Effects of aging on absolute identification of duration.

McCormack, T., Brown, G. D. A., Maylor, E. A., Richardson, L. B. N., & Darby, R. (2002). Effects of aging on absolute identification of duration. *Psychology and aging*, *17(3)*(3), 363-378. https://doi.org/10.1037/0882-7974.17.3.363

Published in:
Psychology and aging

Document Version:
Early version, also known as pre-print

Psychology and Aging
2002, Vol. 17, No. 3, 363–378

0882-7974/02/$5.00 DOI: 10.1037//0882-7974.17.3.363

Effects of Aging on Absolute Identification of Duration

Teresa McCormack, Gordon D. A. Brown, Elizabeth A. Maylor,
Lucy B. N. Richardson, and Richard J. Darby
University of Warwick

Experiments to examine the effects of aging on the ability to identify temporal durations in an absolute identification task are reported. In Experiment 1, older adults were worse than younger adults in identifying a tone's position within a series of 6 tones of varied durations. In Experiment 2, participants were required to identify a tone's position in 9 tones of varied durations. Older adults' performance was again worse than that of younger adults; moreover, they showed a qualitatively different pattern of errors than younger adults. In Experiment 3, in which the tones varied in pitch, the performance of older adults was worse than that of younger adults, but the error patterns of the 2 groups were similar. The results suggest that older adults have distorted memory representations for durations but not for pitch.

Studies of timing abilities in older adults have a long history (e.g., Feifel, 1957; McGrath & O'Hanlon, 1968). Although numerous studies have found age-related impairments (e.g., Craik & Hay, 1999; Licht, Morganti, Nehrke, & Heiman, 1985; Wearden, Wearden, & Rabbitt, 1997), others have found no age effects (e.g., Rammsayer, Lima, & Vogel, 1993; Salthouse, Wright, & Ellis, 1979; Surwillo, 1964), and there is at least one report of more accurate timing in older than in younger adults (Eisler & Eisler, 1994).

These mixed findings may be due to the fact that methods for studying timing abilities have varied considerably. Age-related declines in performance are reasonably robust on tasks involving the estimation in seconds or minutes of relatively long durations (for review, see Block, Zakay, & Hancock, 1998), particularly under conditions in which the attentional demands of the task are high (Craik & Hay, 1999; Vanneste & Pouthas, 1999). However, it is not clear whether such tasks primarily measure *biopsychological timing*, which is defined as "timing based directly on . . . biologically based mechanisms such as internal clocks or oscillators" (Wearden, 1994, p. 217). Some researchers have tried to examine these basic mechanisms more directly, by measuring discrimination of very brief temporal intervals (Rammsayer et al., 1993) or movement production in tapping tasks (Ducheck, Balota, & Ferraro, 1994; Krampe, Mayr, & Kliegl, 2000). Their findings suggest that the functioning of the internal clock is not affected by age. Other recent studies that have examined basic timing processes by using stimulus identification tasks have found evidence of at least moderate age effects (McCormack, Brown, Maylor, Darby, & Green, 1999; Wearden et al., 1997). However, it is possible that this is because stimulus identification tasks rely on long-term memory processes in addition to internal clock processes.

For example, in one of McCormack et al.'s (1999) and Wearden et al.'s (1997) experiments, during a preexposure phase participants were presented with a reference duration. At test, they were required to judge whether presented durations were the same as the reference duration that they had encountered earlier. (In what follows, tasks that involve making judgments with respect to preexposed reference durations are referred to as *duration identification tasks*.) The age effects found in the McCormack et al. and Wearden et al. studies may have been due to age-related changes in the memory processes that deal with representations of reference durations, rather than changes in the functioning of the internal clock itself. The current study examined the nature of such age effects in more detail.

Such effects can be understood within the context of formal models of human and animal timing. Many recent studies of human timing (Allan & Gibbon, 1991; Wearden, 1991b; Wearden & Ferrara, 1995, 1996; Wearden, Rogers, & Thomas, 1997) have used duration identification tasks that have been adapted from tasks used to study animal timing (Church & Deluty, 1977; Church & Gibbon, 1982). The findings of such studies have been interpreted within the framework of scalar expectancy theory (SET), the most widely applied theory of human and animal timing (Church, 1984; Gibbon, 1977; Gibbon, Church, & Meck, 1984; Wearden, 1991a, 1994; Wearden & Lejeune, 1993). According to SET, biopsychological timing is based on the output of an internal clock that generates representations of durations. Memory representations formed from the output of the internal clock are held in working memory and may be transferred to long-term memory. For example, in tasks such as those used by McCormack et al. (1999) and Wearden et al. (1997), preexposed reference durations are assumed to be stored in long-term memory. Decision processes

---

Teresa McCormack, Gordon D. A. Brown, Elizabeth A. Maylor, Lucy B. N. Richardson, and Richard J. Darby, Department of Psychology, University of Warwick, Coventry, England.

This research was supported by Grants G9608199, G9606610N, and G9805928 from the Medical Research Council; Grant R000 23 9002 from the Economic and Social Research Council; and Grant 88/515050 from the Biotechnology and Biological Sciences Research Council. We are grateful to Ian Neath and Nick Chater for discussion of the model described in this article and to Ralf Krampe and Anna Eisler for their helpful comments on a version of this article. We are also grateful to Fiona Anderson for recruiting the older participants in Experiments 2 and 3.

Correspondence concerning this article should be addressed to Teresa McCormack, Department of Psychology, University of Warwick, Coventry CV4 7AL, England. E-mail: T.McCormack@warwick.ac.uk

compare test durations held in working memory with the representations of reference durations retrieved from long-term memory. On any given test trial, the similarity between the representation of a just-presented interval and the representation of the reference interval retrieved from long-term memory controls the response.

With this approach, we can distinguish between at least two different types of possible age-related impairments, which correspond to separate parameters in mathematical models of SET. First, there could be an increase with age in the variability or noise associated with internal clock, memory processes, or both (McCormack et al., 1999; Wearden et al., 1997). An increase in noise with age would lead to a general reduction in overall levels of performance. However, proportional increases in noise will not lead to qualitative differences in the patterns of errors made by participants. Second, there could be changes with age in the extent to which long-term memory representations of reference durations are *distorted* (Meck, 1996). Distortion is assumed to affect the way reference durations are represented in long-term memory, resulting in them being systematically misremembered as either too short or too long. For example, consider a task in which participants are preexposed to a reference duration of 2 s, and are then required to judge whether test durations are the same length as this reference duration. If lengthening in reference memory occurs, the 2-s duration might be misremembered as equivalent to a test duration of 2.1 s. Alternatively, shortening in memory might lead to the reference duration being misremembered as equivalent to 1.9 s. Each type of memory distortion would lead to a distinctive pattern of errors on subsequently presented test durations.

There is some evidence from the animal literature that impaired timing in aged rats is due to the distortion in long-term memory of representations of reference durations rather than to greater noise in timing (Lejeune, Ferrara, Soffie, Bronchart, & Wearden, 1998; Meck, Church, & Wenk, 1986). Specifically, aged rats show a pattern of performance that suggests that they remember reference durations as being too long. In the studies of rats, the increase with age in distortion has been linked with changes in levels of the neurochemical choline (Meck et al., 1986), levels of which are also thought to change with human aging (Bartus, Dean, & Flicker, 1987). Thus, on the basis of the animal findings, changes with age in memory distortion of durations would be predicted for human timing. However, McCormack et al. (1999) and Wearden et al. (1997) were able to model age-related declines in performance on their timing tasks by assuming that there was an increase with age in levels of noise, rather than changes in memory distortion.

In the present study, we used a type of duration identification task that yields a sufficiently rich data set that allowed us to examine the nature of age effects in more detail. The type of task used in the present study is usually described as an *absolute identification* task. Absolute identification tasks have been carried out using a variety of perceptual dimensions, such as duration, pitch, weight, and line length (Berliner & Durlach, 1973; Bower, 1971; Hawkes, 1961; Miller, 1956; Murdock, 1960; Pollack, 1952). In such tasks, participants are preexposed to a series of stimuli that vary along the given dimension (e.g., a series of tones that range from low to high pitch or a series of lines that vary in length from short to long). Each stimulus is usually assigned a number according to its place in the series (e.g., in a six-stimulus series, the stimulus with the lowest value is labeled *1* and the stimulus with the highest value is labeled *6*). Test trials consist of participants judging the number of a presented stimulus, thus assigning it a position within the series.

The absolute identification task is structurally similar to the duration judgment tasks that we have used previously within the SET framework (McCormack et al., 1999) because participants must judge stimuli in terms of previously presented reference durations. An important advantage of using an absolute identification task to study age effects on duration judgments is that the task generates detailed patterns of error data for modeling purposes. If age differences are due to memory distortion, there should be qualitative changes in the pattern of errors, whereas if such differences are due to changes in levels of noise, there should simply be differences in the overall number of errors (see Modeling section). The memory load of the task can also be varied by altering the number of to-be-identified durations in the series.

Experiments 1 and 2 examined age effects on tasks involving absolute identification of durations, whereas Experiment 3 addressed a further issue about the nature of the age-related declines found in previous studies of duration identification. Are the age effects found in previous studies the result of changes in processes specific to timing, or do older adults have global problems in stimulus identification? In Experiment 3, we examined whether patterns of age-related declines are similar on an absolute identification task when pitch rather than duration is the underlying perceptual dimension. As far as we are aware, the stimulus identification tasks used in previous aging studies of duration judgments have not been used to examine whether there are also age-related declines in performance when a dimension other than duration is involved. However, research on clinical populations has already established the possibility of a dissociation between impairments in judgments involving the durations of stimuli versus other stimulus dimensions. For example, Nichelli, Alway, and Grafman (1996) found that patients with cerebellar damage were impaired on a task involving judgments of durations, but not on an analogous task involving judgments of line length.

General Methodology

All of the experiments to be presented used an extreme groups design in which the performance of younger adults (aged approximately 20 years) was compared with that of older adults (aged approximately 70 years). Participants for each experiment were recruited in the same way from large populations of volunteers. No one participated in more than one experiment. To avoid repetition, participant information that is common to the three experiments is summarized here.

