Temporal Production and Secondary Tasks: Application of a Pacemaker-Gate-Counter model

David Timothy Field

Thesis submitted for the award of
Doctor of Philosophy

Department of Psychology
University of Surrey
2001
To Mum and Dad
Abstract

This thesis examines the effects of secondary tasks and click trains upon temporal judgement in the context of a Pacemaker-Gate-Counter (PGC) model. All the timing experiments reported employ a paradigm in which subjects are first trained to reliably reproduce a 2.5 s target interval, and are then required to perform time production with a concurrent secondary task.

Previous research with digit memory loads has shown that varying memory load had no impact upon concurrent time production (Fortin & Breton, 1995; Fortin & Massé, 1999). Here, it is shown that increasing pitch memory load lengthens time production, but that this is not the case for a colour memory task, or a timbre memory task. The effect obtained with pitch is replicated, and it is demonstrated that the effect is directly due to the processing requirements of retaining pitch information. Furthermore, the pitch effect is not due to a difference in attentional requirements between retaining pitch and retaining digits. Finally, it is shown that the lengthening of time production also occurs when a concurrent duration memory load is increased.

In confirmation of previous research (e.g. Fortin, Rousseau, Bourque, & Kirouac 1993), it is shown that when memory-search is performed concurrently with time production, increasing the number of items to be searched causes a lengthening of time production. A novel finding is that the increase in mean time produced is not accompanied by an increase in standard deviation. Furthermore, it is shown that the shortening of mean time production caused by concurrent click trains does not interact with the increase caused by concurrent memory search, and is accompanied by a reduction in standard deviation. These findings are taken to support the separation made in the PGC model between the Pacemaker and Gate components.

Overall, the data presented in this thesis provide a number of constraints upon future theorising within the framework of the PGC model and other similar models.
Acknowledgements

Firstly, I would like to thank my supervisor, John Groeger, who provided critical comments and positive feedback in the right quantities, and at the correct times.

Secondly, I would like to thank Peter Simpson, for many provocative discussions, as well as technical assistance.
# Table of Contents

Chapter 1

1.1 General introduction: aims of chapter 1

1.2 Different kinds of temporal judgement 2

1.3 Basic findings applicable to animal and human timing 6

1.4 Do perceptual timing and motor timing share the same mechanism? 11

1.5 Nontemporal factors affecting duration judgement:
   stimulus characteristics and secondary tasks 19

1.6 The Pacemaker-Gate-Counter (PGC) model 37

1.7 Neuropsychological evidence 46

1.8 Questions addressed in this thesis 58

Chapter 2

2.1 Aims of chapter 60

2.2 Introduction 60

   2.2.1 Fortin et al. (1993) in the context of the PGC model 63

   2.2.2 Short-term memory for pitch 65

   2.2.3 The memory-search process 73

2.3 Experiment 1 78

   2.3.1 Method 79

   2.3.2 Results 83

   2.3.3 Discussion 90
4.2.1 Method 140
4.2.2 Results 142
4.2.3 Discussion 145

4.3 Experiment 7 146
4.3.1 Method 150
4.3.2 Results 152
4.3.3 Discussion 156

4.4 Experiment 8 158
4.4.1 Method 162
4.4.2 Results 163
4.4.3 Discussion 165
4.4.4 Maintaining duration concurrently with time production 167
4.4.5 Method 168
4.4.6 Results 169
4.4.7 Discussion 173

Chapter 5
5.1 Aims of Chapter 176

5.2 Experiment 9 177
5.2.1 Method 182
5.2.2 Results 185
5.2.3 Discussion 188
1.1 General Introduction: Aims of Chapter

The psychology of time is a broad domain, concerned with sufficiently diverse types of temporal judgement that no single theory could account for all the findings concerning the human experience of time. The first aim of the General Introduction is to define which types of temporal judgement this thesis is concerned with, as well as alluding briefly to those which the theory tested here is not concerned with (section 1.2). The theme of the chapter is to present evidence from various types of investigation of the prospective processing of short intervals, and to show that this evidence is consistent with a model, which will be called the Pacemaker-Gate-Counter (PGC) model. The aim of the empirical work presented in this thesis is largely to test the PGC model. The first section of literature considered will be that concerning judgements of duration made by animals and some analogous experiments with human subjects (section 1.3). This will provide the opportunity to describe some of the basic findings relating to the processing of short intervals. In section 1.4, literature suggesting that the system underlying the ability to make temporal judgements also plays a role in the temporal control of motor action will be described. In section 1.5, the issue of how temporal judgement varies due to the non-temporal characteristics of stimuli, as well as the influence of competing demands on processing resources from non-temporal secondary tasks will be considered. Having described a cross section of the relevant data, the PGC model will be presented in detail (section 1.6). Where it is directly relevant to the issues under consideration, neuropsychological evidence will be considered within each of the sections described above. Further neuropsychological evidence concerning the processing of short intervals,
which is largely consistent with the PGC model, will constitute section 1.7 of the General Introduction. Finally, section 1.8 outlines the specific questions addressed in this thesis.

1.2 Different kinds of temporal judgement

Human temporal judgement comes in many forms. Judgement of order, or succession, is one common kind of temporal judgement. This type of judgement is commonly contrasted with the perception or estimation of duration. However, succession and duration are themselves heterogeneous categories covering very different kinds of judgement. In the case of succession, one kind of judgement would simply be based upon general knowledge (e.g., "Was the Boer war preceded by World War I?"). Another kind of succession judgement relies upon episodic memory (e.g., "Was my holiday in Greece before I had my appendix out"). Finally, another class of judgements involves situations such as when an experimental subject is asked to attend to whether a click presented to the left ear occurred before one presented to the right ear. In this case the real time difference is likely to be only a few 10's of msec, and the aim is to establish the threshold for this kind of judgement. Clearly, different cognitive processes underlie these different kinds of judgement. Similar distinctions may be drawn within the duration judgement category. Subjects may be asked to say which of two tones lasted longer, or to reproduce the duration of a single tone by holding down a button. However, judgements may also be made from general knowledge (e.g., "Does it take longer to boil an egg or bake a potato?"), or personal episodic memory (e.g. "How long did it take to play that tennis match?").
A major distinction, which crosses the category boundary drawn between succession and duration described above, is whether temporal judgement is retrospective or prospective. Prospective judgements occur when subjects focus their attention upon making a temporal judgement concerning a target interval as the interval unfolds. The experimental situations outlined above are classic example of prospective temporal judgement. An everyday example might be the experience of waiting in the queue for the cash-point in which attention is drawn to time by the experience of waiting. Retrospective judgements occur when temporal judgements relating to events in the past are made. Critically, the subject is not aware of the need to make an estimate until after the target event is complete. An important difference between the two kinds of judgement is that prospective judgements largely require short-term memory processing, whereas retrospective judgements invariably involve long-term memory. Due to the limitations of short-term memory, prospective judgements of long intervals invariably involve long-term memory and become akin to retrospective judgements. Retrospective judgement is a reconstructive process, being influenced by such factors as the number of events that took place in the target interval. As such, it does not require a dedicated timing process to be assumed. Several authors provide discussions of how this occurs, as well as the other issues touched upon above (e.g., Fraisse, 1984; Block, 1990; Zakay & Block, 1997; Ornstein, 1969; Friedman, 1993). However, there is a growing body of evidence that an internal clock, directly concerned with the subdivision and measurement of time, underlies the prospective judgement of short intervals. The current research is concerned with that mechanism.
What is meant by a short interval? Fraisse (1984) makes a distinction between intervals of 3 sec or less, where time is “perceived”, and longer intervals where it may only be “estimated”. This distinction is adopted here, and the PGC model applies only to what Fraisse calls “time perception”. A number of lines of evidence support this distinction. Firstly, if subjects are required to reproduce intervals varying between a few hundred msec and around five seconds, then a qualitative change in performance is evident at around 3 seconds (Pöppel, 1994; Elbert, Ulrich, Rockstroh, & Lutzenberger, 1991). For values below 3 seconds a slight overestimation typically occurs, whereas an increasing underestimation takes place with longer target values. The point of crossing over at 3 seconds has been referred to as the “indifference interval”. Elbert et al. (1991) measured event-related potentials (ERP) whilst subjects reproduced visual stimuli ranging from 1 to 8 s. A contingent negative variation was evident for intervals of up to 3 s only, suggesting a different mode of processing for short intervals. Secondly, the “indifference interval” appears to correspond to a wider perceptual phenomenon, in which objects may be temporally grouped only if they occur within a few seconds of each other. For example, the Necker cube is an ambiguous figure that may be perceived in two different perspectives, and it is impossible to prevent spontaneous switching between the two percepts. The spontaneous rate of reversal is once every 3 seconds on average (Pöppel, 1994). The same value has been demonstrated for ambiguous auditory percepts (Radilova, Pöppel, & Ilmberger, 1990), and for binocular rivalry (Gómez, Argandona, Solier, Angulu, & Vanquez, 1994).

A third example, which illustrates the qualitative change in temporal processing at 3 seconds, also relates to the issue of the whether time is explicitly represented in motor
co-ordination. If subjects must tap in time with a metronome beat, then for inter-tap intervals of up to 3 seconds, finger taps consistently anticipate the metronome beat by 20 or 30 msec. However, if the metronome interval is extended beyond 3 seconds finger taps become responses to the metronome beats (Mates, Müller, Radil, & Poppel, 1994). This suggests that a motor program can only incorporate a temporal delay of similar magnitude to the "indifference interval", at least in the case of finger tapping. There are two main issues concerning motor co-ordination and timing. Firstly, how can it be known that a motor action whose success is time critical actually requires the nervous system to represent time explicitly? For example, when a cricket ball is delivered the time of release is critical to the bowler's success. However, this time may be linked to the position or velocity of the arm rather than being explicitly represented. In other examples, it is hard to see how successful performance could be achieved without representing time directly. A case in point is that of playing the drums, or simply tapping ones foot in time to an imagined tune. In fact, the often used experimental method of time reproduction, and the variant method used in the experiments reported here, known as "production of a learned interval" are, strictly speaking, tests of motor co-ordination. The second question addressed here, is to what extent time critical motor co-ordination and the prospective temporal judgement of short intervals are subserved by the same mechanism. This issue is discussed in section 1.4, where it is argued that a clear link has been demonstrated. To this extent, the PGC model has application to motor co-ordination.
1.3 Basic findings concerning animal and human timing

What is the psychophysical law relating subjective time to real time? In the case of short intervals with humans, and longer intervals with animals, a simple linear relationship has been demonstrated, and some of the evidence for this will be considered here. Thus, subjective time is, on average, an accurate reflection of real time. This question is complicated somewhat in the human case by the use of strategies, such as chronometric counting. A second important question, is whether a general form of Weber’s law applies to temporal judgement? Here, evidence will be presented that the variability of temporal judgement is directly proportional to the target interval, for both animals and humans. The mean accuracy property of temporal judgement and the Weber property are the two most important empirical findings that any theory of temporal processing must seek to explain (Wearden, 1994).

The Weber property and the mean accuracy property have been demonstrated in rats by Roberts (1981), using the peak procedure with intervals in the range of 20 to 40 s. In the peak procedure each trial commences with a start signal, such as a tone, and subjects receive food for the first lever press after the target interval has elapsed. On a small proportion of trials responding is not rewarded with food. The presence of these trials allows the response rate to be measured for the periods before and after the target interval has elapsed. Roberts (1981) found that the peak response rate corresponded almost exactly to the target interval, and that the standard deviation of the curve increased in proportion to the target interval.
Wearden & McShane (1988) performed experiments analogous to the peak procedure with human subjects, using the method of interval production. In order to render chronometric counting ineffective target durations ranged from 0.5 to 1.3 s, and accurate feedback was provided after each trial. The results were directly comparable with those from rats with longer target intervals. The peak response frequency corresponded to the target interval, and the coefficient of variation was approximately constant as the target interval varied. The model which embodies these properties is known as scalar timing theory (see, for example, Gibbon, Church, & Meck, 1984). The PGC model described in section 1.6 is based upon scalar timing theory, with the addition of an attentional gate to allow the effects of secondary non-temporal tasks upon timing to be modelled.

The coefficients of variation obtained for humans with short intervals by Wearden & McShane (1988) were in the range of 0.11 to 0.14, whereas those found by Roberts (1981) for animals were in the range of 0.20 to 0.30. Human timing was found to be less variable despite the fact that non-timing variance in interval production due to the motor response is proportionally greater when the target interval is short. Wearden (1991) suggests that coefficients of variation in the range 0.11 to 0.14 may be seen as a signature of non-counting based timing in humans. When chronometric counting was employed the coefficient of variation declined as a function of the target interval, reaching a very low value by the time 8 s was reached, and thereafter remaining roughly constant. In the

---

1 An alternative possibility is that the coefficients of variation were higher in the peak procedure because the target intervals were much longer. As discussed in section 1.2, there appears to be a qualitative boundary in human temporal judgement at 3 s. If this is also the case for animals, then coefficients of variation in the range of 0.11 to 0.14 may be seen as a signature of short interval timing in general, rather than timing without counting in humans. However, the peak procedure relies upon the use of long intervals.
experiments reported here, human subjects produce intervals of 2.5s, and coefficients of variation are consistently found to be within the range suggested by Wearden for non-counting based timing. Subjects were not instructed to avoid counting, and clearly counting was a possible strategy. However, the experiments generally required subjects to perform secondary non-temporal tasks, which may have prevented the counting strategy from being used.

While chronometric counting sets humans apart from animals, and may confound some investigations of temporal judgement, it has been demonstrated by Wearden (1991) that the individual count units themselves display the mean accuracy and Weber properties. In an ingenious study, subjects produced intervals of 6 s using chronometric counting, and the number of count units was varied between 5 and 12. Data confirmed that individual count units conformed to scalar timing theory. Firstly, the mean duration of count units was equal to the ideal duration, and secondly, the standard deviation of the count units was directly proportional to the mean. This may be taken to suggest that when humans count out an interval, the underlying process is the same as that for temporal judgement without counting.

There are two other paradigms that have been used with both animals and humans, namely temporal generalisation, and temporal bisection. Both of these paradigms remove the contribution of motor variance to the data, which is present in interval production. These paradigms have produced results which are broadly similar between animals and humans, but which differ in important details. In temporal bisection with animals (e.g., Platt & Davis, 1983) the subject is presented repeatedly with two tones, which represent a short standard duration, and a long standard duration. There are
two response levers: one is reinforced if it is pressed after the short tone is presented, and the other is reinforced if it pressed after the long tone is presented. Once initial training is complete, unreinforced stimuli of various intermediate durations are presented. The probability of a long response may then be plotted as a function of stimulus duration. The paradigm takes its name from the “bisection point”, which is the duration where 50% of responses are “long”. This point is thought to reflect the animal’s subjective midpoint between the short and long standards. Platt & Davies (1983) found that the bisection point occurred at the geometric mean of the two standard durations, i.e. slightly less than the arithmetic mean.

Typically the short and long standards in a bisection experiment with animals are in the range of 2 and 8 s. Wearden (1991) reports an analogous experiment with human subjects, which used stimuli an order of magnitude shorter than those used with animals in order to prevent chronometric counting. The training phase was much shorter than that used with animals, and exemplars of the standards were presented between experimental blocks. Results indicated that the bisection point in humans was closer to the arithmetic mean than the geometric mean. However, Allan & Gibbon (1991) found a bisection point closer to the geometric mean with human subjects. Wearden (1991) suggests that differences between animal and human performance do not arise from the underlying representational system for duration, but from the nature of the decision process.

Church & Gibbon (1982) reinforced responses by rats if a response was preceded by a stimulus of a standard duration (e.g., 4 s), while responses to stimuli of other durations were not reinforced. This is the procedure known as “temporal generalisation”. Responding declined as a function of the distance of the stimulus from the standard, and
the generalisation gradient was symmetrical around the standard, i.e. the response rate was identical for a stimulus 20% longer than the standard and a stimulus 20% shorter than the standard. This result may be taken to suggest that subjective time is a linear function of real time in the rat.

Wearden (1991) reports an experiment in which human subjects perform temporal generalisation with a standard stimulus of 0.4 s. Food reinforcement was replaced with accurate feedback. Results indicated that the generalisation gradient was not symmetrical around the standard; longer stimuli were more likely to be classified the same as the standard than equivalent shorter stimuli were. However, when the nonstandard stimuli were spaced logarithmically around the standard, the generalisation gradient became more symmetrical. Again, it was suggested that the decision process in humans is different from that in animals, but that the underlying clock mechanism is the same.

One possible reason for the differences described above between animal and human performance in bisection and generalisation, is that the stimuli used with human subjects were all shorter than the “indifference interval” described in section 1.2, whereas those used with animals were all longer than the “indifference interval”. There is some evidence of a 3 s boundary in animal time perception, similar to that found in human subjects. Gerstner & Fazio (1995) hypothesised that the duration of vigilance postures in animals reflect the same temporal processing constraints as those identified at the “indifference interval” in humans. Their study found that vigilance postures had an average duration of around 3 s in different species of primates, carnivores, and marsupials. Furthermore, a similar study of humans conducted by Wirtz & Wawra (1986) found a similar value. There is growing evidence that the processing of intervals
shorter than 3 s relies on a system located in the cerebellum, whereas the processing of somewhat longer intervals is associated with the basal ganglia (Wittman, 1999; Hazeltine, Helmuth & Ivry, 1997; but see the discussion of Harrington, Haaland, & Hermanowicz (1998) in section 1.7 for a different view). Having argued that there are good theoretical grounds to expect the differences found between the analogous animal and human experiments described above because of the difference in stimulus durations employed rather than the species difference, evidence is provided by Wearden, Rogers & Thomas (1997) which is supportive of the “species difference” account. Human subjects performed temporal bisection using longer stimulus durations, similar to those used in animal experiments, and chronometric counting was prevented by means of a concurrent digit-shadowing task. However, despite the similarity in stimulus duration to the animal subjects, results were similar to those for humans with short durations (i.e., bisection closer to the arithmetic mean than the geometric mean).

1.4 Do perceptual timing and motor timing share the same mechanism?

Two main points have been made thus far about the timing phenomena that the PGC model seeks to explain. Firstly, it is limited to time perception, that is the prospective processing of durations of 3 s and less. Secondly, it has been argued that the model applies in principle to both animal and human timing. This section asks whether the processes described by the model might underlie the timing of actions, as well as
perceptual timing tasks such as the generalisation and bisection tasks described in section 1.3.

One study that directly addressed the issue of whether the perception of time and the timing of actions are related was conducted by Keele, Pokorny, Corcos & Ivry (1985). In one task, subjects judged the duration of brief perceptual events. In a second task the same subjects produced a series of taps, and attempted to keep the inter-tap interval constant. The second task was performed under a foot tapping condition and a finger tapping condition. The correlation between variability of performance on the two motor tasks, as well as that between the motor tasks and the perceptual task was high (around 0.6). Additionally, the perception task did not correlate with performance on a motor task that did not require precise temporal control. These results may be taken to suggest that a central mechanism supports timing performance on the different tasks. In a similar correlational study, Franz, Zelaznik & Smith (1992) asked subjects to produce repetitive timed movements with the finger, the forearm, and the jaw. In addition, another task involved making repetitive syllable sounds at the same frequency as the tapping tasks. Significant correlations in timing variance were found between all of the tasks, and the authors suggest that the common timing process may also be involved in the production of movements involved in speech. However, it seems likely that the correlation between the syllable task and the tapping tasks arose because of the repetitive timing component of the syllable task. Such repetitive syllable production is not characteristic of normal speech, and therefore these results cannot be taken to suggest that the timing process used in repetitive tapping is also used in speech production. In order to support such a hypothesis, it would be necessary to demonstrate the correlation between
tapping task variability and variability of timing of syllable production in normal speech. Allan (1979) reviews evidence from cross-modal correlational studies, which supports the existence of a central timing mechanism. In particular, intermodal temporal judgements are as highly correlated as intra-modal ones, and transfer effects have been demonstrated between the visual and auditory modalities.

Assuming for the moment that time-critical motor responses do rely upon the same mechanism as perceptual judgements of duration, it seems likely that not all of the variance in such a task could be accounted for by the timing process. In particular, a certain amount of variability might be associated with the execution of movements. This issue was first investigated by Wing & Kristoffersen (1973), who used an ingenious method to partition the variance in a tapping task into timing variance and non-timing variance. Subjects first matched their finger-taps to a pacing signal, and when the pacing signal ended, they attempted to continue the sequence. Two assumptions were made. These were, firstly, that the two sources of variability were independent, and secondly that subjects made no attempts to correct for errors made on previous taps. From these assumptions, differential effects of the two kinds of variability upon tapping may be predicted. Firstly, the influence of the timer will be restricted to individual taps, and secondly that motor variability will cause a negative correlation between successive intervals. As an illustration of the effect of motor variability, consider a pair of taps in which the motor response component dividing the two taps happens to be on the long side. This will simultaneously cause the first interval to be lengthened and the second interval to be shortened. The size of the observed negative correlation between successive intervals was used by Wing & Kristofferson (1973) to estimate motor variability. Timing
variability was then estimated by subtracting the value for motor variance from the overall variability.

The fact that Wing & Kristoffersen (1973) were able to support their assumptions constitutes strong evidence that paced tapping relies on a central timer. However, their method is not applicable to perceptual timing tasks. Ivry & Hazeltine (1995) developed a new method of partitioning variance on timing tasks (called the "slope method") which is applicable to both tapping tasks and perceptual tasks. Because the method is directly applicable to both tasks, it allowed Ivry & Hazeltine (1995) to provide a strong test of the hypothesis that a single timer subserves both processes. Slope analysis is based on the two assumptions that firstly variance on a timing task is composed of duration-dependent and duration-independent variance, and crucially, that duration-independent variance will remain constant as the target interval is varied. Duration-independent variance in a tapping task corresponds to motor implementation variance, while in a perceptual task it arises from the time taken to detect a signal. It has already been established that duration-dependant variance varies in direct relation to the target interval. Ivry & Hazeltine asked the same subjects to perform the Wing & Kristoffersen repetitive tapping task and a time perception task, across a range of target intervals. The slope of the function of variability with target interval served as an estimate of the Weber fraction for the two tasks, while the intercept of the function gave an estimate of duration-independent variance. If the two tasks relied on the same timing system then the slope of the function (Weber fraction) would be identical for both tasks. In their Experiment 1, the Weber fraction was much less for the tapping task. However, in that study the number of presentations of the target interval differed across experimental conditions. In the tapping task, a long series of
target intervals was presented in order to allow the subject to “synchronise”, whereas the perceptual task relied on a single presentation of the target interval. In their Experiment 2 the motor task relied upon a single presentation of the interval, as well as a single production. In their Experiment 3 the number of presentations of the target interval in the perception task was increased to make it comparable to the motor task. Both these changes to the paradigm resulted in identical Weber fractions in both tasks, providing strong support for the correlational studies discussed earlier. The Weber fraction was lower in both tasks when there were multiple presentations of the target interval, suggesting that the internal representation of the target interval can be refined by multiple presentations.

The research described so far considered whether time perception and interval production or timed tapping share the same timing mechanism. A more general question is whether this mechanism might also underlie the timing of motor responses where timing is critical, but where the subject is not solely concerned with the duration between two motor responses. In everyday tasks, timing is critical, but responses are generally timed in relation to some event in the world, rather than each other (a notable exception is playing the drums, which is very similar to the tapping tasks used in Experiments). One study suggests that the answer to the more general question may be positive. In a number of publications, Triesman and colleagues model the pacemaker component of the internal clock as a temporal oscillator. Triesman, Faulkner, Naish & Brogan (1990) showed that high tempo auditory click trains cause biases in temporal judgements, and that the extent of the bias was related to the speed of the click trains. Deviations (or “dip peaks”) from the generally linear relationship between click train speed and the amount of acceleration
of the rate of subjective time were thought to represent resonation with the fundamental frequency of the temporal oscillator. Analysis of the positions of the deviations allowed Triesman et al. (1990) to estimate the frequency of the oscillator as 49.5Hz. Triesman, Faulkner & Naish (1992) showed that the same underlying oscillator pattern can be detected in the timing of motor action by imposing click trains during a choice-reaction time task. This demonstration was particularly impressive because the frequencies of click trains that caused deviations in the choice reaction time experiment were the same as the frequencies that caused deviations in the earlier work with time perception.

Because the choice-reaction time task has fundamentally different temporal requirements to the tapping task — to respond as quickly as possible — this study supports the idea that the timing of motor responses in general shares at least some resources with the perception of duration. Further support for this idea comes from the fact that Triesman et al. (1992) were able to extend the effects of click trains to typing, which is a motor task involving more complex combinations of motor responses than choice-reaction time. In the context of the PGC model, it might be speculated that the pacemaker is shared between perception and motor co-ordination, but that other processes are likely to differ between perceptual timing and the timing involved in motor co-ordination.

**Neuropsychological evidence**

The most important contribution made by neuropsychological evidence in the debate about the extent to which a single timing system subserves motor co-ordination and perceptual timing, has been to suggest that the element in common is functionally and anatomically localised in the cerebellum. In sections 1.6 and 1.7 it will become clear
that not all of the memory and attention systems with a proposed involvement in perceptual timing are relevant to motor co-ordination. However, the pacemaker, which acts to subdivide the continuum of time into units, seems to be common to both types of task.

Ivry & Keele (1989) compared the performance of several patient groups (Parkinson, cerebellar, cortical, peripheral neuropathy, and age-matched controls) on the Wing & Kristoffersen tapping task, as well as a perceptual task with similar target intervals. Because durations were defined in the auditory modality, a control condition was included, in which patients' perception of loudness was assessed. The only group to show deficits on both timing tasks was the cerebellar group, suggesting that the cerebellum has primacy in both perceptual and motor timing. Additionally, they showed no deficit on the control task, suggesting that the results on the other two tasks could not be attributed to any deficit in auditory perception. In a further study, Ivry, Keele, & Diener (1988) compared the performance on the tapping task of patients with medial cerebellar lesions to that of patients with lateral cerebellar lesions. Because the lesions were focal, patients provided their own controls. Performance on the affected hand, ipsilateral to the site of the lesion, was compared to performance with the unaffected contralateral hand. The performance of all patients was more variable with their affected hand compared to their “control” hand. The tapping data was then subjected to the Wing & Kristoffersen method of decomposing variance into that associated with timing and that associated with motor implementation. This revealed a clear dissociation; the patients with medial lesions suffered from increased motor implementation variance, while those with lateral lesions showed increased timing variance. Therefore, both regions of the
cerebellum are important in time-critical motor tasks, whereas only the lateral cerebellum would be expected to make an important contribution to perceptual duration judgements.

Schubotz, Friederici, & Yves von Cramon (2000) conducted an fMRI study, which revealed that time perception and motor timing share a common cortical and subcortical basis. They did not employ a motor timing task themselves, but refer to a substantial amount of literature which indicates that a “motor circuit” is responsible for the timing of motor actions. The “motor circuit” is comprised of the premotor cortex, primary motor cortex, primary somatosensory cortex, medial wall motor areas in cingulate cortex, the putamen (part of the basal ganglia), thalamus, supplementary motor area, and cerebellum. The study was designed to test the hypothesis that a perceptual timing task would activate the same, or similar, areas. In order to provide sufficiently long periods of brain activation for detection by fMRI, the perceptual timing task employed had to be one involving continuous monitoring for approximately 10 s. Therefore, subjects were asked to attend to rhythms and respond only if an abnormality from a fixed pattern was detected. Such an abnormality occurred on only 30% of trials. These trials were excluded from the analysis in order to prevent activations due to motor responses contaminating the data. The authors accept that the task may have also been performed by comparing individual durations, rather than rhythm per se, but argue that this does not matter. In order to control for modality differences, the task was performed in both auditory and visual conditions. In the auditory condition, the baseline condition used for the subtractive methodology consisted of an analogous task, using the same stimuli, in which deviations of pitch had to be discriminated. In the visual baseline condition, colour deviation was monitored. All subjects reported that the visual rhythm
condition was harder, and it resulted in greater activation of the lateral cerebellum. However, there were no significant differences in the foci of activation between modalities, which is supportive of the notion of a central timing mechanism. As predicted, the "motor circuit" was active during the processing of rhythm, suggesting a common cortical and subcortical basis for both time perception and motor timing. Of particular interest is the fact that the cerebellar activations were focused upon the lateral, rather than medial, cerebellum. This confirms the conclusions of Ivry et al. (1988), discussed above.

1.5 Nontemporal factors affecting duration judgement: stimulus characteristics and secondary tasks

The judgement of duration is not neutral with regard to the nontemporal characteristics of stimuli. Under certain circumstances, two stimuli having identical duration will be consistently judged as different from each other. For example, in the "filled duration illusion" (e.g., Thomas & Brown, 1974) a filled duration is perceived as longer than an empty duration. In this section, a number of such factors related to the characteristics of stimuli will be considered. The judgement of duration is also systematically affected by the extent to which subjects are required to process information other than the target interval. As this effect is one of the focuses of the current research, the literature concerning the influence of secondary tasks upon temporal judgement will be reviewed.
The mass of findings concerning the effects of many different stimulus characteristics can seem particularly confusing. This initial impression is confirmed by the fact that many effects are subject to qualification by other factors. For example, Thomas & Brown (1974) found that an empty interval that is divided by markers seems longer than a completely empty interval, and that the more markers there are the greater this effect is. This apparently simple finding was qualified by Adams (1977), who showed that this effect is abolished if the duration boundary markers are presented to one ear and the dividing markers are presented to the other ear. Another question relating to this effect is whether it is actually a variant of the filled duration illusion, or a separate phenomenon.

One effect of stimulus characteristics upon temporal judgement, which turns out to be of great theoretical importance for models of time perception, is the finding that a filled interval defined in the auditory modality is judged as longer than an equivalently defined visual interval (e.g. Goldstone & Lhamon, 1974). This finding will be treated in detail after other findings related to stimulus characteristics have been briefly covered. Goldstone, Lhamon, & Sechzer (1978) found that stimuli of greater intensity are judged as longer than stimuli of less intensity, and that this effect is stronger for vision than audition. This effect might be analogous to that found by Triesman et al. (1990) with click trains; more intense stimuli could cause an arousal of the pacemaker relative to less intense ones. As with the effect of segmenting an interval, the filled duration illusion might be seen as part of the same phenomenon — an empty interval is the lowest intensity stimulus that may be presented, and is therefore judged as shorter than any other
kind of interval. Another variable which affects temporal judgement is stimulus size. Larger visual stimuli are judged as longer (Thomas & Cantor, 1975, 1976). Stimuli that move are judged as longer than stationary stimuli, and faster moving stimuli are judged as longer than slower ones (Brown, 1995). Schiffman & Bobko (1974) found that as a set of stimuli became more orderly, their judged duration became shorter. In order to demonstrate this they used a varying flash pattern of 8 lights to mark out a temporal interval. In this case, it seems unlikely that the manipulation is acting upon the speed of the pacemaker. Rather, a more complex temporal pattern may draw more attention towards temporal processing, causing more temporal information from the pacemaker to be noted.

Returning to the finding that sounds are judged longer than lights, Wearden, Edwards, Fakhiri & Percival (1998) present an investigation of this phenomenon, which is unusual among its peers because it attempts to provide a detailed theoretical explanation of the effect. This was achieved through the application of scalar timing theory, and as the PGC model tested here is largely based upon scalar timing theory, this research is described in detail here. Their Experiment 1 used the temporal generalisation without feedback technique, in which a standard duration is compared with a series of stimuli, including the standard, and subjects make same / different judgements. Standard stimuli of both modalities were compared with stimuli in both modalities. The results of the cross-modal comparison conditions indicated that visual standards were maximally matched to auditory comparison stimuli that were shorter, while auditory standards were maximally matched to visual comparison stimuli that were longer. An important subsidiary finding was that in the within-modality conditions, the generalisation gradient
was flatter for the visual stimuli than the auditory ones, meaning that comparison stimuli were more likely to be confused with the standard in the visual condition. This implies that the representation of the duration of the visual stimuli was more variable than that of the auditory stimuli.

Scalar timing theory proposes a pacemaker, which produces pulses at a particular average rate, and an accumulator where some transformation of the number of pulses is stored. One reason that sounds might be judged longer than lights is if the pacemaker ran faster for auditory stimuli. However, as Wearden et al. (1998) point out there are other possible reasons for the effect. One such possibility is that the time taken to open and close a switch between the pacemaker and the accumulator might vary between the two modalities. If the difference between the time taken to open and close the switch were less for auditory stimuli then this would cause auditory stimuli to be judged longer by a small amount, and this amount would be constant across a range of durations. In contrast to this, if the modality effect is due to a difference in pacemaker speed between the two modalities, then the size of the difference should increase with the magnitude of the target duration. A third possibility is that both effects might be observed.

In a second experiment, Wearden et al. (1998) test the above predictions by asking subjects to verbally estimate the duration of a range of visual and auditory stimuli. Results showed that the function relating estimates to actual duration was steeper for the auditory stimuli. Over the range of durations sampled, this produced an interaction between modality and stimulus length, which was consistent with the hypothesis that the pacemaker runs faster for auditory stimuli. There were no intercept differences, suggesting that differences between modalities in the time taken to operate the switch do
not contribute to modality differences in judged duration. Again, performance was more variable in the visual condition.

