rise to 50% *Long* responses). As noted above, in cases where the BP differs markedly between conditions, the DL may be a better measure of variability than the WF.

Penney et al. (2000) report data from three experiments using bisection in intramodal and crossmodal conditions. Perhaps the simplest results come from their intramodal Experiment 2, where different groups judged auditory or visual comparisons relative to auditory or visual standard durations of 3–6, 4–12, or 2–8 s. The psychophysical function appeared flatter for the visual condition in the 3–6 s case, but not the others, but there was no significant effect of modality on DL overall. In their Experiment 1, auditory or visual stimuli were compared with auditory, visual, or simultaneous audio-visual standards (with the stimuli overlapping in the latter case). Standards were 3 and 6 s. There was no significant effect

of modality, although DL values were slightly higher for visual comparisons in both the intramodal and simultaneous cases. Their Experiment 3 presented the audio-visual standards in two different ways (sequential or simultaneous). In the sequential case the DL was higher for visual comparisons for the 3/6 s standards but not for the 4/12 s ones, but in the simultaneous case there was no significant difference. Overall, then, the data hinted at slightly more variable temporal representations for visual stimuli than auditory ones, but the difference was only occasionally statistically significant.

Similar results were obtained by Wearden et al. (2006), when all durations were less than about 1 s. In their Experiment 1, each block began with standard stimulus presentations, and the participants' task was to judge the comparison durations on the block relative to the standards presented at the start of the block, so auditory and visual comparisons could be compared with either auditory or visual standards, but the different types of trials were segregated in blocks. DLs were not presented, or analysed statistically, but values can be calculated from the WR value tables. DLs were 79.5 ms and 96.4 ms for intramodal auditory or visual comparisons, respectively, and 130.2 for visual comparisons and auditory standards and 87.1 ms for the reverse case. Their Experiment 2 used the partition bisection method of Wearden and Ferrara (1995), so auditory and visual stimuli ranging from 200 to 840 ms in duration were intermixed randomly and had to be classified as *Short* or *Long*, without any standards being presented. Now DLs were 87.6 ms for auditory stimuli and 89.8 ms for visual ones. So, as in Penney et al. (2000) small differences appeared between measures of variability for auditory and visual comparisons, but always in the direction of the visual stimuli being more variable.

The fragility of variability differences between judgements of auditory and visual stimuli in bisection is further supported by a developmental study by Droit-Volet et al. (2004). Children of five or eight years of age, or adults, received bisection tasks with standards of 200 and 800 ms, and comparisons spaced linearly between them. Auditory or visual comparisons were judged against auditory or visual standards. In general, the WF measure reported found higher values for visual comparisons, for both modalities of standard, but only in the five-year-olds. For the other groups there was no significant difference. However, the WF measure results from both the DL and the BP. Using data from their Table 2 (p. 809), DLs can be calculated for the unimodal visual and auditory conditions, although these were not statistically analysed in the article. Values are: five-year-olds – vis/vis 183.8 ms, aud/aud 131.8 ms; eight-year-olds – vis/vis 96.6 ms, aud/aud 79.7 ms; adults – vis/vis: 72.6 ms, aud/aud 65.2 ms. So, DLs were higher for the visual stimuli on average in all age groups, although the differences were small for eight-year-olds and adults.

Wearden et al. (1998) used a temporal generalisation technique (Wearden, 1992) where auditory or visual comparisons were judged as equal in duration, or not, to previously presented auditory and visual standards. Of particular interest

are the intramodal cases, where visual comparisons produced just-significantly flatter generalisation gradients than auditory ones, suggesting more variable temporal representations. Although, as noted above, Klapproth (2003) did not replicate the auditory/visual difference in generalisation gradients found by Wearden et al. (1998), and gradients from the intramodal conditions of his Experiment 1 clearly differed in the direction of a flatter gradient for visual stimuli, as in Wearden et al. (1998).