The younger participants were undergraduate students at the University of Warwick who received course credit for their participation. The older participants were members of a volunteer panel who had been recruited for a study of memory and aging through local newspaper articles and advertisements in libraries, supermarkets, and health centers; they were required to make their own travel arrangements to attend a testing session held at the University of Warwick. Older adults were paid £5 (approximately $7) as a contribution toward their travel expenses. All participants were tested in groups (maximum of 10 per group).

Table 1 shows background information for the older adults in each experiment. With the exception of age, background informa-

Table 1
Background Data for Participants in Experiments 1, 2, and 3

	Young		Experiment 1: Old (*n* = 50)		Experiment 2: Old (*n* = 32)		Experiment 3: Old (*n* = 33)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	19.50[a]	1.43	74.12	4.73	70.30	9.25	68.73	8.19
Fluid ability[b]	102.78[c]	11.64	67.02	16.49	61.63	21.87	64.61	17.09
Speed[d]	71.59[e]	9.64	43.04	11.18	44.84	11.56	42.78	13.17
Vocabulary[f]	18.77[g]	2.83	23.62	3.78	22.88	3.83	22.30	3.45
Current health[h]	4.02[i]	0.80	3.94	0.74	3.90	0.64	3.82	0.68
Hearing (corrected)[h]	4.27[j]	0.70	3.65	0.72	3.84	0.72	3.91	0.72
Eyesight (corrected)[h]	4.31[k]	0.80	4.06	0.56	4.06	0.67	4.18	0.53

[a] *n* = 231. [b] AH4 test, a timed problem-solving test of fluid ability (Heim, 1968). [c] *n* = 147. [d] Digit Symbol Substitution test from the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). [e] *n* = 231. [f] Part 1 of the Mill Hill Vocabulary Test (Raven, Raven, & Court, 1988). [g] *n* = 84. [h] Self-rated on a 5-point scale (1 = *very poor,* 2 = *poor,* 3 = *fair,* 4 = *good,* 5 = *very good*). [i] *n* = 216. [j] *n* = 214. [k] *n* = 210.

tion was unavailable for the younger participants who took part in the current experiments. However, background information was available for more than 200 younger adults drawn from the same undergraduate population; for comparison, their details are shown in the first column of Table 1. The AH4 (Heim, 1968) is a timed problem-solving test of fluid ability or intelligence, equally divided between verbal and arithmetic problems and spatial problems. The scores from the two halves of the test have been combined in Table 1. Vocabulary was assessed by using the first part of the Mill Hill Vocabulary Test (Raven, Raven, & Court, 1988), in which participants are required to select the best synonym for a target word from a set of six alternatives. Speed of processing was measured by using the Digit Symbol Substitution subscale from the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). In terms of the younger adults' standard deviation units, it can be seen from Table 1 that the means for the older adults were between 3.07 and 3.54 *SD*s below the mean for the younger population for fluid ability, between 2.77 and 2.99 *SD*s below the younger adults' mean for speed, but between 1.25 and 1.71 *SD*s above the younger adults' mean for vocabulary. Thus, the present background data are consistent with the pattern of aging typically reported in the literature (e.g., Salthouse, 1991) of age-related decline in fluid intelligence and speed, but higher levels in older than in younger adults of crystallized intelligence (as measured here by vocabulary).

Self-rated measures of participants' current state of health, eyesight (with glasses or contact lenses, if worn), and hearing (with aids, if worn) are also included in Table 1. In terms of younger adults' standard deviation units, the older means were 0.10–0.25 *SD*s below the younger means for health, 0.51–0.89 *SD*s below for hearing, and 0.16–0.31 *SD*s below for eyesight. However, all of the means were generally high, with averages equivalent, or close to, ratings of *good*.

Experiment 1

In this experiment, older adults' performance was compared with that of younger adults on an absolute identification task in which there was a series of six stimuli that varied in duration. To examine whether age effects are consistent over different duration values, we used two stimulus sets that spanned different ranges of durations. There is some debate in the literature regarding the extent to which timing of different duration values exploits different timing mechanisms (see Gibbon, Malapani, Dale, & Gallistel, 1997). For example, Rammsayer (1992, 1994, 1999; Rammsayer & Lima, 1991) has claimed that timing of durations shorter than approximately 500 ms is mediated by different mechanisms than timing of intervals in the range of 1 to 2 s, although other authors have made parallel claims regarding different duration ranges (see Clarke, Ivry, Grinband, Roberts, & Shimizu, 1996; Poppel, 1996). Gibbon et al. have argued that demonstrating that impairments in timing are not dissociable across different duration values would provide crucial evidence for common timing mechanisms across the range of durations.

Method

Participants. There were 43 younger participants (23 women and 20 men) with a mean age of 19.44 years (*SD* = 1.55) and 50 older participants (24 women and 26 men). They were randomly assigned to one of two conditions that differed in terms of the durations of the stimuli used in the task: Twenty-four younger adults and 26 older adults completed a short condition, and 19 younger adults and 24 older adults completed a long condition.

Apparatus and stimuli. The experiment was run on an Apple Macintosh computer, and stimulus presentations were controlled by the Superlab software package. The auditory stimuli were 500-Hz sinusoidal tones generated using the SoundEdit Pro software package, with the first and last 50 ms of each tone ramped. The tones were presented over external speakers attached to the computer. Durations of the tones used in each condition are shown in Table 2. The tones in each series were equally spaced on a logarithmic scale, and the spacing was identical across the two conditions. The longest tone in the short condition was of the same duration as the shortest tone in the long condition. The visual displays that provided feedback during the task were presented on a large (51 cm diameter) black-and-white screen. The correct number for each test stimulus was displayed on the screen as a number approximately 12 cm high. Participants made their responses by writing the number of the tone on a response sheet.

Procedure. Participants were seated at individual desks in a laboratory. The average viewing distance from the computer screen was approximately 2 m, with no participant being more than 3.8 m from the screen. Those with

Table 2
Stimulus Durations (in Milliseconds) in Experiment 1

	Condition	
Tone no.	Short (ms)	Long (ms)
1	250.0	622.1
2	300.0	746.5
3	360.0	895.8
4	432.0	1,075.0
5	518.4	1,290.0
6	622.1	1,548.0

either poor eyesight or poor hearing were encouraged to sit at the front of the room; no participant reported problems in either hearing the auditory stimuli or reading the visual displays.

In both conditions, participants received the test stimuli in random order with the constraint that over the course of the experiment each stimulus was preceded three times by each other stimulus in the set. In order that this constraint could be met, participants received an extra trial on one of the test stimuli: There were 18 trials that involved identifying Tones 1, 2, 3, 5, and 6 and 19 trials that involved identifying Tone 4 (making 109 trials in total).

The procedure for the two conditions (short and long) was identical. Participants were told that they would hear some sounds and would have to make judgments about them based on their duration. They were also told that there was a set of six tones that formed a series from short to long, with Tone 1 being the shortest in the series and Tone 6 being the longest, and that their task was to judge the number of each test tone that was presented. They were instructed to give a response to every trial even if they were unsure.

In the initial exposure phase of the experiment, participants heard the series of tones four times, twice in ascending order (shortest to longest) and twice in descending order (longest to shortest). One second after each tone was played in the exposure phase, the visual display “That was Tone N” appeared on the computer screen for 2 s. The test trials were administered immediately after this exposure to the series of test tones. Before each trial, the experimenter announced the trial number to ensure that participants were completing the correct row on their response sheet. Following this, the display “Ready?” appeared on the computer screen for 2.5 s, and was then replaced by a blank screen. The test tone was played 1.5 s later, and participants made their response immediately. Once all participants had responded and were looking at the computer screen, the experimenter initiated presentation of the correct answer for that trial by pressing a key on the computer keyboard. The display “That was Tone N” then appeared on the screen for 2 s. The experimental session lasted for approximately 50 min in total.

Results

Figures 1A (short condition) and 1B (long condition) show the proportion of correct responses for each tone number. It can be seen that the younger adults appear to perform better than the older adults in both conditions, although the age differences appear to be more consistent across the series in the short condition. Consistent with previous research on absolute identification, the plots of correct performance as a function of stimulus serial position are approximately U shaped, with better performance on stimuli at the start and end of the series.

An initial analysis of variance (ANOVA) on correct performance was conducted with between-subjects variables of age and condition and a within-subject variable of tone number. The main effect of age was significant, $F(1, 89) = 22.48$, $MSE = .05$, $p < .01$, partial $\eta^2 = .20$, with the overall performance of the younger adults better than that of the older adults ($M = .61$ and $M = .52$, respectively). The effect of condition was marginally significant, $F(1, 89) = 3.91$, $MSE = .05$, $p = .051$, partial $\eta^2 = .04$, with performance in the long condition better than that in the short condition ($M = .58$ and $M = .54$, respectively). The interaction between condition and age was not significant ($F < 1$). The main effect of tone number was significant, $F(5, 445) = 33.23$, $MSE = .03$, $p < .01$, partial $\eta^2 = .27$, as was the interaction between tone number and condition, $F(5, 445) = 4.55$, $MSE = .03$, $p < .01$, partial $\eta^2 = .05$ (see Figure 1). There were no other significant interactions (all F values < 1.45).

There was a small difference in the mean ages of the two groups of older adults in the two conditions (mean ages of 75.3 and 72.8 years for the short and long conditions, respectively), and this group difference may have been responsible for the overall marginal effect of condition. Thus, although the interaction between age and condition was not significant, we checked whether the condition effect was significant for the older adults when age differences were taken into account ($F < 1$ for the condition effect of the younger adults). An analysis of covariance (ANCOVA) on overall correct performance was conducted on the data of the older adults only, with a between-subjects variable of condition and age as a covariate. The effect of condition was significant, $F(1, 47) = 5.14$, $MSE = .05$, $p < .05$, with the overall proportions of correct responses having adjusted means of .49 in the short condition and .55 in the long condition.