In a third experiment, Wearden et al (1998) investigated the differences in variability found in their first two experiments. It was proposed that variability differences arose because the switch process was more variable for visual than auditory stimuli. Alternatively, results with visual stimuli might be more variable because when the pacemaker is slower it is more variable. In order to test these hypotheses, additional conditions were included, in which stimulus presentation was preceded by a 5 sec train of 5 Hz clicks, which were similar to those used by Triesman et al. (1990) and in Experiment 10 here. Such click trains are hypothesised to affect the rate of the pacemaker. If the pacemaker speed manipulation also affected variability, then this would suggest that variability differences arose directly because of differences in pacemaker speed. However, if changes in mean judgement due to click trains were not accompanied by changes in variability, then this would suggest that the variability difference between modalities was independent of pacemaker speed, and consequently differences in variability of switch operation would be implicated. Results indicated that click trains increased the subjective length of stimuli in both modalities by an equal amount. Again, this is consistent with the idea that a single clock underlies judgements in both modalities (but that it runs slower for visual stimuli). Click trains interacted with stimulus duration, such that the effect was greater at longer stimulus duration, consistent with the idea that the click trains accelerated the pacemaker. In contrast to the results of the second experiment, intercept differences were found between the two modalities, suggesting that the difference between the time taken to open and close the switch differs across the two
modalities. Returning to the question of the effect of click trains upon variability, results indicated that the changes in mean estimates caused by the click trains were not accompanied by changes in variability. This was taken as support for the idea that the increased variability of judgements of visual stimuli compared to auditory ones arises from the switch rather than the pacemaker speed differences. However, it might be argued that a stronger test of the pacemaker speed hypothesis would have been made by using a range of different click train frequencies in the experiment. In experiment 10 reported here click trains of 5, 13.5 and 25 Hz were employed, and the amount of change in mean time production was related to click train frequency. In particular, use of the 25 Hz condition should have produced a larger effect upon mean judgements, and therefore any associated change in variability would have been more easily detected. In summary, Wearden et al’s (1998) results are consistent with the position that there is a single pacemaker, which runs faster for auditory stimuli than visual ones, resulting in differences in judged duration between the modalities. However, differences in variability of judgement between the modalities do not arise from the difference in pacemaker speed, rather the switch component in the scalar model is more variable in the visual modality.

What happens when subjects are required to judge stimuli presented in both modalities simultaneously? Walker & Scott (1981) provide evidence that under these circumstances the auditory stimulus tends to dominate. They asked subjects to reproduce visual stimuli, auditory stimuli, and auditory-visual compound stimuli of equal lengths. Auditory stimuli were reproduced as longer than visual stimuli, while the compound stimuli were reproduced as the same length as auditory stimuli. In one condition, subjects
were required to reproduce both parts of a compound stimulus separately, and the visual part of the compound was reproduced with the same duration as the auditory part. The latter two findings both support the proposition of a single clock with auditory dominance.

This section began by making a distinction between the effects of stimulus characteristics upon perceived duration and the effects of secondary tasks. This distinction is not always clear in the case of studies concerned with stimulus characteristics because the experimenter cannot control the extent to which subjects choose to actively attend to the nontemporal content of the presented information. Therefore, it is possible that some of the effects associated with stimulus characteristics are actually attentional effects similar to those discussed below for secondary tasks. In the case of studies investigating the effects of secondary tasks, the experimenter is on firmer ground, at least to the extent that the subjects attention is explicitly directed towards particular nontemporal characteristics of the stimuli. Before moving on to discuss research concerned with the effects of secondary tasks, one study will be considered which attempted to contrast a situation in which information was simply presented with another in which the same information had to be explicitly processed (i.e., a concurrent secondary task). Vroon (1970) presented subjects with two intervals, each of which was one minute long. One of the intervals contained 30 tones, while the other contained 60 tones. Subjects who were asked to just listen to the tones thought that the interval with 60 tones was longer. A second group of subjects, who had to process the tones so that they could press one button when a high tone was played, and another button in response to a low tone, thought that the interval with 30 tones was longer. This is an example of the
classical attentional effect — the greater the amount of nontemporal information processing the less attention was available to note the passing of time. However, there was a severe methodological problem with this study, which makes a clear interpretation of the results difficult. Vroon states that subjects were “not acquainted with the purpose of the procedure”. This suggests that the first of the two judgements made by each subject was a retrospective estimate. However, the second judgement was then likely to be prospective, as by then subjects were most likely aware of the need to estimate time.

Secondary tasks

An early attempt to be precise about the relationship between the processing of duration, and the processing of nontemporal information was made by Thomas & Weaver (1975). In a forced-choice procedure subjects had to judge the duration of a luminous zone on a screen. The durations used were very short, in the region of 100 msec. The nontemporal task involved detecting the presence or absence of a target letter in an array, which was projected onto the luminous zone. These tasks were performed separately and together, and in the mixed condition emphasis was placed on one or the other task. Judgement accuracy in the duration task was found to be a function of the amount of attention directed towards the other task. A model was presented in which attention is shared between a nontemporal processor and a temporal processor. Unfortunately, because of the very short visual stimuli used it is likely that subjects were able to make judgements based upon visual persistence (Long & Beaton, 1980). Therefore, it cannot be known whether explicit timing processes were tested in Thomas & Weaver’s procedure.
Using a considerably longer target interval of 42 s, Hicks, Miller & Kinsbourne (1976) manipulated the amount of nontemporal information that was processed concurrently with the timing task. Subjects were asked to sort a deck of cards into 1, 2, or 4 stacks, as fast as possible. They were also told that they would have to estimate the duration of the task afterwards, and asked not to count. More cards were dealt in the easier conditions, but analysis in terms of bits of information transmitted showed that more information was "transmitted" by subjects in the harder conditions. As the card-sorting task became more difficult, subjects increasingly underestimated the target interval, and the amount of underestimation was a linear function of the bits of information processed during the target interval. These results are interpretable within the Thomas & Weaver framework of a temporal processor and a nontemporal processor competing for resources.

Fraisse (1979) exploited the classic finding, which also inspired the famous Stroop experiment, that it takes longer to name patches of colour than it does to read the corresponding words. Subjects were trained on a task where they had to read letters at a rate of one per second (a form of time production). When they could do this, they were transferred to experimental conditions where they had to either read colour words once per second, or name colours once per second. Naming colours resulted in longer productions than reading words. This effect can be explained by postulating that while the secondary task takes place temporal processing ceases or slows. Because the naming task takes longer than the reading task, this interruption occurs for longer, effectively slowing the subjective sense of time passing more in one condition than another.
Another similar study, this time using the Stroop task proper, was conducted by Zakay (1993). Three levels of secondary task difficulty were employed: word reading, colour naming (the standard Stroop), and colour word association. Under prospective conditions, subjects performed one of these tasks for 10 s, and were then asked to reproduce the period of time for which they had performed the task. Verbal estimates declined as a function of secondary task difficulty. In another condition, subjects were asked to perform one of the three tasks until they felt 10 s had elapsed. Now, estimates lengthened as a function of secondary task difficulty. In agreement with Fraisse (1979), and Hicks et al. (1976), both these results suggest an increasing failure to note the passing of time as more processing resources are directed away from time passing. These three studies are all in agreement, despite using widely different target intervals, secondary tasks, and estimation methods, which suggests a strong and generalisable effect of nontemporal processing upon temporal judgement.

Zakay (1998) demonstrated that control by the subject of attention allocation policy can influence prospective time judgements. Subjects performed either the Stroop task or an auditory analogue of the Stroop, concurrently with an estimation of the duration (verbal estimate or reproduction) for which they had performed the task. In one condition, they were instructed to give priority to timing, and in another, the secondary task was prioritised. Target intervals of 12.5 s and 25 s were used in the two different tasks. Error rates on the secondary tasks were sensitive to the attentional allocation policy, as were duration estimates — the target interval was estimated a shorter when the secondary task was prioritised. These findings held for all the combinations of target interval, secondary task, and estimation paradigm.
Casini & Macar (1997) demonstrate that simply expecting a nontemporal event is sufficient to reduce the rate at which temporal information is accumulated, using target intervals approximating to the indifference interval, which are therefore in the range considered by the PGC model. Subjects made a decision as to whether a visual stimulus was long (3.5 s) or short (2.5 s). They were also required to indicate whether a brief increment in light intensity that occurred during the interval was weak or strong. The increment in intensity could occur near the start, in the middle, or near the end of the interval. Subjects were expected to shift their attention back and forth between the two processing tasks until the increment in intensity occurred, and the decision regarding it could be made. It was expected that temporal underestimation would manifest itself as a bias towards responding “short”. Results indicated that there was an increased tendency to respond “short” when the intensity increment occurred later in the interval. Presumably, this was because subjects switched attention back and forth between the two types of task, and while attention was directed towards the nontemporal task pulses from the pacemaker failed to reach the accumulator. Therefore, these results are consistent with those from other studies discussed above. Subjects were also asked to prioritise one task or the other to varying extents, which lead to the same effects as those observed by Zakay (1989) discussed above. In a second experiment, Casini & Macar (1997) addressed the possibility that only part of the interference with temporal judgement was due to the effects of attention switching while waiting for the secondary task stimulus to occur, and that interference also occurred due to the processing of the increment. In order to investigate this a small proportion of trials were included that contained no increment in intensity. Because subjects would be expecting the increment to occur for longest in these
trials, if the shortening of subjective duration was wholly due to expectancy then the "no increment" trials would result in the greatest amount of shortening. In the event, these trials resulted in an intermediate amount of shortening. Therefore, it was concluded that both the expectancy of a nontemporal event, and the processing of that event, make separate contributions to the subjective shortening of duration.

All of the studies discussed above assess the interference of nontemporal processing upon temporal processing. The underlying model of attention assumes that both timing and the secondary task require access to a limited capacity resource (attention). The assumption that prospective timing is an effortful process is a plausible one; if timing were automatic then it would not be subject to interference from other tasks. Nonetheless, Brown & West (1990) attempted to test this assumption more directly. Their task involved the prospective timing of multiple concurrent visual stimuli, in the range of 6 to 16 s. There were four groups of subjects, all of whom were presented with identical stimulus arrays consisting of 4 letters. Each of the four letters persisted for a different amount of time. All subjects participated in a control condition, in which only one stimulus was attended to and reproduced. The first group performed the same task in the experimental phase, as a control for practice effects. The second group had to attend to 2 of the 4 targets, and had to reproduce one of them (selected randomly). The third and fourth groups performed the same task, but had to attend to either 3 or all 4 of the targets. Subjects were asked to use whatever strategy they found most appropriate, and it is therefore likely that chronometric counting was used, especially in the control condition. Results indicated that as the number of attended targets increased, absolute timing accuracy declined sharply. In all of the studies discussed above, directional timing error
was analysed, and a slowing of subjective time was demonstrated. In Brown & West's (1990) study no directional timing error was found, although the significant change in absolute error implies that both overestimation and underestimation became more frequent as the number of target intervals increased. Unsurprisingly given the increase in absolute judgement error, there was also an increase in variability of judgement as the number of attended targets was increased. Thus, this study stands alone in that a subjective shortening of duration did not occur as the amount of concurrent processing was increased. This may be because all of the processing was temporal, rather than being a combination of temporal and nontemporal processing. Because this study is unique, and because the available models of timing do not predict this result, it is hard to come to any firm conclusion. Indeed, it is possible that Brown & West's results were due to the use of chronometric counting in the control condition, and to a progressively lesser extent as the number of concurrent intervals increased. This would be consistent with the increase in variability of judgement as the number of attended intervals increased. It would be of interest to conduct a replication of this study using much shorter intervals, in order to discourage the counting strategy. Experiment 8 of the present research also combines two temporal tasks, although in a different paradigm.

In a number of studies, Fortin & colleagues have made a significant contribution to understanding the nature of the interference between temporal processing and nontemporal processing. The current research takes as its starting point the findings of Fortin, Rousseau, Bourque, & Kirouac (1993), as well as those of Fortin & Breton (1995). Modifications of the paradigm employed by Fortin & colleagues were used throughout the current research. The details of the two studies mentioned above, as well
as the details of the paradigm used are described in Chapter 2. For the purposes of the General Introduction, the more important findings of Fortin & colleagues will be summarised.

Fortin & colleagues have generally used a method called “time production”, in which subjects are given extensive practice in the production of a short interval (typically 2 or 3 s) until they can do this without feedback. In experimental conditions, secondary tasks are generally interpolated between the commencement and completion of time production. Often, the binary response that is required by the secondary task is achieved by the same key-press as the response terminating time production. Another general feature of the paradigm is that difficulty is systematically manipulated within a particular task, rather than using different tasks to achieve different difficulty levels. This is usually achieved by manipulations of the “set size” of memory sets or visual displays. Reaction time control conditions are used to demonstrate the effectiveness of the set size manipulations, and estimate the processing time required by the secondary tasks. Because of the presence of secondary tasks, subjects are unlikely to make use of counting strategies, and the typical target intervals are within the range addressed by the PGC model.

Using the above methods, Fortin et al. (1993) demonstrated that manipulating the difficulty level of a visual search task produced no impact upon time production. This is theoretically important, as it is the first published demonstration of a secondary task that does not interfere with prospective timing. Previous theorists (e.g. Zakay, 1993) often talked of interference with temporal processing as a general phenomenon, which would be produced by any task that required the subjects attention. This position was made
untenable by the visual search experiment. Fortin et al. (1993) were able to demonstrate that secondary tasks requiring processing in short-term memory did produce interference, and suggested that in order for pulses to pass from the pacemaker to the accumulator short-term memory resources were specifically required. To test the generality of their effects, Fortin & Rousseau (1998) employed a more standard reproduction task, and achieved similar results. As predicted by gating models, if the secondary task took place during encoding of the interval reproduction was shortened, while if it took place during the reproduction phase reproduction was lengthened. The finding that memory search tasks produce a systematic interference with timing was generalised to another timing paradigm by Fortin, Duchet, & Rousseau (1996), who achieved the same effects of short-term memory processing upon a continuous reproduction task similar to the Wing & Kristoffersen procedure.

Fortin & Breton (1995) demonstrated that processing in visuospatial short-term memory produced similar effects to those demonstrated earlier with phonological memory tasks. This was achieved using a mental rotation task. Interestingly, in the negative response condition of the mental rotation task interference was less than in the positive condition, despite the fact it took just as long in the reaction time control condition. In the negative response condition, it is known that strategies that are not dependent upon processing in visuospatial short-term memory are effective. Therefore, this pattern of results provides additional support for the view that short-term memory rather than attention is the critical resource required by temporal judgement. Fortin & Breton (1995) also investigated whether interference was related to maintaining information in short-term memory or performing operations upon that information. It was
clearly demonstrated that just maintaining information produced no set-size effect, which allowed Fortin & Breton (1995) to reiterate their theoretical position that only active processing of information would produce interference with temporal judgement with greater confidence. However, it should be noted that all of the tasks used, including the tasks which did not produce set-size effects produced a constant lengthening of time production, whatever the specific processing resource required by the particular secondary task. Fortin & Massé (1999) demonstrated that increasing the difficulty of the short-term memory processing required by the secondary task increases the amount of interference with time production. This included a demonstration that maintenance only with the more difficult task produced no set size effect.

Fortin & Massé (2000) performed a study similar in its aims to that of Casini & Macar (1997). The time production task was modified by the addition of breaks, in which the subjects had to cease timing. Once the break was over subjects resumed timing. In order to achieve this subjects timed the interval while a tone played. When the tone ceased to play this signified a break in time production. The resumption of the tone indicated that subjects should resume timing the interval where they left off. Varying the duration of the break between 2 s and 6 s had no impact upon total time production, using a target duration of 2 s. This suggests that the accumulation process can be stopped and resumed without any detrimental effect. The same result was obtained with rats by Roberts (1981) using the peak procedure with longer target durations. These results imply that one characteristic of the accumulator is that the information stored in it does not decay over short time periods. However, the overall duration of time production was systematically related to the position of the break within time production. When the pre-
break duration was longer, time production was lengthened relative to when the break occurred earlier, consistent with a slowing of the accumulation rate. Fortin & Massé (2000) explain this effect by proposing that subjects divide attention (and presumably short-term memory resources) between timing and expectancy of the break until the break occurs. This is a similar proposition to that made by Casini & Macar (1997). In a further experiment, trials without a break were included, and this resulted in the greatest lengthening of time production, presumably because attention was divided throughout the processing of the interval. A final experiment showed that when subjects were cued that a trial would not include a break the expectancy effect was abolished. An important feature of expectancy effects appears to be that they are involuntary; they occur even when expecting a stimulus that requires very little processing itself (e.g., an empty break).

The only neuropsychological study that directly addresses the issue of the effects of secondary tasks upon temporal judgement was conducted by Casini & Ivry (1999). They assembled two groups of patients and a control group. One group of patients had cerebellar lesions, while the other group had prefrontal lesions. Clinical evaluation revealed that the patients performed similarly to controls on all but the Digit Symbol subtest of the WAIS-R. The experimental tasks involved a duration discrimination task using a 400 msec tone to define the target interval, a pitch discrimination task, and a combined task where the pitch and duration of the tone both had to be discriminated. Based upon previous behavioural findings, such as those discussed above, it was hypothesised that all three groups would be impaired on the dual task relative to the single task situation. The cerebellum was thought of as a temporal pacemaker, and therefore it was predicted that the cerebellar group would be impaired in single task
timing compared to controls. The frontal lobes are thought to subserve memory and attentional processes required for the comparison of one interval with another, and therefore this group was also expected to be impaired on the duration task. The difference between the two patient groups was expected to emerge when attention had to be shared between timing and the pitch discrimination task. For the cerebellar group, the dual task deficit was not expected to be any greater than that of the control group because their lesions should have impaired only the pacemaker function. However, the frontal group was expected to show a relatively greater deficit compared to controls because their lesions should have impaired attentional functions, which are required for switching between two tasks. An adaptive psychophysical procedure was used to estimate bias (point of subjective equality) and acuity (variability) on both tasks. In the dual task condition, which required two responses to be made, responses on the two tasks were found to be uncorrelated. The usual finding that performing a secondary task leads to an underestimation of time passing was replicated, although the effect was weak. The acuity data supported all the predictions described above. However, both patient groups also showed an unexpected deficit on the pitch discrimination task. The authors suggest that "...the increased variability in frontal patients on the duration perception task could be viewed as a manifestation of a failure to fully attend to the temporal information provided by the cerebellum." This suggestion parallels the functional division made by the PGC model between basic timing operations and the memory and attentional processes required for making judgements and comparisons of stimuli. Furthermore, it is interesting to note that the only subtest of the WAIS-R on which patients differed from controls was the Digit Symbol task, which has been linked to attentional function (Lezak, 1995).
1.6 The Pacemaker-Gate-Counter (PGC) model

The PGC model is similar to scalar timing theory, with the addition of a gating mechanism under the control of short-term memory between the pacemaker and the accumulator. Therefore, it is most similar to the attentional-gate model described by Zakay & Block (1997). It consists of a pacemaker, the gating process, a short-term storage capacity, a comparison process, and a long-term reference memory of learned duration values. The relationships between these components are shown in Figure 1.1. The PGC model primarily seeks to address the prospective perceptual timing of durations equal to, or shorter than, the indifference interval. In recognition of the link between perceptual timing and temporal motor co-ordination, the pacemaker component of the model in particular is held to underlie both types of performance. Here, each of the components of the model will be described in detail, and reference will be made to relevant issues discussed earlier in the General Introduction. However, those predictions made by the model that are tested in this thesis are described in Chapter 5 (sections 5.2 and 5.3). The specific predictions are located there because they form the appropriate introductions to the experiments that test them.
Pacemaker

In its simplest conception, the pacemaker is just a unit that produces pulses at approximately fixed time intervals. The rate of pulse production may fluctuate slightly, but a constant average rate is assumed. A number of manipulations have been hypothesised to affect the average pacemaker rate, most notably high tempo sensory inputs, body temperature, and modality differences. The most detailed model of the pacemaker has been described by Triesman et al. (1990). This model has a number of interesting features, which can reconcile findings that seem to require multiple pacemakers operating at different rates with other findings that seem to suggest a single
pacemaker. According to this model, the pacemaker consists of a temporal oscillator, which transmits its output to a calibration unit. The oscillator is thought to consist of a circuit of neurones connected to each other in such a fashion that an output unit oscillates between firing and not firing at a constant frequency. It is proposed that the oscillator is unaffected by normal sensory inputs, in order that it might provide a fixed reference frequency. The calibration unit multiplies the input it receives by a calibration factor, in order to produce the pulses that constitute the final output of the pacemaker. The calibration factor varies based upon the requirements of motor programs, as well as due to inputs from salient external stimuli (e.g. click trains, or a sound associated with danger).

It is worth considering the reasons Triesman chose the two-stage model of the pacemaker described above. If motor co-ordination were to rely upon a single pacemaker, which had a fixed frequency, then this would result in the advantageous characteristic that different effectors could easily be co-ordinated with one another. However, such a system would be inflexible. It would not allow different effectors to operate at different rates, or even to operate at rates that are multiples of each other. In order to avoid this problem Triesman proposes that oscillators might be duplicated at various points within the motor control hierarchy. This would allow separate effectors to operate independently of each other. However, it introduces the problem of how to co-ordinate movements which require different effectors to operate at the same, or related rates. If all the pacemakers in the system ran at the same rate, and a central motor program continuously adjusted the target time values (numbers of pulses before a movement is initiated) for each individual pacemaker Triesman argues that an extremely inefficient system would
result. A better system would involve leaving the central instructions specifying the target time values for each pacemaker unchanged, while directly adjusting the rate of the individual pacemakers (via the calibration units) in order to change the speed of action. To use Triesman’s example, such a system would be particularly advantageous when “...the rate of performance must be continuously modulated from one speed to another, as when a runner gets up speed. Thus a distributed pacemaker system has two, apparently opposed, theoretical requirements: a stable reference frequency, and a flexible, adjustable frequency.”

In section 1.5, findings were described which suggested that the pacemaker might run slower for visual stimuli than auditory stimuli. Using Triesman’s model there is no need to posit separate pacemaker systems for the two modalities in order to explain this effect. Instead, it may be suggested that auditory stimuli cause an increase in the calibration factor relative to visual stimuli. In particular, Walker & Scott’s (1981) finding that the visual component of a compound auditory / visual stimulus is perceived to be of the same duration as the auditory component is easily accommodated. The auditory part of the compound causes an increase in the calibration factor, which would then apply to all stimuli judged (including visual ones) whilst the auditory stimulus persists. This idea might be tested by an experiment in which subjects judged visual stimuli only, while auditory stimuli were also presented for part of the target duration. If the “auditory acceleration effect” is not dependent upon attention then judgements of visual stimuli should lengthen as the duration of the irrelevant auditory stimuli increases as a proportion of the target interval. However, a control condition employing an irrelevant visual stimulus should produce no effect. The finding that visual stimuli are judged more
variably than auditory ones is not accounted for by properties of the pacemaker. Rather, it may be accounted for by memory processes (see below).

For the purposes Experiments 1 - 9 presented in this thesis, it is generally assumed that subjects' pacemakers operate at a constant rate within each experiment. If any stimulus presented should happen to influence the proposed calibration units, then it is hoped that this effect is constant within each experiment.

**Gating process**

The gating process is incorporated into the model in order to account for the effects of nontemporal processing upon temporal judgement. Zakay & Block (1997) refer to the gate as an “attentional gate”. However, this term is potentially misleading in light of the findings of Fortin & colleagues, which suggest that the gating process is associated with short-term memory. Nonetheless, the inclusion of the gate component is partly in recognition of the fact that prospective timing is a process requiring concentration on the part of the subject, and therefore it is hard to avoid the word “attention”. Perhaps this confusion arises because the concepts of attention and short-term memory are not mutually exclusive, in that one rarely operates without the other. In fact, Cowan (1995) argues that the contents of short-term memory may be defined as the subset of long term memory on which attention is currently focused.

The basic idea behind the gate is that it needs to be open for pulses to pass from the pacemaker to the accumulator in short-term memory. When resources are required by nontemporal tasks the gate is closed, and pulse accumulation is interrupted. However, the most precise quantitative evidence available concerning the gating process — that of
Fortin & colleagues, as well as that presented here — suggests that pulse accumulation is not completely halted for the duration of nontemporal processing. One feature of the gate, which is derived from the findings of Zakay (1998) and Casini & Macar (1997) who asked subjects to control the division of attention between temporal and nontemporal tasks, is that it is at least partly under conscious control.

Scalar timing theory proposes a switch between the pacemaker and the accumulator, which takes a certain amount of time to open and close. No function relating to secondary tasks has been proposed for this switch. In their attentional-gate model, Zakay & Block (1997) include both a switch and an attentional gate. However, there is no obvious reason to think that the two processes are separate. For this reason, only the gate is included in the PGC model. This position is in agreement with that of Lejeune (1998), who considers in detail the issue of whether a gate and a switch are required, and concludes that only one of the two is needed. The opening and closure of the gate must each have a mean duration and a certain amount of variability associated with them, which is assumed to add to the total variability of temporal judgement. However, the latency associated with the opening and closing of the gate is not investigated here. If the latencies associated with gate operation due to the interpolation of secondary tasks could be estimated, and these values were found to be similar to the estimates of the latencies for the switch provided by Wearden et al. (1998), then this would provide preliminary evidence supporting the assumption that the two mechanisms are one and the same. Ways of estimating this latency are discussed in Chapter 6.
Short-term memory for duration

In order to make perceptual judgements of duration the pulses from the pacemaker must be kept somewhere in some form. The component that achieves this function is usually referred to by models of time perception as a "counter" or an "accumulator". Both of these terms capture the idea that what is stored must be some nontemporal transformation of the output from the pacemaker. Values in the accumulator may be maintained once they have reached their final values, as demonstrated by the fact subjects can remember one value of duration in order to compare it with another one presented immediately after. Therefore, it is possible to speak of short-term memory for duration. In fact, because of the peculiar nature of time as a "stimulus", short-term memory is implicated even in the perception of a single interval. If pulses from the early part of the interval were not maintained in memory during the latter part then the interval would become truncated. For these reasons, the PGC model uses short-term memory for duration as a theoretically broader term than "counter" or "accumulator".

Any short-term memory system is most easily defined by its capacity limit. It is possible that the indifference interval might reflect a capacity limit in short-term memory for duration. In terms of the loss of information, it appears that interference plays a greater role than decay. This proposition is supported firstly by the lack of a decay function for interruptions in time production of up to 6 s found by Fortin & Massé (2000), and secondly by the fact substantial interference effects in short-term memory for duration were found by Deutsch (1986). These effects took the form of directional

---

2 There was no decay function in terms of judgement bias. Data concerning variability was not reported.
judgement biases, and are described in detail in sections 3.2 and 4.4.

It is possible that visual stimuli produce more variable judgements than auditory stimuli because short-term memory for duration is less efficient in the visual modality. Note that such an effect would not preclude the switch latency difference identified by Wearden et al. (1998). This could occur because there are two different memory systems, one for each modality, or alternatively there might be a single system, with access to this system being less efficient in the visual case. In the auditory modality, Ivry & Hazeltine (1995) found that for both production and perception tasks, increasing the number of presentations of the target stimulus prior to subject performance caused a reduction in variability. It is hard to see how the switch latency modality difference could account for this. This suggests that the representation of a target interval can be refined within short-term memory for duration, perhaps by an averaging process. However, it is possible that multiple presentations result in a representation being formed in long-term memory for duration. In fact, it is hard to define the boundary between the two forms of memory.

Long-term memory for duration

The various animal experiments discussed earlier demonstrate that values of duration can be represented, and associated with reward, in long-term memory. In the human case, examples from everyday life are easy to find. Upon seeing a traffic light turn to orange, a driver is able to anticipate the coming change to green. In order to make use of values of duration in long-term memory it is generally necessary to also employ the short-term prospective estimation system. In the traffic light example, the long-term
memory value would be compared with the current contents of short-term memory for duration.

Church & Gibbon (1982) propose that long-term memory for duration is based upon populations of individual traces, each of which corresponds to a single exposure to a target stimulus. The PGC model adopts this proposition because it allows the model to predict performance consistent with Weber's law. The reasons for this are discussed in section 5.2, where the interaction between various parts of the model is considered. It is also worth noting that there are possible modality differences in long-term memory for duration. In fact, if short-term memory were to differ between modalities, then these differences (in variability) would be reflected in long-term memory. An experiment to test this idea might involve a training phase using visual stimuli, and a test phase involving auditory stimuli. If variability of judgement is determined by properties of the long-term store then the variability associated with this experimental condition should be similar to that associated with a condition in which both the training and test phases use visual stimuli, and greater than the variability associated with a within modality auditory condition.

Comparison Process

One interval in the short-term memory store can be compared with a second as the pulses that define the second interval accumulate. Therefore, the comparison process is either a continuous one, or it occurs at frequent discrete intervals. Such a continuously updating process is likely to require attention, and this is perhaps the reason why attention, as well as short-term memory is needed for time estimation. Whether the
comparison process has a latency associated with it is unknown. Values in the short-term memory store may also be compared with values in the long-term store. Wearden (1991) has suggested that the decision process may differ between humans and animals.

1.7 Neuropsychological evidence

Traditional studies of temporal judgement are limited to inferring properties of underlying systems from observed behaviour. For example, it is clear from behavioural data that Weber's law applies to temporal judgement, which places an important constraint upon theoretical models of timing. Neuropsychological data is potentially extremely important, because behavioural measures are accompanied by corresponding data concerning the brain. Up until this point in time, neuropsychological data has been most useful in confirming the kind of divisions of function that are made in the PGC model, and in localising them to particular regions of the brain.

In section 1.4, neuropsychological evidence was presented which supports the idea that perception and motor co-ordination share the same pacemaker. This evidence also suggested that the pacemaker is localised in the cerebellum. In section 1.6 another study was described, which suggested that the memory and attention functions required by timing are located in the frontal lobes, as well as providing further evidence that the cerebellum is the basic source of temporal information. In this section various other studies will be described that support the position already described. Additionally, some studies will be considered that implicate the basal ganglia in timing, and which therefore suggest that multiple timing systems might operate in parallel.
**Further evidence of cerebellar timing**

Mangels, Ivry & Shimizu (1998) compared the pattern of timing deficits in a group of patients with focal, unilateral neocerebellar lesions with those of another group, who suffered from focal prefrontal lesions and age matched controls. Adaptive psychophysical procedures were used to estimate sensitivity on two temporal discrimination tasks, one centred on 400 msec and one centred on 4s. The cerebellar group was impaired at both target durations, whereas the prefrontal group only exhibited deficits at the longer target interval. However, the prefrontal patients were also the only group to exhibit deficits on a pitch discrimination task. Furthermore, the prefrontal patients benefited from strategic support provided by the Experimenters (e.g., counting strategies), whereas the cerebellar group did not. The authors interpretation of this pattern of results is that the cerebellar group suffered from a fundamental deficit in the performance of their internal clock (pacemaker), and therefore their deficits were pervasive, but limited to timing tasks. In contrast to this, the prefrontal group suffered from general deficits in attention and the executive control of action, and therefore their timing problems were only evident with the longer target interval. Underlying this reasonable interpretation is the assumption that the 4 s interval would place a greater load upon the attentional and short-term memory systems associated with timing than the 400 msec interval. These findings are supportive of the divisions of function made in the PGC model.

Tesche & Karhu (2000) studied the timing responses of the cerebellum using magnetoencephalography (MEG). This scanning technique has far better temporal
resolution than PET or fMRI, which allows the response of the cerebellum to individual stimuli to be measured. Electrical stimulation of subjects’ fingers occurred at regular intervals (0.5, 1, 2 or 4 s) in a control condition. In an experimental condition, an occasional stimulus was omitted on a random basis. In both conditions, the responses of the cerebellum as well as the somatosensory cortex were monitored. In the control condition, both regions responded to each stimulation. However, when a stimulus was omitted the somatosensory cortex did not respond, suggesting that its involvement was essentially a reactionary one. The cerebellum responded to the omitted stimulus at the point in time at which it would normally have occurred. In fact, the response of the cerebellum slightly anticipated the expected time of the stimulus. An important feature of this demonstration of the anticipatory timing ability of the cerebellum is that no motor response was made during the experiment, and therefore the cerebellar involvement cannot be due to its known role in motor co-ordination. In one condition subjects were asked to read a book, which meant they were deploying less attention to the electrical stimulations. This manipulation reduced the magnitude of the cerebellar response, indicating a possible role for the cerebellum in the attentional effects on timing described in section 1.5. In fact, the cerebellum may have a wider role in attention and the temporal co-ordination of mental events, analogous to its role in the temporal co-ordination of motor responses (Ivry, 2000). Ivry describes evidence that patients with acquired cerebellar disorders perform poorly on problem solving and verbal retrieval, as well as observing that neuroimaging studies have consistently found task-related activation in the cerebellum despite careful matching of the motor requirements of the control and experimental tasks.
Jeuptner, Rijntjes, Weiller, Faiss, Timmann, Meuller & Diener (1995) performed a PET study of subjects performing a two-alternative forced choice judgement, in which they had to indicate if a second interval was longer or shorter than a comparison interval. The target interval was 300 msec long, and the comparison intervals were either 200 or 400 msec long. Activation was observed bilaterally in the cerebellum, as well as in the basal ganglia, temporal lobe, prefrontal cortex, and cingulate cortex. Hazeltine, Helmut & Ivry (1997) criticise this study because the control condition simply involved the consecutive presentation of two identical intervals, and therefore the observed activations may have occurred because of the general requirements of making a sensory judgement. Hazeltine et al.'s caution proved to be well founded as a subsequent PET study using another sensory judgement task as a control condition showed a different pattern of results. Maquet, Lejeune, Pouthas, Bonnet, Casini, Macar, Timsit-Berthier, Vidal, Farrara, Degueldre, Quaglia, Delfiore, Luxen, Woods, Mazziotta, & Comar (1996) used the generalisation procedure described earlier with a light intensity judgement task as well as a visual short duration task. No significant differences in activation were found between the two tasks. These results may simply be indicative of the wider role of the cerebellum in cognition alluded to earlier, whereas the work of Tesche & Karhu (2000) demonstrates unequivocally the role of the cerebellum in timing.

Temporal processing in the basal ganglia

Demonstrations of the timing capacity of the basal ganglia seem to parallel those of the cerebellum, providing a problem for models of timing. Before considering how the
two sets of findings might relate to each other the evidence relating to the basal ganglia will be briefly described.