Wearden and Bray (2001), in their Experiments 1 and 3, presented either auditory or visual stimuli using two slightly different 'episodic' temporal generalisation methods. Here, two stimuli were presented on each trial, and the task was to judge whether or not they had the same duration. The stimuli changed from trial to trial (Experiment 1) or were even random in duration (Experiment 3). No statistical comparisons between auditory and visual conditions were made in that article, but inspection of data in the figures (e.g., their Figs 2 and 4) suggested that, if there was any difference in generalization gradients at all, it was in the direction of visual stimuli producing slightly flatter gradients.

Overall, then, data from bisection and temporal generalisation studies concur in finding small, but sometimes non-significant, differences in variability, with visual stimuli producing data with slightly higher variability.

At first sight, verbal estimation methods (where people assign verbal labels in conventional time units like ms to presented stimuli) appear to yield a very clear picture. Coefficients of variation (standard deviation/mean) are markedly and significantly lower for auditory stimuli than visual ones (Wearden et al., 1998, 2006). However, as mentioned above, this difference may reflect the large difference in mean estimates found in these studies, rather than any difference in absolute variability. Unfortunately, the original data from Wearden et al. (1998) and Wearden et al. (2006) are no longer available to allow standard deviations to be calculated, but Fig. 2 shows standard deviations from Williams et al. (2019), which were not reported in the original article, but were calculated here from raw data from all 52 participants. The upper panel shows standard deviations plotted against stimulus duration, the lower panel standard deviations plotted against mean estimate. ANOVA found no significant effect of modality overall, but an effect of stimulus duration, and a significant modality by stimulus duration interaction, presumably reflecting the fact the standard deviations were higher for the visual stimuli but only at the longer durations. Up to around 800 ms, standard deviations seem very similar. In any case, the difference between auditory and visual variability when the standard deviation is used is much less clear than when the coefficient of variation is analysed, so it seems that, as in the case of verbal estimation of the duration of filled and unfilled intervals (Wearden et al., 2007), much of the difference in coefficient of variation between estimates of different stimulus types comes from differences in mean estimates. If the average standard deviation over the 10 durations estimated was calculated for each

participant, auditory and visual averages were correlated ($r = 0.42$, $p < 0.01$), so if an individual produced variable estimates for one modality they tended to do so for the other one as well.

Bratzke and Ulrich (2019) employed intramodal and crossmodal reproduction of time intervals of 800 and 2400 ms. The method used involved first presenting an auditory or visual target stimulus then, after a short delay, a stimulus (either auditory or visual) was presented again and the participant terminated it with a single press when they judged that it had the same duration as the target. When both stimuli were in the same modality, standard deviations of the times reproduced were clearly higher for visual stimuli for both target durations (see their Fig. 3, 1222), and the difference was larger for the longer duration. In contrast, Walker and Scott's (1981) reproduction Experiments 1 and 3 found no significant effect of modality on standard deviation.

One procedure that may provide consistent evidence for variability differences between modalities is the threshold discrimination procedure. Here, two stimuli are presented on each trial, one a standard, and the other a comparison, and the participant must judge which lasts longer. A correct response reduces the difference between the two on the next trial; an incorrect response increases the difference. The procedure continues until performance seems stable, and the last 10 or 20 trials are used to determine the smallest difference detectable, the discrimination threshold. Although there is no published quantitative model of this procedure, common sense suggests that more variable time representations should make the task more difficult, as the representations of the durations of the two stimuli on the trial would tend to overlap more if more variable from trial to trial, thus be more difficult to discriminate, and this would lead to a higher threshold for stimuli producing greater variability.