Subsequent analyses examined participants’ erroneous responses. Figures 2A (short condition) and 2B (long condition) show the distributions of responses given to each tone (e.g., for Tone 1, the proportion of responses were 1, 2, 3, and so on). It can be seen that these response gradients peak at the correct number and decline with distance from the correct number. In both conditions, the response gradients appear to be steeper in the younger adults. Group differences in the steepness of response gradients were examined by calculating for each participant the proportion of erroneous responses that was one number away from the correct one. For the short condition, the mean proportions of errors of this type were .91 ($SD = .02$) and .82 ($SD = .02$) for the younger and older adults, respectively; for the long condition, the means were

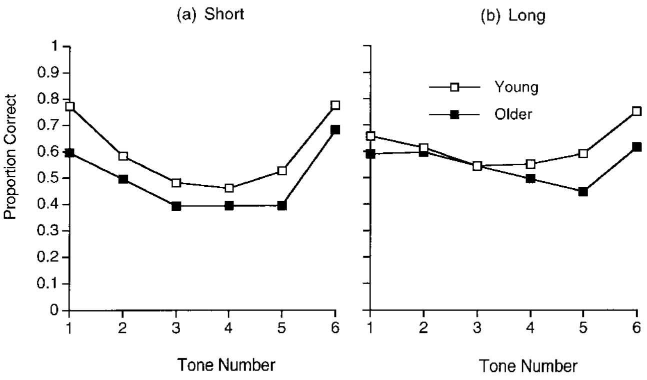
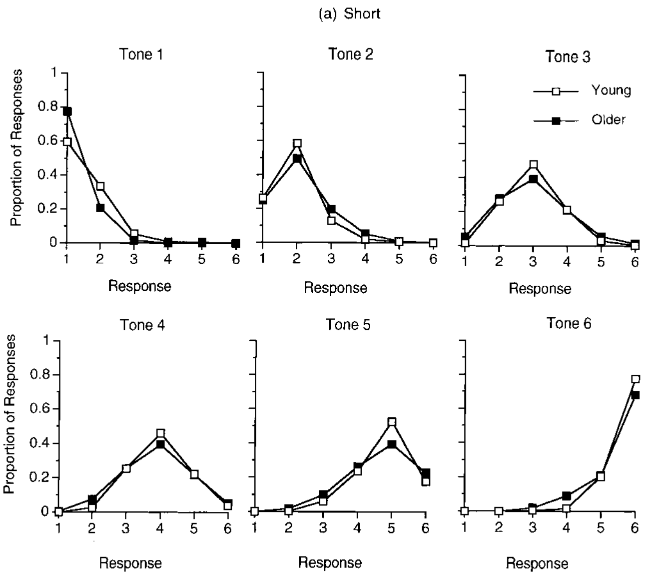


Figure 1. Proportion of correct responses in Experiment 1 as a function of age and tone number for the (a) short condition and (b) long condition.



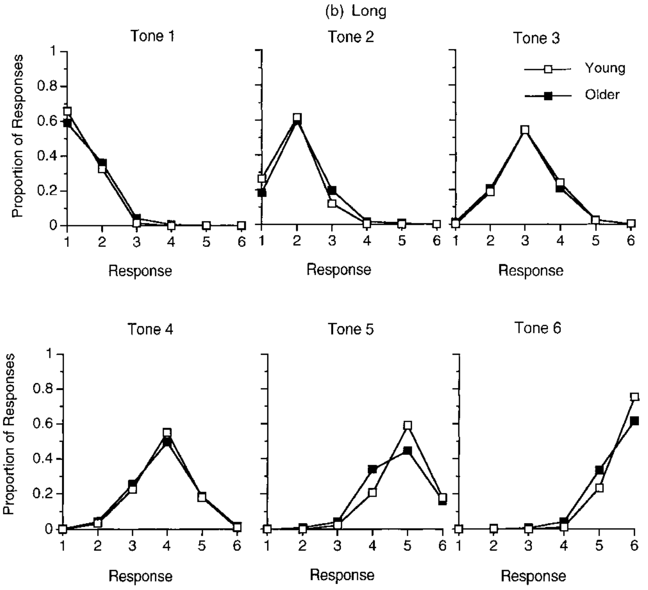


Figure 2. Response gradients for each tone as a function of age for the (a) short condition and (b) long condition of Experiment 1.

.95 (*SD* = .01) and .91 (*SD* = .01) for the younger and older adults, respectively. Two *t* tests (one for each condition) on these data showed that a greater proportion of younger adults' than older adults' erroneous responses were just one number away from the correct one: $t(48) = 3.50$, $p < .01$, for the short condition, and $t(41) = 2.62$, $p < .05$, for the long condition.

Finally, we examined whether the average response given to each tone differed between the groups. It can be seen from Figures 2A and 2B that the response gradients are typically symmetrical, with the erroneous responses of participants being approximately evenly distributed between numbers greater and smaller than the tone's actual number. Figures 3A (short condition) and 3B (long condition) show the average response given to each tone. It can be seen that the average response to each tone is typically close to the tone's true number, with participants showing no consistent tendency to give a lower or higher response than a tone's actual number. The tendency of the older groups to have a higher average response for Tone 1 and a lower average response for Tone 6 reflects the lower levels of correct responding in the older group, because errors for these tones are numbers toward the middle of the range. Two ANOVAs (one for each condition) examined these data, with a between-subjects variable of age and a within-subject variable of tone number. The main effect of age was not significant for either of the conditions (both *F* values < 1), although the interaction between age and tone number was significant for both conditions: $F(5, 240) = 6.01$, $MSE = .06$, $p < .01$, partial $\eta^2 = .11$, for the short condition, and $F(5, 205) = 5.31$, $MSE = .05$, $p < .01$, partial $\eta^2 = .12$, for the long condition.

Discussion

The central finding of Experiment 1 was that performance on the duration identification task was poorer in older than in younger adults. However, declines in performance with age, although significant, were small, with performance remaining reasonably accurate in the older group. Our findings are consistent with previous reports of modest age-related declines on tasks involving judgments of the durations of stimuli (McCormack et al., 1999; Wearden et al., 1997), although the present study used a different task to examine this decline. The age differences found in this experiment are considered in more detail in the Modeling section.

The duration identification task yielded serial position curves and response gradients that were generally similar to those found in tasks involving other stimulus dimensions, although we note that the serial position curves for the short condition were more U shaped than those for the long condition. There was also some evidence that the short condition was more difficult for older participants than the long condition. The possibility that older adults find shorter durations more difficult to identify is explored in the second experiment.

Experiment 2

In Experiment 2, we increased the number of tones to be identified from six to nine to examine the possibility that age

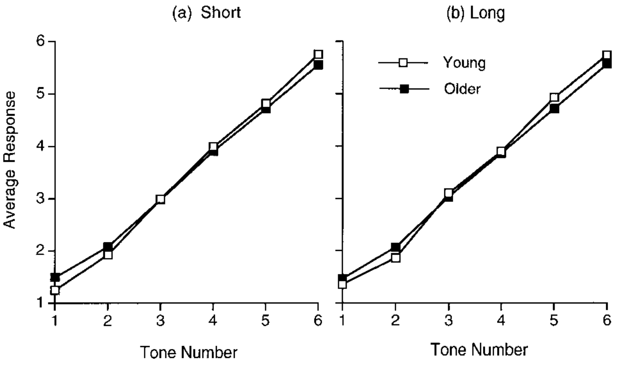


Figure 3. Average response to each tone as a function of age for the (a) short condition and (b) long condition of Experiment 1.

effects will be more marked when there are more stimuli to be identified. The tones also spanned a larger range of durations (from 250 ms to just over 2 s). One advantage of using a larger series of tones that covers a wider range of durations is that it enables a within-subject comparison of performance on the identification of short and long tones. If older adults do find shorter durations more difficult to identify, we might expect to find the serial position curves of this group to be asymmetrical, with poorer performance on tones at lower serial positions.

Method

Participants. The participants were 21 younger adults (12 women and 9 men), with a mean age of 19.53 years ($SD = 0.65$), and 32 older adults (20 women and 12 men).

Apparatus and stimuli. The apparatus and stimuli were identical to those used in Experiment 1, except that there was a series of nine tones in this experiment, the durations of which are shown in Table 3. All of the tones in the series had a pitch of 500 Hz, and their durations were evenly spaced along a logarithmic scale. Participants made their response by circling a number on a response sheet that contained 135 rows of the numbers 1–9.

Procedure. Both the initial exposure phase and the testing phase of the experiment were similar to those used in Experiment 1. Tones were presented in a random order, with each tone in the series presented 15 times during the test, making a total of 135 test trials. Participants assigned a number from 1 to 9 to each test tone. As in Experiment 1, participants received feedback during testing, with the correct answer for each trial displayed on the computer screen immediately after all participants had made their response.

Results

The mean proportion of correct responses to each tone is shown in Figure 4. It can be seen from the figure that the older adults made fewer correct responses than did the younger adults. An ANOVA was conducted on these data, with a between-subjects variable of age and a within-subject variable of tone number. The main effect of age was significant, $F(1, 51) = 27.28$, $MSE = .07$, $p < .01$, partial $\eta^2 = .35$, with the overall performance of the younger adults better than that of the older adults ($M = 0.56$ and $M = 0.43$, respectively).[1] The main effect of tone number was also significant, $F(8, 408) = 58.09$, $MSE = .02$, $p < .01$, partial $\eta^2 = .53$, with better performance on stimuli at the start and end of the series. The interaction between age and tone number was significant, $F(8, 408) = 4.29$, $MSE = .02$, $p < .01$, partial $\eta^2 = .08$. Post hoc comparisons showed that the age differences were significant on all tones except for Tone 1 and Tone 7 (a significance level of $p < .05$ was set for these and subsequent analyses).