Meck (1996) reviews a number of pharmacological and lesion studies with rats using the peak procedure with intervals of around 30 s. Drugs that affect the dopaminergic and cholinergic systems produce substantial shifts in the position of the peak in response rate. It is argued based on the time course of the drug effects that the two classes of drugs produce these effects via different mechanisms. Dopamine effecting drugs produce initially large effects, which decline over time. During the peak procedure, the majority of trials are reinforced, and non-reinforced trials are used to collect data. The reinforced trials allow the rat time to adjust its responses to the effects of the drug, in order to receive food. Therefore, the effects of the dopamine influencing drugs are consistent with an acceleration of clock speed, which allows the animal to adjust its long-term memory of the target interval. In contrast to this pattern, the drugs that affect the cholinergic system produce effects on timing that develop slowly and are long lasting. It is hypothesised that these drugs effect the rat's long-term memory of the target interval. Meck (1996) describes a model in which a dopamine-dependant pacemaker-counter system is located in the basal ganglia. This system is complimented by an acetylcholine-dependant temporal memory and attention system located in the frontal cortex. Meck presents evidence of further dissociation of function within the pacemaker-counter system. Lesions of the dorsal striatum prevent the learning of intervals of 60 s, whereas lesions to the substantia nigra impair performance with intervals between 20 s and 60 s. Meck hypothesised that the output from the substantia nigra is integrated in a dorsal striatum based counter for longer intervals.
Based upon the above evidence it would be easy to conclude that the cerebellar timing system has primacy for intervals of a few seconds, while the basal ganglia system dominates for longer intervals. Such a position is suggested by Wittman (1999). However, there is also evidence from studies of Parkinson’s disease that the basal ganglia system operates for short time intervals. Harrington, Haaland, & Hermanowicz (1998) tested nondemented Parkinson’s patients and a control group on a paced finger tapping task with two different intervals, and a duration perception task with the same two target intervals (300 msec or 600 msec). An adaptive testing procedure was used to estimate bias and variability in the perceptual task. They also included a pitch perception task to control for auditory processing deficits and a rapid tapping task to control for limitations of movement rate. The Wing & Kristofferson procedure was used to partition variance on the tapping task into motor and timing components. Medicated patients were used, as medication has previously been found to return the motor variance of Parkinson’s patients to levels similar to that of controls. The Parkinson’s group proved to be impaired relative to controls on the timing component of the tapping task, as well as more variable in the perception task. The deficit was not general in nature, as demonstrated by the lack of a difference between the patients and controls on the analogous pitch task.

The research conducted by Harrington et al. (1998) with Parkinson’s patients is very similar in method and results to that of Ivry & Keele (1989) and Ivry et al. (1988) with cerebellar patients, which was discussed in section 1.4. Because the intervals used were similar, the suggestion that the two systems operate for target intervals in different ranges is not plausible. Furthermore, both systems seem to be involved in motor timing and perceptual timing. Therefore, the possibility must be raised that the two systems
somehow operate together. The nature of the interaction between the two systems is an important focus for future research.

**Modality differences**

Modality differences in behavioural data were discussed in section 1.5. In section 1.4 a scanning study conducted by Schubutz et al. (2000) was described. The main focus of this research was to show that perceptual timing tasks produce activations in the “motor circuit”. However, the rhythm perception task used by Schubotz et al. (2000) was performed in both the auditory and visual modalities, and there were no differences in foci of activation between the two tasks. The behavioural task was atypical, and unique in the timing literature. A different result was obtained by Jäncke, Loose, Lutz, Specht & Schah (2000) using the well understood paced finger tapping task. Subjects performed the tapping task with either auditory or visual pacing stimuli, and fMRI was used to measure activations during the continuation phase. Behavioural data indicated that tapping was more variable when pacing stimuli were visual. During both continuation conditions, a fronto-parietal network was active. However, other brain areas involved differed between the two modalities. It was concluded that PFT in the context of auditory stimuli relies more upon brain areas involved in internal motor control, while PFT in the context of visual stimuli relies more on brain areas subserving the processing and imagination of visual stimuli. The cerebellum was involved in both conditions, though the foci of activations within the cerebellum differed between conditions. The reason for the differing conclusions of the two studies regarding modality differences probably arises from the differences in the tasks used. As the task used by Jancke et al. (2000) is
one known to be a good index of timing ability in the normal population, as well as with patient groups, it seems reasonable to favour their conclusions.

Other findings

Harrington, Haaland & Knight (1998) examined the role of the cortex in timing, using similar methods to those used by Harrington et al. (1998) with Parkinson's patients. Patients had either focal right hemisphere damage (RHD) or focal left hemisphere damage (LHD). A control group was also used for each of the patient groups, which was matched to the patient groups in terms of age, education, and sex. The tasks were duration perception with short auditory intervals, using an adaptive procedure to measure thresholds, pitch perception, also using an adaptive procedure, and a nonspatial attention task. This latter task was analogous to the Stroop task, in that a cue presented before a reaction time trial could be neutral, valid, or invalid, and a measure of “cost” was derived by calculating the difference between the valid cue trials and the invalid cue trials. The results of the pitch discrimination task indicated that both patient groups were impaired relative to their control groups. This suggested that any deficits on the duration perception task might be due to basic problems in processing stimuli or making decisions, rather than specific timing deficits. For this reason, analysis of the duration perception task was performed twice: once with all patients included, and once with only those patients included who scored within 1 standard deviation of the control group on the frequency perception task. There were no differences between the groups on the duration task in terms of the point of subjective equality. However, differences in difference thresholds (variability) were found. Including all the participants, both the RHD and the
LHD groups were impaired relative to their respective control groups, but the difference was much greater in the RHD group. Analysis of the restricted sample revealed that only the RHD group was impaired relative to its control group. Both groups were impaired on the nonspatial attention task relative to their controls. Perhaps the most interesting findings emerged from the correlations between performance on the different tasks. Firstly, sensitivity in the frequency task was positively correlated with sensitivity in the duration task, but this correlation was much stronger in the LHD group. This suggested that the more limited deficits of the LHD group on the duration task were likely to be due to nontemporal aspects of the task. Importantly, the measure of cost in the attention task was not correlated with performance on the pitch task. However, it was correlated with performance on the duration task, but only for the RHD group. Taken together, these data suggest that the specific attentional processes required by temporal judgement are right hemisphere functions. Harrington et al. (1998) went on to relate lesion location to timing performance. RHD patients with impaired timing had damage to premotor cortex, prefrontal cortex, or inferior parietal cortex. These areas were typically not damaged in RHD patients with unimpaired timing. However, all of the unimpaired LHD patients did have damage to these areas. The authors conclude that a right hemisphere prefrontal-inferior parietal network is implicated in timing.

Nichelli, Venneri, Molinari, Tavani, & Grafman (1993) examined time estimation in patients with memory disorders. Specifically, amnesiacs and Alzheimer’s disease patients were compared to controls. Unlike in the two Harrington et al. (1998) studies, there were no tasks included to control for nontemporal aspects of the timing tasks, which involved repeated reproduction of a 1 s interval and verbal estimation of longer intervals.
with a digit-shadowing task designed to suppress counting. The amnesic patients performed normally on the short interval task in terms of accuracy (bias) and variability. However, they were impaired in terms of their accuracy with the long intervals, most of them showing a strong tendency to underestimate. The Alzheimer's disease patients were impaired in terms of variability on both tasks. The authors interpret their results within the framework of scalar timing, suggesting that the amnesic patients were only impaired when the storage demands of the task exceeded their short-term memory range, and that the Alzheimer's disease patients were impaired in terms of clock, memory and decision mechanisms.

Casini & Macar (1996) utilised slow cortical potential recordings from frontal areas of the brain to support the hypothesis that attentional resources are required for duration processing. Casini & Macar (1996) begin with three observations. Firstly, that a wide range of converging evidence suggests that the prefrontal cortex is involved in tasks that require sustained attention. Secondly, that sustained attention is often implicated in temporal processing, especially in the context of the effects of secondary tasks upon temporal processing. Thirdly, that there is a conceptual problem in the suggesting that attention is directed towards time; namely that time as a stimulus does not exist in the way that for example, intensity or amplitude do, and so what can be the focus of attention? In order to address this question Casini & Macar (1996) aim to show that the prefrontal activation associated with a temporal task are similar to those associated with a more traditional attention demanding nontemporal task. The authors point out that a positive result would demonstrate not that the prefrontal cortex is directly concerned with processing time (in the way that the Tesche & Karhu (2000) study demonstrates directly
cerebellar involvement in representing time in the brain), but that attention is required for some aspects of the process to occur. The temporal task used involved viewing a visually presented row of letters, which lasted for either 3 or 4 s, and then when a second row of letters appeared, subjects had to switch them off when they thought the elapsed interval was equal to the duration of the first row of letters. This was effectively a standard reproduction paradigm, and subjects were asked not to count. The other task used was an anagram task. Because a subtractive methodology was employed, a control condition was employed, in which subjects simply read the word “rien” (French for “nothing”). Recordings were only taken from frontal areas, and so the question of which brain areas other than frontal ones are involved in timing was not addressed. The data supported the primary hypothesis that prefrontal cortex would be involved in both tasks, and therefore support the idea that attention may be directed towards time, as well as more substantial stimulus attributes. There were two intriguing additional findings. Firstly, the accuracy of performance was predicted by features of the contingent negative variation that appeared during the presentation of the standard. This suggests that the encoding phase of a reproduction task is more critical in determining performance than the reproduction phase, and that encoding requires attention. Secondly, the level of activity in the prefrontal cortex was greater when performance was inaccurate. This counter intuitive finding suggests that the relationship between level of attention, efficiency of performance, and level of prefrontal activity is not a simple one.

Finally, a single unit study of cells in the prefrontal cortex of monkeys attempted to directly implicate the prefrontal cortex in the processing of duration (Niki & Watanabe, 1979). Monkeys were trained using a differential reinforcement of long
latencies paradigm with a target interval of 2 s. This is similar to the peak procedure, described earlier, but without the trials on which reward is withheld in order to observe the full distribution of responses around the target interval. Therefore, the procedure simply involves rewarding the animal if it waits 2 s or more to release a lever after a start signal. Certain cells were found whose firing activity changed 1 to 1.5 s before the lever was released, and then ceased immediately. As in Casini & Macar (1996) the accuracy of performance was predictable from the early part of a trial, as firing changes started later on trials when the lever was released later. Other cells were identified that responded to the presence of the stimuli, the expectation of reward, and the motor response. Therefore, the authors argued that the activity of the first class of cells reflected time estimation processes rather than other requirements of the task.

Conclusions

Neuropsychological evidence has added a new dimension to the investigation of temporal judgement. So far, a wide network of brain areas has been implicated, including the cerebellum, basal ganglia, prefrontal cortex, and the motor areas and other parts of the right hemisphere. Neuropsychological evidence has not contradicted behavioural models, such as scalar timing theory, which can often accommodate neuropsychological evidence. It seems that the pacemaker function is subserved by either the cerebellum, the basal ganglia, or both — this issue remains to be resolved. Short-term memory, attentional, and long-term storage functions required by timing are more likely to be attributable to the cortex.
1.8 Questions addressed in this thesis

The General Introduction has provided a general (if selective) overview of the diverse literature relating to the prospective judgement of time. The findings described have been broadly consistent with the PGC model of timing presented in section 1.6. The aim of the empirical chapters is to explore several specific questions within the framework of the PGC model. The experiments presented in Chapter 2 aim to establish a paradigm appropriate for the testing of predictions relating to the effects of secondary tasks upon temporal judgement. The paradigm is that of Fortin & colleagues, however, the choice of secondary tasks differs in order to add generality to the findings of Fortin & colleagues. The second experiment of Chapter 2 produced an unexpected result, namely that a concurrent pitch memory load caused systematic interference with time production. This result stood in contrast to the negative result with a digit memory load obtained by Fortin & Breton (1995). Chapter 3 explores this effect further, by asking whether it was due to the memory load itself, or some other aspect of the pitch task. The issue of whether maintaining a pitch memory load requires attention is also addressed, in an attempt to find an explanation of the effect within the framework of the PGC model. In Chapter 4, the possibility that other types of memoranda also produce a lengthening of time production similar to that found with pitch is investigated. To anticipate the findings, negative results were obtained with colour and timbre; however, an effect was obtained with a concurrent memory load consisting of values of duration. The type of effect found for the duration memory load allows some suggestions to be made about the properties of the short-term memory component of the PGC model. In Chapter 5 specific
predictions made by the PGC model due to the interaction of the pacemaker with the gating process are investigated. Firstly, the possible effects of secondary tasks on the variability of time production are addressed. Secondly, the possibility of an interaction between the shortening of time production caused when the pacemaker is accelerated (using click trains) and the lengthening of time production caused by secondary tasks is investigated.
2.1 Aims of Chapter

The aim of the two experiments that are reported in this chapter was to establish a paradigm appropriate for the testing of the predictions made by the PGC model. The dual task paradigm used by Fortin et al. (1993) and Fortin & Breton (1995) was chosen for this purpose because it allows precise control of the processing demands made by a secondary task concurrent with timing. The original experiments reported by Fortin & colleagues used the Sternberg digit memory-search as a secondary task. Here, this task is replaced by a pitch memory-search. The aim of doing so was to support the conclusions made by Fortin & colleagues concerning the involvement of short-term memory in time judgement, by demonstrating a similar pattern of interference with timing using a different short-term memory task. Literature concerning processes involved in memory-search, as well as literature concerning short-term memory for pitch, is briefly reviewed. That part of the design and procedure which is common to the experiments presented in later chapters is described in detail, as are certain aspects of data analysis which apply generally to all the time production experiments in the thesis.

2.2 Introduction

Numerous studies have shown that increasing concurrent secondary task difficulty produces an increase in the amount by which temporal judgements are biased (e.g., Hicks et al., 1976; Vroon, 1970; Thomas & Cantor, 1978; Zakay, 1993). The secondary tasks used have varied considerably, including card sorting, tone discrimination, and the Stroop
task. These tasks are often complex, having varying perceptual, central, and motor processing requirements. Because of this complexity, it has been hard to specify exactly which features of the secondary tasks produce interference with timing. Unsurprisingly, given the wide range of interfering secondary tasks, authors such as Zakay (1993) have argued that interference with timing is modulated by the general difficulty level of a secondary task.

Fortin et al. (1993) made the methodological innovation of attempting to use secondary tasks where only one aspect of processing difficulty varied in a particular experiment. This resulted in the critical finding that in some cases variation in secondary task difficulty produces no impact upon time judgement bias, suggesting that the general difficulty account of secondary task interference with timing is wrong. The paradigm used involved training subjects to press two keys, separated by their subjective estimate of a 2 s interval. This was referred to as a time production trial. Feedback was provided on the accuracy of performance, and sufficient trials were provided for performance to stabilise. In an experimental trial, the Sternberg (1966) memory-search task was performed concurrently with time production. First, a memory set of between one and six digits was presented serially. Then the subject initiated time production. The probe item was presented 0.5 s after the start of time production. Time production was terminated by the recognition response. Time production increased linearly, as a function of memory set size, on average by 24 msec per item beyond the first. The reason the increase was linear, and had the magnitude it did, is related to properties of the Sternberg task. The Sternberg task is discussed in Section 2.2.3.
The learned time production with concurrent memory-search paradigm has a number of attractive features. Firstly, the perceptual processing requirements concurrent with timing remain constant and minimal as search difficulty is increased. Secondly, the motor requirements of both tasks were combined into the same response. Therefore, the motor requirements of the secondary task also remain constant as secondary task difficulty increases. This means that the increase in time production with set size can be clearly attributed to the central processing requirements of the secondary task. Fortin et al. (1993) argue that the critical feature of the Sternberg task that leads to interference with timing is the need to make mental comparisons within short-term memory.

The Fortin paradigm is easily extended to other forms of decision task where set size can be manipulated, and this is precisely what Fortin et al. (1993) proceeded to do. This allows them to test their theory that secondary task interference with timing occurs because of the short-term memory demands of such tasks by systematically reducing those demands, while keeping task difficulty constant. In a second experiment, a memory set size of one was used throughout, and the number of items in the display varied between one and five. This reduced short-term memory load, and introduced a visual search component. Under these circumstances, the slope of time production with set size was reduced to 11 msec (from 24 msec in the first experiment). In a third experiment, the short-term memory requirements of the secondary task were reduced still further. This was achieved by using feature absent visual search as a secondary task. Triesman & Souther (1985) found that in a visual display where distractors possessed a particular feature and this feature was absent from targets, reaction time increased as function of the number of distractors. Because the features that define a target are constant from trial to
trial, the short-term memory requirements of this task do not increase as the number of distractors increases. When this task was combined with time production, increasing the number of distractors had no effect at all upon time production. This supports the argument that the amount of interference with time production a secondary task causes is determined by the short-term memory demands of that secondary task. The Fortin paradigm was adopted for the present research because of the opportunity it gives to systematically vary the processing demands of a secondary task, both within a particular task through manipulations of set size, and between processing resources by changing the task.

2.2.1 Fortin et al. (1993) in the context of the PGC model

The PGC model proposes that during time production training a population of memory traces is accumulated, and that this population is the representation of the target interval in long-term memory. Each memory trace is the counter value at which a response was made on a particular trial. In an experimental trial, a single trace is drawn from the population at random. This provides a target value for the accumulation process that is supported by the short-term memory for duration module. Time production begins with a motor response (in the form of a button press). Once the button press is complete, pulses from the pacemaker begin to accumulate in the counter. Once the number of pulses accumulated is equal to the target value minus a small quantity to allow for the latency of the terminating button press, time production is terminated. Clearly, part of the duration of time production, as well as part of its variability, may be attributed to motor
responses. However, the motor response is kept as constant as possible throughout all the experiments reported in this thesis, insuring comparability between experiments in terms of the non-motor (timing) aspects of the task. When time production is lengthened by the requirement to perform a concurrent secondary task, this is because pulse accumulation has been temporarily interrupted or slowed down. This is represented in the model by the gate component. The findings of Fortin et al. (1993) suggest that in order to keep the gate “open” short-term memory resources are required. Actively attending to time passing is an ongoing short-term memory task. An active visual search has no impact upon concurrent timing because it does not place a demand upon short-term memory resources.

The theoretical importance of replicating Fortin et al. (1993)

Fortin et al. (1993) provide the first published demonstration of a secondary task where varying difficulty produces no impact upon prospective temporal judgement. They also make the claim that those tasks that do produce interference are ones where short-term memory requirements vary on a trial to trial basis. Conversely, it is claimed that non-interfering tasks will be ones where the short-term memory requirements are minimal and constant from trial to trial (e.g., feature absent visual search). The case that the defining feature of tasks that will produce interference with timing is the short-term memory requirements of those tasks rests, in part, on the demonstration that the Sternberg task systematically interferes with timing. However, the Sternberg task may not be taken as completely representing the domain of short-term memory. It may, on the other hand, be taken as a representative phonological short-term memory task. It is shown in section 2.2.2 that there are functional dissociations within short-term memory, which reflect the
different kind of materials that may be remembered. Given that not all secondary tasks interfere with timing, it is logically possible that not all short-term memory tasks will interfere with timing. The aim of Experiment 1 was to replicate the interference result of Fortin et al. (1993) using a short-term memory task that required demonstrably different resources to the phonological resources required by the Sternberg task. A successful replication would increase the likelihood that Fortin et al.’s (1993) conclusion concerning the critical involvement of short-term memory resources in timing was correct. The memory task chosen for this purpose was recognition memory for pitch.

2.2.2 Short-term memory for pitch

Cruse & Jones (1976) report a pitch analogue of the Sternberg task, and this was adopted here. The stimuli consisted of eight pure tones, spaced out equally on an exponential scale of frequency, none of which corresponded to notes on the western musical scale. One interesting feature of pitch memory tasks, which stands in contrast to the case of phonological memory, is the ineffectiveness of rehearsal. Attempting to rehearse absolute pitch information results in no benefit to recognition performance (Massaro, 1970; Wickelgren, 1969; Peynircoglu & Palmer, 1992).

Apart from the lack of benefit from rehearsal, a number of other lines of evidence suggest a functional separation between pitch memory and phonological memory. For example, Deutsch (1970) showed that recognition memory for a standard tone could be

---

1 In the case of words, rehearsal benefits recognition less than it does recall (Groeger, 1999; Cowan, 1997). Recall tasks are not possible with pitch, unless the subjects possess perfect pitch. This suggests that accurate rehearsal is impossible in principle for most people.
degraded by subsequent sounds, this effect being strongest for other tones, and extremely weak for speech. The procedure involved presentation of a standard tone followed 5 s later by a test tone, which could be identical or differ by a semitone. Subjects made a same/different judgement. If the 5 s interval was empty then performance was perfect. If it was filled with spoken numbers, which subjects were instructed to ignore, the error rate rose to 2.4%. Asking subjects to recall the numbers caused the error rate to rise to 5.6%, while making the tone judgement had no effect on the recall of numbers. Filling the interval with tones, which were to be ignored, caused the error rate to rise to 32.3%. Further research (Deutsch, 1972) showed a quantitative relationship between memory interference and the similarity of the standard tone and intervening tones. The effect of an intervening tone upon recognition performance was also shown to depend upon its position within the sequence of intervening tones (Deutsch, 1975), the effect being greater if a tone was closer in the sequence to the standard. Additionally, the effect of an interfering tone on recognition memory could itself be modulated by the pitch relationship of that tone with a third tone (Deutsch & Feroe, 1975). Deutsch proposes a model to accommodate these findings, in which pitch information is stored in a tonotopically organised array. Lateral inhibition takes place between elements of the array, the amount of inhibition being a combined function of pitch distance and temporal distance.

Deutsch’s evidence for a separate pitch memory store has been complimented by recent research. Pechmann & Mohr (1992) contrasted the performance of musicians and non-musicians in the experimental conditions of Deutsch (1970), as well as an interpolated visual material condition. Non-musicians showed some interference from
non-tonal material when that material had to be processed and massive interference from
tonal material. Musicians only showed interference from tonal material, and the
magnitude of interference was almost identical to that of the non-musicians. Therefore,
the musicians expertise manifested itself in immunity from the attentional demands of
secondary tasks only. The authors argue that because the tonal memory store appears to
be unaffected by expertise, this suggests that it is a relatively low level functional
module\(^2\).

Other researchers have shown that within non-verbal acoustic memory, pitch
appears to be a distinct module. In particular, Semal & Demany (1991) have shown that
varying the timbre (spectral composition) of intervening material has no impact upon
pitch recognition judgements. Whereas, Starr & Pitt (1997) have shown that recognition
memory for timbre was sensitive to interference from the timbre of intervening stimuli,
and that varying the pitch of the intervening stimuli had no impact upon timbre
recognition. These findings may be taken to support a functional dissociation of pitch
from timbre in auditory short-term memory. Anourava, Rämä, Alho, Koivusalo, Kahnari
& Carlson (1999) used n-back memory tasks for auditory spatial location and pitch. Both
tasks employed three tones, that varied in pitch and location (which was simulated using
interaural intensity differences). Thus, subjects could be asked to perform either kind of
task with the same stimuli. During the 3125 msec interstimulus interval of the n-back task
distractor stimuli were presented. These consisted of a pair of tones, which varied in

\(^2\) Pechmann & Mohr's (1992) research does not address the issue of how exactly the expertise of musicians
would be expected to manifest itself. It might be the case that asking musicians to perform pitch
recognition is analogous to asking chess players to recall arbitrary lay outs of chess pieces — a situation in
which they perform at the same level as novices (de Groot, 1965, Chase & Simon, 1973). It is possible that
musicians expertise would be more evident with respect to memory for melody.
either pitch or simulated spatial location. Distractor tones varying in pitch were shown to interfere with the pitch n-back task only, whereas the spatial distractors only interfered with the spatial n-back task. The authors interpret this as indicative of a dissociation between auditory memory for location and pitch. Finally, Semal & Demany (1993) used the Deutsch paradigm to establish that varying the amplitude of individual interpolated tones had no effect upon pitch recognition.

An alternative perspective

Jones, Macken & Harries (1997) examined the possibility that the interference in memory for the test tone observed in the Deutsch paradigm is not caused by the similarity of the test tone and the interpolated material. Instead, they propose that the cause is the extent to which the interpolated material is perceptually integrated with the test tone. If the test tone is perceptually similar to, and therefore naturally integrated with, the interpolated material, as in the case of interpolated tones in Deutsch (1970), then it will be hard to distinguish from that interpolated material in memory. However, if the test tone is not perceptually grouped with the interpolated material, as is the case for the spoken digits in Deutsch (1970), then it will stand apart from that interpolated material in memory. Under the latter circumstances recognition performance would be predicted to be superior. The evidence presented in favour of this alternative view will now be considered.

Jones et al. (1997) introduced a number of additional factors into the Deutsch paradigm, each of which was thought to encourage the grouping of the test item into a distinct perceptual stream from the interpolated items. Firstly, the interval between the
test tone and the first item of the interpolated sequence was increased. This resulted in a reduction of the error rate compared to the standard interpolated tones condition. However, this result would also be predicted by the Deutsch model, which states that amount of interference of one tone on another is determined by their temporal distance, as well as their pitch distance.

Jones et al. (1997) also showed that if the pitch of spoken speech varies in a similar way to that of the interpolated tones, it too causes interference with pitch recognition. This merely confirms that the pitch characteristics of speech are not retained once it has been transformed into an abstract phonological code (see Frankish, 1996). Rather, the pitch characteristics would be retained in the pitch memory store. This is a confirmation of the results of Semal, Demany, Ueda, Halle (1995), who performed a similar experiment. Subjectively, the effect of a series of spoken digits, each of which is spoken at a different pitch, is of a number of different speakers. In a conversation, the pitch memory system contributes to our ability to know who is saying what, and to link it with their previous contributions to the conversation, whereas phonological memory is concerned with language, and by extension, the information content of conversation.

In another experiment, Jones et al. (1997) almost double the number of interpolated tones, while reducing the duration of individual tones so as to keep the overall length of the interpolated sequence the same as before. They predict that the temporal cue introduced by this manipulation will result in the segregation of the interpolated material from the standard, leading to an improvement in recognition of the standard. A simplistic interpretation of the memory interference account predicts that recognition performance should be worse under these circumstances, because there are
more interfering tones. Results indicated that recognition performance did indeed improve, in line with the perceptual grouping hypothesis. In fact, the interference hypothesis also predicts this improvement. Firstly, the amount of interference between two tones is critically dependent upon their temporal relationship, and the amount of interference falls off rapidly as temporal distance increases. Therefore, only a portion of the Jones et al. (1997) interpolated sequence was ever likely to have much of an effect upon the standard. Increasing the number of interpolated tones in the non-critical portion of the interpolated sequence should have only minimal effect. However, within the critical part of the interpolated sequence, there were still more tones, suggesting the possibility of more interference. This possibility seem unlikely in light of Deutsch & Feroe (1975), who showed that the effect of one tone in the interpolated sequence upon the standard was itself modulated by other tones within the sequence. With such a large number of tones, it is likely that in the Jones et al. (1997) experiment the effects of individual tones upon each other were cancelling out the effects of individual tones upon the standard. Therefore, it is far from clear that the interference account would predict an increased error rate when the number of interpolated tones is increased concurrently with a shortening of tone length.

There are two general points to be made about the perceptual grouping account. Firstly, it seeks to explain the general interference effects between categories of stimuli observed by Deutsch and others, but it does not offer an account of the specific interactions within the pitch memory system detailed by Deutsch. For example, Deutsch (1975) found that repeating the test tone within the interpolated sequence resulted in improved recognition performance, and that the amount of improvement was dependant
upon the position of the repeated tone within the sequence. Because repeating the test tone makes the interpolated sequence more similar to the test tone, it should result in a greater tendency for the perceptual grouping of the test tone and the interpolated sequence. The grouping account predicts that this should result in less distinctiveness in memory for the test tone, and therefore worse recognition performance, this being the opposite of the obtained result. Secondly, Deutsch selected her subjects on the basis that they could perform the task at 100% accuracy when there were no interpolated materials. Jones et al. (1997) used a criterion of 70% correct. Therefore, generally higher error rates would be expected in the Jones et al. (1997) work. More importantly, patterns of interference have been shown to interact in some respects with ability levels by Pechmann & Mohr (1992). This makes direct comparison of the two sets of findings difficult.

In conclusion, sufficient doubts have been raised about the work of Jones et al. (1997), that Deutsch's model of pitch memory, and the separation of pitch from phonological memory, may be adopted as a working assumption by the current work. As discussed above, it is essential to the logic informing the choice of secondary tasks here, and in later chapters.

Neuropsychological evidence

There are several reported cases of amusia without aphasia in the neuropsychological literature (Peretz, Kolinsky, Tramo, Labreque, Hublet, Demeurisse, & Belleville, 1994; Peretz, Belleville, & Fontaine, 1997). In amusia without aphasia
language processing and the processing of environmental sounds are preserved, while perception and memory for basic aspects of music such as pitch and melody are impaired.

Zatorre & Samson (1991) compared the performance of three lobectomy groups and a control group on Deutsch’s task. A left frontal group showed no impairment, whereas right frontal and right temporal groups were impaired. A PET study designed to reveal brain areas involved in comparing the pitch of two tones was carried out by Zatorre, Evans, Meyer, & Gjedde (1992). Two foci of activation were observed in the right ventrolateral frontal cortex. These were within the pars opercularis and the pars triangularis. Perry, Petrides, Zatorre, & Evans (1994) carried out a PET study designed to reveal brain areas involved in the retention and rehearsal of pitch information. Bilateral activation of the auditory cortex was found, but this activation was most extensive in the right superior temporal plane posterior to Heschl’s gyrus. For an extensive review of the neuropsychological evidence concerning the processing of pitch, and other aspects of music, see Oscar, Marin & Perry (1999).

The strong right hemisphere bias found for pitch memory stands in contrast to the left hemisphere bias for phonological material. For example, Awh, Smith, & Jonides (1995) report PET evidence that the storage and rehearsal of phonological material activated mainly left hemisphere brain areas. In particular, Brodmann’s area and area 40 in parietal cortex were implicated in storage, while left hemisphere areas involved in rehearsal were Broca’s area, the premotor area, and the supplementary motor area.

---

3 It is logically possible that these activations were associated with the need to make a comparison rather than being specific to the processing of pitch. A control condition might have involved the comparison of values of luminance, or loudness.
Similar findings were obtained by Awh, Jonides, Smith, Schumacher, Koepppe, & Katz (1996), again using PET.

Summary

The behavioural evidence provided by Deutsch and others suggests that the processing of pitch information in short-term memory is carried out by a different system from the one involved in the processing of phonological material. Jones et al. (1997) argue that Deutsch’s findings were an artefact of perceptual streaming. However, the evidence presented for this is at best equivocal. Given that the available neuropsychological evidence points to a physical separation between the two kinds of short-term memory, it seems reasonable to conclude that the dissociation is real. Therefore, generalising Fortin et al. (1993)’s finding, that memory-search with phonological material systematically affects temporal judgement, to a pitch memory-search task, will strengthen the case for the critical involvement of short-term memory resources in temporal judgement.

2.2.3 The memory-search process

Fortin et al. (1993) used the Sternberg (1966) paradigm as a secondary task. On each trial, this involved visually presenting subjects with between one and six digits in a serial fashion. This memory set was randomly selected from the set of digits, nought to nine. A probe item was then presented, and a recognition judgement required. On fifty percent of trials, the probe item was the same as one of the memory-set items. Such
positive probes were randomly selected from the memory-set. Negative probes were randomly selected from those digits not part of the memory set on a particular trial.

Combining the Sternberg task with time production resulted in a linear relationship between memory-set size and time production. Each additional item lengthened time production by 24 msec on average. Performing the memory-search in a reaction time condition, without concurrent timing, resulted in a slope of 36 msec, a result that is very close to that reported by Sternberg (1966). The reason the slope is less for concurrent time production is discussed in section 2.3.3. Often, it is also found that negative probes are responded to slightly slower than positive probes, although the two slopes are very similar. Error rates are typically extremely low, although they may rise slightly as set size increases. Two main features of the memory-search task are discussed here. Firstly, the reason the function is similar for positive and negative probes is considered, and secondly, the reason the function of reaction time with set size is usually linear is discussed.

One obvious process model of performance on the Sternberg task would suggest that the probe item is compared with each member of the memory set in turn, and that in the case of positive probes, the search process terminates when a match is made. This model would predict, incorrectly, a set size with reaction time slope for positive probes equal to half the slope for negative probes. Therefore, the first thing to note about this kind of memory-search task is that the search process is exhaustive, rather than self-terminating. Sternberg (1966) contrasts a serial exhaustive scanning process with a parallel exhaustive process, and concludes that only the serial exhaustive process can fit the data. The serial exhaustive search hypothesis supposes that each memory item is
compared with the probe in turn, and that the average time taken to compare the probe with one item is constant. This results in the linear property of the functions. The intercept of the function will then represent the sum of perceptual processes, motor processes, and any other fixed processes involved in the task. The slope represents the average time taken to compare each item with the probe.

Sternberg (1966) claims that this serial scanning process is a basic cognitive operation. This view is supported by the fact that when the memory sets at each set size are fixed, the slope of the function is the same as when the memory-sets vary. This means that the scanning process is unaffected by practice, or by the amount of time the items have been in memory for.

Another major model of memory-search data is the “trace-strength discrimination” model (Baddeley & Ecob, 1973). This is based upon the observation that in the procedure where the memory-set varies from trial to trial, a recency effect may be observed. This takes the form of faster reaction times to items in the final positions of the memory set. It was also found that reaction time is less to a positive probe if the probed item was repeated during the presentation of the memory set. Both of these observations imply violations of the basic premise of the serial scanning model — that each item is compared with the probe in turn, and that this takes about the same amount of time for each item. Baddeley & Ecob (1973) propose that there is no search process. Instead, members of the positive set have a fixed capacity pool of activation divided between them. As the number of members of the positive set increases, there is less activation available to each member of the set. When the probe item (e.g., “5”) is presented, a binary decision can then be based upon the level of activation of “5” in memory. The less
activation the probed item has, the harder this level is to discriminate from baseline levels, and so the time taken to make the discrimination increases. A repeated item has a relatively higher level of activation, and is therefore easier to discriminate from the baseline. For the same reason, more recently presented items are responded to more quickly.