Using a method like this, Grondin et al. (1998) found that filled auditory intervals resulted in lower thresholds than filled visual ones. Stauffer et al. (2012) also used threshold determination of auditory and visual stimuli, with a 100 ms standard. Thresholds were significantly higher for the visual stimuli than for auditory ones (20.2 vs. 8.8 ms). Williams et al. (2019), using a 700 ms standard and both auditory and visual stimuli, likewise found much higher thresholds for visual stimuli than auditory ones, around twice as high (196.8 vs. 103.2 ms). Similarly, Ulrich et al.'s (2006) discrimination procedure, although slightly different from the threshold determination method, yielded consistently higher DL values for the visual than auditory conditions in their intramodal (what they called 'congruent') cases, indicative of the visual stimuli producing more variable temporal representations.

However, all these results may need some qualification according to Rammsayer et al. (2015). These authors performed a threshold determination study with auditory and visual stimuli, but with standards of either 50 ms or 1000 ms. In both cases, but particularly in the 50 ms case, thresholds were

significantly higher for visual than auditory stimuli. However, when performance with the 50 ms standard was used as a covariate to analyse data from the 1000 ms standard condition, the effect disappeared. Rammsayer et al. (2015) argued that judgements with the shorter durations were performed by an automatic mechanism, whereas those with the longer standard were subject to cognitive control, and that the modality effect occurred only for the automatic process. Even when longer durations were employed, Rammsayer et al. suggest that the auditory/visual difference was due to ‘input from the sensory-automatic timing system’ assumed to operate at short intervals. They further argue that apparent differences between the processing of auditory and visual stimuli should diminish as they get longer.

Our Table 1 presents the duration ranges used in the studies we discuss and, in support of Rammsayer et al.’s argument, the one study most clearly failing to find auditory and visual differences (Brown & Hitchcock, 1965) used the longest durations, up to 17s, when, presumably, any effect of the short-term automatic mechanism proposed would be negligible. On the other hand, many studies use time values longer than the longest one employed by Rammsayer et al. (1000 ms) and still find auditory/visual differences, so overall this argument remains difficult to confirm or reject with existing data. In addition, an alternative explanation of Brown and Hitchcock’s very unusual result is chronometric counting, as discussed earlier.

8. Potential Explanations of Auditory/Visual Differences in Duration Judgements

Our conclusions can be simply summarised. Auditory stimuli have longer subjective durations than visual ones when this is assessed with a range of different procedures and stimuli. Effects may be modulated by age in some cases, and by intensity. The auditory/visual difference appears to be multiplicative, that is, to get larger as stimulus durations become longer. Differences in variability in terms of standard deviations or slopes of psychophysical functions appear small, and often non-significant statistically, although variability is usually higher for visual stimuli than auditory ones. Relative measures of variability, such as coefficient of variation, or Weber fraction, may exhibit substantial auditory/visual differences, but this may be largely due to differences in mean.

How can these effects be explained? The most favoured explanation (Bratzke & Ulrich, 2019; Penney et al., 2000; Ulrich et al., 2006; Wearden et al., 1998) has been some form of pacemaker-counter model, generally based on ideas such as those of Treisman (1963) and Gibbon et al. (1984), where pulses from a pacemaker, which runs more quickly for auditory than visual stimuli, are accumulated in a counter. The attraction of this kind of model is perhaps no surprise, as it immediately produces (a) the auditory/visual mean effect in the correct direction, and (b) a

multiplicative effect. It may also deal with any effects of intensity that are found if pacemaker rate varies with stimulus intensity.

Wearden et al. (1998) proposed that the auditory/visual differences they found were explainable by different pacemaker speeds for the two modalities, as well as differential variance of onset and offset of the stimuli, but they did not carry out any formal modelling. Wearden (2015), however, showed that underlying pacemaker speed differences would lead to differences in verbal estimates using his model which took account of quantisation (the tendency to use a limited number of rounded estimates) in verbal estimation.