Table 3
Stimulus Values in Experiments 2 and 3

	Stimulus dimension	
Tone no.	Duration (ms; Experiment 2)	Pitch (Hz; Experiment 3)
1	250.0	350.0
2	325.0	376.6
3	422.5	405.2
4	549.3	436.0
5	714.0	469.2
6	928.2	504.8
7	1,206.7	543.2
8	1,568.7	584.5
9	2,039.3	628.9

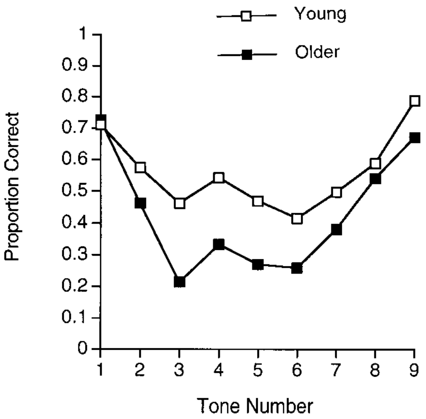


Figure 4. Proportion of correct responses in Experiment 2 as a function of age and tone number.

Figure 5 shows both groups' response gradients for each tone. It can be seen from the figure that, as in Experiment 1, the response gradients were less steep for the older adults than for the younger adults. For each participant, we calculated the proportion of erroneous responses that were one number away from the tone's actual number. The mean proportions of errors of this type were .87 ($SD = .03$) and .74 ($SD = .02$) for the younger and older adults, respectively. A *t* test on these data showed that a greater proportion of younger adults' than older adults' erroneous responses were just one number away from the correct number, $t(51) = 3.66$, $p < .01$.

Examination of the response gradients in Figure 5 suggests a further, yet more striking, difference between the response distributions of younger and older adults. When older adults made an error, they were more likely than were younger adults to give a response number that was lower than the tone's actual number. Indeed, for some tones, older adults' response gradients peak at a lower number than the tone's actual number. For example, older adults gave more "2" than "3" responses to Tone 3. By contrast, the response gradients of the younger adults peak at the correct response for all tones.

Figure 6 shows the average responses given to each tone. It can be seen from the figure that the older adults gave a lower average response to most tones than the younger adults, whose average responses were closer to the correct numbers. An ANOVA was conducted on these data, with a between-subjects variable of age and a within-subject variable of tone number. The main effect of age was significant, $F(1, 51) = 18.28$, $MSE = .65$, $p < .01$, partial

---

[1] Within the old group, the correlation between age and overall correct performance was also significant, $r(32) = -.45$, $p < .01$. To examine possible effects of hearing ability, we also examined the correlation between hearing (self-rated on a 5-point scale) and performance with this group. This correlation was not significant, $r(32) = .15$, $p = .21$, and the correlation between age and performance remained significant when hearing was partialled out, $r(29) = -.43$, $p < .02$.

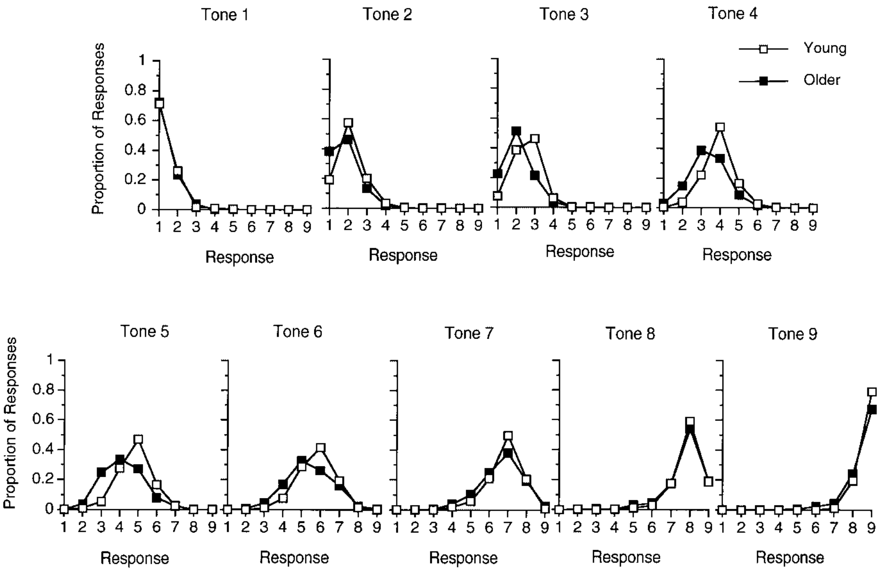


Figure 5. Response gradients for each tone in Experiment 2 as a function of age.

$\eta^2 = .26$, as was the interaction between age and tone number, $F(8, 408) = 6.30$, $MSE = .09$, $p < .01$, partial $\eta^2 = .11$. Further post hoc comparisons showed that the average response of the older adults was significantly lower than that of the younger adults for Tones 2, 3, 4, 5, and 6. Such a tendency to give lower average responses is not simply an effect of overall poorer performance in the older group because a reduction in accuracy of performance should lead to a flattening of the plot of average responses, with higher average responses to tones of lower numbers and lower average responses to tones of higher numbers.

To check whether these age effects were consistent across the experimental session, and, for example, were not induced by test fatigue, performance on the first half of the experiment (Trials 1–67) was compared with performance on the second half (Trials 68–135). An ANOVA was conducted on the proportion of correct responses to each tone, with a between-subjects variable of age and within-subject variables of test half and tone number. Neither the main effect of half, $F(1, 51) = 2.20$, $MSE = .05$, $p = .14$, partial $\eta^2 = .04$, nor the interaction between half and age ($F < 1$) was significant, and the three-way interaction between age, half, and tone number was also not significant, $F(8, 498) = 1.20$, $MSE = .03$, $p = .30$, partial $\eta^2 = .02$. An ANOVA was also conducted on the average response given to each tone, with a between-subjects variable of age and within-subject variables of half and tone number. Although the main effect of half was significant, $F(1, 51) = 5.53$, $MSE = .27$, $p = .02$, partial $\eta^2 = .10$, with the average response to each tone slightly higher in the second than in the first half, there was no significant interaction between age and half ($F < 1$), and the three-way interaction between age, half, and tone was also not significant ($F < 1$).

Thus, there was an effect of age on correct performance that was consistent across the testing session. The age difference in error patterns (i.e., the tendency for older adults to give a lower response to each tone) also did not vary significantly across the session. There are at least two possible explanations for this pattern of errors in the older adults. First, older adults may have given a lower response to each tone because of an age change in the way tones in the series are remembered. Specifically, older adults may have had distorted representations of tone durations in long-term memory, such that they remembered tones as being longer than they really were. For example, if Tone 2 was represented in long-term memory as being longer than it really was, older participants would be likely to give an erroneous “2” response to (the longer) Tone 3 at test, and so on.

The second possibility is that the older adults’ pattern of responding simply reflects a response bias, such that this group preferred to use lower response numbers. If this is correct, then inspection of the number of times that older adults used each response number should reveal that lower response numbers were used more often than higher response numbers, despite there being an identical number of trials of each type. The number of times that

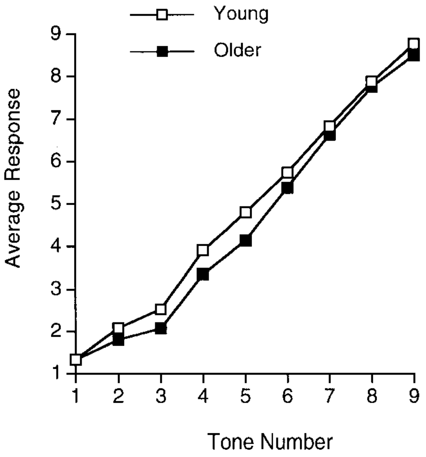


Figure 6. Average response to each tone in Experiment 2 as a function of age.

each response number was produced across all test trials is shown in Figure 7. (If participants used each number equally often, then the mean number of each response should be 15.) It is clear that older participants were indeed generally more likely to choose lower response numbers than larger ones, whereas younger adults used each response number approximately equally often. Thus, it is possible that older participants tended to give lower responses to tones simply because of a response bias in favor of choosing lower response numbers more often. This explanation, in terms of the causal effects of a response bias, can be contrasted with the first explanation (distorted representations), according to which a response bias could emerge purely as a consequence of distorted reference memory representations of the stimuli.

To rule out this possible explanation, further analyses examined the way in which each response number was distributed across tones (e.g., the number of times the response “2” was given to Tone 1, to Tone 2, to Tone 3, and so on. This can be contrasted with the previously described response gradients, which show the pattern of responses for any given tone, rather than how any given response was distributed across tones). If the explanation for the older adults’ pattern of errors is that they simply preferred to use lower response numbers more often, if we look at any given response, we would not expect there to be age differences in the symmetricality of the distribution of that response across tones.

An example may make this clearer: The response “2” could be used very frequently (e.g., such that it is given more frequently to Tone 3 than is the response “3” itself), but errors involving this response could nevertheless be evenly distributed between Tone 1 and Tone 3. If, however, older participants systematically give a lower response because of memory distortion, each response number should have been given to a tone that had a greater number than the correct tone (e.g., the response “2” should be more likely to have been given to Tone 3 than to Tone 1, or even than to Tone 2). Such a tendency would be expected to be present regardless of the number of times each response was actually given: A similar prediction can be made for the least frequent response of the older adults, which was “6” (i.e., we would expect “6” to have been frequently given as a response to Tone 7).

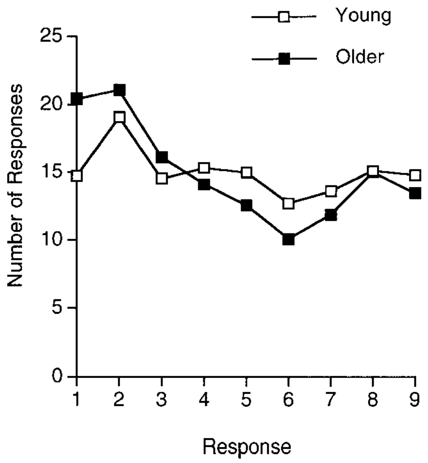


Figure 7. Mean number of responses of each type as a function of age in Experiment 2.