The "trace-discrimination model" is attractive, but it too is not without its problems. If some of the pool of activation is required for an additional memory load, which is itself not relevant to the search task, then this should affect the reaction time with set size function. This is because the irrelevant memory load will require some of the limited pool of activation. However, Darley et al. (1972) demonstrated that an irrelevant memory load did not affect the slope of the memory-search function. The trace-strength discrimination model attempts to explain serial position effects through the matching process. However, Jensen & Lisman (1998) point out that, in studies such as these, serial position effects only occur when the retention interval between the last memory-set item and the probe is less than 1.5 s. They propose that this is the case because perceptual processes involved in identifying the probe item take less time if the same item has been viewed recently — a form of repetition priming. Therefore, serial position effects are not necessarily a product of the search process itself. Jensen & Lisman (1998) provide a neurophysiological model of how the serial scanning might occur. The model proposes that theta oscillations, in the range of 4-10 Hz, and gamma oscillations, in the range of 20-80 Hz, underlie performance on memory-search tasks.
Both these components are observed in the EEG record during memory scanning, and may be localised to the prefrontal cortex\textsuperscript{4}.

\textit{Implications for the use of memory search as a secondary task}

The fact that a task known to produce a linear reaction time function also had a linear effect upon time production (Fortin et al., 1993), when used as a secondary task, implies that the interference may be modelled in an additive fashion. This is an advantageous property of memory scanning as a secondary task. By measuring the reaction time function for a particular memory-search task, it is possible to make a precise estimate of the processing time required by that task. By changing the nature of the memory-set items, it is possible to affect the slope of the reaction time with set size function. Knowledge of both the reaction time and concurrent time production slopes for a given memory-search task is extremely useful for testing predictions made by the PGC model. Another useful feature of memory-search as a secondary task is that, as observed by Sternberg (1966), practice does not affect the slope of the search function. This is a great advantage for the present research, where there are always a large number of trials performed by each subject.

\textsuperscript{4}A full discussion of this model is beyond the scope of this thesis. However, one interesting proposition is that the gamma oscillations are nested within the theta oscillations. Each gamma cycle may represent an item in memory. If responses can only occur at the end of each theta cycle, then this would explain the paradox of exhaustive search for positive probe items.
2.3 Experiment 1

Thus far, the Fortin et al. (1993) proposition that temporal processing is critically dependent upon the availability of short-term memory resources has been described. The evidence for this position is based upon interference with timing from a phonological short-term memory task. It has been argued that it would be advantageous to extend the evidence for this position, by replicating the effect with a different short-term memory task. Short-term memory for pitch is a good candidate, because available evidence suggests a functional separation between it and phonological memory. Process theories of memory search were briefly considered, in order to outline why a linear relationship between number of memory items and search time is expected.

In Experiment 1 the pitch analogue of the Sternberg task, first used by Cruse & Jones (1976), will be used. Two main experimental conditions will be employed: pitch memory search reaction time, and pitch memory search concurrent with time production.

The aims of Experiment 1 may be summarised as follows:

- To measure the slope of the function of reaction time with set size for pitch memory search
- To measure the slope of the function of time production with set size, when time production is performed with a concurrent pitch memory search.
- To assess whether the requirement to perform time production concurrently with memory search interferes with the memory search process, to be assessed by a comparison of the error rates in the two main experimental conditions.
To assess an aspect of the paradigm unreported by Fortin & colleagues: whether introducing the secondary task qualitatively changes the nature of the timing process; if this is the case then timing variability should increase with set size.

To investigate a second feature of the paradigm unreported by Fortin & colleagues: whether the representation of the target interval remains available to subjects throughout the 192 experimental trials, in which feedback on the temporal aspect of performance is not provided.

2.3.1 Method

Subjects

Eight subjects participated in the reaction time condition. One of these was excluded due to a very high error rate. Fifteen subjects participated in the concurrent time production condition. One of these was excluded due to a very high error rate, while another subject was excluded because his time productions were considerably greater than the target value. Subjects were unpaid volunteers.

Stimuli

The memory-set consisted of 8 pure tones of 500 msec duration (259 Hz, 360 Hz, 500 Hz, 695 Hz, 966 Hz, 1343 Hz, 1866 Hz, and 2591 Hz) presented in mono using Hi-TEX CHP-727V headphones. The volume level was set to the maximum comfortable level during piloting, and this level was maintained throughout the experiment. The
physical frequency of each stimulus was 1.38 times that of the previous one. This resulted in perceptually equal steps between each of the stimuli (see sections 4.3 & 4.3.2). Memory-set size varied between 1 and 4, individual items being separated by 500 msec pauses. The memory-set items were drawn randomly from the overall set on each trial. The probe was positive (present in the memory-set) on fifty percent of trials. Positive probe items were selected randomly from the current memory-set, while negative probes were selected randomly from members of the overall set which were not part of the current memory set.

Design & Procedure

There were two within subjects variables in both the reaction time condition and the concurrent time production condition, probe type (positive or negative) and memory-set size (1-4). Subjects participating in the reaction time condition received four practice blocks of 48 trials each, in order to allow any initial practice effects to take place and stabilise their performance. Within each block, the sequence of trials was randomised. There were two separate four-block sessions of experimental trials. Each testing session lasted 45 minutes. There were 384 experimental trials per subject in the reaction time condition, for a total of 2688 trials.

In the concurrent time production condition subjects were first trained to produce a 2.5 s target interval. They were not informed of the duration in seconds, but were shown a visual stimulus that remained on screen for the appropriate amount of time. To begin training, there were two blocks of 48 training trials in which feedback was given after each trial, followed by two blocks of 24 in which feedback only occurred after all
the trials were complete. After this time production training session, subjects were introduced to the combined time production and memory search condition, and completed one block of 48 trials as practice. Memory for the target interval was then refreshed with two blocks of 24 time production practice trials with full feedback, and two blocks of 24 trials in which feedback occurred only after all the trials were complete. Finally, there were four experimental blocks, each consisting of 48 trials of the combined task. The overall testing time was 80 minutes. Each subject experienced 192 experimental trials, for a total of 2496 trials.

Procedure in individual trials

In the reaction time condition, subjects initiated each trial by pressing the space bar. One second after this occurred the first memory-set item was presented via the headphones. Remaining members of the memory-set were then presented, separated from each other by 500 msec pauses. As soon as the last item in the set was presented a fixation dot appeared. This cued the subject to initiate the presentation of the memory probe item by pressing the middle of the three mouse buttons: 500 msec later the probe item started, and continued to sound until a response was made. Subjects pressed the left mouse button to make a positive response and the right mouse button to make a negative response. Feedback on the accuracy of the response was then given. Reaction time was measured from the onset of the probe item. Instructions to subjects stated that responses should be made as quickly and accurately as possible.

In the concurrent time production trials, the sequence of events was similar to that in the reaction time condition. The only difference was that subjects withheld their
response to the probe, until they felt that the time elapsed from the point when they had initiated the probe item was equal to the learned interval. The computer recorded each time production as the time between the middle button response, and the left or right button response. This differed from the procedure for memory search only trials, where reaction time was measured from probe onset. Feedback was given on the memory search component of the task only. The sequence of events in a single concurrent time production trial is shown in Figure 2.1.

Time production training started with the demonstration of the standard interval. In order to perform a practice trial, subjects pressed the middle mouse button and waited for an amount of time that they thought was equal to the target, before pressing either the left or right mouse button. Which button they pressed depended upon the direction of an arrow, which appeared on the screen 500 msec after the subject initiated the trial. The reason for the dual response was to make the training as similar as possible to the experimental trials. In the feedback training condition, the computer informed the subject
of the extent of their over or underestimation after each trial. Errors of 10% or less in each direction were defined as hits. Averaged feedback was given at the end of each block. In the no-feedback training condition, only feedback at the end of each block was available.

2.3.2 Results

Note on standard deviations

Throughout this thesis, two kinds of standard deviation are reported. Figures in tables that are reported in brackets are between subject standard deviations, i.e. each subject provides a single score, and the standard deviation of the mean is calculated. Whenever the standard deviation is treated as a dependent measure the within subject standard deviation is used. In this case, a standard deviation is calculated from the raw data for each individual subject. The mean of these individual standard deviations is then reported, along with an analysis of variance. Within subject standard deviations are reported separately in tables and graphs. Whenever error bars are plotted on graphs, they represent one standard error.

Reaction time condition

The overall error rate was 8.3% in the reaction time condition, compared to 7.5% in the concurrent time production condition. The details of this are presented in Table 2.1, from which it can be easily seen that the error rate increases with memory-set size. Error rates in the reaction time condition were assessed using a 2 (positive versus negative
probe) * 4 (memory-set size) ANOVA. The effect of set size was highly significant, $F (3, 18) = 17.17, p < .001$. At set size 1 the higher error rate occurred for negative probes, at set size 2 error rates were equal for the two probe types, and at set sizes 3 and 4 the rate was higher for positive probes. This interaction was reliable, $F (3, 18) = 3.72, p < .05$.

After errors were excluded, outliers of more than three standard deviations were removed from each subject’s data, for each of the eight conditions of the experiment separately. This amounted to 1.6% of the remaining reaction time data.

Table 2.1. Percentage memory search error rates, with standard deviations, in Experiment 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Probe Type</th>
<th>Memory Set Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Reaction time</td>
<td>Positive</td>
<td>1.2 (2.4)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>4.5 (4.4)</td>
</tr>
<tr>
<td>Time production</td>
<td>Positive</td>
<td>0.6 (1.6)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>2.9 (5.2)</td>
</tr>
</tbody>
</table>

Figure 2.2 shows the effects of memory-set size on reaction time. Averaging across positive and negative trials the slope in reaction time was 65 msec per item beyond the first. Slope was simply calculated as the reaction time for the highest set size, minus the reaction time for the lowest set size, divided by the number of steps — three in this case. The resulting value is regarded only as a figure enabling comparison between different experimental conditions. The same procedure was adopted by Fortin et al. (1993). The alternative was to fit a particular function. However, function fitting was regarded as inappropriate because the error rate increases with set size, and therefore the
number of guessed trials contaminating the data increases with set size. The effects of guessing on processing time are unknown, and would therefore have confounded attempts to fit functions. The effect of set size on reaction time was significant, $F(3, 18) = 15.61, p < .001$. The effect of set-size interacted with probe type such that positive items were responded to more quickly at the lower set sizes only, $F(3,18) = 10.94, p < .001$.

**Time production concurrent with memory search**

The mean error rate in the concurrent time production condition was 7.5%, compared to 8.3% for the reaction time condition. Inspection of Table 2.1 shows that the error rate increased with set size. This was the only significant effect on the error rate, $F(3, 36) = 22.02, p < .001$. After errors were removed from the data set, 0.7% of the raw data was classified as outlying, and therefore excluded, using the same criteria as for the reaction time condition.

![Figure 2.2. Reaction time as a function of memory-set size and probe type in Experiment 1.](image)
In the final block of time production practice without feedback, the mean time produced was 2.76 s, with an average per subject standard deviation of .26 s. Due to a programming error, data showing the progression of learning in the time production training procedure is unavailable for Experiment 1. However, the training procedure was identical in all of the time production experiments presented in this thesis. Therefore, training data from three experiments is analysed and discussed in Appendix 1. Performance on the timing component of the combined task was quite stable across the four experimental blocks, indicating that the learned interval remained available to subjects. Table 2.2 shows the mean times produced in each of the four experimental blocks, as well as the average standard deviation of performance. The trend towards shorter time productions as the experiment progressed was not significant, and the size of the drift was only 3.4% of the size of the target interval. The variability of timing actually fell as the experiment progressed. This effect did not reach significance in the omnibus ANOVA, although the linear contrast was significant, $F(1, 12) = 5.70, p < .05$. If memory for the learned standard had deteriorated then the standard deviation would have been expected to increase rather than decrease as the experimental session progressed.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time production (seconds)</td>
<td>2.60</td>
<td>2.66</td>
<td>2.64</td>
<td>2.61</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.26</td>
<td>0.24</td>
<td>0.23</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Table 2.3 presents statistics describing the response distribution for each of the subjects in the experiment. The data was collapsed across experimental conditions to produce these statistics. There was a wide range of mean times produced, the shortest mean time production being 1.97 s, while the longest was 3.14 s. Each of these extreme values deviates from the target duration by approximately 20% of the magnitude of the target duration. There was also a wide range of variability of performance, the smallest coefficient of variation being 0.04, while the largest was 0.18. Values of skew and kurtosis were almost universally very low, indicating that the time production data was normally distributed. This indicates that it is appropriate to calculate means, and to apply parametric statistical procedures to the data. Subject number four had an exceptionally high value of kurtosis (3.45). Inspection of a histogram revealed that a single extreme value of 4.00 s was responsible for this. This value was 2.97 standard deviations from the mean time production for subject number four within its experimental condition, and hence just avoided exclusion as an outlying value. Excluding it from the calculation procedure for the kurtosis statistic causes that value of that statistic to drop to 1.10, while the value for skew dropped from 1.43 to 1.01.
Table 2.3. Mean time productions (seconds) for individual subjects, along with statistics describing the distribution of performance.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Coefficient of Variation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.06</td>
<td>0.22</td>
<td>0.07</td>
<td>0.20</td>
<td>-0.02</td>
</tr>
<tr>
<td>2</td>
<td>3.03</td>
<td>0.21</td>
<td>0.07</td>
<td>-0.05</td>
<td>0.34</td>
</tr>
<tr>
<td>3</td>
<td>2.50</td>
<td>0.29</td>
<td>0.12</td>
<td>0.30</td>
<td>-0.19</td>
</tr>
<tr>
<td>4</td>
<td>2.07</td>
<td>0.36</td>
<td>0.18</td>
<td>1.43</td>
<td>3.45</td>
</tr>
<tr>
<td>5</td>
<td>2.20</td>
<td>0.36</td>
<td>0.16</td>
<td>0.30</td>
<td>-0.41</td>
</tr>
<tr>
<td>6</td>
<td>2.29</td>
<td>0.37</td>
<td>0.16</td>
<td>0.39</td>
<td>-0.03</td>
</tr>
<tr>
<td>7</td>
<td>2.87</td>
<td>0.11</td>
<td>0.04</td>
<td>0.07</td>
<td>0.46</td>
</tr>
<tr>
<td>8</td>
<td>2.83</td>
<td>0.30</td>
<td>0.11</td>
<td>0.73</td>
<td>0.92</td>
</tr>
<tr>
<td>9</td>
<td>2.60</td>
<td>0.26</td>
<td>0.10</td>
<td>0.24</td>
<td>0.07</td>
</tr>
<tr>
<td>10</td>
<td>3.14</td>
<td>0.13</td>
<td>0.04</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>11</td>
<td>2.86</td>
<td>0.24</td>
<td>0.08</td>
<td>0.30</td>
<td>0.14</td>
</tr>
<tr>
<td>12</td>
<td>1.97</td>
<td>0.28</td>
<td>0.14</td>
<td>0.09</td>
<td>-0.06</td>
</tr>
<tr>
<td>13</td>
<td>3.01</td>
<td>0.25</td>
<td>0.08</td>
<td>1.11</td>
<td>0.68</td>
</tr>
</tbody>
</table>

The main hypothesis under investigation was that increasing the processing time required by the memory search task would lead to a lengthening of concurrent time production. This was found to be the case; increasing the number of items to be searched caused a lengthening of time production — by 32 msec per item beyond the first averaging across the two probe types. This is shown in Figure 2.3, along with the average within subject standard deviation of timing for each experimental condition. The effect of set size on time produced was reliable, $F(3,36) = 4.46$, $p < .01$ (one tailed). A one tailed test was selected here because lengthening (as opposed to shortening) of time production
was predicted to occur as set size increased. The interaction of positive versus negative probe type with set size did not reach significance. Inspection of Figure 2.3 suggests that the standard deviation of time production increased with set size. However, this effect was not reliable, $F(3,36) = 1.14$.

![Graph](image)

**Figure 2.3.** The effect of memory-set size upon mean time production, and average within subject standard deviation of time production, in Experiment 1. The left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean standard deviation of time production.
2.3.3 Discussion

As predicted, performing pitch memory search concurrently with temporal judgement caused a systematic bias in temporal judgements, and this was related to the processing time required by the memory search task. However, the magnitude of the bias was not equal to the known processing time required by the memory search task, as established in the reaction time condition.

A number of other features of the data were also of interest. Firstly, subjects were able to learn to reproduce a temporal interval, and the long-term memory for that interval remained available to them for the subsequent forty-five minutes of the experiment. Secondly, performing the timing task did not interfere with the availability of pitch memory or the accuracy of memory search. Thirdly, as the difficulty of the pitch memory search task increased the timing process was not qualitatively changed.

Turning first to the main hypothesis under investigation, the slope of the function of set size with reaction time allows the processing time required by the pitch memory search to be estimated at 67 msec per item. When this task was combined with time production, increasing the processing time required by the search task caused a lengthening of time production. This was the same result as obtained by Fortin et al. (1993) with digits. However, the slope of the function was only 32 msec per item — approximately half that obtained in the reaction time condition. The fact that pitch memory search does interfere with temporal processing strengthens the argument that the gating process by which information from the pacemaker reaches the counter is mediated by short term memory resources. What are the implications of these results for the gate
component of the PGC model? Clearly, the gating process is not a complete one — while secondary, nontemporal, processing takes place some temporal information from the pacemaker still reaches the accumulator. This is a result broadly in agreement with that of Fortin et al. (1993) who, using digits, found slopes of 36 msec in reaction time and 24 msec in concurrent time production. Both these sets of results indicate that subjective time does not stop passing while resources are directed to another processing task, although it does slow considerably.

There was no reliable change in the standard deviation of time production. However, there is a suggestion in Figure 2.3 that one might be detected in an appropriately designed experiment. Discussion of the variability of temporal judgements, and the predictions the PGC model makes about variability, is deferred until Chapter 5, where two experiments investigating this issue directly are presented.

Timing performance was extremely consistent across the four experimental blocks of the combined task, the only significant change being a reduction in variability of performance as the experiment progressed. If the representation of the target interval had deteriorated over time, then variability would have increased. The reduction in variability may be attributed, in part, to a decrease in the variability associated with the motor component of the time production task. The stability of performance indicates that the training procedure was successful in providing subjects with an enduring representation of the target interval

The memory search error rate was similar whether it was performed on its own, or as part of the combined task. This suggests that the requirements of the timing task did not result in any qualitative change to the way the memory search was performed.
Equally, there was no evidence that the memory search had a severely disruptive effect on temporal judgement. If this had occurred, then as the difficulty of the memory search increased, variability of temporal judgement would have grown considerably. In fact, the increase in variability was proportional to the growth in the mean, and non-significant.

In conclusion, the data supports the proposition made by Fortin et al. (1993) that the gating process is under the control of short-term memory. Additionally, a number of features of the paradigm have been analysed which were not reported by Fortin et al. (1993). Specifically, temporal judgement was found to be stable across the experimental blocks, and it was shown that increases in the difficulty of the secondary task did not lead to any large increase in the variability of temporal judgement.

2.4 Experiment 2

Fortin & Breton (1995) point out that there are two possible causes of the interference with temporal judgement observed in Fortin et al. (1993), and by extension in Experiment 1. The first source of interference is the requirement to somehow compare the probe item with the memory-set, in order to decide whether the probe item is positive or negative. The second possibility is that the need to maintain the memory set through the first part of time production, until the probe item is presented, somehow slows the rate of accumulation of temporal information. Fortin & Breton (1995) tested the possibility that retaining digits in short-term memory would interfere with time production by asking subjects to maintain between one and six digits throughout a two-second time production. The memory probe item was presented after time production was
complete, and therefore the comparison process could no longer interfere with temporal judgement. Under these circumstances, increasing memory load had no effect upon temporal judgement. Fortin & Breton (1995) also demonstrated the converse effect, namely that tasks which involve a comparison process, but very little short-term memory load, do produce systematic interference with temporal judgement. In one experiment this was achieved using a rhyme-judgement task, in which the number of syllables was systematically increased — but all the information required by the comparison process was presented simultaneously. In another experiment, the same effect was demonstrated again using a mental rotation task, in which the amount of rotation required was varied. Thus, increasing the short-term maintenance load of a task was shown to produce no impact upon time production, while increasing the processing load with a low, and constant, maintenance requirement had a systematic effect upon time production.

The aims of Experiment 2 may be summarised as follows:

- To replicate, using the pitch memory set, the finding that when memory load only is concurrent with time production, increases in set size do not affect mean time production (Fortin & Breton, 1995)
- To extend the null result with respect to increases in set size of the concurrent memory load to variability of time production, data on which were not reported by Fortin & Breton (1995)
- To replicate Fortin & Breton’s (1995) finding that the error rate on the memory search component of the task is higher under these conditions than under those in Experiment 1
2.4.1 Method

Subjects

Twelve unpaid subjects, aged between 23 and 30 took part in the experiment.

Stimuli

The pitch memory set was used again. The stimuli, and the procedure for selecting memory-sets and probe items was the same as described in Experiment 1.

Design

The memory for the pitch of the tones had to be retained throughout the entire time production of 2.5 s. This may be contrasted with the situation in Experiment 1, where the memory-set only had to be maintained for the first 500 msec of time production. The memory probe item was presented after time production was complete. Therefore, two functions were collected for each subject: time production with memory-set size of concurrent memory load (4 levels), and reaction time by set size (4 levels) and probe type (2 levels).

Procedure

Time production training was identical to that in Experiment 1 in all but one respect. Each time production now began and ended with the middle button, whereas in Experiment 1 time production had ended with either the left or the right button. This change to the training procedure was made because time production in the experimental
condition now began and ended with the middle button. Due to this change, the arrow, which had appeared during each time production practice in Experiment 1, was removed.

In the experimental blocks, each trial began with the presentation of the memory set. Time production was initiated by the subject pressing the middle button. The task was then to wait for an interval equal to their subjective impression of the target interval, before pressing the middle button again. Immediately after this second key-press, the memory probe was presented, and the subject responded with a left click for a positive response, or a right click for a negative response.

2.4.2 Results

The overall error rate in the memory search was 17.6%, which was approximately twice that in Experiment 1. The details of this are presented in Table 2.4, from which it can be seen that the error rate increased with memory set-size. This was the only significant effect on the error rate, $F(3, 33) = 19.60, p < .001$. After trials on which a memory search error occurred were excluded, outlying data points of more than three standard deviations were removed. This amounted to 1.4% of the memory search data, and 0.5% of the time production data.
Table 2.4. Percentage memory search error rates in Experiment 2.

<table>
<thead>
<tr>
<th>Probe Type</th>
<th>Memory Set Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.2 (5.0)</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.0 (9.9)</td>
</tr>
</tbody>
</table>

In the final block of time production practice without feedback, the mean time produced was 2.70 s and the average standard deviation of performance was .20 s. Table 2.5 shows the mean times produced in each of the four experimental blocks, as well as the average within subject standard deviation of performance. Again, time production was stable across the four experimental blocks. The slight decline in time produced across the four blocks was not significant. The reduction in the standard deviation of performance as the experiment went on did not reach significance in the omnibus ANOVA. However, the linear contrast was reliable, $F(1,11) = 5.04, p < .05$.

Table 2.5. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of Experiment 2.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(seconds)</td>
<td>2.63</td>
<td>2.58</td>
<td>2.57</td>
<td>2.50</td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.29</td>
<td>.27</td>
<td>.26</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.4 shows the effect of increasing the number of items to remember on time production, as well as the average standard deviation of time productions at each set
size. Taking set size 1, rather than the lower point at set-size 2, as the starting point the slope of the function was 22 msec per item beyond the first. This unexpected effect was reliable, F (3,33) = 4.8, p > .01. The standard deviation of performance was not reliably affected by increasing the memory set size.

Figure 2.4. Time production and variability of performance as a function of pitch memory load in Experiment 2. The left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean standard deviation of time production.

Figure 2.5 shows the response times to the memory probe. As expected, the effect of set size is significant, F (3, 33) = 4.35, p > .01 (one tailed), there being no reliable interaction, or difference between positive and negative trials. A one tailed test was selected because reaction time was predicted to lengthen (not shorten) with set size. The slope of the function was 61 msec, compared to 67 msec in the reaction time condition of Experiment 1.
2.4.3 Discussion

The main result of Experiment 2 was that increasing pitch memory load had a systematic effect upon temporal judgement. This was an unexpected effect, in light of Fortin & Breton's (1995) finding that increasing a memory load consisting of digits had no systematic effect upon temporal judgement. Also of interest, was the increase in the memory search error rate relative to Experiment 1. Fortin & Breton (1995) also found an increased error rate under memory load only conditions with digits (6.5%), compared to the concurrent search condition (less than 3.0%). The magnitude of this increase is very similar to that observed here.

In Experiment 1 the memory-set only had to be maintained for the first 500 msec of time production. In Experiment 2 it had to be maintained throughout the full 2500
msec time production. One interpretation of the higher error rate is that the memory set had longer to decay. However, this seems unlikely, as no significant trace decay occurs for a single tone over the course of five minutes (Butler & Ward, 1988). Another interpretation is that the tones interfere with each other in memory, and that this interference is an ongoing process. This account would be consistent with Deutsch's theory, described in section 2.2.2. A third possibility is that the requirement to concurrently perform temporal judgement reduces the amount of resources available for maintaining the memory set. This possibility could be tested by conducting a control condition, in which the memory load had to be maintained for a 2500 msec interval controlled by the experimenter, rather than by the subject. If removing the need to perform timing during retention of the memory set lead to a decrease in the error rate, then this would suggest that concurrent timing had lead to the increased error rate.

Given that the memory search task was made more difficult by the conditions of Experiment 2, and that the error rate increased with the memory-set size, it might be expected that subjects would increasingly shorten their time productions as the memory load was increased. Such a strategy would have reduced the error rate on the memory search task. However, this pattern is not observed in the results. Instead, time production lengthens with the increase in set size. Did this occur because the pitch memory system shares resources with the short-term memory resource that supports the counter component of the PGC model? On the face of it, this seems unlikely. Firstly, there is no obvious reason why such disparate functions as pitch memory and temporal memory would be supported by the same system. Secondly, if this were the case it is unclear that the effect of such a double load upon the same system would be a lengthening of time
production. Without any obvious model of how the two sorts of information might interact with each other, it seems equally plausible that increasing pitch memory load would cause a shortening of time production. Another possibility, if the two memory buffers were strongly linked, is that increasing pitch memory load would cause a large increase in the variability of temporal judgement. Inspection of figure 2.4 shows that this did not occur. In addition, the general variability of time production in Experiment 2 is almost identical to that in Experiment 1, where the time the pitch information had to be maintained for was much less.

In summary, these results challenge the conclusion of Fortin & Breton (1995), that only tasks involving active processing of material in short-term memory interfere with temporal processing. Under some circumstances, just maintaining a memory load can cause interference. Maintaining a memory load may be an active process, especially if rehearsal is involved. However, it does not involve actively making comparisons between the memoranda, which was required in Experiment 1. These results are not as easily explained by the PGC model as the results of Experiment 1. In Experiment 1, the idea of a gate being closed to temporal information during a memory search process that known to take a few hundred milliseconds seemed plausible. In contrast, the effect that was observed in Experiment 2 must be a more diffuse one. There is no portion of the 2500 msec of temporal production to which the effect may be localised, and therefore it must be regarded as a slowing of the accumulation process rather than an interruption. Hence, the gating metaphor seems inappropriate here. The experiments presented in Chapters 3 and 4 further explore the unexpected effect of pitch memory load on temporal judgement.
3.1 Aims of Chapter

The main aim of this chapter is to explore possible reasons why Experiment 2 produced a result that contradicted Fortin & Breton (1995), who found that just maintaining a memory-set composed of digits had no impact upon temporal judgement. In Experiment 2, it was found that just maintaining a memory-set composed of pure tones varying in pitch had a systematic impact upon temporal judgement. Experiment 3 tests the possibility that the unexpected effect was not caused by the memory load per se, but was instead a function of the duration of presentation of the memory set. Experiment 4 removes the need to maintain the memory-set during time production, in the expectation that if the results of Experiment 2 were due to processes involved in maintaining the memory set, this will abolish the effect. Experiment 5 examines the possibility that maintaining pitch information is substantially more demanding of attentional resources than maintaining digits, in an effort to account for the differential effects of these two tasks upon temporal judgement.

3.2 Experiment 3

A series of pure tones presented via headphones is an extremely salient event; the tones are hard to ignore. Visually presented digits, on the other hand, would easily be ignored by a poorly motivated subject. As well as being necessarily aware of the pitch of the tones, a subject is likely to register the temporal characteristics of the memory set. A
pattern of salient tones, broken up by silent intervals, is likely to draw attention towards the temporal characteristics of the sequence of events. Thus, it is possible that each pitch based memory set will form a kind of "perceptual gestalt", grouped according to temporal characteristics. This seems less likely in the case of a set of visually presented digits, such as those used by Fortin & Breton (1995) because of the absence of the transitions between silence and sound.

If subjects are drawn to register temporal properties of the pitch memory set, then this raises the possibility that this aspect of the memory set, rather than the requirement to remember pitch information, is the cause of the subsequent effect upon the production of a learned interval. Deutsch (1986) provides a precedent for the occurrence of this kind of effect. Subjects were presented with a standard duration (S), followed by a retention interval of a few seconds. Memory for S was assessed by presentation of a test stimulus (T), which had to be categorised as the same in duration as S, shorter than S, or longer than S. With an empty retention interval performance on this task was extremely good. However, when the retention interval is filled with "to be ignored" interpolated durations (I), then judgement of the relationship between S and T was systematically biased. If the duration of I was longer than S, then this produced a tendency to respond that T was short, whereas values of I shorter than S produces a tendency to judge long. Deutsch (1986) explains these results in terms of memory interference; I interferes with S in short-term memory.

It is possible that a similar kind of memory interference occurred in Experiment 2. It is hypothesised that, on each trial, the subject retrieves a target value for the learned duration from long-term memory, and maintains it in short-term memory. If, as discussed
above, the temporal characteristics of the memory set are processed, and its overall duration noted, then the target value for the time production may be subject to biasing by the duration value of the memory-set. The duration values of the memory sets in Experiment 2 were 500 msec, 1500 msec, 2500 msec, and 3500 msec. Given the target value of 2500 msec, it is possible that an artificial shortening was produced at set size’s 1 and 2, and an artificial lengthening was produced at set size 4. In the present experiment, the duration of presentation of the memory set was manipulated independently of the number of items to be remembered. Thus, the aim was to establish whether the results of Experiment 2 occurred due to processes involved in maintaining the pitch memory load, the temporal characteristics of the memory load, or both factors.

The aims of Experiment 3 may be summarised as follows:

- To replicate the unexpected lengthening of time production associated with increases in memory-set size that occurred in Experiment 2.
- To test the hypothesis that the set size effect in Experiment 2 may have occurred due to the fact that memory set presentation time co-varied with set size, rather than being caused by the processing requirements of retaining the memory-set

3.2.1 Method

Subjects

Sixteen unpaid subjects ranging in age between 18 and 30 took part in the experiment. The majority of the subjects were undergraduates, who received course credit in return for their participation.
The aim of Experiment 3 was to further explore the results of Experiment 2. Therefore, the same pure tones were used as in Experiments 1 and 2. The procedure for selecting memory sets and probe items was also identical. In Experiments 1 and 2 individual tones always had a duration of 500 msec. Here, the duration of the tones was varied systematically (see below).

Design

Because of the requirement to keep the testing time per subject under control, only memory-set sizes two and four were used. Memory sets of three different durations were employed: 1.5 sec, 2.5 sec, and 3.5 s. In order to produce these values at both set sizes the durations of individual tones were manipulated. Conditions were always arranged such that the duration of the pause between tones was equal to the duration of a tone — as it had been in Experiment 2. The durations used are shown in Table 3.1, broken down by set size and set duration. The durations of items in the top left, and bottom right cells of Table 3.1 are identical to the 500 msec value used for individual set items in Experiments 1 and 2.

Time production took place under six experimental conditions: memory-set size (2) * duration of memory set (3). Subsequent reaction times to the memory probe took place under twelve experimental conditions: memory-set size (2) * probe type (2) * duration of memory set (3).
Table 3.1. Durations of individual memory-set items (secs).

<table>
<thead>
<tr>
<th>Duration of Memory-Set</th>
<th>1.5 sec</th>
<th>2.5 sec</th>
<th>3.5 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set size 2</td>
<td>0.50</td>
<td>0.83</td>
<td>1.17</td>
</tr>
<tr>
<td>Set size 4</td>
<td>0.21</td>
<td>0.36</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Procedure

The time production training procedure was identical to that used in Experiment 2. This involved first providing subjects with trial by trial feedback on the accuracy of their productions, and making feedback available at the end of blocks only. During experimental trials, the memory set had to be remembered throughout time production, again, as in Experiment 2. When the memory probe item was presented, immediately after time production was complete, it continued to sound until a response was made. Thus, the durations of the memory set items were irrelevant to the recognition judgement. This procedure was the same as that used in Experiment 2.

Each subject underwent 20 trials in each of combinations of probe type and memory set duration, for a total of 240 trials. This meant there were 40 trials per cell of the time production design, because probe type was not apparent to subjects until after time production was complete. The trials were divided into four blocks of 60 trials each. The total number of trials analysed was 3840.
3.2.2 Results

The overall error rate was 18.6%, which was comparable to the 17.6% error rate in Experiment 2. The details of this are shown in Table 3.2. By far the largest factor influencing the error rate was memory-set size, \( F(1, 15) = 141.07, p < .001 \). Additionally, more errors occurred when the duration of the memory set was shorter, \( F(2, 30) = 12.95, p < .001 \). The effect of duration on error rate was greater at set-size 4 than at set-size 2, \( F(2,30) = 3.59, p < .05 \). The effect of duration was also greater for positive probes than negative ones, \( F(2, 30) = 7.05, p < .01 \). This pattern was reflected in a significant three-way interaction, \( F(2,30) = 5.56, p < .01 \). After trials on which a memory search error occurred were excluded, outlying data points of more than three standard deviations were removed. This amounted to 1.9% of the memory search data, and 0.8% of the time production data.