Penney et al. (2000) modelled their data using a model of bisection (the memory mixing sample known exactly model) derived from the principles of Scalar Timing Theory (Gibbon et al., 1984), which involved a number of parameters. Not only did they generally assume that the auditory pacemaker ran faster than the visual one, but they also had parameters representing memory variability (which was the same for auditory and visual stimuli, consistent with their general lack of findings of differential variability for the different modalities, discussed earlier), as well as another representing the relative contribution of auditory and visual stimuli to the ‘memory mixture’ used as the basis for the standards employed in the task.

Perhaps the most elaborate attempt to use a pacemaker-counter model to account for judgements of stimuli in different modalities comes from Ulrich et al. (2006), some of whose data were described earlier. As mentioned above, their task involved deciding which of two stimuli presented on a trial was longer, with one of the stimulus durations adjusted depending on the response, and the stimuli could have the same, or different, modalities. Their modelling assumed that auditory stimuli were associated with a faster and less variable pacemaker than visual stimuli. In intramodal trials (which they called ‘congruent’) the psychophysical function would be flatter for visual stimuli as a consequence of higher underlying pacemaker variability. The different pacemaker rates would naturally produce the usual auditory/visual difference in crossmodal (‘incongruent’) trials. However, their model made some less expected predictions such as that, on crossmodal trials, the order in which the auditory and visual stimuli were presented made a difference to the DL, and this was found to be the case.

The modelling was in good qualitative accord with the data obtained in three experiments, although it fared less well in terms of quantitative fit. A similar problem was encountered when the same model was used to deal with reproduction data in Bratzke and Ulrich (2019). One possibility is that some other process is needed as well as the basic pacemaker-counter model, for example, the idea of Wearden et al. (1998) that the onset and offsets of visual stimuli introduced more duration-independent variance than the onsets and offsets of auditory ones.

Although at the time of writing pacemaker-counter models may not enjoy the vogue they once had, competitors may struggle to equal them as explanations of crossmodal effects. For example, another potential idea is that the auditory/visual difference usually found results from more rapid initial processing ('prior entry') for auditory stimuli compared to visual ones. Even if this is true effects might be expected to be (a) very small, and (b) duration-independent, so slopes from verbal estimation studies, for example, would be expected to be the same, rather than different as actually found.

One obvious question which arises in the present context with a pacemaker-counter approach is whether there is a single pacemaker with variable speed for the different modalities, or different pacemakers for auditory and visual stimuli. Stauffer et al. (2012) attempted to address this difficult issue by presenting participants with tasks involving the discrimination of the duration of filled and unfilled intervals, as well as temporal generalisation, and a rhythm perception task, with all tasks being presented with both auditory and visual stimuli. They then used structural equation modelling to explore three models of how the covariance of performance on the different tasks was generated. The three models involved (a) a single modality-independent source of timing, (b) separate timing mechanisms for auditory and visual stimuli, and (c) a hierarchical model with both modality-independent and modality-specific components. They concluded that the third model fitted data best. This type of approach may be the best way of resolving what Fillippopoulos et al. (2013) called the *modality paradox*, '... if stimuli in different modalities are timed by a common mechanism, why are there any marked differences in duration judgements at all ... on the other hand, if stimuli in different modalities are timed by completely different mechanisms, why are time judgements from stimuli of different sorts so similar in general form?' (p. 708). The present article has identified a number of differences between time judgements of auditory and visual stimuli, but at the same time, data suggest that performance on tasks like bisection, verbal estimation, interval production and reproduction, discrimination tasks, and category judgements are similar in form for auditory and visual stimuli.

However, evidence from Motala et al. (2018) may argue against any sort of modality-independent timing. These authors used a rate aftereffect procedure: for example, exposure to a fast rate of sensory stimulation makes a moderate rate appear slower, and exposure to a slow rate has the opposite effect. No evidence was found for any crossmodal adaptation effects when the adapting stimulus was in one modality and the test stimulus in another one, suggesting independent timing of the rate of auditory and visual signals. However, this procedure is very different from that used in the experiments reviewed above, where the focus is on stimulus duration rather than rate, so Motala et al.'s conclusion needs to be viewed with caution.