Figure 8 shows the distributions of each response number across all tones (i.e., for each response number, the number of times it was given to each tone). It can be seen from this figure that older adults are more likely than younger adults to produce a given response to a tone higher than the tone for which that response is correct. For some response numbers, the peak of the older adults’ distribution of the response across tones is actually displaced to the higher tone.

Thus, the tendency for older adults to give a lower response to tones does not seem to be due to a response bias in favor of choosing lower response numbers. The uneven distribution of responses of each type shown in Figure 7 may itself be a result of changes in memory processes with age.

Discussion

The findings of this experiment were consistent with those of Experiment 1 in that there were significant age differences in performance on the duration identification task. The age differences in this experiment were more marked than those in Experiment 1, and, although the older group’s performance was well above chance for all tones, they did particularly badly on the middle serial positions (20–30% correct). As in Experiment 1, the serial position curves for both younger and older adults were approximately U shaped. There was no evidence from this experiment that the older adults found identification of shorter tones more difficult than longer tones. However, the older group did show a pattern of errors that was qualitatively different from that of younger adults: When older adults made an error, it tended to be to give a lower number in response to the test tone than the tone’s actual number (e.g., they often gave the response “2” to Tone 3). This finding was not predicted on the basis of the error patterns in Experiment 1, although the age differences in the previous experiment were relatively small compared with those found in Experiment 2, in which there were more tones in the series.

The older adults’ pattern of errors may be due to changes with age in memory for the tones in the series, such that this group remembered the tones as being longer than they really were. The possibility that the performance of older adults can be explained in terms of such a distortion process is explored in the Modeling section. The idea that reference durations can become systematically misremembered in this way through some distortion process is one that has occurred in the context of accounts of both human (McCormack et al., 1999) and animal timing (Meck, 1983, 1996; Meck, Church, & Olton, 1984). Of importance, the shifts of the peaks of the response gradients observed in the current study are reminiscent of the peak shifts observed in studies of timing in aging rats in which the peak procedure is used (Lejeune et al., 1998; Meck et al., 1986). In accounting for such peak shifts, animal researchers have also assumed that older rats distort durations in long-term memory, such that reinforced reference durations are remembered as being too long (for review, see Meck, 1996).

One possible alternative explanation for the current findings lies in the spacing of the tones. The tones in the series were equally spaced along a logarithmic scale; thus, if the tone durations were plotted on a linear scale, the spacing between the tones would increase across the series. If durations were represented psychologically along a linear rather than a logarithmic scale (Gibbon &

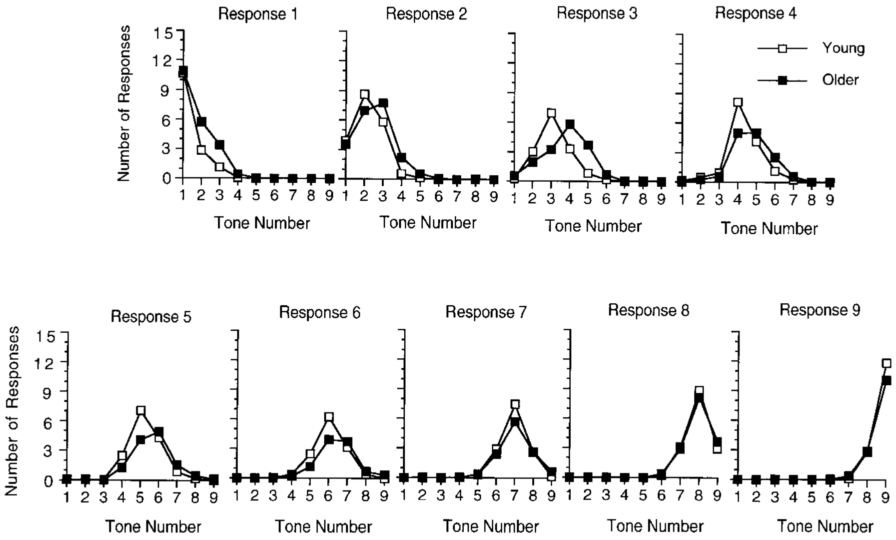


Figure 8. Distribution gradients for each response in Experiment 2 as a function of age.

Church, 1981), then particular errors patterns might be predicted: Participants might be more likely to confuse a tone with one of a lower rather than a higher serial position because any given tone will be closer to this lower tone than to the one above it on the internal psychological scale. Thus, perhaps the error patterns of the older adults could potentially be explained in this way because the erroneous responses of this group tended to be a response lower than the correct number. The issue as to whether durations are represented logarithmically rather than linearly is one that has been the subject of much debate (see Gibbon & Church, 1981), and has yet to be resolved (Staddon & Higa, 1999). However, an explanation of the older adults' data in terms of linear rather than logarithmic representation of duration seems difficult for two reasons. First, such an account would not predict that a lower response should actually be given more often than the correct one, as is seen in the older adults (i.e., although it would predict asymmetries, it would not predict peak shifts). Second, the error patterns of the younger adults are not consistent with such an account (see Figure 8). Any asymmetries in the response gradients of the young group were small and certainly less marked than would be expected if the tones were perceived as unevenly spaced.

Experiment 3

In Experiment 3, participants received an absolute identification task that was identical to that of Experiment 2, except for the stimulus dimension: The tones in the series varied in pitch rather than in duration. This experiment enabled us to address two questions raised by our previous findings. First, do age differences in levels of performance on absolute identification extend to tasks that involve dimensions other than duration? Second, do older adults show a similar tendency to give a lower response to tones in an absolute identification task when a dimension other than duration is used? To address the first question, we chose the dimension of pitch because it allowed the experimental procedure to be as close as possible to that used in Experiment 2. If there are age differences on pitch identification, it would suggest that there may be a general age-related decline in absolute identification performance, rather than a specific impairment in temporal processing in older adults. We note that such a finding has consequences for the interpretation of the findings from previous recent studies, including our own, that have established age-related declines in judging duration (McCormack et al., 1999; Wearden et al., 1997).

However, age-related changes in performance may be qualitatively different for duration identification than for pitch identification. In particular, there is no reason to believe that aging may lead to representations of pitch that are systematically distorted in memory, whereas the suggestion that memory representations of durations may be distorted in certain populations has occurred previously in the timing literature (Lejeune et al., 1998; McCormack et al., 1999; Meck et al., 1986). Thus, there are grounds for expecting different error patterns in older adults in pitch identification than in duration identification.

Method

Participants. Twenty young adults (10 women and 10 men), with a mean age of 20.31 years (*SD* = 3.70), and 33 older adults (22 women and 11 men) participated in the experiment.

Apparatus and stimuli. These were identical to those used in Experiment 2, except that the nine tones in the series differed in terms of their pitch rather than their duration. The duration of all of the tones in the series was 1,000 ms, with the first 50 ms and last 50 ms ramped. The pitches of the tones, which were equally spaced on a logarithmic scale, are listed in Table 3.

Procedure. The structure of the task was identical to that used in Experiment 2, with each tone appearing 15 times during the test phase, giving a total of 135 experiment trials. Participants were told that they would hear some sounds and would have to make judgments about them based on their pitch. They were told that there was a set of nine tones that formed a series from low pitch to high pitch, with Tone 1 being the lowest in the series and Tone 9 being the highest, and that their task was to judge the number of each test tone that was presented. All other aspects of the procedure were identical to those used in Experiment 2.

Results

The proportion of correct responses to each tone is shown in Figure 9. As in Experiment 2, there appear to be age differences in performance. An ANOVA on these data found a significant effect of age on correct performance, $F(1, 51) = 16.32$, $MSE = .14$, $p < .01$, partial $\eta^2 = .24$, with the overall performance of younger adults better than that of older adults ($M = .54$ and $M = .39$, respectively).[2] There was also a significant effect of tone number, $F(8, 408) = 21.41$, $MSE = .02$, $p < .01$, partial $\eta^2 = .30$, with better performance on stimuli at the start and the end of the series. The interaction between age and tone number was not significant, $F(8, 408) = 1.42$, $MSE = .02$, $p = .21$, partial $\eta^2 = .03$.

Figure 10 shows the response gradients for each tone as a function of age. It can be seen that the gradients of the older adults appear to be less steep than those of the younger adults, and a *t* test on the proportions of erroneous responses that were one number away from the correct number revealed a significant effect of age, $t(51) = 3.37$, $p < .01$. The mean proportions of errors of this type were .74 ($SD = .03$) and .61 ($SD = .02$) for the younger and older adults, respectively. It can be seen from Figure 10 that the response gradients are generally symmetrical, and there is no evidence in the older group of a tendency to underestimate tones' numbers. Unlike in Experiment 2, all of the gradients peak at the correct response for each tone. Figure 11 shows the average response given to each tone by both groups. It can be seen from the figure that the average responses are similar for both groups, and an ANOVA on these data found no significant effect of group and no significant interaction between group and tone number (both *F* values < 1).

Discussion

As in the previous experiments, older adults performed significantly worse than younger adults in the absolute identification task. Age differences in the pitch identification task were of a similar overall magnitude to those in the duration identification task of Experiment 2. Approximately U-shaped serial position curves were again obtained for both younger and older participants. However, unlike in the duration identification task, the error patterns of older adults were similar to those of younger adults: Older adults did not show a tendency to give a lower response to tones. This difference between the pitch and the duration identification tasks is unlikely to be due to differential task difficulty because levels of performance were similar across tasks. To confirm this, we examined whether there was a significant effect of task when the data from the older group in Experiment 2 were compared with the data from the older group in Experiment 3. The main effect of task was not significant, even when age and fluid ability were included in the analysis as covariates.

Thus, the pattern of older adults' errors in the pitch and duration tasks was different, despite similar overall levels of performance on the tasks. This finding suggests that older adults' tendency to give a lower response to tones in the duration identification task of Experiment 2 cannot be attributable to either the experimental design or the use of nine response categories.