Table 3.2. Percentage pitch memory search error rates in Experiment 3, in which the memory probe was presented after the completion of time production.

<table>
<thead>
<tr>
<th>Set size 2</th>
<th>Duration of Memory-Set</th>
<th>1.5 sec</th>
<th>2.5 sec</th>
<th>3.5 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td></td>
<td>17.2 (11.5)</td>
<td>15.6 (7.5)</td>
<td>5.3 (4.6)</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td>36.3 (9.1)</td>
<td>22.8 (10.0)</td>
<td>12.8 (7.1)</td>
</tr>
<tr>
<td>Set size 4</td>
<td>Positive</td>
<td>24.7 (14.6)</td>
<td>25.9 (9.8)</td>
<td>15.0 (8.2)</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td>6.6 (11.5)</td>
<td>5.3 (8.6)</td>
<td>22.8 (13.8)</td>
</tr>
</tbody>
</table>
In the final block of time production practice, the mean time produced was 2.51 sec, and the average standard deviation of performance was 0.26 s. Again, there was no change in mean time production as the experimental session progressed, despite the testing session being longer than in Experiments 1 and 2. The details of this are shown in Table 3.3. On this occasion, there was no significant reduction in the variability of performance across experimental blocks.

Table 3.3. Mean time production, and within subject standard deviation of time production in each of the four experimental blocks of Experiment 3.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time production (seconds)</td>
<td>2.47</td>
<td>2.41</td>
<td>2.39</td>
<td>2.37</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>.31</td>
<td>.32</td>
<td>.30</td>
<td>.29</td>
</tr>
</tbody>
</table>

Figure 3.1 shows how mean time production, and average standard deviation of time production, was affected by memory-set size, and the duration of presentation of the memory-set. The effect of number of items to be remembered was significant, F (1,15) = 3.80, p < .05 (one-tailed). A one tailed test was selected here because, on the basis of Experiment 2, increasing set size was expected to lengthen (not shorten) time production. Time production was 44 msec longer at set size 4 than set size 2. In Experiment 2 the slope of the effect was estimated at 22 msec per additional item. Thus, the findings of Experiment 3 are in exact quantitative agreement with those of Experiment 2. The effect of the duration of presentation of the memory set, which seems clear in Figure 3.1, was
not reliable, \( F (1,15) = 2.20, p > .1 \). There were no significant effects upon the standard deviation of performance.

The mean reaction times to the memory probe item were 889 msec at set size 2, and 927 msec at set size 4. This difference was reliable, \( F (1,15) = 9.33, p < .01 \). There were no other reliable effects upon the reaction time.

Figure 3.1. Mean time production and mean standard deviation of time production, as a function of number of items to be remembered, and duration of memory set, in Experiment 2. The left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean standard deviation of time production.
3.2.3 Discussion

Maintaining a memory set composed of pure tones was once again shown to systematically affect temporal judgement, in that the amount of interference was related to the number of items that had to be remembered. As a replication of Experiment 2 these results are particularly impressive, because of the quantitative agreement between the two sets of results. This was reflected in the effect of the memory loads on time production, as well as the overall error rate on the memory search task.

Manipulating the duration of presentation of the memory set did not have a significant effect upon time production. However, inspection of Figure 3.1 suggests that a significant effect might be obtained in an appropriately designed experiment. Therefore, the results of Experiment 3 are equivocal in their support for Deutsch (1986). One way to make a reliable result more likely would be to use a shorter target value for the learned time production. In this way the variation in the duration of the biasing stimulus would be much greater in relation to the learned standard, and might have a proportionally greater effect. However, the relevant conclusion here is that, with the stimulus configuration used in Experiment 2, the effect on time production was not due to the temporal characteristics of the memory set.
3.3 Experiment 4

Experiment 3 ruled out the possibility that the effect of increasing pitch memory load on temporal judgement in Experiment 2 was an artefact of stimulus presentation times. Another possible explanation of the results of Experiment 2 is that the process of maintaining pitch information requires a resource that is also required by temporal judgement. Assuming for the moment that this is the case, Fortin & Breton's (1995) finding that remembering between one and six digits has no systematic effect upon temporal judgement implies that this resource must be one that is not required to retain digits. One possibility is that such a resource would be required in order to represent pitch information. Alternatively, the resource in question may be involved in the active maintenance of pitch information.

Experiment 4 aimed to contrast the above hypothesis with the alternative possibility, that some other, unspecified feature of the task requirements, lead to the effect of increasing the number of tones to be remembered in Experiment 2. In order to achieve this the paradigm was modified such that the response to the memory probe was made before time production. Therefore, there was no longer a requirement to try to remember pitch information concurrently with time production. If a set size effect upon time production were to be found then it could not be because of any active process of remembering pitch. Such a set size effect could perhaps be due to some after effect of perceiving pitch. If, however, the set size effect were to be abolished by this modification, then this would support the hypothesis that the set size effects in Experiments 2 and 3 were caused by the requirement to remember pitch information.
The aims of Experiment 4 may be summarised as follows:

- To remove the requirement to remember the memory-set during time production, in the expectation that this would abolish the unexpected set size effect upon time production that was found for remembering pitch in Experiment's 2 and 3.
- To make a comparison of the error rate on the memory search task when timing was no longer concurrent with search, with the error rates found in Experiments 1 and 2

3.3.1 Method

Subjects

10 subjects, who varied in age between 18 and 30, participated in the experiment. The majority of these were undergraduates, who received course credit in return for their participation.

Stimuli

The tones used to form the memory-set were identical to those used in Experiments 1 – 3, as were the procedures for selecting individual memory-sets and probe items.
Design and Procedure

Time production training was identical to that in Experiments 1-3. The experimental trials were identical to reaction time trials in Experiment 1, with the exception that the response to the memory probe also defined the start of time production. Therefore, the sequence of events in a single trial was as follows: The memory set was presented, the subject requested the presentation of the memory probe by pressing the middle button, and 500 msec later this occurred. The subject then pressed the left button to indicate that the memory probe was present in the memory set, or the right button to indicate that it was absent. This button press also signified the start of temporal interval production. Once the subject judged that sufficient time had elapsed, the middle button was pressed to indicate the end of time production. Finally, feedback on the memory search component of the task was given. Therefore, two functions were collected from each subject: reaction time by set-size and positive versus negative probe, and time production by set-size and positive versus negative probe.

Each subject underwent 24 trials in each of the 8 cells of the full design, for a total of 192 trials. The trials were divided into four blocks of 48 trials each. The total number of trials analysed was 1920.

3.3.2 Results

The average memory search error rate was 4.4%, which was the lowest rate obtained thus far. Inspection of table 3.4 reveals that the error rate increased with memory-set size. This was the only effect on the error rate to reach significance, F (3, 27)
After trials on which a memory search error occurred were excluded, outlying data points of more than three standard deviations were removed. This amounted to 1.25% of the memory search data, and 0.78% of the time production data.

Table 3.4. Percentage pitch memory search error rates in Experiment 4, in which the probe item was presented before time production.

| Memory Set Size |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | Probe Type      | 1   | 2   | 3   | 4   |
|                 | Positive        | 0.4 (1.3) | 2.5 (2.9) | 3.8 (4.1) | 6.7 (6.0) |
|                 | Negative        | 2.5 (2.9) | 3.3 (5.5) | 5.8 (5.6) | 9.7 (9.7) |

In the final block of time production practice without feedback, the mean time produced was 2.60 s and the average standard deviation of performance was .19 s. Table 3.5 presents the mean times produced in each of the four experimental blocks, as well as the average within subject standard deviation of performance. The slight shortening of time production across the four experimental blocks reached significance, $F (3, 27) = 4.47, p < .05$. There was no reliable change in standard deviation as the experiment progressed.

Table 3.5. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of Experiment 4.

| Experimental Block |
|--------------------|-----------------|-----------------|-----------------|-----------------|
|                    | 1               | 2               | 3               | 4               |
| Time production    | 2.98            | 2.87            | 2.81            | 2.85            |
| (seconds)          |                 |                 |                 |                 |
| Standard deviation | .31             | .27             | .27             | .27             |
Turning to the memory search phase of the experiment, response times to the probe item are shown in Figure 3.2. The effect of set size was significant, $F(3, 27) = 31.58, p < .001$. The figure suggests that the main source of this result is the large difference between set sizes one and two. A repeated contrast confirmed that this was the case: the reaction time at set size two was reliably different from that at set size one, $F(1, 10) = 32.77, p < .001$, while the differences between set sizes three and two, and four and three, did not approach significance.

Figure 3.2. Reaction time to the memory probe in Experiment 4.

Figure 3.3 presents the time production data. There were no effects of set size or positive versus negative probe type on either the mean time produced, or the variability of time production.
3.3.3 Discussion

The main result was that when time production followed memory search, time production showed no relationship with the size of the memory set. Of secondary interest was the particularly low error rate on the memory search component of the task.

In Experiment 1, the memory search process was concurrent with time production, while in Experiment 2 the memory set had to be maintained throughout time production. Under both these circumstances, time production was lengthened by the number of items in the memory set. The question addressed by the present experiment...
was whether the set size effect in Experiment 2 was due to an active maintenance process, or whether alternatively, it was, in some unspecified way, a consequence of simply having perceived the tones. Because the effect of set-size on time production was abolished here, it may be concluded that the unexpected effect in Experiment 2 is associated with active processing involved in maintaining the representation of pitch information over time.

The memory search error rate was approximately half that obtained in both the reaction time and concurrent time production conditions of Experiment 1. Because of the relatively low numbers of subjects, it is hard to draw any firm conclusion about this. Yet more caution is required in interpreting the error rates, because in many of the experimental conditions of all the experiments reported so far, the standard deviation of the error rate is actually greater than the mean. This indicates large individual differences in a number of possible factors, such as motivation, or the efficiency of pitch memory. Having stated the reasons for caution, it might be speculated that a low error rate was obtained here because placing a 2.5 s time production after each memory search increased the amount of time in the testing session when there was no memory load. Increasing the temporal gap between the point when one memory set could be “forgotten”, and the next one was introduced, may have acted to reduce the amount of interference in memory from the redundant memory set. This would be consistent with Deutsch’s model of pitch memory in which pitch information is recorded along a temporal dimension, as well as a pitch dimension (see section 2.2.2).
3.4 Experiment 5

The lengthening of time production with set size in Experiment 2 was caused by the processing requirements of retaining the memory set. Whatever doubt existed on this point was dispelled by the results of Experiments 3 and 4. However, these experiments were not designed to shed light on the issue of why retaining digits produced no set size effect (Fortin & Breton, 1995), whereas retaining pitch did produce a set size effect. Experiment 5 attempted to answer this question by ascertaining the effects of retaining digits, and of retaining pitch, upon a third nontemporal task. The aim was to obtain a dissociation between the effects of pitch and digits upon the third task. If such a dissociation were to be achieved, an analysis of the processing requirements of the third task would reveal the critical difference between retaining pitch and retaining digits. It might then be hypothesised that this difference was also the reason for the differential effects of retaining the two types of memoranda upon temporal processing.

In order to choose the nontemporal secondary task it was necessary to generate a hypothetical account of how the requirements of retaining digits differ from those of retaining pitch. One possibility is that pitch memory search, and retaining pitch information, is simply more difficult than is the case with analogous digit tasks. This is suggested by a comparison of the results of Experiment 1 with those of Fortin et al. (1993). The slope of the reaction time function for digits was 36 msec, whereas it was 67 msec for pitch. These differences were associated with a higher error rate in the pitch task. When memory search was combined with time production, the digit memory search task produced a slope of 24 msec, whereas the pitch task produced a slope of 32 msec.
again with a higher error rate. Because the pitch task is harder, when retention only of pitch information was combined with time production, perhaps this was inherently a much harder task than retaining digits. According to this account, varying the number of items in the pitch memory set between one and four was equivalent to varying the number of items in a digit memory set between, for example, six and ten. Therefore, the task difficulty account suggests that if Fortin & Breton (1995) had used these higher set sizes, they would have achieved results consistent with those of Experiment 2.

Fortin & Massé (1999), and Fortin & Champagne (1998) present data that rules out the task difficulty explanation. In both sets of experiments the memory search task used consonants instead of digits, and subjects had to either process item and order information during time production, or they had to maintain this information only. Fortin & Massé (1999) obtained a slope of 100 msec in a reaction time control condition, in which the probe consisted of an item, as well as a digit referring to the serial position of the item. A positive trial was now one where both pieces of information were correct. Therefore, in terms of reaction time slopes, the difficulty of this task exceeds that of the pitch memory task employed here. When this processing task was combined with time production a slope of 57 msec was obtained, compared to only 21 msec when only item information was processed. In a second experiment, processing of item information only during time production\(^1\) was compared with processing of item information and maintenance of order information. Maintaining order information and searching item information did not increase the slope of the time production function with set size,

\(^1\) This condition was identical to Fortin et al. (1993), except that the memoranda were consonants.
compared to just searching item information. Additionally, Fortin & Champagne (1998) showed that maintenance only of both item and order information concurrent with time production produced a flat set size function, similar to that of Fortin & Breton (1995). Therefore, in a task that was more difficult than the pitch memory task, maintenance only produced no set size effect, whereas maintenance only did produce an effect with pitch. This pattern of results is inconsistent with the difficulty account, which may therefore be disregarded.

An alternative to the task difficulty explanation is that there is some qualitative difference between processing involved in retaining digits, and that involved in retaining pitch. Such a difference might arise because most people are extremely familiar with processing digits, but much less familiar with processing absolute pitch information. It is likely that more attention will be required to maintain the less familiar material, leaving relatively less attention available to other tasks. Pechman & Mohr’s (1992) findings, which were described in section 2.2.2 provide some evidence in favour of this account. In particular, non-musicians’ memory for a standard tone was disrupted by pitch unrelated secondary tasks, if the processing involved was active, whereas musicians’ memory for the test tone was unaffected by pitch unrelated secondary tasks. It was suggested that musicians’ processing of both kinds of information was relatively automatic, whereas non-musicians had to give relatively more attention to the pitch task.

If the amount of attention that is required to retain the memory set is the crucial difference between the pitch memory task and the digit memory task then the PGC model is well placed to explain the unexpected effect in Experiment 2. If attention must be divided between the timing task and retaining the pitch information, then it is likely that it
will be switched back and forth between the two tasks. The process of switching is embodied in the gate component of the model. The accumulation of temporal information from the pacemaker would be temporarily interrupted while attention is directed towards maintaining the memory set. The greater the amount of pitch information to be retained, the longer attention will linger on this aspect of the task, and the magnitude of the real time interruptions to temporal processing will increase. In contrast to this, it is suggested that a sequence of digits may be maintained in memory for the period of time production without directing attentional resources towards it, at least for lists of up to six memoranda.

As a test of the above theory, an experiment was envisaged in which the two sorts of memory load were to be maintained, while a secondary task was performed that was known to require attention. According to the theory, maintaining pitch information should produce greater interference with such a secondary task than maintaining digits. The Stroop task was particularly suited to the requirements of the experiment because the conflict condition clearly requires more attention than the condition in which the colours of words unrelated to the concept of colour are identified. Thus, employing a Stroop secondary task provided the experiment with a control condition. Furthermore, it has already been demonstrated by Zakay (1993) that the Stroop task uses resources also required by temporal processing. Zakay’s subjects underestimated time in passing when concurrently performing the Stroop task, and the extent of underestimation was sensitive to the difficulty of the Stroop task.
MacLeod (1991) provides a comprehensive review of the extensive literature on the Stroop task. Of the many findings relating to the Stroop task, only a few will be mentioned here. Firstly, the time taken to read a word is less than that taken to name a colour, when these two things are separate stimuli. This has often lead to the assumption that reading words is a more automatic process than naming colours, and hence that it requires less attention than colour naming. When subjects are asked to name the ink colour of the word “red” printed in green ink, there is a cost to response time, relative to naming the ink colour of a colour-unrelated word, such as “when” in green ink. According to one account of this effect, attention is required to suppress the relatively automatic response to the word, and bring forward the relatively less automatic response to the colour. Hence, the processing time for the conflict stimulus is greater than that for just naming colours, and much greater than that for reading words. Conversely, if subjects are asked to read words, while ignoring incongruent colours (the so called “reverse Stroop”), there is no cost to processing time relative to reading words. This is taken as further support for the relative automaticity of reading words. A relatively automatic task will not suffer interference from a less automatic one. The critical feature of the Stroop task for the present investigation is that in a cue-conflict trial attention must be deployed in order to resolve the conflict in favour of the less automatic processing task. In the control trials, where the ink colour of a colour-unrelated word must be identified, the attentional requirement is much less.

Stuss, Shallice, Alexander & Picton (1996) discuss the role of attention in the executive control of action, in the general context of a consideration of the function of the
prefrontal cortex. Attention is not viewed as a unitary process, rather it is first subdivided into a spatial allocation system, and a system concerned with the executive control of attention. The former function is localised to the parietal lobe, while the latter system is associated with the prefrontal cortex. The executive system is itself functionally subdivided into the subsystems of sustaining, concentrating, sharing, suppressing, switching, preparing, and setting. The authors consider that the Stroop task may be used as neuropsychological test of suppressing, and that it indexes some resources relevant to the processes involved in sustaining, and concentrating. Following this perspective, it may be suggested that the degree to which retaining a memory set is an active process, requiring attentional resources will determine the extent to which it, and the Stroop task, interfere with each other.

**Hypotheses**

Following the logic described above, an interaction was predicted such that maintaining pitch information would lead to a set size effect on reaction time to Stroop conflict trials, but not on Stroop control trials. In contrast to this, it was predicted that maintaining digit information would not produce a set size effect upon either kind of Stroop trial.

The aim of Experiment 5 may be summarised as follows:

- To establish whether the dissociation between the set size effect of a pitch memory load on concurrent time production, and the lack of such an effect when a phonological memory load is maintained (Fortin & Breton, 1995), is due a difference in the attentional requirements of the two tasks.
3.4.1 Method

Subjects

Sixteen subjects, varying in age between 18 and 39 were tested. Each subject participated in all of the experimental conditions. As an incentive, cash prizes were awarded to the three subjects achieving the lowest error rates.

Stimuli

The alphanumeric characters 0-9 were used in the digit condition of the memory task. They were presented serially, each digit appearing for 1.2 secs. This value was chosen to correspond to that used by Fortin et al. (1993). The pitch memory set was identical to that used in Experiment 1. The trial by trial procedures for selecting both kinds of memory set, and a memory probe, was also the same as in Experiment 1.

Each Stroop control word was selected randomly from a set of three. These were “hat”, “went”, and “speak”. Fifty percent of trials were control trials. The colour of a control word was selected randomly from the set of three colours red, blue, and green on a trial by trial basis. Each Stroop conflict word was also selected from a set of three. The colour words employed were “red”, “blue”, and “green”. On each trial the conflicting ink colour was selected randomly from the two colours which did not correspond to the meaning of the word. The response to the Stroop stimulus was made with the left hand. For this purpose, three keys of the computer keyboard were labelled with the words corresponding to the three colours used in the experiment. McClain (1983) found that manual responding reduced the magnitude of the Stroop effect compared to oral
responses. However, the effect was still reliable. Manual responses to keys labelled with words produced larger Stroop effects than manual responses to keys labelled with colour patches, which is the reason the former procedure was adopted here.

It was discovered during piloting that some subjects were able to reduce or even abolish the Stroop effect by defocusing their eyes, or squinting. In this way, they could perceive the colour without reading the word. In order to counteract this the screen position at which the stimulus was presented was randomised on a trial by trial basis. This had the effect of requiring the subjects engage in saccadic eye movements, and prevented the defocusing/squinting strategy. It is likely that this increased the attentional requirements of the Stroop task. However, any such increase would be equal for control and conflict trials, as well as for both the digit and tone experimental conditions.

Design & Procedure

The experiment was divided into two testing sessions, the order of which was counterbalanced. The two sessions took place on different days. The first session began with 100 Stroop trials. The purpose of this was to familiarise subjects with that component of the task. This was followed by one block of 48 memory search trials, using either the pitch or the digit memory set. Again, the purpose of this was to make sure subjects were familiar with this component of the experiment. The final practice block of 48 trials combined the two tasks. Subjects maintained the memory set while performing the Stroop task, and then the memory probe item was presented. Once the dual task practice block was complete, subjects moved on to the experimental blocks. In the digit condition, there were 6 blocks of 48 trials, while in the pitch condition there were 4
blocks of 48 trials. The number of trials was calculated to allow 12 trials in each of the cells of the full experimental design, which was Stroop (2) * Memory probe type (2) * Memory set size (4 or 6). While reaction time to the memory probe was assessed in the full 16 or 24 experimental conditions, memory probe type did not exist as an independent variable for the Stroop reaction times. Therefore, there were 24 trials per cell for the Stroop part of the design. In the second testing session subjects first performed one block of the memory search task that they had not previously encountered, did one combined task practice block, and then moved on to the experimental blocks. Each subject performed 192 trials in the pitch condition, and 288 trials in the digit condition. There were a total of 3072 trials in the pitch condition, and 4608 trials in the digit condition.

Procedure in individual trials

Each trial of the combined task was initiated by the subject pressing the middle of the three mouse buttons. This caused the first item of the memory set to be presented. As soon as the memory set was complete, a fixation dot appeared to indicate that this was the case. Next, the subject initiated the Stroop trial by pressing the middle mouse button again, and 500 msec later the Stroop colour word was presented. Subjects responded to this as quickly as possible with the fingers of their left hand, using the labelled keys. 500 msec later, the memory probe item was presented, and subjects responded by pressing the left mouse button if the probe was present in the memory set, or the right button if it was not present. Finally, feedback was provided separately on each of the two responses.
3.4.2 Results

The standard Stroop effect was obtained. However, before this is described, error rates on both experimental tasks will be reported. In the digit condition, the mean error rate on the Stroop task was 1.84%, with a standard deviation of 2.63%. There were no effects of control versus conflict Stroop trials, or of memory-set size upon this value. In the pitch condition, the mean error rate on the Stroop task was 2.96%, with a standard deviation of 4.13%. Again, this value did not vary with set-size, or Stroop trial type. The mean error rate on the digit memory search task was 7.12%. In table 3.6 this is broken down by memory-set size, probe type, and the type of Stroop trial that preceded the recognition judgement. The effect of set size was significant, $F(5, 75) = 9.10, p < .001$. Set-size interacted with positive versus negative probe type, such that the slope of the increase in error rate was steeper for the positive trials, $F(3, 27) = 4.037, p < .01$. The mean error rate in the pitch memory search task was 10.58%. The details of this are presented in Table 3.7. The largest effect on the error rate was due to increasing memory-set size, $F(5, 45) = 19.20, p < .001$. On average, there was a higher error rate on trials with a negative probe item, and this difference was very close to significance, $F(1,15) = 4.47, p < .1$. Again, there was an interaction such that the slope of the increase in error rate with set size was greater for positive than for negative items, $F(4,45) = 4.79, p < .01$. On average, the error rate was higher when a memory search trial was preceded by a conflict Stroop trial than when it was preceded by a control Stroop trial. This effect was marginally significant, $F(1,15) = 3.75, p < .1$
Table 3.6. Percentage error rates in the digit memory search condition of Experiment 5.

<table>
<thead>
<tr>
<th>Memory-set size</th>
<th>St control</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.6 (4.0)</td>
<td>5.2 (9.0)</td>
<td>3.6 (8.6)</td>
</tr>
<tr>
<td>2</td>
<td>5.7 (11.7)</td>
<td>4.2 (8.6)</td>
<td>9.4 (18.2)</td>
</tr>
<tr>
<td>3</td>
<td>2.6 (5.0)</td>
<td>3.1 (8.5)</td>
<td>4.7 (8.6)</td>
</tr>
<tr>
<td>4</td>
<td>7.8 (10.7)</td>
<td>5.2 (8.0)</td>
<td>6.3 (10.8)</td>
</tr>
</tbody>
</table>

Table 3.7. Percentage error rates in the pitch memory search condition of Experiment 5.

<table>
<thead>
<tr>
<th>Memory-set size</th>
<th>St control</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.1 (6.0)</td>
<td>2.6 (5.9)</td>
<td>11.5 (9.6)</td>
</tr>
<tr>
<td>2</td>
<td>4.2 (7.5)</td>
<td>13.0 (11.0)</td>
<td>14.6 (10.3)</td>
</tr>
<tr>
<td>3</td>
<td>2.1 (3.7)</td>
<td>5.7 (8.5)</td>
<td>14.1 (12.4)</td>
</tr>
<tr>
<td>4</td>
<td>9.9 (9.2)</td>
<td>12.5 (13.3)</td>
<td>13.5 (10.0)</td>
</tr>
</tbody>
</table>

All trials on which a decision error occurred in either task were excluded from the raw data. Outlying values in the Stroop task were identified within each cell of the design separately for each subject. Data points lying three or more standard deviations from the
mean reaction time of the condition they occurred in were excluded. This procedure resulted in the exclusion of 1.3% of the remaining Stroop data in the digit condition and 1.2% of the remaining Stroop data in the pitch condition. Identifying outliers using the three standard deviation criteria was inappropriate for the memory search data because there were only 12 observations for each subject in each cell of the design. Instead, outliers were identified by the experimenter. One value of 13.7 s in the digit condition was treated as an error and excluded. Other extreme values were considered to be ones lying between 3 and 6 secs. All of these data points were allowed to remain in the data set, but were reduced to 3 secs. The values of 0.5% of the digit memory search data, and 0.3% of the pitch memory search were altered in this way.

The reaction times in the Stroop task, while maintaining digit information, are presented in Figure 3.4. Reaction times on Stroop conflict trials were longer than those on Stroop control trials, by 53 msec on average, $F (1,15) = 6.85, p < .01$ (one-tailed). A one tailed test was selected here because reaction time in the Stroop conflict condition is typically found to be slower than that in the control condition MacLeod (1991). There was an unpredicted effect of memory-set size on Stroop reaction time, $F (5, 75) = 2.40, p < .05$. However, because the effect occurred for both types of Stroop trial this does not suggest any specific sharing of resources between performance on the Stroop conflict trials and the maintenance of digits. What is demonstrated by this, is that the Stroop task is sensitive the effect of secondary processing loads.
Turning to the effect of maintaining pitch information on Stroop reaction time, it was predicted that a set size effect would occur for the Stroop conflict trials, but not the Stroop control trials. In fact, there was no set-size effect on either kind of Stroop trial. This is shown in Figure 3.5. However, there was a reliable Stroop effect, with an average magnitude of 42 msec, $F(1,15) = 3.19, p < .05$ (one-tailed).
After the response was made to the Stroop colour word, the memory probe item was presented. Figure 3.6 presents the response times to the memory probe for the digit memory set, collapsed across Stroop condition. There were highly significant effects of memory probe type (positive versus negative), $F(1,15) = 31.10, p < .001$, and memory-set size, $F(5, 75) = 16.76, p < .001$. There was also an unpredicted effect of the type of Stroop trial preceding the memory probe. This is shown in Figure 3.7, in which response times to the memory probe item is collapsed across probe type. On average, response times to the memory probe were 47 msec slower if the probe was preceded by a conflict Stroop trial than if it was preceded by a control Stroop trial, $F(1,15) = 4.88, p < .05$.

![Figure 3.6](image.png)

Figure 3.6. The effect of probe type, and digit memory-set size on reaction time to the memory probe in Experiment 5
Figure 3.7. The effect of preceding Stroop trial type, and digit memory-set size on reaction time to the memory probe in Experiment 5

Figure 3.8 presents the response times to the memory probe for the pitch memory set, collapsed across Stroop condition. There were significant effects of memory probe type (positive versus negative), $F(1,15) = 4.74, p < .05$, and memory-set size, $F(3, 45) = 17.89, p < .001$. Again, there was an unpredicted effect of the type of Stroop trial preceding the memory probe. This is shown in Figure 3.9, in which response times to the memory probe item is collapsed across probe type. On average, response times to the memory probe were 40 msec slower if the probe was preceded by a conflict Stroop trial than if it was preceded by a control Stroop trial, $F(1,15) = 10.34, p < .01$. 

Figure 3.8. The effect of probe type, and pitch memory-set size on reaction time to the memory probe in Experiment 5
3.4.3 Discussion

There are two main points to be made about the results of Experiment 5. Firstly, the hypotheses that maintaining pitch information required greater attentional resources than maintaining digits was not supported. Secondly, there was an interesting, and completely unexpected, effect of Stroop conflict trials relative to Stroop control trials upon subsequent memory search trials. Each of these main results will be addressed in turn.

The “attentional resources” hypothesis

It was predicted that, if maintaining pitch information required greater attentional resources than maintaining digits, performance on Stroop conflict trials would be affected by maintaining pitch, but not by maintaining digits. In the event, no set-size effect of
maintaining pitch information upon Stroop response time was found. However, an unpredicted set size effect of maintaining digits upon Stroop response time occurred. If this effect had been stronger on the Stroop conflict trials than on the Stroop control trials then this might have been interpreted as a reversal of the hypothesised attentional difference between digit maintenance and pitch maintenance. Because the effect on both kinds of Stroop trial was equal, a different explanation of this unexpected set size effect is required.

One plausible reason why maintaining digits produced a set size effect on Stroop trials is related to the fact that both digits and Stroop words are represented in a phonological code. It was established in section 2.2.2 that maintaining pitch does not require resources involved in phonological representation, and therefore this account also predicts the negative result in the pitch condition. There are two possible mechanisms of interference. Firstly, maintaining the digit memory load might slow the process of reading the word. Secondly, maintaining digits might interfere with the process of outputting the response, which also involved the phonological code because the response keys were labelled with words. One way to test this phonological interference hypothesis would be to introduce an extra experimental condition, in which the Stroop task is replaced with an analogous non-phonological task in which two responses compete with each other. This would remove the similarity of representation between the two tasks, and should therefore abolish the set size effect. An appropriate Stroop-like task would be the shapes version of the “Navon-Stroop” (Navon, 1977; Stirling & Coltheart, 1977). In the “Navon-Stroop” a large “H” may be composed of small copies of the letter “H” or the letter “S”. When required to respond to the small letters subjects suffer interference from
the large letter if it is incongruent. In the shapes version of this task, which might be used in the experiment suggested here, an arbitrary shape is made up of either small copies of itself or a different small arbitrary shape. Again, the large shape interferes with identifying the small shape if the two are incongruent.

Returning to the ongoing issue of why there was a set-size effect of maintaining pitch information upon time production in Experiment 2, which does not occur when digits are maintained, a positive answer is still no nearer. However, Experiment 5 has ruled out the possibility that the divergent result occurred because maintaining pitch information shares an attentional resource with timing that is not common to maintaining digits and timing. In Chapter 4 an attempt will be made to shed light upon this issue by isolating the unexplained effect more precisely. The principle question will be whether digits represent a special case of non-interference, or pitch represents a special case of interference.

**Interference with memory search by the Stroop task**

In both the pitch and digit memory search tasks reaction time to the memory probe was longer if the preceding Stroop trial involved response competition than if it was a control trial. The slope of the set-size function was unaffected by this, implying that the basic memory search process was delayed, not fundamentally altered. It is not possible to provide an explanation of this effect based only upon the results of this experiment. However, a number of experiments may be suggested that test three alternative explanations.
In the first proposed experiment, congruent Stroop trials would be included, keeping all other conditions identical to those here. In a congruent Stroop trial the word “red” is presented in red ink. Reaction times to congruent trials are typically faster than those to control trials, and, because the presence of such trials increases the utility of reading the word, the magnitude of interference in conflict trials increases relative to the control words (MacLeod, 1991). In one outcome of this experiment, the congruent trials would lead to faster subsequent memory search reaction time than the control trials, and the size of the interference from the conflict trials would increase. Such an outcome would suggest that the effect observed here was not specific to the Stroop paradigm. Rather, it might suggest that the latency of a particular instance of a speeded response is positively correlated with that of its immediate predecessor. Therefore, the effect observed here occurred simply because reaction time to Stroop conflict trials is typically longer than that to control trials. This interpretation might be confirmed by a second experiment, in which the Stroop task was replaced by a choice-reaction time task. The number of alternative responses would be varied between, for example, two and five. Unlike in the Stroop task, the type of processing required here varies only in quantity, not in quality. If varying the “set size” in the choice reaction time task lead to a variation in memory search reaction times, then this would confirm the “correlated response latency” account described above.