In spite of the historical interest in pacemaker-based models, there are, however, some alternatives to them, and we will briefly describe two of these. So far as we know, at the time of writing, oscillator-based models like the striatal beat-frequency (SBF) model of Matell and Meck (2004), have not been applied to the problem of subjective duration differences between auditory and visual stimuli, but Oprisan and Buhusi (2011) addressed an issue which is, at least in some respects, similar, that of pharmacological effects on timing, in particular the effect of dopaminergic agonists and antagonists. Meck (1983) demonstrated, using a bisection procedure with rats, that dopamine agonists and antagonists had apparently different effects on subjective time. In terms of pacemaker-counter theory, as used by Meck, agonists made the pacemaker of the clock run more quickly, antagonists made it run more slowly. Oprisan and Buhusi (2011) developed a variant of the SBF model, which differed from the original mainly by having a more physiologically plausible account of the operation of the neurons the model contains, and showed that their model could simulate the effects of both dopamine agonists and antagonists well (see their Fig. 2, for example). The model is complex, and the reader should consult the original article for details, but the basic method for producing the results was a change in the frequency of the oscillators, depending on the drug administered, a step clearly reminiscent of changes in ‘clock speed’ as originally proposed by Meck (1983). The task simulated by Oprisan and Buhusi (2011) was very different from those used in the articles reviewed in the present work, so their solution may not be directly applicable to the effects we report. However, assuming different oscillator frequencies for auditory and visual stimuli may be a promising approach.

Roseboom et al. (2019) developed a model which made time judgements based on the number of changes in videos analysed by a perceptual classification network. Some versions of the model were found to mimic human judgements of the duration of the videos very well. It is not, however, clear how such a model could make judgements of the duration of the kind of unchanging visual stimuli used in many of the experiments reviewed in the present article, particularly very short ones. Likewise, it is unclear how this model could be extended to unchanging auditory stimuli, without some assumptions of what ‘changes’ in constant stimuli actually were, and assuming that there were more of these when the stimulus was auditory rather than visual.

Even if pacemaker-based models seem to have some advantages over potential competitors, the question arises of why modality-specific pacemakers should exist and, specifically, why the pacemaker for auditory stimuli seems to run faster than that for visual ones. The facile answer is that this depends on properties of the auditory and visual system. While this is almost certainly true, the challenge for cognitive neuroscientists interested in time perception is to explain in quantitative detail how properties of the auditory and visual systems produce the range of effects noted in the present article. Given that the data are reasonably consistent

and seem reliable across a range of procedures and stimuli, auditory/visual differences in duration judgements would seem to be an excellent testbed for developing and exploring neuroscientific models of timing.