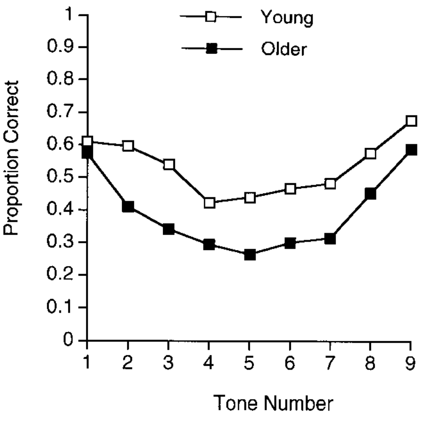


Figure 9. Proportion of correct responses in Experiment 3 as a function of age and tone number.

Modeling

The aim of the modeling work was to examine whether a simple two-parameter model of absolute identification can be applied to duration identification, and to explore whether psychologically meaningful parameters of such a model can be varied to capture the age differences that we have found in these tasks. The main emphasis in the modeling work is on the findings of Experiment 2, because this experiment yielded the most important qualitative age differences, although we also examine whether a similar model can be applied to data from the other two experiments. We note that previous models of perceptual identification have not been applied to duration judgment tasks, and, conversely, models of timing within the SET framework have not been applied to the type of identification task used in the current study.

We hope that one outcome of the current modeling work is that models of timing can be aligned more closely to other models of absolute identification that have been applied to data collected by using similar paradigms and other stimulus dimensions (e.g., Berliner & Durlach, 1973; Lacouture & Marley, 1995). This bears on the important theoretical question of whether temporal duration behaves, psychophysically, similarly to other dimensions such as length or intensity. The model we develop is an exemplar model, similar to, but simpler and less sophisticated than, models already developed to account for absolute identification and categorization performance using nontemporal dimensions. The central assumption is that memories of stimulus values (e.g., duration or pitch) are stored as exemplars, and that presented test tones are compared with exemplars in memory by using simple distance-based similarity metrics. Models of categorization typically assume that the psychological similarity between two items will be some reducing (normally exponential; Shepard, 1957) function of the distance between the representation of the two items in psychological space.

Three questions must be answered for such a model to proceed. First, how are the psychologically scaled representations of temporal durations related to actual durations? Second, what is the relation between the psychological similarity of two durations and the distance separating the representations of those durations on

---

[2] Within the old group, the correlation was not significant between age and overall correct performance, $r(33) = -.13$, $p = .24$, or between hearing and correct performance, $r(33) = .06$, $p = .26$.

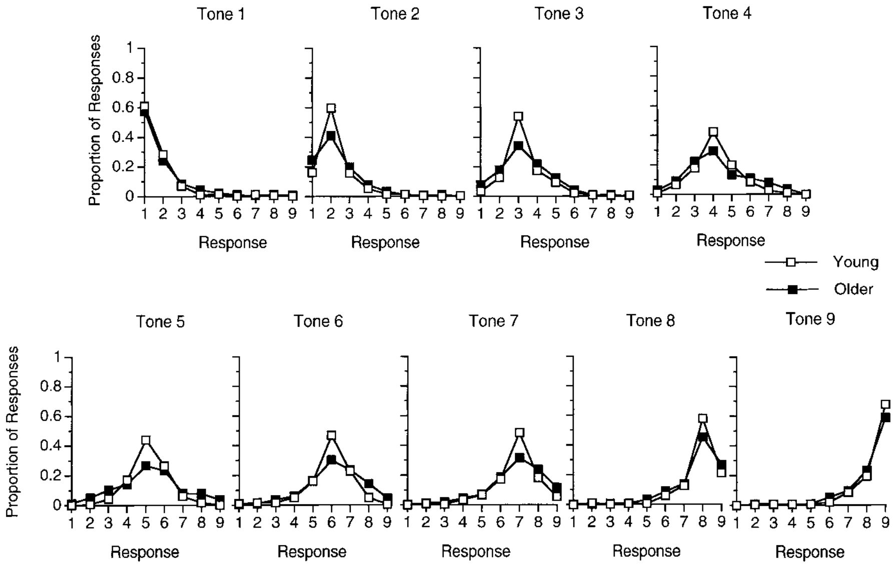


Figure 10. Response gradients for each tone in Experiment 3 as a function of age.

the internal psychological scale? Third, how are response probabilities generated from the relevant psychological similarity values? The answers to these three questions provide the main assumptions of the model. First, we assume that temporal durations, like magnitudes on other dimensions, are represented psychologically in memory as logarithmically transformed versions of the actual durations. Thus, $M_i = \mathrm{Log}(T_i) + D$, where M_i is the memory representation of the ith stimulus, T_i, and D is distortion. D is always assumed to be zero for presented test tones. A positive value of D corresponds to T_i being represented in reference memory as longer than it actually is; a negative value corresponds to memory representations shorter than the true durations. The D parameter is proportional; for example, an increase of 20% of each temporal duration in memory would be equivalent to a parameter value of 0.1823. Because the parameter is proportional, its value remains constant whatever the units in which time is measured.

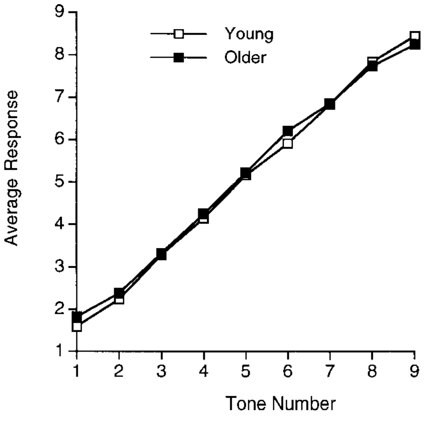


Figure 11. Average response to each tone in Experiment 3 as a function of age.

Second, we assume that the psychological similarity of two durations is a negative exponential function of the distance between their representations in psychological space, such that representations far apart in psychological space are less psychologically similar than are representations closer to one another in psychological space. This assumption is made by many models of absolute identification and categorization (e.g., Nosofsky, 1986; Shepard, 1957). Thus,

$$\eta_{i,j} = e^{-c|M_i - M_j|}, \tag{1}$$

where $\eta_{i,j}$ is the similarity between T_i and T_j, $|M_i - M_j|$ is the distance between the representations of T_i and T_j in psychological space, and c governs the slope of the generalization function. The use of an exponential similarity-distance metric, when coupled with the logarithmic timing assumption, is equivalent to the claim that ratios of absolute durations are relevant in determining stimulus confusability (Brown, Neath, & Chater, 2001). More specifically, the confusability of the two durations will be the ratio of the shorter duration to the longer, the ratio raised to the power c. Many models of timing, including those within the SET framework, assume linear rather than logarithmic temporal representations (Gibbon & Church, 1981); the present model (logarithmic underlying representation coupled with exponential similarity) is formally equivalent to the use of linear underlying representations coupled with ratio-determined similarity.

Third, we use a version of the Luce similarity choice model (Luce, 1963) to produce response probabilities. Specifically,

$$P(\mathrm{R}_j|\mathrm{S}_i) = \frac{\eta_{i,j}}{\sum_{k=1}^{n} \eta_{i,k}}, \tag{2}$$

where $P(R_j|S_i)$ is the probability that response j is given to stimulus i, n is the number of stimuli (here 9), and $\eta_{i,j}$ is the similarity of the psychological representation of stimulus i to the memory representation of stimulus j. A more detailed justification of the model, as well as a description of its application to a range of other paradigms, is given in Brown et al. (2001).

There are two free parameters in the simple version of the model described previously. The distortion parameter, D, governs the amount of distortion of stored exemplars in memory. The parameter c that governs the slope of the generalization function can be interpreted psychologically in terms of the noisiness of timing processes. We note that, unlike many previous accounts of absolute identification performance (e.g., Kornbrot, 1978; Nosofsky, 1985), the model does not include response bias parameters for individual stimuli; response biases are instead incorporated directly into the model. This works in the following fashion: Response probabilities in the model are first calculated on the assumption of no response bias. Each response probability R_j is then divided through by the experimentally derived summed response probability,

$$\sum_{j=1}^{n} \mathrm{R}_j,$$

with the result that the overall response probabilities produced by the model directly reflect those observed in the data. The response probability matrix is then renormalized so that the stimulus presentation probabilities return to unity, and the process is repeated iteratively until both stimulus probabilities and response probabilities accurately reflect the relevant experimental probabilities.[3]

Modeling of Data From Experiment 2

Our goal in modeling the data from Experiment 2 was to capture both the quantitative and the qualitative differences between the groups. Specifically, the aim of the modeling was to simulate the overall group differences in levels of correct performance and to reproduce the distinctive error pattern produced by the older adults (their tendency to give a lower response to a tone than the tone's actual number). The two parameters in the model have different effects on error patterns. Varying the c parameter in the model will affect the psychological confusability of duration representations, leading to changes in overall levels of performance. The result of varying this parameter is to change the steepness of the response gradients (such as those plotted in Figure 5), but it should have no effect on the symmetry of the gradients or on the location of their peaks. By contrast, varying D, the distortion parameter, affects overall levels of performance and the shapes of the response gradients. Given the data from Experiment 2, it would seem likely that D must be varied to capture the group differences in error patterns shown in Figure 5. What is less clear is whether it is necessary to vary both c and D to obtain a good fit to the data.

In statistical comparisons of model fits, our principal concern was to compare the adequacy of models with and without a distortion parameter included. This was done by computing log-likelihood ratios. Because of the nature of the experiment, the modeling of mean data, and the response bias corrections, no single method of estimation is ideal. We, therefore, estimated log-likelihood ratios in three different ways; in practice, all three methods gave rise to identical patterns of significant versus nonsignificant results. First, we estimated likelihood ratios from the residual sums of squares (Borowiak, 1989), ignoring all response probabilities less than .05. Second, as the assumptions of equal variance and independence assumed by this method are not fully met, we also calculated log likelihood directly from estimated frequencies (estimated rather than actual frequencies were used because of the response bias corrections). Third, we calculated log likelihood by using the G^2 statistic (see Read & Cressie, 1988); here, all zero values were replaced with the value .01. Although the resulting chi-square values differed somewhat according to method, in no case was a comparison significant according to one method but not another, or vice versa. We quote R^2 values throughout because they are straightforward to interpret.