Alternatively, in the experiment where congruent Stroop trials are included, memory search response times may not be found to vary depending on whether they are

---

2 Weak evidence against this account is provided by the marginally significant increase in the memory search error rate when a pitch memory search trial follows a conflict Stroop trial, compared to a control Stroop trial. This is not predicted by the “correlated response latency” account.
preceded by congruent trials or control trials. This would suggest that the interference effect observed here is somehow specific to the Stroop conflict condition. One possibility is an interpretation in terms of task-switching costs. Wylie & Allport (2000) asked subjects to perform several trials of the colour naming Stroop task consecutively, followed by several word naming (“reverse Stroop”) trials, and then more Stroop trials e.t.c. In the situation where the type of trial had just switched, reaction time was increased. It was thought that this might represent proactive interference from a conflicting task. It might be the case that switching from the conflict condition of the Stroop task to memory search incurs a greater switching cost than switching from the control condition. Such task-switching costs persist when the interval before the switch trial is 1 s or more (e.g., Rogers & Monsell, 1995; Wylie & Allport, 2000). In the current experiment, the interval between the response to the Stroop trial, and the presentation of the memory probe item was always 500 msec. If this interval were varied, then the task-switching account would predict that this should not affect the magnitude of the effect. However, if reducing the interval increased the magnitude of the effect, and increasing it abolished the effect, then this would be inconsistent with the task-switching hypothesis. Such a result would be consistent with the possibility that performing in the Stroop conflict condition, and memory search shared a specific resource, and that this resource was affected by something analogous to a refractory period.
4.1 Aims of Chapter

In Chapter 3 it was demonstrated that the unexpected effect of retaining pitch information on temporal judgement, obtained in Experiment 2, was due to the processing requirements of remembering that information. It was also shown that retaining pitch information did not differ from retaining digits in terms of the attentional requirements of the two tasks. Therefore, no specific explanation was found of the contradiction between the findings of Fortin & Breton (1995) with digits, and those of Experiment 2 with pitch. The approach in the present chapter is to attempt to further isolate the unexplained effect. Logically, it is possible that digits are a unique case, or that pitch is a unique case, or that they represent two classes of memoranda, which differ in their effects upon temporal judgement. Therefore, one experiment replicates the conditions of Experiment 2 using an arbitrary visual memory set, while another experiment uses a memory set varying in an auditory dimension other than pitch. The final experiment in the present chapter broadens the enquiry by attempting to maximise the interference with a learned time production task that is caused by a secondary memory load task. This experiment uses a memory set that is itself composed of values of duration. Hence, subjects are required to perform two tasks involving temporal processing simultaneously. Furthermore, this paradigm allows some properties of the short-term memory component of the PGC model to be determined.
4.2 Experiment 6

The pitch memory set was sampled from a continuum, and might be thought of as pre-categorical. There are an infinite possible number of such pitch memory sets, and there is no reason to think that subjects could use categories or labels to easily remember the differences between different sets of items. However, the digit memory set is drawn from the set of integers, which are discrete rather than continuous, and certainly post-categorical. This was also true of the memory set composed of consonants, which was used by Fortin & Masse (1999). The consonant memory set produced results compatible with those of Fortin & Breton (1995), rather than those of Experiment 2. It is possible that the distinction between pre and post categorical is the critical variable that determines whether a concurrent memory load will interfere with temporal judgement.

The aim of the present experiment was to replicate the conditions of Experiment 2 using a memory set varying on a continuous sensory dimension other than pitch. If maintenance only of another such memory set produced similar interference to that in Experiment 2, then this would constitute preliminary evidence in favour of the “pre and post categorical” hypothesis. The sensory dimension chosen was colour. Colour is not a simple function of one physical dimension in the way that pitch is a function of physical frequency, except where variation only occurs within one colour category. For this reason a memory set was used that was composed of eight different shades of the colour grey.

There is a long history of research investigating various aspects of colour memory. An early example is that of Collins (1932), who found very little memory decay in 5 seconds for blue and yellow, but rather more for red and green. More recently,
Vartanov, Manukyan, Sokolov, & Tsakonas (1995) used a multidimensional scaling technique on judgements of the similarity of stimulus pairs to assess the perceptual colour space. The perceptual colour space was also assessed with retention intervals of 4, 15, or 60 seconds. It was found that the basic shape of the colour space remained the same as the retention interval increased, but that the accuracy of judgements was reduced. Fargo (1996) used a retroactive interference paradigm for assessing colour memory analogous to that used by Deutsch (1970) with pitch. Same / Different judgements were made, with a five second retention interval filled with other colour stimuli, and a retroactive interference effect was demonstrated whether or not the task-irrelevant stimuli were within the same colour category as the to-be remembered item. A proactive interference paradigm was also employed, in which the irrelevant stimuli were presented before the to-be-remembered stimulus. In this case, the effect only occurred for task-irrelevant stimuli within the same colour category as the to-be-remembered item. Fargo (1996) attributes this difference to the fact subjects were able to use a verbal recoding strategy when colours from many categories were present, whereas it is claimed that the within category effects demonstrate the existence of visually based memory for colour.

The aims of Experiment 6 may be summarised as follows:

- To investigate whether the set size effect of maintaining a memory set upon concurrent time production, which was found in the case of a pitch memory set, but not in the case of a phonological memory set (Fortin & Breton, 1995), also occurs for other memory sets based upon perceptual dimensions
• To demonstrate that the difficulty of a memory task, defined in terms of reaction time with set size slope and error rate, does not determine whether or not that task will produce a set size effect upon time production

4.2.1 Method

Subjects

Eleven subjects, aged between 21 and 30 took part in the experiment. As an incentive, the three subjects with the lowest error rates were awarded cash prizes.

Stimuli

In order to create an appropriate analogue of the pitch memory set, eight shades of the colour grey were used. Perceptual equidistance of each stimulus from the next in the series was insured by using the Munsell® colour notation system. The Munsell® notation system defines colours along the dimensions Value (brightness), Chroma (saturation), and Hue (dominant wavelength). Each of these attributes varies between 1 and 10. In order to produce grey, Chroma and Hue are set to zero, and Value is set to be greater than zero. The eight settings of Value used were 1.1, 2.2, 3.3, 4.4, 5.5, 6.6, 7.7, and 8.8. Using these settings, the darkest and lightest greys used were as far away from black and white as the members of the set were from each other. The stimuli were displayed as circles in the centre of the computer monitor. The background consisted of many small grey squares. These were arranged in a new random pattern for each subject. The greys used
for the background were different from those used in the memory set, but their mean Munsell® value was equal to the mean Munsell® value of the memory-set. The values employed were 1.3, 2.4, 3.7, 4.6, 5.3, 6.4, 7.5, and 8.4.

In order to display colours on a modern computer monitor a value is passed to each of the three colour guns (red, green, and blue, hereafter RGB). Because monitors differ considerably, and each monitor has a range of settings of brightness and contrast etc, it is not possible to specify exact RGB values that will correspond to particular Munsell® values. Rather, the values for each monitor must be calculated based upon calibration with a colorimeter. Once the gamma function for each of the colour guns is obtained, then the appropriate RGB values to correspond to particular Munsell® values may be calculated. A full discussion of this is provided by Hunt (1987). The monitor used in the present experiment was situated in a completely dark room for both the calibration and the experiment. The brightness control was set to minimum, and the contrast was set to maximum. These settings produce the best approximation of black.

It should be noted that while the procedure described above was followed in order to create the stimuli used in the experiment, all that was really required was eight stimuli that would produce a set size effect in the context of a memory search task. To minimise the difference in difficulty level between different memory sets of the same set size, it is helpful if the set members are subjectively equally spaced. This criterion would most likely have been achieved if arbitrary equally spaced RGB values had been chosen. Therefore, in the event, the procedure employed in order to make the colours used conform to Munsell® values was not, strictly speaking, necessary to satisfy the aims of the experiment. However, it did result in perceptually equal steps between the stimuli.
Each memory set item was presented for 1.5 s. There was no interval between successive items. The trial by trial procedure for selection of the memory set, and probe item, was identical to that used in Experiment 2.

Design and Procedure

The time production training procedure and the arrangement of experimental trials was identical to that in Experiment 2. This meant that in the experimental trials, subjects were presented with the memory set, started time production, waited appropriately, ended time production, were presented then with the memory probe item, and responded by pressing the left button if they thought the probe item was positive or the right button if they thought it was negative. Finally, feedback was given on the memory search component of the task only. The critical feature of the design was that subjects had to retain the memory set for the entire duration of time production.

4.2.3 Results

The overall error rate in the memory search component of the task was 21.7%, compared to 17.6% in Experiment 2. The details of this are presented in Table 4.1, from which it can be seen that the error rate increased with memory set-size. This was the only significant effect on the error rate, $F(3, 30) = 16.20, p < .001$. After trials on which a memory search error occurred were excluded, outlying data points of more than three standard deviations were removed. This amounted to 2.18% of the remaining memory search data, and 0.97% of the remaining time production data.
Table 4.1. Percentage memory search error rates in Experiment 6.

<table>
<thead>
<tr>
<th>Probe Type</th>
<th>Memory Set Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>6.4 (6.3)</td>
</tr>
<tr>
<td>Negative</td>
<td>10.6 (9.0)</td>
</tr>
</tbody>
</table>

In the final block of time production practice without feedback, the mean time produced was 2.53 s and the average standard deviation of performance was .21 s. Table 4.2 presents the mean times produced in each of the four experimental blocks, as well as the average within subject standard deviation of performance. The slight shortening of time production across the four experimental blocks did not reach significance. There was no reliable change in standard deviation as the experiment progressed.

Table 4.2. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of Experiment 6.

<table>
<thead>
<tr>
<th>Experimental Block</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>Time production</td>
</tr>
<tr>
<td>(seconds)</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
</tbody>
</table>

Figure 4.1 shows the relationship between the number of items in the memory set and time production. Contrary to the hypothesis, increasing the number of items to be remembered had no reliable effect upon time production, $F(3, 30) = 1.54, p > .1$, or variability of timing, $F(3, 30) = 0.71, p > .1$. 

143
Response times to the memory probe, which was presented after time production was complete, are shown in Figure 4.2. The effect of memory-set size was highly significant, $F(3, 30) = 12.70, p < .001$. The slope of the function was 116 msec per item, which was considerably steeper than either the 67 msec slope for pitch in the reaction time condition of Experiment 1, or the 61 msec slope in Experiment 2.
4.2.3 Discussion

Increasing the number of items in the “grey memory set” did not lead to an increase in concurrent time production such as that found with pitch in Experiments 2 and 3. Therefore, the hypothesis that any type of memory task involving the maintenance of absolute values of a continuum will lead to systematic interference with temporal judgement is not supported.

However, caution is required in the interpretation of these results because many of the subjects reported using verbal strategies to assist in remembering the items. If the verbal strategy was the dominant way of performing the task then the analogy with the pitch memory task would break down, and the hypothesis would not have been fairly tested. In this case, the processes involved in the memory set would be similar to those involved in remembering digits or consonants. However, it is unlikely that the verbal strategy was dominant. When asked how they had performed the task, most subjects
reported using verbal labels. They were typically able to list between five and eleven of these, and none of the subjects could correctly state that there were eight memory-set items. Furthermore, only two subjects reported verbally labelling the probe item before making the recognition decision. The other subjects all felt that the judgement was made at the visual level, implying the existence of a visual memory for the memory-set. Based upon this evidence it is possible to conclude only that a verbal strategy was used to support visual memory. Therefore, the validity of the experiment may be accepted with a cautionary note. Experiment 7 will provide a less ambiguous test of the “pre versus post categorical” hypothesis.

The memory search task produced a higher error rate, and a steeper reaction time function, than that produced by the pitch memory set under equivalent conditions in Experiment 2. In this sense, it was a more difficult task. Despite the relative increase in the difficulty level of the secondary task there was no set size effect upon time production. Therefore, the results of this experiment constitute further evidence against the “difficulty account” of the unexplained effect of pitch memory load on time production, which was considered in section 3.4.

4.3 Experiment 7

Experiment 6 provided evidence that the distinction between memory tasks based on pre-categorical sensory dimensions, and post-categorical verbal memory tasks, is not a useful one in accounting for the results of Experiment 2. However, doubts remain as to whether the hypothesis was tested fairly because subjects in Experiment 6 tried to use
verbal strategies. The first aim of Experiment 7 was to use a different memory set, which would more clearly satisfy the pre-categorical requirement.

Assuming for the moment that the results of Experiment 6 were valid, other possible reasons for the results of Experiment 2 may be considered. The non-interfering memory-sets used by Fortin & Breton (1995), and Fortin & Masse (1999) were both visually presented, although it is assumed that the process of remembering the items involved an abstract phonological code. The non-interfering memory set employed in Experiment 6 was also visual in nature. However, the interfering memory set employed in Experiments 2 and 3 was auditory in nature. It is possible that this modality difference is the critical dimension determining the interference or non-interference of a memory load with temporal judgement. Therefore, the second aim of Experiment 7 is to test the "modality" hypothesis by replicating the conditions of Experiment 2 with a memory set varying in some other dimension of the auditory modality. If the "modality" hypothesis is correct then this memory set will also produce a set size effect upon time production.

In order to satisfy both aims of the experiment a memory set was required that was both demonstrably pre-categorical, as well as in the auditory modality. A positive result, in which a set size effect occurred, would require further experimentation, in order to reject either the "modality" or the "pre versus post categorical" hypothesis. However, a negative result, in which no set size effect occurs, would effectively rule out both hypotheses.

One obvious choice of stimulus dimension for the present experiment was loudness. However, it was not possible to successfully pilot such a task. Instead, a memory for timbre task was constructed, based upon tasks used in the existing literature,
which is briefly reviewed below. One advantage of a timbre based memory task is that the pitch of the memory items may be held constant. Therefore, there is no pitch memory load, and any set-size effect that occurs cannot be due to an uncontrolled pitch component in the task.

*Short-term memory for timbre*

Timbre is that quality of a sound by which it may be distinguished from another sound of identical pitch and intensity. For example, two musical instruments may play the same note at the same intensity, and yet they are still easily distinguishable from each other. Subjective pitch is a direct, if non-linear, function of physical frequency. Subjective timbre is not simply determined in this way. Various attributes of a sound can affect its timbre, and these factors may interact in complex ways. These factors include the attack and decay times of various components of a sound, as well as the spectral composition\(^1\). For the purposes of the present experiment, varying the attack and decay times was undesirable because their presence would encourage subjects to use temporal cues to distinguish the memory set items from each other. This would represent a situation in which two temporal tasks were being performed concurrently, making interference extremely likely to occur. This issue, which is separate and distinct from those raised by Experiment 2, is addressed in Experiment 8. In order to create a memory set comparable to that used in Experiment 2 some aspect of a sound must be varied which

---

\(^1\) Pure tones, such as those used in Experiment 2, are composed of a single waveform. Complex tones are composed of multiple frequencies of sound. The perceived pitch of a complex tone corresponds to its fundamental frequency, which is always the lowest frequency component of the sound. Other waveforms present in the sound are typically harmonics of the fundamental. “Spectral composition” refers to variations in these harmonics.
is not critically dependant upon temporal cues for its perception, as was also the case for pitch. Fortunately, the spectral composition of a sound may be varied while attack and decay times are kept constant, resulting in a systematically varying percept.

Two previous investigations of short-term memory for timbre have used the method of varying timbre by varying the spectral composition of a sound. Semal & Demany (1991) used the Deutsch interpolated tones paradigm. Subjects made same / different pitch judgements on a pair of tones separated by a retention interval of 4.3 s. Interpolated tones were either remote in pitch to the test tones, or similar in pitch. Likewise, they were either similar in timbre, or markedly different in timbre. Tones closer in pitch to the test tones produced greater error rates in the pitch judgement task, but the manipulation of timbre had no effect. The authors concluded that the memory system specialised for storing pitch does not store timbre. A complimentary set of experiments was carried out by Starr & Pitt (1997). In this research the same / different judgement was made based on the timbral similarity of the test tones. The interpolated tones were varied in terms of timbre and fundamental frequency (pitch). Varying the timbre of the interpolated tones affected the error rate, but there were no effects of the pitch manipulation. Following the logic of the Deutsch experiments, this evidence may be seen as suggestive of a distinct storage capacity for timbre.

For the purposes of the present experiment a memory-set composed of eight items was generated by varying the spectral composition of a tone with a fixed fundamental frequency. Clearly, it was desirable to make the memory search task with the new tones as comparable as possible to the pitch memory search task. It was hoped that this might be achieved if the perceptual distance of each set member from the next in the series was
comparable to that for the pitch memory set. Therefore, as part of piloting, a group of subjects performed a similarity judgement task on all the possible pairs of tones within each memory set.

The aims of Experiment 7 may be summarised as follows:

- To establish that a memory set composed of tones varying in timbre was perceived on a continuum of timbre analogous to the continuum of pitch used for the memory set in Experiments 1 to 3.
- To test the hypothesis that the set size effect of maintaining pitch upon concurrent time production that occurred in Experiments 2 and 3 might also occur for other memory sets of an auditory nature.

4.3.1 Method

Subjects

Twelve subjects, aged between 21 and 33 took part in the main experiment. As an incentive, the three subjects with the lowest error rates were awarded cash prizes. Twelve unpaid subjects aged between 21 and 45 took part in the similarity judgement task.

Stimuli

All of the eight tones in the timbre memory set had a fundamental frequency of 200 Hz. Therefore, they were identical to each other in pitch. Each tone had two harmonic components. The frequency of the harmonics was systematically varied in order to produce a range of sounds that sounded like different voices of an early electronic synthesiser. However, it should be noted that the sounds were all subjectively
close to their immediate neighbours, and there was no obvious categorical distinction between any of the sounds. Table 4.3 presents the frequencies of the harmonics that were used to create the memory set. The physical steps in the harmonics were equal from tone to tone. Thus, the first tone consisted of the fundamental plus the 1st and 2nd harmonics, the second tone consisted of the fundamental plus the 4th and 5th harmonics etc. The amplitude (intensity) of each of the two harmonics was equal to that of the fundamental.

Table 4.3. The frequency of the harmonics (Hz) which were combined with a 200 Hz pure tone to produce the memory set for Experiment 7.

<table>
<thead>
<tr>
<th>Tone</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Harmonic</td>
<td>400</td>
<td>1000</td>
<td>1600</td>
<td>2200</td>
<td>2800</td>
<td>3400</td>
<td>4000</td>
<td>4600</td>
</tr>
<tr>
<td>2nd Harmonic</td>
<td>600</td>
<td>1200</td>
<td>1800</td>
<td>2400</td>
<td>3000</td>
<td>3600</td>
<td>4200</td>
<td>4800</td>
</tr>
</tbody>
</table>

Each memory set item was presented for 500 msec, with 500 msec pauses between items, as in Experiment 2. The trial by trial procedure for selection of the memory set, and probe item, was identical to that used in Experiment 2.

*Design and Procedure*

The time production training procedure and the arrangement of experimental trials in the main experiment was identical to that in Experiment 2. This meant that in the experimental trials, subjects were presented with the memory set, started time production, waited appropriately, ended time production, were presented then with the memory probe item, and responded by pressing the left button if they thought the probe item was positive or the right button if they though it was negative. Finally, feedback was given on
the memory search component of the task only. The critical feature of the design was that subjects had to retain the memory set for the entire duration of time production.

Half the subjects who did the rating procedure rated the pitch stimuli first, while the other half rated the timbre stimuli first. For each type of stimulus, there was a short practice session. There were 28 pairs of stimuli to be rated for each stimulus type, and each subject provided three judgements of each pair. The order of presentation of the pairs was randomised, as was the order of presentation of stimuli within each pair. Each stimulus was presented for 500 msec, with a 500 msec pause between stimuli. Responses were made on a scale of 1 to 7 using the numeric keypad. The response 1 was labelled as “most similar”, while the response 7 was labelled as “least similar”.

4.3.2 Results

Similarity ratings

For each subject the mean rating across each of the seven pairs where the physical difference was one step was calculated. The same procedure was adopted for the six pairs where the physical difference was two steps and the five pairs where the physical difference was three steps etc, finishing with the single pair with a physical difference of seven steps. This procedure was conducted separately for the pitch and timbre sets. Figure 4.3 presents the mean similarity ratings plotted against physical distance for the two sets of stimuli. The effect of pair distance was highly significant, $F (6, 66) = 94.74, p < .001$. Additionally, the timbre stimuli were rated as more similar to each other by a small amount than the pitch stimuli were, $F (1, 11) = 6.85, p < .05$. Inspection of Figure
4.3 reveals that the change in subjective similarity corresponding to a one step change in physical similarity was equal for the two types of stimuli, and that the relationship between pair distance and subjective similarity was linear. This was reflected in a non-significant interaction between pair distance and stimulus set, F < 1. This suggested that memory search tasks with the two types of stimuli would be of roughly equivalent difficulty. If anything, the fact that the timbre stimuli were, on average, subjectively more similar to each other than the pitch stimuli suggested that the timbre memory search task would be more difficult.

![Graph showing similarity ratings of two sets of stimuli](image)

Figure 4.3. Similarity ratings of the two sets of stimuli, as a function of physical similarity (pair distance). 1 represents a judgement of “most similar”.

**Timbre memory load concurrent with timing**

The overall error rate in the memory search component of the task was 18.7%, compared to 17.6% in Experiment 2. The details of this are presented in Table 4.4, from
which it can be seen that the error rate increased with memory set-size. This was the only significant effect on the error rate, $F (3, 33) = 18.70, p < .001$. After trials on which a memory search error occurred were excluded, outlying data points of more than three standard deviations were removed. This amounted to 1.44% of the remaining memory search data, and 0.32% of the remaining time production data.

Table 4.4. Percentage memory search error rates in the timbre memory search task used in Experiment 7.

<table>
<thead>
<tr>
<th>Memory Set Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>9.0 (8.7)</td>
<td>15.6 (11.8)</td>
<td>20.5 (14.8)</td>
<td>24.3 (13.6)</td>
</tr>
<tr>
<td>Negative</td>
<td>9.7 (7.0)</td>
<td>19.1 (12.9)</td>
<td>19.4 (9.1)</td>
<td>31.9 (11.4)</td>
</tr>
</tbody>
</table>

In the final block of time production practice without feedback, the mean time produced was 2.57 s and the average standard deviation of performance was .16 s. Table 4.5 presents the mean times produced in each of the four experimental blocks, as well as the average within subject standard deviation of performance. The slight shortening of time production across the four experimental blocks reached significance, $F (3, 33) = 3.50, p < .05$. There was no reliable change in standard deviation as the experiment progressed.
Table 4.5. Mean time production, and average within subject standard deviation of
time production in each of the four experimental blocks of Experiment 7.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time production (seconds)</td>
<td>2.64</td>
<td>2.59</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.22</td>
<td>0.22</td>
<td>0.19</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Figure 4.4 shows the relationship between the number of items in the memory set
and time production. Contrary to both the “modality” hypothesis and the “pre versus post
categorical hypothesis”, increasing the number of items to be remembered had no reliable
effect upon time production, F (3, 33) = 0.62, $p > .1$, or variability of timing, F (3, 33) =
0.70, $p > .1$.

Figure 4.4. Time production and variability of performance as a function of timbre memory load in Experiment 7. The
left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean
standard deviation of time production.
Response times to the memory probe, which was presented after time production was complete, are shown in Figure 4.5. The effect of memory-set size was highly significant, $F(3, 33) = 8.45, p < .001$. The slope of the function was 41 msec per extra item, which was less than either the 67 msec slope for pitch in the reaction time condition of Experiment 1, or the 61 msec in Experiment 2.

![Figure 4.5. Reaction time to the timbre memory probe, which was presented after time production, in Experiment 7.](image)

### 4.3.3 Discussion

The results of the similarity judgement task suggested that the timbre memory search task would be comparable to its pitch analogue. This was born out by the error rate, which was very similar to that of the pitch task in Experiment 2. However, the slope of the reaction time function was less than that of the pitch memory task. There is no obvious process theory which would predict this difference, although it is consistent with the view that pitch and timbre are distinct from each other within auditory short-term
memory. Despite the slope difference, it seems reasonable to suggest that the two memory tasks were comparable in difficulty on the basis of the error rates.

The "modality" hypothesis suggested that the timbre memory task would produce a set size effect upon time production, similar to that for pitch, on the basis that there may have be some unspecified linkage between auditory short-term memory and the memory processes involved in timing. The failure to obtain a set size effect upon time production with timbre suggests that the effect is a highly specific interaction between short-term memory for pitch and temporal judgement. The apparent dissociation between the effect of pitch, and the lack of an effect of timbre is remarkable because the two experiments were identical in every respect apart from the nature of the tones to be remembered. Because the two experiments were so similar, the difference in the results must be accounted for by the involvement of a resource in the pitch task that is not involved in the timbre task, and this resource must be shared with timing. However, it is unclear what this resource might be, or what the functional or adaptive significance of such an overlap between two apparently unconnected functions could be.

In terms of the PGC model, the effect of remembering pitch information upon time production may be seen as interference with the process of pulse accumulation. The only other circumstance under which the PGC model predicts lengthened time production would be if the pacemaker were slowed down. There are no theoretical grounds to think that remembering pitch information — and not remembering timbre — would cause this to happen. The possibility of interference due to an overlap of memory buffer systems was discussed, and found to be implausible, in section 2.4.3. Nothing in the results of Experiments 3 – 7 suggests a change to that conclusion. If the effect of pitch memory
load upon time production is to be seen as an interruption — or slowing — of the rate of pulse accumulation, then this suggests the operation of gate component of the PGC model. However, this conclusion is not supported by the result of Experiment 5, which suggested that remembering pitch information did not require a greater amount of attention than remembering digits. Further suggestions relating to this unexplained effect will be made in the general discussion.

4.4 Experiment 8

Remembering pitch information produces a set size effect upon time production, which is hard to explain through a sharing of memory buffers. However, the idea of a memory buffer overlap suggests another experiment. What would happen if subjects were required to remember one or more event durations while performing the production of a learned interval?

The PGC model proposes that a counter or accumulator, which is a form of short-term memory, is involved in the production of learned intervals. Pulses from the pacemaker accumulate in the counter for comparison with values from long-term memory. The PGC model also proposes that when two intervals are presented successively, for comparison with one another, the counter acts to keep the first interval available for comparison with the second. Therefore, if the production of a learned interval were performed concurrently with the retention of information about one or more test intervals, the PGC model predicts the disruption of either, or both, tasks.
Furthermore, because a definite sharing of resources is predicted, this interference is likely to be greater than that caused by the pitch memory task.

The nature of the interference caused by remembering the duration of one or more intervals while performing a learned time production might provide a clue to the capacity or structure of short-term memory for duration. If short-term memory for duration were subject to a capacity limit in terms of duration itself, then the number of items remembered concurrently with learned time production should have no effect upon time production. Rather, the total magnitude (in terms of duration) of the concurrent memory load would determine the amount of interference. As an illustration, remembering one value of 1200 msec would produce more interference than a set composed of 200 msec, 400 msec, and 300 msec items. Such a result would be consistent with a model in which short-term memory for duration was like a jug, and the pulses from the pacemaker were like the water that may fill the jug. According to this view, event duration may be represented as the actual duration of some specific set of processes in the brain. The alternative possibility is that remembering more items will produce more interference, independent of magnitude, as was the case with pitch. Such a result would suggest that short-term memory for event duration is item based. It would also make the view that duration is represented in short-term memory as the actual processing time of a neural process look less plausible. An alternative view is that some transformation of event duration is represented. Such transformed values could also form the basis of long-term memory for duration, and short-term memory for duration could then be seen as an activated subset of long-term memory. Note that according to this view the pacemaker still produces a representation of duration as duration, because it acts to segment time.
However, the output of the pacemaker may be transformed into a non-temporal analogue of duration within short-term memory.

Before addressing the question of what will happen when two separate temporal tasks are combined, it is necessary to measure performance on the two tasks when they are performed individually. Performance on the learned time production component of the task can be compared to that from previous experiments involving nontemporal secondary tasks. However, there was no such baseline available for the duration memory-search task. Therefore, the first part of Experiment 8 was a reaction time condition.

Deutsch's (1986) research on recognition memory for duration was described in section 3.2. To recap, Subjects are required to judge whether a test duration defined by two short blips is shorter, the same as, or longer than a standard duration. Irrelevant durations interpolated between the standard and test durations cause systematic biases of judgement. Interpolating values in the range of the standard, but a little shorter, seemed to "shorten" the memory of the standard, resulting in a tendency to judge "long" incorrectly. Furthermore, interpolating values in the range of the standard, but a little longer, seemed to "lengthen" the memory of the standard, resulting in a tendency to judge "short" incorrectly. In another condition, the interpolated duration was in the range of half the standard duration. As all of the interpolated stimuli were now shorter than the standard, it might seem likely because of the previous results that all of the judgements would be biased in the "long" direction. In fact, the comparison judgements were biased in the direction of double the interpolated duration. Thus, when the interpolated duration was slightly less than half the length of the standard there was a tendency to judge "long", but when the interpolated duration was slightly more than half the standard the bias was
towards “short”. This occurred irrespective of the fact that the interpolated duration was
much shorter than the standard. Deutsch (1986) suggests that this surprising result
occurred because temporal patterns are distorted by the listener so they appear closer to an
ideal pattern with a simple hierarchical relationship between the elements. It was also
suggested that the process of distortion takes place in short-term memory. These results
might be taken to imply a highly economical system of representation in short-term
memory for duration. In any given situation, one reference duration might be specified in
absolute terms, while all other relevant values would be represented as simple multiples
of the reference duration.

The participants in Deutsch’s (1986) research were music students. Duration in
western music is represented hierarchically, and Deutsch’s subjects were trained in this
hierarchical representational system. Therefore, her results may have been the result of
learning rather than inherent properties of the short-term memory for duration system.
Here, an attempt was made to replicate Deutsch’s remarkable finding, using a sample of
musically untrained subjects. However, the paradigm used was memory search, rather
than the interpolated stimuli paradigm. The hypothesis was tested by creating two
memory sets. In one memory set the relationship between the elements conformed to a
hierarchy, while in the other it did not. In the hierarchically organised memory set,
representing multiple items should result in less distortion of items than in the non-
hierarchical set. This is because Deutsch’s (1986) theory leads to the suggestion that the
items in the non-hierarchical set will become distorted to appear as if they are
hierarchically related. The consequences of such distortion should be a higher recognition
error rate for the non-hierarchical memory set.
The aims of the first part of Experiment 8 may be summarised as follows:

- To establish whether subjects can perform an item based memory search task, analogous to the pitch memory search task, in which the critical feature of the memory-set items is duration.
- To contrast the error rate on a version of the task in which the memory set items are hierarchically organised with a version in which they are not, the aim of this being to test Deutsch's (1986) model of memory for duration.

4.4.1 Method

Subjects

Nine unpaid subjects, varying in age between 20 and 34, took part in the reaction time condition.

Stimuli

All of the stimuli used were 2000Hz pure tones, presented using headphones. The values of duration, in milliseconds, used for the eight items in the hierarchical memory set were 18, 36, 72, 144, 288, 576, 1152, and 2304. The corresponding items in the non-hierarchical memory set were 19, 43, 78, 164, 290, 610, 1081, and 2305. These values were chosen so that that average duration of an item was equal for both sets. Additionally, the average gap size between adjacent set members was fixed. Therefore, the average discriminability of items from each other should have been approximately equal for the two sets. During the presentation of a memory set, individual items were
separated by 576 msec pauses. This value was the same as the value of one of the hierarchical set items. By choosing this value, the possibility of memory distortion being caused by the duration of the pause was avoided, at least for the hierarchical set. After a memory set was presented, a fixation dot appeared. The probe item was initiated by the subject pressing the middle mouse button. This caused the immediate onset of the probe item. Subjects were instructed not to respond until the probe item ceased to sound. Reaction time was measured from the offset of the probe item. Memory sets and probe items were selected in the same way as for previous experiments. However, trials on which the probe item was negative, and two or more steps longer than the longest set member, were excluded from the analysis. This was because it was possible to make the recognition decision prior to probe offset, resulting in an artificially fast reaction time.

Design and procedure

There were two testing sessions, one for the hierarchical memory set, and one for the non-hierarchical memory set. The order of testing was counterbalanced. Within each session there was one practice block, and four experimental blocks. Each block consisted of 48 trials, and within each block the order of trials was randomised. Therefore, the total number of experimental trials in each condition was 1728.

4.4.2 Results

The reaction time condition of Experiment 8 had two aims. Firstly, to compare error rates across the hierarchical and non-hierarchical memory sets, and secondly to
establish that a memory-search task was possible using values of duration. Before error rates were calculated, trials in which the probe item was negative, and two or more steps greater than the longest memory set item, were excluded\(^2\). The percentage error rates for the two conditions are shown in Table 4.6. It was predicted that the error rate would be higher for the non-hierarchical memory set. In fact, the error rate was higher for the hierarchical memory set (27.0% versus 24.7%). However, this difference was not reliable, F (1, 8) = 2.1, \(p > .1\). The error rate increased with set-size, F (3, 24) = 61.20, \(p < .001\), and was higher for negative probes than positive probes, F (1, 8) = 46.83, \(p < .001\).

<table>
<thead>
<tr>
<th>Set-Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>2.8 (2.9)</td>
<td>25.3 (18.2)</td>
<td>2.8 (2.9)</td>
<td>19.4 (7.0)</td>
</tr>
<tr>
<td></td>
<td>25.0 (8.3)</td>
<td>30.7 (9.2)</td>
<td>14.4 (8.4)</td>
<td>25.6 (8.2)</td>
</tr>
<tr>
<td></td>
<td>21.3 (8.9)</td>
<td>43.4 (13.5)</td>
<td>25.5 (8.2)</td>
<td>39.6 (11.3)</td>
</tr>
<tr>
<td></td>
<td>22.7 (10.2)</td>
<td>45.1 (9.9)</td>
<td>27.7 (10.2)</td>
<td>42.7 (14.2)</td>
</tr>
</tbody>
</table>

After errors were excluded, the issue of outlying values was addressed. In some cases, there were insufficient data points to use the usual three standard deviation criteria. As an alternative strategy, individual values greater than 4000 msec were reduced to 4000 msec. This procedure resulted in the alteration of 27 values — 1.1% of the remaining data. Additionally, 1 value of less than 100 msec was deleted from the data set.

\(^2\) This resulted in the removal of 39.2, 19.2, 11.6, and 5.3 % of the data for set sizes 1 to 4 respectively.
The response times to the memory probe item, for both the hierarchical and non-hierarchical memory sets are shown in Figure 4.6. There was no difference between the response time functions of the two memory sets, and no effect of probe type upon response time. Averaging across the two types of memory set, the slope of the function was 104 msec per item beyond the first. This effect was highly significant, $F(3, 24) = 16.87, p < .001$.

![Figure 4.6. The effects of duration memory set size, and memory set type on reaction time in Experiment 8.](image)

4.4.3 Discussion

The error rate was very high for both the hierarchical and non-hierarchical memory sets. In particular, there was a very steep rise from set size 1 to 2. This could be taken to suggest that short-term memory for duration can only operate effectively to maintain a single item. However, reaction time did increase steeply as a function of the
number of items to be searched. This confirms that information about a number of items was maintainable, although it was probably far from precise.