References

- Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learn. Motiv.*, 22, 39–58. doi: 10.1016/0023-9690(91)90016-2.
- Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *Am. J. Psychol.*, 74, 17–26. doi: 10.2307/1419819.
- Bobko, D. J., Thompson, J. G., & Schiffman, H. R. (1977). The perception of brief temporal intervals: power functions for auditory and visual stimulus intervals. *Perception*, 6, 703–709. doi:10.1068/p060703.
- Bratzke, D., & Ulrich, R. (2019). Temporal reproduction with and across senses: Testing the supramodal property of the pacemaker-counter model. *J. Exp. Psychol. Hum. Percept. Perform.*, 45, 1218–1235. doi: 10.1037/xhp0000667.
- Brown, D. R., & Hitchcock, L. Jr. (1965). Time estimation: Dependence and independence of modality-specific effects. *Percept. Mot. Skills*, 21, 727–734. doi: 10.2466/pms.1965.21.3.727.
- Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behav. Proc.*, 74, 244–250. doi: 10.1016/j.beproc.2006.09.012.
- Droit-Volet, S., Tournet, S., & Wearden, J. (2004). Perception of the duration of auditory and visual stimuli in children and adults. *Q. J. Exp. Psychol. A*, 57, 797–818. doi: 10.1080/02724980343000495.
- Fillippopoulos, P. C., Hallworth, P., Lee, S., & Wearden, J. H. (2013). Interference between auditory and visual duration judgements suggests a common code for time. *Psychol. Res.*, 77, 708–715. doi: 10.1007/s00426-012-0464-6.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Ann. N. Y. Acad. Sci.*, 423, 52–77. doi: 10.1111/j.1749-6632.1984.tb23417.x.
- Goldfarb, J. L., & Goldstone, S. (1964). Properties of sound and the auditory-visual difference in time judgment. *Percept. Mot. Skills*, 19, 606. doi: 10.2466/pms.1964.19.2.606.
- Goldstone, S. (1968). Production and reproduction of duration: Intersensory comparisons. *Percept. Mot. Skills*, 26, 755–760. doi: 10.2466/pms.1968.26.3.755.
- Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1959). Intersensory comparisons of temporal judgments. *J. Exp. Psychol.*, 57, 243–248. doi: 10.1037/h0040745.
- Goldstone, S., & Lhamon, W. T. (1974). Studies of auditory-visual differences in human time judgment: 1. Sounds are judged longer than lights. *Percept. Mot. Skills*, 39, 63–82. doi: 10.2466/pms.1974.39.1.63.
- Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. *Bull. Psychon. Soc.*, 12, 83–84. doi: 10.3758/BF03329633.
- Grondin, S. (2003). Sensory modalities and temporal processing. In H. Helfrich (Ed.), *Time and Mind II: Information Processing Perspectives* (pp. 61–77). Gottingen, Germany: Hogrefe and Huber.
- Grondin, S., Meilleur-Wells, G., Oulette, C., & Macar, F. (1998). Sensory effects on judgments of short time-intervals. *Psychol. Res.*, 61, 261–268. doi: 10.1007/s004260050030.
- Guyau, M. (1890). *La Genèse de l'idée de Temps*. Paris, France: Alcan.
- Hirsch, I. J., Bilger, R. C., & Deatherage, B. H. (1956). The effect of auditory and visual background on apparent duration. *Am. J. Psychol.*, 69, 561–574. doi: 10.2307/1419080.