We focus initially on the apparent effects of distortion observed for older participants in Experiment 2. Thus, in our first simulation of the data from Experiment 2, both c and D were allowed to vary between groups. The resulting model fits are shown in Figure 12A (younger adults) and 12B (older adults). These are plots of the responses given to each tone and are fits to the data plotted in Figure 5. Good fits were obtained in both cases, with an overall R^2 value of .97. The values of the distortion parameter D that led to the illustrated fits were $-.016$ and .109 for younger and older participants, respectively.

It can be seen that, consistent with the account provided previously, the characteristic shifts in responding (such that the older participants gave lower responses to tones) could be produced by changes in the distortion parameter. The c parameter varied little, with similar values for the younger participants ($c = 4.5$) and the older participants ($c = 4.7$). Indeed, holding the c parameter constant across age group reduced R^2 by less than .01. A further simulation, in which the distortion parameter was not allowed to vary, led to a significantly worse fit, $\chi^2(1) = 36.4$, $p < .05$.[4] Values of c were 3.37 (older participants) and 4.32 (younger participants). A similar result was obtained with all three methods of model comparison. More important, when D was not allowed to vary away from a default value of zero, the characteristic error patterns were not observed in the modeling of the older group's performance. In summary, it is clear in the context of the model that systematic distortion of stored exemplars in memory plays a key role in characterizing the qualitative shift in performance between younger and older participants.

Modeling of Data From Experiments 1 and 3

The findings of the modeling of the data from Experiment 2 raise two important questions. First, can we also account for the small but significant age effect in Experiment 1 by varying only the distortion parameter? Second, how can the age effects in Experiment 3 (pitch identification) be captured?

In Experiment 1, older adults performed significantly worse than younger adults, but, unlike in Experiment 2, the tendency for this group to give a lower response to tones was not particularly

---

[3] This method has the advantage of reducing the number of parameters required, although it needs to be used with caution if response preferences are large or systematic.

[4] Reported η^2 values were based on estimates from residual sums of squares.

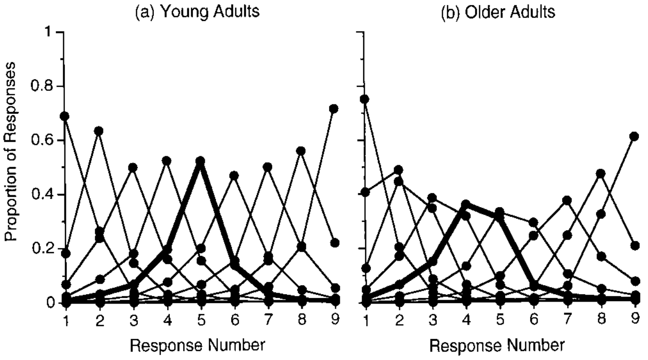


Figure 12. Model fits of the response gradients shown in Figure 5 for (a) younger adults and (b) older adults. Tones 1–4 and 6–9 are indicated by the nonboldfaced lines. The response gradient for Tone 5 is the line shown in boldface.

marked (see Figure 3, although some of the gradients are not perfectly symmetrical). Nevertheless, we explored whether the group differences in overall performance could be captured by varying only the *D* parameter. In these simulations, we focused on the data from the short condition because the age effects were more obvious in this condition. An excellent fit to the data could be obtained by allowing only the *D* parameter to vary between groups ($R^2 = .98$). *D* was estimated at 0.008 for the younger group and 0.064 for the older group; *c* was 6.46 for both groups. Although *D* estimates varied in a qualitatively similar way to the simulation of Experiment 2, consistent with the data, peak shifts are not apparent in the response gradients produced by the model.

Why does varying *D* lead to peak shifts in the simulation of Experiment 2, but not in that of Experiment 1? The explanation lies in the fact that there were fewer durations in the series in Experiment 1. Having fewer durations in the series means that representations of adjacent durations in the series are, in effect, relatively more distant from each other. This means that even if a tone is distorted to some extent in memory, it is less likely that it will be reliably confused with another position in the series, with the result that peak shifts are unlikely (although, even with small numbers of tones, peak shifts will emerge if the degree of distortion is increased sufficiently). However, although varying distortion provides a good fit to the data from Experiment 1, the data can be simulated about equally well by allowing only the *c* parameter to vary between 8.63 (younger group) and 6.13 (older group), with *D* held constant at 0.059 ($R^2 = .98$). (Allowing both *c* and *D* to vary between groups did not lead to a significantly better fit; values of *c* were 8.94 and 6.01, and values of *D* were 0.064 and 0.055 for the younger and older groups, respectively.) Thus, it would appear that for this data set, the choice of the *D* parameter to model group differences can only be made in the light of the simulations of Experiment 2, in which varying that parameter is necessary to capture the data from the older group.

In Experiment 3 (pitch identification), the older adults were performing at a level similar to that of Experiment 2. However, although the response gradients in Experiment 3 were flatter in the older adults, their symmetry and the locations of their peaks were similar in younger and older adults (see Figure 10). Thus, it would seem likely that group differences could be captured by varying *c* rather than *D*. A simulation in which only *c* was allowed to vary between the groups ($D = 0$ for both groups; $c = 14.45$ and 9.43 for the younger and older groups, respectively) provided a good fit to the data ($R^2 = .98$ for younger participants and .97 for older adults). Although adding distortion into the model led to a better fit statistically for both younger, $\chi^2(1) = 4.2$, $p < .05$, and older, $\chi^2(1) = 5.2$, $p < .05$, adults, only an additional 0.5% of the variance was accounted for, and the estimated value of the *D* parameter was small at −.009 and −.01 for younger and older participants, respectively (values of *c* were 15.86 and 10.6). Perhaps more important, there was no statistical benefit in allowing *D* to be estimated separately for younger and older adults, $\chi^2(1) < .5$. As before, identical results were found for each of the three model-comparison techniques. Therefore, in contrast to Experiment 2, the data from Experiment 3 can be successfully modeled without assuming group differences in the *D* parameter.

Psychological Interpretation of Model Parameters

In psychological terms, the *c* parameter is a measure of the psychological confusability of representations of tones in the series and, thus, of their distinctiveness in memory. The extent to which representations of different durations or different pitches are distinctive could reflect noise in either perceptual processes or memory processes. Existing SET models of timing typically assume noise in one or both of these processes, and it is generally accepted that further research is necessary to pinpoint sources of variability in timing (see Wearden, 1999).

In line with existing theorizing in the SET framework, we assume that the *D* parameter used in our modeling of duration identification reflects memory rather than clock processes. The *D* parameter in the current model corresponds to the *k* memory parameter in SET models and reflects the veridicality of long-term memory representations of reference durations. As in SET models, we assume that such representations can be distorted such that they are systematically remembered as too long, as in the current model of the older adults' data, or too short.

It is fair to say that the processes underlying the distortion of durations in long-term memory are as yet poorly understood. Meck (1996) has argued that the distortion parameter reflects an inaccuracy in the way durations become stored in long-term memory. Under normal circumstances, the translation of working memory representations of reference durations into long-term memory representations is assumed to be veridical. However, Meck has suggested that this translation process can be affected by neuropharmacological changes, resulting in distortion effects. It remains to be established whether it is correct to describe distortion as a problem with memory storage. An alternative explanation could be that reference durations are initially stored veridically in long-term memory, but become distorted over time as a result of a forgetting process. Regardless of how it is characterized, considerable evidence from animal studies suggests that memory distortion of some form does occur under certain conditions (for review, see Meck, 1996), and it seems likely that the *D* parameter assumed in the current model reflects the same type of memory distortion as is observed in animal studies.

General Discussion

The main finding of Experiments 1 and 2 was that older adults performed significantly worse than younger adults on duration identification. The age differences found in these experiments were consistent across different ranges of durations. However, the findings of Experiment 3 (pitch identification) indicate that the effects of aging are not limited to duration identification. Although these findings suggest that older adults have global difficulties with the perceptual identification of stimuli, there was an important difference between the pitch and the duration identification tasks in the error patterns of older adults. In the duration identification task of Experiment 2, older adults had a systematic tendency to give a lower response to a tone than the tone's actual number (e.g., they tended to give response "2" to Tone 3). This was not the case for older adults performing the pitch identification task or for younger adults performing either the duration or the pitch identification task. This error pattern was also not apparent in the older group performing the duration identification task of Experiment 1, although this may have been because age differences were less marked in this experiment because of its smaller memory load (there were fewer durations in the series).

Performance on the duration identification task of Experiment 2 was modeled by using a simple two-parameter model of absolute identification. To capture the error pattern of the older adults, it was assumed that there is memory distortion in the long-term memory representations of the reference durations in the series, such that they are remembered as being longer than they really were. Varying distortion in this way was sufficient to model the important age effects in the data; namely, the flatter response gradients and the peak shifts in the gradients of the older group. The small age effects in Experiment 1 could also be captured in this way, although, unlike in Experiment 2, the data from this experiment could be fitted equally well by assuming age changes in the noise associated with timing processes.

Thus, we have shown that a simple model of absolute identification and categorization can account extremely well for data from the duration identification task. The model, while closely aligned to those normally applied to categorization and absolute identification tasks, nevertheless embodies some of the main assumptions of models within the SET framework. Similar to SET, it is assumed that performance depends on the similarity between test stimuli and stored representations of the reference durations in the series. Furthermore, the Weberian properties instantiated in SET are also present in the current model. However, models in the SET framework commonly assume that duration is represented psychologically on a linear scale, whereas the current model assumes logarithmic timing. Although aspects of animal performance on timing tasks do appear to be best captured by assuming linear timing (Church & Gibbon, 1982; Gibbon & Church, 1981), the issue as to whether human timing is linear or logarithmic is one that has yet to be resolved in the literature. As noted earlier, the model we have described is, in any case, formally equivalent to one in which durations are represented linearly and confusability is a power function of duration ratios.