The findings of Deutsch (1986) suggested that the error rate would be higher for the non-hierarchical memory set because distortion occurs between items in short-term memory, in order to make them conform to a hierarchical representational scheme. The data provide no support for this theory, but neither do they falsify it. The failure to find a difference in error rates in the predicted direction may have been due to a number of factors. One likely possibility is that the overall error rate, and the variability associated with the error rate of the task was simply so high that it disguised any additional effects of memory distortion specific to the non-hierarchical memory set. Another possibility is that the differences between the two memory sets were too small to allow the effect to emerge. The non-hierarchical memory set was constrained to have the same average item value, and the same average interval between items as the hierarchical set. If these constraints were relaxed, in order to allow greater differences between the two sets, then the predicted effect might emerge. A final possibility is that Deutsch’s effect was a consequence of the fact her subjects were trained in western tonal music. This hypothesis might be explored using the current paradigm by providing one group of subjects with some form of training on the hierarchical memory set, while training a second set of subjects on the non-hierarchical memory set. It might be predicted that the effects of training relative to an untrained control group would be greater for a subsequent memory search task with the trained set than the untrained set. Furthermore, this effect should occur independently of which set was trained.
4.4.4 Maintaining duration concurrently with time production

The first part of Experiment 8 established that subjects can perform a duration memory search task, albeit with a high error rate, and that the time to make a recognition judgement lengthens as a function of the number of comparisons that must be made. The second part of Experiment 8 investigates what happens when the duration memory set must be maintained whilst producing a learned interval, as was performed with pitch in Experiment 2. Only the hierarchical memory set was used. Because of the high error rate found for the memory search task, only set sizes 1 to 3 were employed in the combined task.

In the General Introduction, research performed by Brown & West (1990) was described. Briefly, subjects were required to attend to multiple stimuli, and then reproduce one of them. As the number of attended stimuli increased, reproduction became more variable and absolute error of reproduction (only) increased. Because the paradigm employed here is so different, particularly in terms of the memory processes, it is hard to derive predictions for the current experiment from Brown & West (1990).

Based on the considerations discussed earlier, it was predicted that substantial interference with time production would occur due the requirement to remember values of duration concurrently. Whether this interference would be related primarily to the number of items concurrently remembered, or to the magnitude of the concurrent memory load was considered as an exploratory hypothesis.

The aims of the second part of Experiment 8 may be summarised as follows:
To establish whether performing two duration based tasks simultaneously would result in severe disruption to the time production task, to be assessed by comparing variability of time production here with that in other experiments.

To compare the magnitude of the interference with mean time production of retaining values of duration while performing time production with that of concurrently remembering pitch, observed in Experiments 2 and 3.

To investigate whether interference with concurrent time production caused by retaining values of duration would be primarily related to the number of items retained or their total magnitude (msec).

### 4.4.5 Method

#### Subjects

Fourteen subjects, aged between 18 and 33, took part in the experiment. They received course credit in return for their participation.

#### Stimuli

The hierarchical memory set was used. The details of this memory-set are described in Section 4.4.1.
Design and procedure

The time production training procedure was identical to that used in earlier experiments. There were four blocks of 48 experimental trials. Therefore, each subject received 32 trials in each condition of the experiment. This number of trials was greater than the usual 24 trials because a high error rate was anticipated. The total number of trials in the experiment was 2660.

A single trial involved the serial presentation of the memory set, followed by a fixation dot indicating that the subject could start time production when ready. The probe item was presented immediately after the subject ended time production, as in Experiment 2. The recognition response was made once the probe item had ceased to sound. The main aim of this procedure was to require subjects to retain the memory-set throughout time production. Thus, they were required to perform two duration based tasks concurrently. Trials on which the probe item was negative, and two or more steps longer than the longest set member, were excluded from the analysis of the recognition judgement data. However, they were only excluded from the time production analysis if the recognition judgement was incorrect.

4.4.6 Results

The overall error rate on the memory search task was 27.3%, which was very similar to that in the reaction time condition. The details of this are presented in Table

---

3 This resulted in the removal of 38.2, 19.2, and 12.5 % of the data at set sizes 1,2, and 3 respectively.
4.7. The effect of set size was highly significant, $F(2, 26) = 23.38, p < .001$, as was that of probe type, $F(1, 13) = 13.85, p < .01$. After trials on which a recognition error occurred were excluded, 1.5% of the remaining time production data, and 2.2% of the reaction time data was excluded using the 3 standard deviation criteria for outliers.

Table 4.7. Percentage error rates in the concurrent time production condition of Experiment 8, in which values of duration were maintained concurrently with temporal processing.

<table>
<thead>
<tr>
<th>Memory set size</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>11.8 (7.1)</td>
<td>23.7 (10.6)</td>
<td>23.7 (10.5)</td>
</tr>
<tr>
<td>Negative</td>
<td>25.4 (8.6)</td>
<td>39.2 (11.9)</td>
<td>40.2 (5.3)</td>
</tr>
</tbody>
</table>

In the final time production training block, the mean time produced was 2.68 sec, and the average standard deviation of timing was .26 s. The mean times produced in each of the four experimental blocks, as well as the average within subject standard deviations of time production are shown in Table 4.8. The shortening of time production across the four experimental blocks was significant, $F(3, 39) = 4.20, p < .05$. There was no reliable change in variability of timing as the experiment progressed.

Table 4.8. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of Experiment 8.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time production</td>
<td>2.63</td>
<td>2.61</td>
<td>2.54</td>
<td>2.49</td>
</tr>
<tr>
<td>(seconds)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.27</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
</tr>
</tbody>
</table>
The hypothesis that interference with time production due to the need to concurrently maintain a memory set would be determined by the total magnitude of that memory set, rather than the number of items, was assessed first. In order to do this the magnitude of each memory set was first calculated, by simply adding up the values, in milliseconds, of the individual items. The resulting variable was correlated with values of time production, for each subject separately. The results of this analysis are shown in Table 4.9, inspection of which shows that there were small, but reliable, correlations for three of the fourteen subjects.

The alternative hypothesis, that interference with time production from the secondary memory task would be determined by the number of items remembered, was assessed in the same way as for pitch, timbre, and greyscale previously. As the number of items remembered increased time production lengthened\(^4\). This is shown in Figure 4.7. The effect was larger than that caused by remembering pitch information. The difference between set size 1 and 2 was 19 msec, and the difference between set size 2 and 3 was 62 msec, compared to approximately 22 msec per additional item in Experiment 2. This effect was highly significant, \(F (2, 26) = 8.87, p < .001\). The effect of increasing the number of items on the variability of timing approached conventional significance levels in the omnibus ANOVA, \(F (2, 26) = 2.87, p > .05\). However, the linear contrast was reliable, \(F (1, 13) = 5.00, p < .05\).

\(^4\) This result is inconsistent with that of Brown & West (1990), who did not find any bias in mean temporal judgement using reproduction or production methods when multiple intervals were attended to. However, their paradigm was not directly comparable to that employed here.
Table 4.9. The correlation of magnitude of duration memory load with time production in Experiment 8. Significance levels are for 2 tailed predictions.

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.22</td>
<td>0.09</td>
<td>0.13</td>
<td>0.08</td>
<td>0.08</td>
<td>0.24</td>
<td>-0.03</td>
</tr>
<tr>
<td>N</td>
<td>136</td>
<td>135</td>
<td>133</td>
<td>128</td>
<td>134</td>
<td>132</td>
<td>145</td>
</tr>
<tr>
<td>Significance</td>
<td>&lt;.05</td>
<td>&lt;.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.11</td>
<td>0.15</td>
<td>0.05</td>
<td>0.24</td>
<td>-0.09</td>
<td>0.07</td>
<td>-0.11</td>
</tr>
<tr>
<td>N</td>
<td>138</td>
<td>127</td>
<td>144</td>
<td>146</td>
<td>140</td>
<td>138</td>
<td>143</td>
</tr>
<tr>
<td>Significance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.01</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.7. Time production and variability of performance as a function of duration memory load in Experiment 8. The left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean standard deviation of time production.
Response times to the memory probe, which was presented after time production was complete, are shown in Figure 4.8. The effect of memory-set size was significant, $F(2, 26) = 7.64, p < .01$, while the effect of probe type was highly significant, $F(2, 26) = 19.27, p < .001$. The difference in reaction time between set size 1 and set size 3 was 229 msec, compared to 213 msec for the equivalent conditions in the reaction time only condition reported earlier.

![Figure 4.8. Response times to the duration memory probe, which was presented after time production in Experiment 8](image)

4.4.7 Discussion

The requirement to remember values of duration while also producing a learned interval biased subjects towards producing increasingly longer intervals as the number of items in the concurrent memory load was increased. There was very little evidence that the total magnitude of duration remembered determined the extent of biasing. This
pattern of results is consistent with the idea that short-term memory for duration operates on transformed values of duration, rather than representing duration as the processing time of particular neural events.

The nature of the disruption of time production was systematic. Both the mean time produced and the standard deviation of time production was increased by increasing the number of items concurrently remembered. However, overall variability did not change compared to that observed in Experiments 2, 6 and 7, where a similar procedure was used with different memoranda. A more serious disruption might have manifested itself in a considerable increase in variability of timing as the number of concurrently remembered items was increased. The disruption to timing was systematic and incremental rather than catastrophic.

The magnitude of interference was greater than that produced by the pitch memory task in Experiments 2 and 3. If the two effects had been similar in magnitude, then it might have been suggested that the two kinds of information were very closely linked within short-term memory. Because the interference is smaller when the concurrent memory load is pitch, rather than duration, based, this suggests that the two memory systems are separate, but functionally linked at some level. This possibility will be taken up in the General Discussion, where literature published after the current research was completed will be considered.

How can the PGC model account for the current findings? Based on the available evidence only speculation is possible. As was the case for pitch, the gate component of the model does not provide a convenient explanation of this effect. This is because the items were maintained for the entire duration of time production. The gate was included
in the model to account for the effect of discrete interruptions to the timing process, such as those associated with concurrent memory search (as opposed to maintenance). Memory search is known to take only a few hundred milliseconds, as opposed to maintenance, which takes place throughout time production. Therefore, it is not easy to think of concurrent memory loads as interrupting the flow of pulses from the pacemaker to the counter.

However, it is possible that placing one demand upon short-term memory for duration reduces its efficiency to deal with a second demand. Assuming resources are required to keep one value of duration “online”, fewer resources are then available to register a second duration. This can be conceptualised within the model as a failure to register a proportion of the pulses from the pacemaker. This would result in more real time elapsing before the target value for a learned time production was reached, consistent with the present data. This situation is functionally equivalent to a slowing of the pacemaker rate. The effects of changes in the rate of the pacemaker are discussed in Section 5.3. Briefly; pacemaker rate changes are associated with changes in mean time production and the associated variability of performance. However, changes in mean time production associated with the operation of the gate are not accompanied by changes in variability — see section 5.2. The significant change in variability found here supports the suggestion that the effect on mean time produced was due to a slowing of the accumulation rate, rather than an interruption of the accumulation process.
5.1 Aims of Chapter

The experiments that were reported in Chapters 3 and 4 were an attempt to clarify the unexpected results of Experiment 2. As such, they were not designed to test particular predictions made by the PGC model. The main aim of the experiments reported in Chapter 5 was to directly test the PGC model. The PGC model predicts that when time production is lengthened due to a concurrent secondary task, such as memory search (but not maintenance), variability of timing will not increase. However, the results of Experiment 1 were ambiguous regarding this prediction. Experiment 9 was designed to provide a stronger test of the prediction. By increasing the size of the effect of the secondary task on mean time produced, any associated changes in variability would be magnified compared to those in Experiment 1. Additionally, a control condition was included that was designed to demonstrate that the time production method is sufficiently sensitive to detect differences in variability of performance between two target intervals that differ only by a small amount. Experiment 10 used a click train manipulation, first devised by Triesman et al. (1990), to change the speed of the pacemaker during time production. Within the same experiment, the processing time required by a concurrent memory search task was varied. Including both manipulations in the same experiment provided a test of the assumption that the pacemaker and gate components of the model are separate. There were three main predictions made by the PGC model in Experiment 10. These were firstly, that the effects of click trains would not interact with those of the secondary task. Secondly, that changes in mean time production due to changes in pacemaker speed would be associated with changes in variability of performance and
lastly, as in Experiment 9, that changes in mean time production due to the secondary
task would not be associated with changes in variability.

5.2 Experiment 9

*The PGC model and timing variability*

The pacemaker is assumed to produce pulses at a rate that fluctuates randomly
around a fixed value. When timing is the only task undertaken the pulses flow into the
counter in an uninterrupted fashion. The counter stores some constantly updated
transformation of the number of pulses accumulated. The updating value may be
compared with the final value of other previously encountered intervals. The previously
encountered values may be ones also maintained within the counter / short-term memory
system from an earlier part of the experimental trial, or they may be values reactivated
from long-term memory. In the case of the production of a learned interval, the target
value is one from long-term memory.

Because the generation and accumulation of pulses is a linear process, the model
can account for the linear relationship between real time and subjective time, which is
typically found for short intervals (e.g. Wearden & McShane, 1988). The random
fluctuations in the output of the pacemaker tend to increasingly cancel each other out as
the length of a target interval grows. Thus, the model as described thus far predicts
incorrectly that timing variability will decrease as the target interval lengthens. In fact,
Weber's law has been found to apply to temporal judgement (e.g. Wearden & McShane,
1988). For short intervals, the variability of timing grows as a linear function of the target value.

The PGC model proposes, in accordance with Church & Gibbon (1982), that the Weber property of temporal judgement is accounted for by the noisiness of the memory processes involved in timing. Consider first, the paradigms employed in animal work, and here, in which an interval is learned over many trials. It is assumed that on each learning trial a single trace is laid down in long-term memory. Eventually, a population of traces is formed that represents the target interval. The population is assumed to have the properties of a Gaussian distribution. Furthermore, it is assumed that the standard deviation of the distribution will be a constant fraction of the mean across a range of target values. On an experimental trial, a single trace is retrieved at random from the population to serve as the target value in short-term memory for that trial. Across many trials, this random process results in the Weber property of temporal judgement.

Why does the population of memory traces for a target interval form a Gaussian distribution? Based upon the properties of the pacemaker and counter as they have been described, if encoding and retrieval were noise-free processes, then Gaussian distributions would not be formed. If the assumptions were made that retrieval is a noisy process, and encoding into long-term memory is not noisy, then Weber’s law could be derived without Gaussian memory distributions. However, in the case of human performance, temporal judgements can be made without learning trials. A single presentation of a target stimulus, followed by a comparison stimulus is sufficient for a judgement to be made. Over many trials using different target values, perhaps also with different subjects for each trial, Weber’s law would be demonstrated. As retrieval from
long-term memory is not required in such a paradigm — both values are present in short-term memory — retrieval is not a sufficient account of timing variability. The alternative assumption, adopted here, is that encoding into short-term memory is the noisy process. This has two consequences. Firstly, the likely results of the experiment described above are accounted for. Secondly, variability in short-term memory will be reflected in long-term memory, resulting in populations of memory traces with the properties required by the model.

*The gate component of the PGC model*

The gate is included in the model to represent the idea that processing resources are required in order to give access to short-term memory for duration (the counter). Thus, if time were being ignored in favour of another task, pulses from the pacemaker would not reach the counter. In experiments like those of Fortin et al. (1993), a secondary task is performed part way through the production of a learned interval. This results in subjects underestimating the amount of real time that has passed, and consequently producing a longer interval. It is assumed that the gate is closed (or much narrowed) during the time taken by the secondary task. Therefore, increasing the time taken by a secondary task results in a lengthening of time production. What effect, if any, should this have upon the variability of timing? In an experiment such as Experiment 1, it may be assumed that the time taken to close and open the gate is constant across experimental conditions, because all the experimental conditions require one interruption of timing. Therefore, the variability due to the operation of the gate may not be estimated in the kind of experiments reported here. However, the effects of the duration of gate closure
may be observed using the paradigm employed here. In terms of the model, a time production trial that has been lengthened due to real time interruption by a secondary task still uses a target value retrieved from long term memory in the same fashion as for an equivalent trial without a secondary task. Therefore, a population of such trials should display the same variability as a population of trials without a secondary task, plus the variability associated with the closure and reopening of the gate. This latter source of variability is held constant when the processing time required by a secondary task, such as memory search, is increased. Therefore, the observed lengthening of produced intervals with secondary task processing time should not be accompanied by any change in variability. Another way to think of this, is that the effect of a secondary task in experiments like those reported here is analogous to adding a constant to a distribution — it changes the mean, but not the standard deviation.

The present experiment

Fortin et al. (1993) do not provide information on how the variability of time production varied with set-size. Inspection of Figure 2.3 reveals that in Experiment 1, the standard deviation of time production appears to increase, as the processing time required by the memory search task increases. This is clearly in contravention of the prediction made by the PGC model described above. However, the effect did not approach conventional significance levels. The aim of the present experiment was to resolve the ambiguity created in Experiment 1. This was achieved by employing a secondary task that would produce a much larger change in mean time produced, thereby accentuating any associated changes in variability. To achieve this aim, the processing time required
by the secondary task must be increased. One way to increase the difficulty of, and hence processing time required by, the memory search task is to require subjects to verify the serial position of the probe item in the memory set. This may be achieved by presenting a probe item, as well as a digit referring to the position of the probe in the memory-set. Fortin & Massé (1999) adopted this procedure using consonants as memory-set items. Concurrent processing of item-information alone resulted in a set-size with time production slope of 21 msec per item beyond the first. Concurrent processing of item and order information resulted in much increased slope of 57 msec. Based upon the item information only slope of 32 msec in Experiment 1, a slope of approximately 60 to 70 msec was expected for concurrent processing of item plus order information.

A timing only control condition was also conducted, involving a new group of subjects. Two target values were planned, one of which would be the mean time produced for the smallest set-size under dual task conditions, while the other target value would be the mean time produced at the largest set-size. The purpose of this was to verify that a change in standard deviation was detectable in principle, using the time production procedure, with target intervals that only differed by a small amount.

The aims of Experiment 9 may be summarised as follows:

- By requiring subjects to process order information, and to make the decision regarding the probe item concurrently with time production, it was hoped that a greater slope of set size with time production would be observed here than in Experiment 1

- To investigate whether changes in mean time production caused by increasing the set size of the memory search task are accompanied by changes in
variability of timing; a positive result here would suggest that the attentional
gating process may be characterised as “flickering”, while a negative result
would suggest that the number of gate operations is unaffected by the duration
for which attentional resources are directed away from timing
• The control condition aimed to establish that the time production paradigm
was sufficiently sensitive to detect small changes in variability of timing
where these occur

5.2.1 Method

Subjects

Twenty-two subjects ranging in age between 22 and 40 took part in the main
experiment. Three of these subjects were excluded due to very high memory search error
rates. Subjects were paid up to £10 depending upon timing accuracy and the number of
correct memory searches. Nineteen unpaid subjects took part in the control condition.
Two of these were excluded because they produced longer values for the short target
interval than they did for the long target interval. A third subject was excluded because
his time productions exhibited a distinctly bimodal distribution for one of the target
intervals.

Stimuli

The memory-set consisted of the same 8 pure tones that were used in Experiment
1. However, only set sizes 2, 3 and 4 were used, because order judgements are
meaningless at set-size 1. The presentation of the probe was accompanied by a visually
presented digit, which referred to a serial position within the memory-set. If the probe was present in the memory-set, and the digit referred to the correct serial position of that item in the set, then subjects made a positive response. If the probe was not present in the set, or the probe was present but the digit referred to an incorrect serial position for the item, then subjects made a negative response. Blocks were arranged so that fifty-percent of trials were positive, with twenty-five percent of trials consisting of each of the two types of negative probe.

Design

The training procedure for the main experimental task was the same as that used in Experiment 1. There were 240 experimental trials per subject, divided into five blocks of 48. Therefore, a total of 4560 experimental trials were analysed. For the control condition the short interval was 2200 msec., and the long interval was 2290 msec, these values being chosen to correspond to the mean times produced at set sizes two and four in the main experiment. The order in which the two intervals were trained and tested was counterbalanced. In order to familiarise themselves with the interface subjects began with 20 feedback trials and 20 no-feedback trials using the first target interval. These trials were not analysed. After two more blocks of 40 trials with feedback, blocks of 20 trials without and with feedback were interleaved, until 80 trials without feedback were completed. These 80 trials without feedback formed the data for the first target interval. The procedure was then repeated for the second target interval, omitting the familiarisation phase. A total of 1280 trials were analysed for each target interval.
Subjects were informed that the second target interval was slightly different to the first, but were unaware of the direction of the difference.

**Procedure**

The procedure in individual trials of the main experiment was identical to that in Experiment 1, with the exceptions regarding the probe item noted earlier. The critical feature of this procedure is that the memory probe item is presented during time production, and when production is ended, the recognition response is also made. This means that retention of the memory material is concurrent with timing for the first 500 msec of time production, and then memory search is concurrent with timing, and the subject completes time production while remembering the result of the memory search. In the control condition feedback training trials were identical to those used in Experiment 1. First, subjects were shown an example of the target interval, which they reproduced by making two button presses. Feedback was then given as to the extent of under or overestimation, and then the subject started another reproduction. In order to space the no-feedback trials (i.e. the trials in which the data was collected) out in time in a similar fashion to the time productions in the main experiment a set of between two and four tones was presented prior to each time production. Subjects were instructed to ignore these tones. A fixation cross was presented to indicate that presentation of the set was complete.
5.2.2 Results

The mean memory search error rate was 13.5%. Error rates broken down by experimental condition are shown in Table 5.1, inspection of which shows that errors were more frequent in the two probe present conditions than in the probe absent condition, $F(2, 36) = 19.19, p < .001$, and that errors increased with set-size, $F(2, 36) = 33.85, p < .001$. There was also a significant interaction, indicating that the increase in error rate with set size was steepest for the positive response condition, $F(4, 72) = 5.03, p < .01$. 0.7% of the remaining data were excluded using the 3 standard deviation criteria for outliers.

Table 5.1. Percentage digit memory search error rates in Experiment 9.

<table>
<thead>
<tr>
<th>Probe Type</th>
<th>Memory Set Size</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Positive</td>
<td>5.1 (6.1)</td>
<td>17.2 (10.3)</td>
<td>24.5 (11.5)</td>
</tr>
<tr>
<td>Negative Order</td>
<td>11.1 (12.8)</td>
<td>17.6 (12.9)</td>
<td>23.7 (9.1)</td>
</tr>
<tr>
<td>Absent</td>
<td>4.2 (6.1)</td>
<td>7.9 (6.1)</td>
<td>10.0 (8.0)</td>
</tr>
</tbody>
</table>

In the final block of time production practice without feedback the mean time produced was 2.51 sec, with a mean standard deviation of .18 sec. Table 5.2 shows the mean times produced in each of the five experimental blocks, along with mean standard deviations. The small trend towards shorter time productions in later blocks was not
significant. However, the reduction in the variability of time production as the experiment progressed was statistically reliable, $F(4,72) = 2.50, p = .05$.

Figure 5.2. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of Experiment 9.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>time production (seconds)</td>
<td>2.30</td>
<td>2.27</td>
<td>2.23</td>
<td>2.24</td>
<td>2.19</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.27</td>
<td>0.25</td>
<td>0.25</td>
<td>0.23</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Figure 5.1 shows the effect of set-size, as well as probe type, on mean time production and the standard deviation of time production. The effect of increasing the number of items to be searched was to add an average of 43 msec per additional item to time production. This effect was reliable, $F(2,36) = 10.64, p < .001$, as was that of the type of memory probe, $F(2, 36) = 4.56, p < .05$. However, neither the effects of set-size, $F(2,36) = .46$, or probe type, $F(2,36) = .08$, on standard deviation of time production were significant. This was precisely the pattern of results predicted by the PGC model.

In the control condition data were first screened for outlying values, using the 3 standard deviation criteria for the 2.2 sec and 2.29 sec target intervals separately. This resulted in the exclusion of 0.55% of the data for the 2.2 sec interval, and 0.94% of the data for the 2.29 sec interval. The mean times produced for the two target intervals were 2.213 sec for the shorter target, and 2.385 sec for the longer interval. This difference was not tested statistically, because data from the only two subjects whose mean productions did not reflect the ordinal difference between the two target intervals was excluded. The
mean standard deviations for the two intervals were .189 sec for the short interval and .223 sec for the long interval. This difference was reliable, \( t(15) = -2.21, p < .05 \).

Whether there were 2, 3, or 4 tones in the set subjects listened to before each trial had no effect upon time production. The average difference between the long and short intervals was 172 msec, whereas the target difference was only 90 msec. It was initially thought that this may have occurred because subjects first trained and tested on the long interval went on to underestimate the short interval, and vice versa. This exaggeration would be advantageous as a way of keeping the two representations distinct in memory. However, inspection of the data revealed no evidence that this had taken place.

Figure 5.1. The effects of digit probe type and memory-set size on mean time production and variability of timing in Experiment 9. The left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean standard deviation of time production.
5.2.3 Discussion

Increasing the processing time required by the secondary task, compared to Experiment 1, increased the observed lengthening of time production. However, while Fortin & Massé (1999) more than doubled their set-size with time production slope by requiring subjects to process order information, here the slope increase was only one-third. A much greater slope was predicted, and therefore the results are less definitive than expected concerning the change or lack thereof, in standard deviation. Despite this, there was clearly no change in standard deviation with set-size. This confirms the prediction made by the PGC model, that lengthening of time production due to secondary task interference would not be accompanied by changes in variability. Experiment 10 will provide another chance to confirm this prediction.

The fact that the standard deviation of time production does not increase as the duration of the gating process increases allows some inferences about the nature of the gating process to be made. It may be assumed that the opening of the gate, and the closing of the gate, contribute some variability to time production. Because variability does not increase as a function of secondary task processing time, this implies that the gate only opens and closes once, regardless of the amount of secondary task processing. Because the time production with set size slope is typically less than the reaction time with set size slope, this suggests that the gating process is "leaky", and still allows some pulses through. The possibility that is not consistent with the data is that of an "all or nothing" gate, which "flickers", because such a process would lead to increasing variability with the number of switches made.
The control condition was designed to confirm that the measurement procedures employed were sensitive enough to detect any changes in standard deviation due to increases as small as 90 msec in mean time produced. Using two target intervals, separated by only 90 msec, a change in standard deviation of performance was detected. However, across subjects, the mean difference between productions of the 2 target intervals turned out to be double that intended. In particular, the longer interval was overestimated by 95 msec on average, while the shorter interval was only overestimated by 13 msec. The reason this occurred is unclear, but it does reduce the value of the control condition.

5.3 Experiment 10

A basic assumption of the PGC model is that secondary task manipulations, like those employed in Experiments 1 and 9, have their effects upon temporal judgement only through the operation of the gate. Therefore, it is implied that such manipulations do not affect pacemaker rate. Equally, those manipulations reported in the literature discussed below, which are assumed to affect pacemaker rate, are implicitly assumed to have no effect upon the gate. The aim of Experiment 10 was to combine both kinds of manipulation in the same experiment, as a test of the prediction made by the PGC model that they will have separate and additive effects upon mean time production. Experiment 10 also provided an opportunity to replicate the results of Experiment 9 concerning the gate and variability of timing, as well as providing the opportunity to assess the effects of pacemaker rate upon variability.
One focus of research concerned with timing mechanisms of the brain has been on manipulations that affect the speed of the proposed pacemaker. Wearden & Penton-Voak (1995) review evidence that suggests increasing body temperature accelerates the pacemaker, and therefore speeds up the rate of subjective time. They also describe a small amount of evidence that the converse effect may be achieved by reductions in body temperature. These effects are explained by the fact that the internal clock ultimately consists of biochemical processes, which are temperature dependent. Triesman, Faulkner, Naish & Brogan (1990) model the pacemaker as a temporal oscillator. They showed that high tempo auditory click trains cause biases in temporal judgements, and that the extent of the bias was related to the speed of the click trains. Deviations from the generally linear relationship between click train speed and the amount of acceleration of the rate of subjective time were thought to represent resonation with the fundamental frequency of the temporal oscillator. Analysis of the positions of the deviations allowed Triesman et al. (1990) to estimate the frequency of the oscillator as 49.5Hz. Triesman, Faulkner & Naish (1992) showed that the same underlying oscillator pattern can be detected in the timing of motor action by imposing click trains during a choice-reaction time task. Triesman, Cook, Naish & Macrone (1994) replicate Triesman et al. (1990) while also obtaining confirmatory EEG evidence for the fundamental frequency of the oscillator. Penton-Voak, Edwards, Percival & Wearden (1996) successfully extend the effects of click trains to four more time judgement paradigms, namely temporal generalisation, pair comparison of duration, verbal estimation, and production of short durations. However, it has not yet been established that manipulations intended to affect the rate of the internal clock act
independently of the effects of secondary tasks, which are mediated by the proposed attentional gate. Therefore the current research will combine a manipulation of the processing time required by a secondary task with an attempt to systematically modify the speed of the clock through click train speed. Increasing the demands of a secondary task effectively slows the rate of subjective time, while increasing the rate of click trains effectively speeds up the rate of subjective time. If the internal clock and attentional gate components of the model are functionally independent units then these manipulations should result in two main effects and no interaction.

Here, click trains were used to modify the rate of the internal clock during production of the learned interval. Three different rates were chosen from the values used by Triesman et al. (1990), on the basis that they each caused a different amount of acceleration of the clock. If the underlying representation of the learned standard remains constant, the standard deviation of time production should vary as a function of clock speed. As an illustration, on a particular trial the representation of the learned interval drawn from the population of traces in long term memory may happen to be from the short end of the distribution, while on another trial a representation may be drawn from the centre of the distribution. At the natural clock speed the real time taken to accumulate time pulses equal to the 2 retrieved target values will differ by a certain amount. At an accelerated clock speed, both target values will be shortened in real time, and the difference in real time between the 2 target values will be reduced. For this reason, over many trials with an accelerated clock, reduced variability of timing should be observed.
Because of the large number of trials required by the experimental design, testing was split into 2 sessions. This provided the opportunity to test subjects retention of the target interval over a period of several days.

The aims of Experiment 10 may be summarised as follows:

- To replicate Fortin et al (1993) who found that combining digit memory search with concurrent time production resulted in a set size slope of 24 msec per item
- To confirm the main result of Experiment 9, namely that changes in mean time production caused by an increase in the set size of a concurrent memory search task are not accompanied by changes in variability of timing
- To obtain a shortening of time production by the introduction of concurrent click trains, and to test the prediction that this shortening should be independent of the lengthening caused by the memory search task
- To test the prediction that the shortening of time production caused by click trains will be accompanied by a reduction in variability of time production
- To establish, through the interaction of click trains with the memory search task, whether the attentional gating process operates in an all or nothing fashion or whether some temporal information can still accumulate while secondary task processing takes place
- To establish whether subjects can retain a representation of the target interval in memory over a number of days
5.3.1 Method

Subjects

Seventeen subjects, aged between 18 and 28 took part in the experiment. They received course credit in return for their participation.

Stimuli

Three rates of click trains were employed, 4.5Hz, 13.5Hz, and 25Hz. Individual clicks were 1 msec bursts of white noise. Each click was separated from the next by a short period of silence. Click trains were presented using loudspeakers, at the loudest volume found generally comfortable in piloting. A sound level meter was used to measure the intensity of the click trains at head level. Intensity increased with the frequency of the click trains, the values being 69, 72, and 74 dB respectively\(^1\). There was also a no click train condition, for comparison with the click train conditions.

The memory-set was composed of the digits 0-9, presented visually. Thus, the memory set was identical to that used by Fortin & Breton (1993). Only set-sizes 2, 4, and 6 were used. The reason that digits replaced tones was in order to allow the click trains to be presented in the auditory modality.

---
\(^1\) Burle & Casini (2001) point out that the effect of click trains is greater when the intensity is higher, and that levels below 67 dB do not produce an effect. The levels used here are all above 67 dB, and the possible effect of the increase in intensity operates in the same direction as the possible effect of frequency.
Design

Time production training took place before the first testing session as in Experiment 1. This consisted of blocks of reproductions of the target interval with trial by trial feedback, interleaved with blocks where feedback was not presented until the end of the block. The number of days between testing sessions varied unsystematically between 1 and 14. The second session started with a block of 24 time production trials without feedback, the purpose of which was to assess long-term memory for the target interval. There then followed 4 blocks of 24 time production practice trials, consisting of 2 blocks with feedback and 2 without feedback, the purpose of which was to refresh subjects memory for the target interval. In the main experiment, there were 4 within subjects variables: memory-set size (3), probe type (2), click trains (4), and testing session (2). Each testing session consisted of 4 blocks of 72 trials. The sequence of trials was randomised within each block.

Procedure

Trial by trial procedures were identical to those in Experiment 1, with the exception that click trains were also presented on three quarters of trials. Thus, the sequence of events began with the serial presentation of the memory set, after which the subject commenced time production. The memory probe was presented 500 msec into time production, and subjects terminated time production by making their response to the memory probe. On trials where click trains were presented the first click occurred simultaneously with the offset of the last memory-set item. Click-trains ceased when the
subject made the final button press, which indicated the end of the time interval and the response to the memory probe.

5.3.2 Results

The mean memory search error rate was 6.8%. Error rates broken down by set size, probe type, and click train conditions are shown in Table 5.3. There was no effect of testing session, or click trains on the error rate. Therefore, the presence of click trains, relative to the silent condition, did not disrupt subjects ability to perform the memory search component of the task. There was a highly significant effect of set size upon the error rate, $F (2,32) = 44.15, p < .001$, as well as an effect of probe type, $F (1, 16) = 4.96, p < .05$. Because 4 subjects produced somewhat different mean intervals in their second experimental session, it was necessary to screen the data of each session for outliers separately. There were only 12 observations per cell of the design in each testing session, which was considered insufficient to screen for outliers for each subject, and in each cell of the design, separately. Therefore outliers of 3 standard deviations or greater were eliminated from each session of each subject's data, ignoring experimental conditions. This resulted in the exclusion of 0.8% of the remaining data.
Table 5.3. Percentage error rates in the digit memory search task, in which the probe item was presented during time production.