- Indraco, A., Spence, C., Vatakis, A., & Harrar, V. (2016). Combined effects of motor response, sensory modality, and stimulus intensity on temporal reproduction. *Exp. Brain Res.*, 234, 1189–1198. doi: 10.1007/s00221-015-4264-2.
- Klapproth, F. (2003). Notable results regarding temporal memory and modality. In H. Helfrich (Ed.), *Time and Mind II: Information Processing Perspectives* (pp. 79–96). Gottingen, Germany: Hogrefe and Huber.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt's *The Experimental Study of the Time Sense* (1868) and its legacy. *Eur. J. Cogn. Psychol.*, 21, 941–960. doi: 10.1080/09541440802453006.
- Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cogn. Brain Res.*, 21, 139–170. doi: 10.1016/j.cogbrainres.2004.06.012.
- Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *J. Exp. Psychol. Anim. Behav. Proc.*, 9, 171–201. doi: 10.1037/0097-7403.9.2.171.
- Motala, A., Heron, J., McGraw, P. V., Roach, N. W., & Whitaker, D. (2018). Rate after-effects fail to transfer crossmodally: Evidence for distributed sensory timing mechanisms. *Sci. Rep.*, 8, 924. doi: 10.1038/s41598-018-19218-z.
- Oprian, S. A., & Buhusi, C. V. (2011). Modeling pharmacological clock and memory patterns of interval timing in a striatal beat-frequency model with realistic, noisy neurons. *Front. Integr. Neurosci.*, 5, 52. doi: 10.3389/fnint.2011.00052.
- Penney, T. P., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and memory processes. *J. Exp. Psychol. Hum. Percept. Perform.*, 26, 1770–1787. doi: 10.1037/0096-1523.26.6.1770.
- Rammsayer, T. H., Bortner, N., & Troche, S. J. (2015). Visual-auditory differences in duration discrimination of intervals in the subsecond and second range. *Front. Psychol.*, 6, 1626. doi: 10.3389/fpsyg.2015.01626.
- Roseboom, W., Fountas, Z., Nikiforou, K., Bhowmik, D., Shanahan, M., & Seth, A. K. (2019). Activity in perceptual classification networks as a basis for human subjective time perception. *Nat. Commun.*, 10, 267. doi: 10.1038/s41467-018-08194-7.
- Stauffer, C. C., Haldemann, J., Troche, S. J., & Rammsayer, T. H. (2012). Auditory and visual temporal sensitivity: evidence for a hierarchical structure of modality-specific and modality-independent levels of temporal information processing. *Psychol. Res.*, 76, 20–31. doi: 10.1007/s00426-011-0333-8.
- Szelag, E., Kowalska, J., Rymarczyk, K., & Pöppel, E. (2002). Duration processing in children as determined by time reproduction: implications for a few seconds temporal window. *Acta Psychol.*, 110, 1–19. doi: 10.1016/S0001-6918(01)00067-1.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the 'internal clock'. *Psychol. Monogr.*, 77, 1–31. doi: 10.1037/h0093864.
- Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Cross-modal temporal discrimination : Assessing the predictions of a general pacemaker-counter model. *Percept. Psychophys.*, 68, 1140–1152. doi: 10.3758/BF03193716.
- Van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain. *Front. Psychol.*, 4, 388. doi: 10.3389/fpsyg.2013.00388.
- Walker, J. T., & Scott, K. J. (1981). Auditory – visual conflicts in the perceived duration of lights, tones, and gaps. *J. Exp. Psychol. Hum. Percept. Perform.*, 7, 1327–1339. doi: 10.1037/0096-1523.7.6.1327.
- Wearden, J. H. (1991). Human performance on an analogue of an interval bisection task. *Q. J. Exp. Psychol. B*, 43, 59–81. doi: 10.1080/14640749108401259.

- Wearden, J. H. (1992). Temporal generalization in humans. *J. Exp. Psychol. Anim. Behav. Proc.*, 18, 134–144.
- Wearden, J. H. (2015). Mission: Impossible? Modelling the verbal estimation of duration. *Timing Time Percept.*, 3, 223–245. doi: 10.1163/22134468-03002051.
- Wearden, J. H. (2016). Modelling chronometric counting. *Timing Time Percept.*, 4, 271–298. doi: 10.1163/22134468-00002070.
- Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory? Episodic temporal generalization and bisection in humans. *Q.J. Exp. Psychol. B*, 54, 289–309. doi: 10.1080/713932763.
- 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. *Q.J. Exp. Psychol. B*, 51, 97–120. doi: 10.1080/713932672.
- Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. *Q.J. Exp. Psychol. B*, 48, 289–310. doi: 10.1080/14640749508401454.
- Wearden, J. H., & Jones, L. A. (2013). Explaining between-group differences on timing tasks. *Q.J. Exp. Psychol.*, 66, 179–199. doi: 10.1080/17470218.2012.704928.
- Wearden, J. H., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled-duration illusion. *J. Exp. Psychol. Hum. Percept. Perform.*, 33, 716–729. doi: 10.1037/0096-1523.33.3.716.
- Wearden, J. H., Todd, N. P. M., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? *Q.J. Exp. Psychol.*, 59, 1709–1724. doi: 10.1080/17470210500314729.
- Williams, E. A., Yüksel, E. M., Stewart, A. J., & Jones, L. A. (2019). Modality differences in timing and the filled-duration illusion: Testing the pacemaker rate explanation. *Atten. Percept. Psychophys.*, 81, 823–845. doi: 10.3758/s13414-018-1630-8.