<table>
<thead>
<tr>
<th>Clicks</th>
<th>Probe Type</th>
<th>Memory Set Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>Positive</td>
<td>3.4 (6.3)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>2.7 (3.9)</td>
</tr>
<tr>
<td>4.5 Hz</td>
<td>Positive</td>
<td>2.9 (4.4)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>3.4 (5.6)</td>
</tr>
<tr>
<td>13.5 Hz</td>
<td>Positive</td>
<td>3.2 (4.8)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>2.7 (5.3)</td>
</tr>
<tr>
<td>25 Hz</td>
<td>Positive</td>
<td>3.9 (7.3)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>1.2 (2.4)</td>
</tr>
</tbody>
</table>

In the final block of time production practice before the first experimental session, the mean time produced was 2.70 sec, with a mean standard deviation of .21 sec. Table 5.4 presents the means and standard deviations of time production for the 4 experimental blocks of the first testing session. There was no significant change across blocks in either mean time produced or standard deviation of time production. In the time production block without feedback at the start of the second testing session the mean time produced was 2.81 sec, with a mean standard deviation of .29 sec. This demonstrates that the group considered as a whole retained the interval well over a period of some days. The mean time intervals produced by each subject in the first experimental session were correlated with those produced in the second session, $r (17) = .69, p < .01$. This suggests that
individual subjects were consistent across the 2 testing sessions. As noted above, inspection of the data reveals that only 4 subjects produced a noticeably different interval in the second testing session. Table 5.5 presents the means and standard deviations of time production for the 4 experimental blocks of the second testing session. Again, there were no significant effects of block number.

Table 5.4. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of the first testing session of Experiment 10.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>time production</td>
<td>2.75</td>
<td>2.79</td>
<td>2.77</td>
<td>2.72</td>
</tr>
<tr>
<td>(seconds)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>.31</td>
<td>.30</td>
<td>.29</td>
<td>.30</td>
</tr>
</tbody>
</table>

Table 5.5. Mean time production, and average within subject standard deviation of time production in each of the four experimental blocks of the second testing session of Experiment 10.

<table>
<thead>
<tr>
<th>Experimental Block</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>time production</td>
<td>2.71</td>
<td>2.71</td>
<td>2.67</td>
<td>2.67</td>
</tr>
<tr>
<td>(seconds)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>.28</td>
<td>.27</td>
<td>.28</td>
<td>.28</td>
</tr>
</tbody>
</table>

Figure 5.2 shows the effects of memory-set-size and click trains on mean time production and the standard deviation of time production. Each additional item in the memory search task lengthened time production by an average of 17.75 msec. The highest tempo click trains resulted in an 82 msec shortening of time production relative to
the no click train condition. The data were analysed using a 2 (testing session) * 2 (memory probe type) * 3 (set-size) * 4 (click trains) ANOVA. There were no reliable effects of, or interactions involving, session or memory probe type. The effect of set-size on time production was significant, $F(2, 32) = 10.61, p < .001$, however, there was no effect of set-size on the standard deviation, $F = .151$. The effect of click trains on mean time produced was significant, $F(3, 48) = 6.39, p < .001$. A repeated contrast showed that the 4.5Hz condition was significantly different from the silent condition, $F(1, 16) = 13.08, p < .01$; the 13.5Hz condition did not differ from the 4.5Hz condition, and the 25Hz condition differed significantly from the 13.5Hz condition, $F(1,16) = 3.33, p < .05$.

Most importantly in terms of support for the PGC model, no interaction occurred between set-size and click-trains, $F = .34$. The effect of click-trains on the standard deviation of performance approached significance, $F(3,48) = 1.60, p > .1$. 


Figure 5.2. The effect of performing digit memory search concurrently with time production, and click trains, upon mean times produced, and variability of timing. The left ordinate provides the scale for mean time production, while the right ordinate provides the scale for the mean standard deviation of time production.

Considering testing session 2 alone, the dissociation predicted by the PGC model between the effects of click trains and secondary task on the variability of performance was evident. Figure 5.3a shows that there was clearly no effect of increasing memory set size upon variability, $F = .001$, while Figure 5.3b shows that variability of performance was decreased for click train conditions compared to no click trains, and that variability was related to click train rate, $F (3, 48) = 2.22, p < .05$ (one tailed). A one tailed test was
selected here because it was predicted that variability of time production would be reduced by the presence of click trains, which is what occurred.

**Figure 5.3a**

**Figure 5.3b**

Figure 5.3. In the second testing session, increasing memory search time had no effect upon standard deviation of time production. However, the standard deviation was affected by the click train manipulation.

5.3.3 Discussion

Five separate predictions of the PGC model were confirmed to varying extents by the results of Experiment 10. Each of these will be dealt with in turn.

Firstly, performing a secondary task concurrently with timing lead to an underestimation of time passing and, therefore, time production was lengthened. The memory search secondary task used was identical to that used by Fortin et al. (1993), and
previously by Fortin & Rousseau (1987). Here, a slope with time production of 18.75 msec per item was obtained. This was in reasonable quantitative agreement with the slopes of 22.5 msec and 24 msec obtained in the two Fortin studies. The PGC model explains these findings by proposing that the flow of pulses from the pacemaker to the counter is interrupted during secondary task processing because the gate closes, in order to allow attention to be directed elsewhere.

Secondly, and in confirmation of the results of Experiment 9, the change in mean time production caused by increasing the processing time required by the secondary task was not accompanied by a change in variability of timing. This apparently counter intuitive result is actually predicted by the operation of the gating process, which adds a real time constant to the subjective target interval. The variability of a population of such target intervals is actually determined by the long-term memory representation of the target interval.

Thirdly, introducing click trains caused an overestimation of time passing, compared to the no click train condition, and this lead to a shortening of time production. Furthermore, the extent to which time production was shortened was related to the frequency of click trains. The model proposes that these results occur because auditory pathways project to the proposed site of the pacemaker in the cerebellum. Therefore, the rate of the click trains is able to influence the rate of the pacemaker. A mathematical model of how this might occur is provided by Triesman et al. (1990).

Fourthly, a novel finding here is that changes in the rate of subjective time due to acceleration of the pacemaker are also accompanied by changes in variability of timing. The PGC model predicts this because the pacemaker is independent of the memory
processes that determine variability of timing. An accelerated pacemaker causes a proportional reduction in the real time required to accumulate pulses equal to any single retrieved target value. The real time shortening of a large target value from a particular memory distribution is greater in absolute terms than the shortening for a target value from the short end of the same distribution. Therefore, accelerating the pacemaker is equivalent to "squashing together" the memory distribution of target values for a particular interval. The empirical support for this conclusion was relatively weak, and was only clear at all in testing session 2. It is possible that this was the case because the participants were highly practised by the time they took part in the second session. One effect of practice might be to reduce the amount of variability of timing attributed to uncontrolled variables, such as motor response errors and distraction. Thus, the effect of click trains on variability of timing may have been swamped by "error variance" in session 1. This is particularly likely given that changes in variability are necessarily much smaller, and harder to detect, than the accompanying changes in the mean, because the coefficients of variation in the experiments reported here are typically of the order of around .1. The issue of increasing the power of the time production paradigm to detect such small changes is taken up in the General Discussion.

Fifthly, the effects on mean time production attributed to the operation of the gate did not interact with those attributed to the acceleration of the pacemaker. Rather, the effects were separate and additive. This supports the basic proposition made by the model that the pacemaker and gate are functionally distinct. This proposition is further supported by the differential effects upon the variability of timing discussed above.
One issue concerning the gating process is whether pulse accumulation ceases completely while nontemporal processing takes place, or whether it slows down. The observation that the slope of time production with set size is always less than the slope of reaction time with set size in the Experiments reported here, as well of those of Fortin & colleagues, supports the idea of a partial gating process. However, if it is assumed that the click train manipulation did not affect the slope of the memory-search with set size function then the results of Experiment 10 may also be taken to imply an “all or nothing” gate, for the following reasons. Firstly, in the absence of click trains it may be taken that there is a one for one relationship between subjective time and real time. Therefore, the real time taken for one pulse to accumulate is equal, on average, to the subjective duration, in terms of target values, that it represents. When the pacemaker is accelerated, the real time taken for one pulse to accumulate is less than the subjective duration, in terms of the target value, that it represents. Therefore, an accelerated pacemaker compared to an unaccelerated pacemaker when the target number of pulses for a time production is the same, results in a shorter production. This logic may be extended to the operation of the gate. In the case of a partial gate, more pulses will pass through the gate during a given period of nontemporal processing when the clock is accelerated. This effect would be multiplicative; the longer the processing time of the nontemporal task the greater the difference in pulses accumulated during nontemporal processing between the two pacemaker conditions. Such a process would manifest itself in an interaction between

---

2 There was no reaction time control condition in Experiment 10. However, the Sternberg effect (memory search with digits) has been replicated many times, and invariably produces a slope of about 37 msec (Fortin et al., 1993; Experiment 5 here). The observed slope with time production in Experiment 10 was 17.75 msec.
click trains and set size, such that the slope of time production with set size would be less at higher pacemaker speeds. This did not occur in Experiment 10. An “all or nothing” gating process would predict the lack of an interaction. This is because when the interruption of the accumulation process is complete, the gating process simply acts to add a constant to time production, regardless of pacemaker speed. Overall, the results of Experiment 10 represent a paradox, as they seem to suggest both an “all or nothing” gate and a partial gate. This issue will be taken up in the General Discussion.

Finally, some preliminary evidence was obtained concerning the durability of long term memory for duration. When asked to perform production of the target interval without feedback, after one or more days without practice, the group as a whole produced values close to the target. Furthermore, the values produced by individuals in the two separate experimental sessions were highly correlated with each other. Many of the predictions made by the PGC model rest heavily upon the proposed characteristics of long-term memory for duration. However, based upon the evidence presented in this thesis, the reality of those characteristics is based entirely upon inference. In the General Discussion, possible ways of addressing the issue of long-term memory for duration directly will be explored.
6.1 General Discussion: Aims of Chapter

The first aim of the General Discussion is to summarise the main findings in this thesis. These are principally those relating to the effects of memory loads on concurrent time production, as well as the findings relating to the separation of the gating process and the pacemaker, and those relating to the variability of temporal judgement. As part of the summary, the unresolved issue of the effect of maintaining pitch information upon time production will be considered in the light of findings published after the empirical work reported here was completed. Thirdly, modifications and improvements to the paradigm that might be implemented in future research are considered. Finally, two specific issues for further research within the framework of the PGC model are discussed. The first issue concerns possible reasons why interference with time production from memory-search tasks is typically less than would be predicted from reaction time data. Secondly, many of the predictions made by the PGC model depend heavily upon the assumed properties of long-term memory for duration, but no tests of these assumptions have been made in this thesis. Here, ways of testing these assumptions are suggested.

6.2 Summary of findings

Fortin et al. (1993) showed that time production was systematically lengthened by increasing the set size of a concurrent digit search task. In Experiment 1, this finding was replicated using a pitch memory search task. This replication increased the generality of
Fortin et al.'s conclusions, because pitch is distinct from phonological information in short-term memory. Fortin & Breton (1995) found that increasing the number of digits to be remembered (not searched) had no effect upon concurrent time production. Experiment 2 was an attempt to replicate this finding using the pitch memory-set. Unexpectedly, a set size effect was found, and this clearly required further investigation. In the event, several experiments (summarised below) were required to explore this effect, and by their nature, these experiments were not tests of predictions made by the PGC model. Therefore, due to the unexpected result of Experiment 2 less time than originally anticipated was available for tests of the model.

Experiments 3 and 4 confirmed that the results of Experiment 2 were not an artefact of the procedure. Experiment 3 tested the hypothesis that the lengthening of time production with set size observed in Experiment 2 was due to the duration of presentation of the memory set rather than the requirement to remember pitch information. This hypothesis was not confirmed, and importantly, the results of Experiment 2 were replicated with a new sample. In Experiment 4, the need to remember the tones during time production was removed, by presenting the memory probe before time production. Because the tones were still presented on each trial, it was possible that a set size effect upon time production might still occur. Such a set size effect would have implied that the effect in Experiment 2 was not related to the need to remember the tones, but was instead an after effect of having processed them. In the event, no set-size effect was found, and this suggested that the set size effect in Experiment 2 was a direct consequence of the requirement to remember the tones. Because the requirement to remember digits does not produce this effect, it was suggested that there might be some difference between the
processing requirements of the two tasks, which lead to the differential effects upon time production.

In Experiment 5 it was argued based upon the findings of Fortin & Massé (1999) that the differential effect could not be explained by a difference in difficulty between the two memory-search tasks, as indexed by reaction times and error rates. Rather, some specific resource that is required by remembering tones and by time production, and is not required by remembering digits, might be the source of the dissociation. Because remembering pitch information is less practised in most people than remembering digits, it was argued that remembering pitch might be more demanding of attention than remembering digits. This hypothesis was attractive in the light of research described in Chapter 1, which suggested a joint role of attention and short-term memory in tasks such as time production. The Stroop task was chosen as a means of increasing attentional load, and it was performed concurrently with the retention of either digits or pitch. Results indicated that there was no dissociation between the two tasks in terms of the amount of attention required to maintain the memoranda, and therefore the hypothesis was rejected. An interesting subsidiary finding was that Stroop conflict trials had an intercept effect upon reaction time to the subsequent memory probe item, for both digits and pitch. Further experiments, designed to elucidate this effect were suggested.

The approach adopted in Experiments 6 and 7 was to use different memoranda, in an effort to establish whether the effect of remembering pitch upon time production was unique. The alternative possibility was that some general principle determines whether maintenance of a given type of stimulus in memory interferes with time production or not. One possibility was that any non-verbal material would produce interference, and
this idea was tested in Experiment 6 by using a memory-set composed of shades of the colour grey. As in the case of Fortin & Breton (1995) with digits, no set-size effect was obtained. This suggested that the verbal / non-verbal distinction was not a useful one. However, some subjects used verbal strategies, and therefore the results were not conclusive. In Experiment 7, a timbre memory set was employed, and evidence was described which indicated a dissociation between timbre and pitch in auditory short-term memory. The Experiment was identical to Experiment 2 in every other respect, and yet no set-size effect was obtained. This result suggested that the interference effect of remembering information upon time production must be highly specific to pitch.

There is no obvious functional reason for the pattern of results described above. Neither is it possible to provide a process model of the interaction between remembering pitch information and time production. However, research published after the experiments reported here were conducted suggests that short-term memory for pitch and duration are separate modules, and that judgements based upon the combination of information from the two modules are prone to a high degree of error. As this research, conducted by Thompson, Hall & Pressing (2001), suggests a possible way elucidate the findings reported here, it will now be described. Thompson et al.'s (2001) research was conducted against the theoretical background of feature-integration theory (FIT), which lead to the suggestion that the processing of music might involve two distinct stages. In the first stage, it is thought that individual dimensions such as pitch, duration, timbre, location, and loudness are registered in functionally independent channels. In the second stage, the individual dimensions are recombined to form the unified percept. One of the predictions of FIT is that illusory conjunctions will occur in short-term memory when
judgements are required that are based upon the combination of features from different dimensions. In the visual domain, Triesman & Schmidt (1982) asked subjects to detect the presence or absence of a conjunction of colour and shape in a brief visual presentation. Illusory conjunctions occurred when, for example, subjects were presented with a red X and a green C. and remembered a red C or a green X. In their first experiment Thompson et al. (2001) presented subjects with two target tones to remember, which varied in pitch and duration, followed by a probe tone. Subjects had to judge whether the probe tone was identical in pitch and duration to either of the target tones. There were five kinds of trials; these were match trials, trials where both targets differed from the probe in pitch, trials where both targets differed from the probe in duration, trials where both features were inconsistent with both targets, and "switch trials" where the probe tone consisted of the duration of one target combined with the pitch of the other. A false positive response on a switch trial would imply that an illusory conjunction of pitch and duration had occurred. As this was expected to occur, it was predicted that the error rate would be highest in the switch condition, but that the match condition would yield a higher proportion of "match" responses than the switch condition, as subjects were expected to have some ability to accurately recombine the features. Note that "match responses" are correct in the match condition and incorrect in the switch condition. These predictions were confirmed, and replicated in a second experiment, which used sequences of 7 tones.

The results of Thompson et al. (2001) may not seem to be pertinent to those of Experiment 2, because in Experiment 2 there was no requirement to integrate pitch information and duration information, rather the requirement was to keep them separate.
However, if some aspect of the integration process occurred automatically, and demanded a shared processing resource, then the lengthening of time production with increasing set-size might be explained. A first step towards testing this speculative theory would be to use the Thompson et al. (2001) paradigm to investigate the possibility of illusory conjunctions in stimuli consisting of duration plus timbre information, as well as stimuli consisting of duration plus digit information. If both investigations failed to produce evidence of illusory conjunctions, or the size of the effect was much smaller than with pitch and duration, then this would be consistent with the “automatic integration” hypothesis.

Returning to the Experiments presented in the earlier chapters of this thesis, Experiment 8 extended the investigation of the effects of concurrent memory loads on time production to the case of duration itself. In a reaction time condition, Deutsch’s (1986) hierarchical theory of short-term memory for duration was tested by contrasting a hierarchically organised memory-set with a non-hierarchical one. No support for this theory was found, although the experiment was not a critical test of the theory. Maintaining a memory-set composed of values of duration produced a larger set size effect upon time production than was the case for pitch. There was no effect of the magnitude of the concurrent memory load upon time production, which suggested that short-term memory for duration has a capacity limit in terms of items rather than duration itself. While the average variability of time production was in line with that of the earlier experiments, the variability of time production increased with set size. This suggested that the interference was systematic, rather than general. It was thought that the gating process was not involved in the effects found, as Experiments 9 and 10 showed that
varying the time for which the gate is closed or narrowed does not produce changes in variability of time production. Rather, it was suggested that maintaining information in short-term memory for duration reduces the efficiency with which new pulses from the pacemaker may be accumulated. This is formally equivalent to slowing the pacemaker, which is consistent with the observed increases in variability.

In Experiment 9 the gating process was investigated. The set size with processing time slope was increased by requiring subjects to process order as well as item information. The aim of this was to magnify any changes in variability of timing associated with changes in the mean. As predicted by the PGC model, changes in the mean were not accompanied by changes in variability. A control condition, in which there was no secondary task, was used to demonstrate that the time production method was sensitive enough to detect changes in variability between target intervals that only differ by a small amount. In Experiment 10, click trains were used to modify the rate of the pacemaker during time production, while subjects also carried out a concurrent digit search task. No interaction was found between the effects of the manipulation aimed at the gate and the manipulation aimed at the pacemaker, which supported the functional separation between the two processes made by the PGC Model. Furthermore, the pacemaker manipulation affected the variability of timing, whereas the gate manipulation did not. This dissociation was also predicted by the PGC model.

Research published after Experiment 10 was conducted has supported the main result concerning the lack of an interaction between pacemaker and gate effects up mean time production. Burle & Casini (2001) trained subjects to produce a 1.1 s interval by holding down a button with the thumb of their right hand. The fingers of the left hand
were used to perform a reaction time task, in which subjects had to indicate which of two LED's was lit. The two tasks were performed singly, and in combination. Click trains were employed during all tasks at 25 Hz. Rather than varying the frequency of click trains, as was done in Experiment 10 here, Burle & Casini (2001) varied intensity (62 dB versus 77 dB). Results indicated that produced intervals were shortened by the louder clicks relative to the quieter clicks, and lengthened by the inclusion of the reaction time task relative to single task conditions. There was no interaction between the two effects. Burle & Casini (2001) take this as support for an “all or nothing” gating process, like that described in section 5.3.3. They point out that click trains would have interacted with the secondary task if the gating process were partial. Therefore, the results of Burle & Casini (2001) and Experiment 10 agree to a certain extent. However, Experiment 10 provided the additional opportunity, which was not available in the Burle & Casini paradigm, to investigate the nature of the gate through the comparison of reaction time and time production slopes. The results of this comparison suggested the opposite conclusion, namely that the gating process is partial. Burle & Casini (2001) demonstrate an increase in timing variability under dual task conditions. The equivalent comparison in Experiment 10 would be to compare variability at the end of training with the average variability of timing across set sizes1. Burle & Casini's (2001) analysis of timing variability is not equivalent to, or comparable with, the analysis of variability as a function of set size in Experiments 9 and 10. In fact, it is unclear whether the increase in timing variability under dual task conditions found in the Burle & Casini paradigm is due

---

1 This analysis was not performed because it would be confounded by practice effects. In order to make this comparison, trials without a secondary task would have to be included in the Experimental blocks. This possibility is discussed in section 6.4.
to the central processing requirements of the reaction time task, or alternatively, whether it is due to the requirement to make an additional button press during time production in the dual task situation.

6.3 Possible modifications to the paradigm

Here, two potential additions to the experimental paradigm are discussed. The first addition involves providing feedback on time production during the experimental session. The principal aim of this is to reduce extraneous variability in the data, thereby improving the power of the paradigm as a tool for investigating changes in variability of timing. Changes in variability of timing are, of necessity, much smaller than changes in mean time production, and therefore harder to detect. The second addition was inspired by anonymous reviewers comments on a paper submitted to Quarterly Journal of Experimental Psychology, and involves adding trials without concurrent nontemporal processing to the experimental blocks. The aim of this is to estimate the variability associated with the gate.

The average time productions of individual subjects are, as would be expected, distributed around the means reported here. If feedback on the time production part of the task were provided during the experimental blocks then between subjects variability in the data would be reduced. Furthermore, the slight tendency towards shortened time production as the experimental session progressed might be eliminated. However, if the feedback provided were as fine grained as that used in the training phase then this would interfere with the effects of the experimental manipulations. The difference between time
production at set size 1 and time production at the maximum set size used typically varies between 100 and 200 msec. Therefore, if feedback was given only on trials where time production was 1 s too short or too long, in the form of a warning to this effect, it would be unlikely to interfere greatly with the effects of set-size manipulations. This might be verified empirically by replicating one of the experiments presented here, but with the addition of feedback.

The second possible modification involves the addition of trials without concurrent nontemporal processing during the experimental blocks. It is desirable that such trials should be as similar as possible to trials with concurrent nontemporal processing. Therefore, subjects would be required to attend to the presentation of a memory set as usual, following which they would be informed by a distinct visual or auditory signal that they would not be required to make any recognition judgement on this trial. An irrelevant stimulus would then be presented instead of the memory probe item. However, time production would be performed as usual. This would allow data to be collected for “set size zero”. Presumably, mean time production would be shorter than at set size 1. This should also be the case for memory-load only experiments, as Fortin & Breton (1995) found that for digit retention time production was lengthened in the experimental blocks relative to the training blocks, even though no set size effect was found. Of greater theoretical interest is the standard deviation of time production at “set size zero”. Presumably, because the gate would not operate there would be one less source of variance in these trials, and so standard deviation should be less than in trials with the inclusion of nontemporal processing. It was demonstrated in Experiments 9 & 10 that increasing the duration of nontemporal processing does not increase the standard
deviation of time production. Presumably, this is because the gate only opens and closes once, regardless of the duration of nontemporal processing. Comparison of the “set size zero” condition with the other set sizes would allow the contribution made by a single incidence of the gating process to timing variability to be estimated.

6.4 The Gate: all or nothing, or variable width?

As discussed earlier the results of Experiment 10 may be taken to suggest either that the gating process is “all or nothing”, or that some pulse accumulation still occurs during concurrent nontemporal processing. Here, experiments are suggested that may resolve this apparent paradox.

The “all or nothing” interpretation relies heavily on the assumption that the objective duration of gate closure is unaffected by the click train manipulation. This assumption would be easily tested by a control condition, in which the digit memory search task was performed under single task condition, but with the addition of click trains. Triesman et al. (1992) showed that click trains have an effect upon choice-reaction time. Therefore, it is quite possible that an effect would be observed, which could take three possible forms. Firstly, an intercept effect could occur, consistent with an effect of click trains upon motor response timing. Secondly, click trains might interact with set size, suggesting that they actually accelerate the scanning process. Thirdly, both effects might be observed. If the second effect were to occur, then this would be surprising and extremely interesting in itself. However, it would undermine the “all or nothing” interpretation of the gating effects in Experiment 10 because gating duration as well as
pacemaker speed would have been shown to vary with click train frequency. The “all or nothing” account could probably accommodate the intercept effect, because this would be constant across set sizes.

The “variable width” account relies on the observation that the slope of time production with set size is always less than that of reaction time with set size for the same memory-search task. However, this comparison has only ever been made using a single target duration in the time production task. The PGC model predicts that the slope of set size with time production should not vary with the duration of the target interval, and this prediction is made by both possible kinds of gating process. This is because the gating process occurs only for as long as is required by nontemporal processing, and this duration is unrelated to the target interval. To confirm this assumption an experiment might be conducted in which multiple target intervals were used, perhaps in different testing sessions. If the assumption was not confirmed, and the slope of time production with set size varied across target durations, then this would allow a new interpretation of why set size with time production slopes are less than set size with reaction time slopes. Specifically, if the slope of time production with set size were to become shallower as target duration increased then this would imply the existence of a corrective or compensatory mechanism in human timing. If evidence of such a mechanism were uncovered then the gating process itself might be “all or nothing”, as is argued by Burle & Casini (2001), but overall time production could still be less than the total of the target interval plus nontemporal processing, which is exactly what was observed in Experiment 10.
6.5 Long-term memory for duration and modality differences

The PGC model follows Church & Gibbon (1982) in proposing that long term memory for duration is a trace based system, in which individual traces correspond to individual learning trials. Normally distributed populations of traces are thought to be formed, with accurate means, and standard deviations that are a fixed proportion of the mean, resulting in the Weber property of subjective time. Many of the predictions made by the PGC model depend crucially upon these assumptions. However, the experiments presented in this thesis did not test these assumptions. In fact, it is very hard to assess long-term memory for duration in humans, because subjects can perform adequately in the test phase of an experiment without using whatever memory was gained in training. For example, Wearden & Ferrara (1995) showed that bisection performance with trained standards did not differ from bisection performance without trained standards.

One issue related to long-term memory for duration, which might prove to be experimentally tractable, is whether there are separate stores for each modality. This may be addressed using a categorical perception (CP) learning paradigm. In a learning phase, participants would be presented with a range of subjectively equally spaced durations between (e.g.) 500ms and 1500ms, in either visual or auditory form. They would be required to categorise each of the stimuli individually according to their membership of one of two categories, short and long. The training phase would continue until performance had reached a threshold. In a follow up condition pairs of the same stimuli will be presented to subjects, who will be required to make a same / different judgement. According to categorical perception theory, the proportion of correct judgements should
be higher where the category boundary was trained than for other stimulus pairs of equal
distance (Goldstone, 1994), even though the perceptual distance between each stimulus
was equated prior to the training phase. One half of the participants would perform the
same / different judgements under same modality conditions, while the second half of the
participants would perform under switched modality conditions. If there are separate
long-term stores for duration in each modality then the CP effect should be stronger
under same modality conditions than switched modality conditions. This would occur
despite the fact that modality of the stimuli provides no information to assist the
same/different judgement, because the only dimension on which the stimuli vary is their
duration. Such a finding would imply that duration encoding in memory is strongly
linked to nontemporal characteristics of the stimuli. However, if the cross modal CP
effect were as strong as the within modality effect then this would suggest that values of
duration are encoded in memory independently of the nontemporal characteristics of
stimuli. The experiment would also provide a demonstration of CP effects in a new
domain. It has been objected that CP effects at category boundaries may arise as a result
of verbal strategies (Roberson & Davidoff, 2000). Such a strategy is of little use with
duration, provided the duration values of adjacent stimuli of the set are close enough that
chronometric counting is not a useful way to distinguish them from one another.
Therefore, establishing whether it is possible to obtain CP effects in duration perception
is of considerable theoretical interest.

Another planned experiment addresses the possibility of modality effects in short-
term memory for duration. Deutsch (1986) established a paradigm in which participants
were presented with a standard duration (S), followed by a retention interval of a few
seconds. Memory for S was assessed by presentation of a test stimulus (T), which had to be categorised as the same as, shorter than, or longer than, S. When the retention interval was empty, performance on this task was extremely good. However, if the retention interval is filled with “to be ignored” interpolated durations (I), then judgement of the relationship between S and T will be systematically biased. The exact form of the biasing observed suggested that the short-term memory for duration system was predisposed to represent intervals as having a simple mathematical relationship with each other. This paradigm may be exploited to investigate modality differences by using values of I which are either in the same modality, or a different modality, to S and T. If the amount of biasing produced by same modality interference is greater than that produced by between modality interference then this will support the notion of modality specific short-term memory for duration systems. An alternative hypothesis is that the auditory interference condition will always produce greater interference, regardless of the modality of S and T. This would suggest that audition has a general primacy in temporal processing. Such a finding would be consistent with the findings of Walker & Scott (1981), described in section 1.5, who used compound stimuli, and the greater accuracy of PFT performance under auditory conditions found by Jancke et al. (2000), described in section 1.7.

A possible confounding factor in both the experiments described above is fact that auditory stimuli are judged as longer than visual stimuli (e.g., Wearden et al., 1998). In the categorical perception experiment this might produce a shift in the position of the category boundary, but as the predictions focus on the strength of the categorical effect (i.e., how much better performance is at the trained boundary), this should not present a problem. In the short-term memory experiment, it might prove necessary to use stimuli
that are subjectively, rather than objectively, of equal duration. Furthermore, auditory interfering stimuli might accelerate the internal clock relative to visual ones, therefore causing an overestimation of subsequent visual stimuli. Burle & Casini (2001) report that for click trains, intensities of less than 67 dB were insufficient to produce biasing effects in time production. Therefore, if the auditory interfering stimuli used were presented at a low intensity, it is hoped that this effect might be avoided. The best approach to this problem might be to design an additional study, in which such biasing effects were quantified. This would then allow an appropriate choice of auditory interfering stimuli to be made.

6.6 Conclusion

In summary, this thesis has contributed to knowledge of human temporal judgement in a number of ways. Firstly, the unexpected effects of a concurrent pitch memory load upon temporal judgement were explored, although not explained. It may be productive for future research to focus upon this effect. Secondly, the findings concerning the operation of the hypothesised gating process were in line with the predictions made by the PGC model. Thirdly, the effects of click trains were dissociated from those of secondary tasks, in terms of their effects upon mean time production and variability of time production. These findings must be taken into account by future theorising within the framework of the PGC model and similar models in the literature. Finally, a number of proposals were made for further experiments, which attempt to explore further issues in human temporal judgement.
References


Appendix 1

Analysis of time production training procedure

Two aspects of time production that can be assessed: mean accuracy, and variability. Percentage hits (where a hit is defined as within 10% of the target value) is a measure which combines these two aspects of performance. In order to score a high percentage of hits, a subject had to deliver time productions with a mean value close to the target value of 2.5 sec, and the variability of scores about the mean had to be low. Figure 1 shows how the percentage hits achieved, in Experiment 10, increased as training progressed. Performance in the feedback and no-feedback blocks is shown separately. However, it must be remembered that the two kinds of observation are not independent, because the two types of trials were actually interleaved in blocks. The clear improvement in performance under the feedback condition is highly significant, F (3, 48) = 18.19, p < .001 (one tailed). The overall improvement under the no-feedback training condition also reaches significance, F (3, 48) = 2.74, p < .05 (one tailed).

![Figure 1. Proportion of time production trials within 10% of the target interval in Experiment 2 of Chapter 5.](image-url)
Figure 2 shows how the percentage hits achieved, in Experiment 8, increased as training progressed. The change in performance across blocks was significant for the feedback trials, $F(3,36) = 4.00, p < .01$ (one tailed). Only the quadratic contrast reached significance for the no-feedback trials, $F(1,12) = 6.04, p < .05$. Figure 3 shows the same information for Experiment 2. The improvement in the feedback condition was significant, $F(3, 33) = 4.31, p < .01$ (one tailed). The pattern of results for the no-feedback condition resulted in a significant main effect, $F(3, 33) = 6.25, p < .01$. A repeated contrast revealed that the increase in accuracy between the first and second blocks was significant, $F(1, 11) = 6.15, p < .05$, as was the decline that occurred in the third block, $F(1, 11) = 24.36, p < .001$. The small upturn in the fourth block did not result in a performance level reliably different from that in the third.

![Figure 2](image_url)  
Figure 2. Proportion of time production trials within 10% of the target interval in Experiment 3 of Chapter 4.
Discussion

It is evident that performance improved as a function of the number of training blocks. However, in the three sets of data presented, learning is much more clearly demonstrated in the feedback condition than the no feedback condition. The learning functions for the no feedback condition are not monotonic increasing ones. A number of factors may have contributed to this, and because the training procedure was designed pragmatically, rather than experimentally, it is hard to disentangle the various possibilities. Firstly, the blocks of the different kinds of trials were interleaved with each other. So, for example, two feedback blocks were followed by two no-feedback blocks, which were followed by two more feedback blocks. Therefore, changes in performance across the four feedback blocks were not observed independently of the effects of the no-
feedback blocks. Secondly, a number of the subjects who showed relatively high aptitude for the timing task had already reached an optimal level of performance, in terms of hits, before the dual task practice block. When, after this, it was explained to them that they would now return to practising the task they had already mastered, it is likely their motivation dropped. Inspection of the data reveals a number of such cases, where a ceiling effect occurred early on, and performance then declined. This seems to be the most likely explanation of why the learning functions for the no-feedback condition are non-monotonic. Despite the downturn in performance at the end of training, the absolute level of the performance is very impressive. Across the three sets of data reported, subjects achieved roughly 50% hits on average in the last no feedback block. This means that half of their productions were within 250 msec of the target value of 2500 msec. Although it is not logically possible to define what the chance performance level would be in this task, it seems intuitively obvious that subjects were not performing in anything approximating to a random fashion.