The flatter response gradients that we found for older adults in both pitch and duration identification tasks resemble the flatter response gradients found in studies of aging and spatial memory (Allen, Kaufman, Smith, & Propper, 1998), with older adults having flatter transposition gradients in memory for spatial locations. Similarly, Maylor, Vousden, and Brown (1999) have found that transposition gradients in a serial order memory task become flatter with age. The age differences in steepness of response gradients in these previous studies have been assumed to be due to noisier or less distinctive memory representations in older adults. The idea that the processes involved in perceptual and memory representation are noisier in older adults has a long history (Gregory, 1957; Welford, 1958; see Allen et al. for a recent formal expression of this idea), and age-related declines in human timing have previously been modeled by assuming changes in noise in perceptual or memory processes (McCormack et al., 1999; Wearden et al., 1997).

In the current model, this suggestion would have been captured by varying the similarity function (the c parameter). By contrast, we found in our modeling work that changes in the response gradients in the duration identification tasks could be adequately captured only by allowing variation in levels of memory distortion. It is also possible to account for the effects of aging on timing in rats by assuming only age changes in memory distortion (Lejeune et al., 1998; Meck et al., 1986). However, we note that varying either distortion or the similarity function will result in a flattening of the response gradients, although varying only distortion will lead to peak shifts, and that it would be necessary to assume more marked age changes in the similarity function if the distortion parameter were constrained to vary only between certain values. Furthermore, it is necessary to vary the similarity function rather than the distortion parameter to capture the age differences in performance on the pitch identification task of Experiment 3, in which group differences in the steepness of the response gradients occurred in the absence of significant peak shifts.

Considerable evidence from the animal timing literature suggests that the type of memory distortion effect found in the current study of duration identification is linked to acetylcholine functioning in the frontal cortex (for review, see Meck, 1996). Lesioning of the frontal cortex in rats leads to a shift in response gradients that suggests that reference durations are remembered as too long, and choline antagonists have a similar (though reversible) effect (Meck, 1983, 1994; Meck & Church, 1987; Meck et al., 1986). The choline system is known to be affected by aging (Bartus, Dean, Beer, & Lippa, 1982; Bartus et al., 1987). Thus, the current results are consistent with much existing neuropharmacological research on animal timing, and previous claims regarding the link between cholinergic functioning and age-related memory decline (Bartus et al., 1987).

An alternative way of viewing the results of these experiments is in terms of the predictions of neuropsychological models of categorization rather than models specific to timing. We note that, given the similarity of the model we have described to standard models of categorization, our findings are consistent with the predictions of other theorists that there should be age effects on categorization. Ashby, Alfonso-Reese, Turken, and Waldron (1998) have argued, on the basis of their neuropsychological model of category learning, that there should be an age-related decline in categorization (because of the effects of aging on levels of dopamine; Gabrieli, 1995; van Domburg & ten Donkelaar, 1991), although they have noted the absence of relevant data. Although the current tasks are identification rather than categorization tasks, the age-related decline that we have found is consistent with Ashby et al.'s prediction, given the established relationship between stimulus identification and categorization performance (Nosofsky, 1986).

In conclusion, we have shown that there is an age-related decline in duration identification. Although there were also age effects on pitch identification, older participants showed a pattern of performance unique to duration identification, indicating that representations of reference durations are distorted in memory. This finding is consistent with previous research on aging and animal timing. Our results are a further demonstration that aging affects basic timing processes and, more broadly, stimulus identification.

References

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

Allen, P. A., Kaufman, M., Smith, A. F., & Propper, R. E., (1998). A molar entropy model of age differences in spatial memory. *Psychology and Aging, 13,* 501–518.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 422–481.

Bartus, R. T., Dean, R. L., Beer, B., & Lippa, A. S. (1982). The cholingeric hypothesis of geriatric memory dysfunction, *Science, 217,* 408–417.

Bartus, R. T., Dean, R. L., & Flicker, C. (1987). Cholinergic psychopharmacology: An integration of human and animal research on memory. In H. L. Meltzer (Ed.), *Psychopharmacology: The third generation of progress* (pp. 219–232). New York: Raven Press.

Berliner, J. E., & Durlach, N. I. (1973). Intensity perception: IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America, 53,* 1270–1287.

Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology and Aging, 13,* 584–596.

Borowiak, D. S. (1989). *Model discrimination for nonlinear regression models.* New York: Marcel Dekker.

Bower, G. H. (1971). Adaptation-level coding of stimuli and serial position effects. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 175–201). New York: Academic Press.

Brown, G. D. A., Neath, I., & Chater, N. (2001). *A ratio model of scale-invariant memory and identification.* Manuscript submitted for publication.

Church, R. M. (1984). Properties of the internal clock. In J. Gibbon & L. G. Allan (Eds.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and time perception* (pp. 566–582). New York: New York Academy of Sciences.

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

Church, R. M., & Gibbon, J. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes, 8,* 165–186.

Clarke, S., Ivry, R., Grinband, J., Roberts, S., & Shimizu, N. (1996). Exploring the domain of the cerebellar timing system. In M. A. Pastor & J. Artieda (Eds.), *Time, internal clocks, and movement* (pp. 257–280). Amsterdam: Holland Elsevier Science.

Craik, F. I. M., & Hay, J. F. (1999). Aging and judgments of duration: Effects of task complexity and method of estimation. *Perception & Psychophysics, 61,* 549–560.

Duchek, J. M., Balota, D. A., & Ferraro, F. R. (1994). Component analysis of a rhythmic finger tapping task in individuals with senile dementia of the Alzheimer's type and in individuals with Parkinson's disease. *Neuropsychology, 8,* 218–226.

Eisler, A. D., & Eisler, H. (1994). Subjective time scaling: Influence of age, gender, and Type A and Type B behavior. *Chronobiologia, 21,* 185–200.

Feifel, H. (1957). Judgment of time in younger and older persons. *Journal of Gerontology, 12,* 71–74.

Gabrieli, J. (1995). Contribution of the basal ganglia to skill learning and working learning in humans. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 277–294). Cambridge, MA: Bradford.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84,* 278–325.

Gibbon, J., & Church, R. M. (1981). Time-left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 87–108.

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

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Towards a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology, 7,* 170–184.

Gregory, R. L. (1957). Increase in "neurological noise" as a factor in aging. In *Proceedings of the Fourth Congress of the International Association of Gerontology* (pp. 314–324). Merano, Italy.

Hawkes, G. R. (1961). Absolute identification of duration. *Perceptual and Motor Skills, 13,* 203–209.

Heim, A. W. (1968). *AH4 Test.* Windsor, England: NFER-NELSON.

Kornbrot, D. E. (1978). Theoretical and empirical comparison of Luce's choice model and logistic Thurstone model of categorical judgment. *Perception & Psychophysics, 24,* 193–208.

Krampe, R. T., Mayer, U., & Kliegl, R. (2000, April). *Timing variability of movements in older adults: Slower clock or executive control problem?* Poster presented at the Cognitive Aging Conference 2000, Atlanta, GA.

Lacouture, Y., & Marley, A. A. J. (1995). A mapping model of the bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383–395.

Lejeune, H., Ferrara, A., Soffie, M., Bronchart, M., & Wearden, J. H. (1998). Peak performance in young adult and aged rats: Acquisition and adaptation to a changing temporal criterion. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 51B,* 193–227.

Licht, D., Morganti, J. B., Nehrke, M. F., & Heiman, G. (1985). Mediators of brief time intervals in elderly domiciled males. *International Journal of Aging and Human Development, 21,* 211–225.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). New York: Wiley.

Maylor, E. A., Vousden, J. I., & Brown, G. D. A. (1999). Adult age differences in short-term memory for serial order: Data and a model. *Psychology and Aging, 14,* 572–594.

McCormack, T., Brown, G. D. A., Maylor, E. A., Darby, R. J., & Green, D. (1999). Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology, 35,* 1143–1155.

McGrath, J. J., & O'Hanlon, J. F. (1968). Relationships among age, intelligence, and rate of subjective time. *Perceptual and Motor Skills, 26,* 1083–1088.

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

Meck, W. H. (1994). Cholinergic functioning and the internal clock. *Behavioral Pharmacology, 5,* 27.

Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive Brain Research, 3,* 227–242.

Meck, W. H., & Church, R. M. (1987). Nutrients that modify the speed of the internal clock and memory storage processes. *Behavioral Neuroscience, 101,* 457–464.

Meck, W. H., Church, R. M., & Olton, D. S. (1984). Hippocampus, time and memory. *Behavioral Neuroscience, 102,* 54–60.

Meck, W. H., Church, R. M., & Wenk, G. L. (1986). Arginine vasopressin innoculates against age-related increases in sodium-dependent high affinity choline uptake and discrepancies in the content of temporal memory. *European Journal of Pharmacology, 130,* 327–331.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review, 67,* 16–31.

Nichelli, P., Alway, D., & Grafman, J. (1996). Perceptual timing in cerebellar degeneration. *Neuropsychologia, 34,* 863–872.

Nosofsky, R. M. (1985). Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics, 37,* 89–91.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Pollack, I. (1952). The information of elementary acoustic displays. *Journal of the Acoustical Society of America, 24,* 745–749.

Poppel, E. (1996). Reconstruction of subjective time on the basis of hierarchically organized processing system. In M. A. Pastor & J. Artieda (Eds.), *Time, internal clocks, and movement* (pp. 165–185). Amsterdam: Elsevier Science.

Rammsayer, T. (1992). Effects of benzodiazepine-induced sedation on temporal processing. *Human Psychopharmacology, 7,* 311–318.

Rammsayer, T. (1994). A cognitive-neuroscience approach for elucidation of mechanisms underlying temporal information processing. *International Journal of Neuroscience, 77,* 61–76.

Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 52B,* 273–286.

Rammsayer, T. H., & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics, 50,* 565–574.

Rammsayer, T. H., Lima, S. D., & Vogel, W. H. (1993). Aging and temporal discrimination of brief auditory intervals. *Psychological Research, 55,* 15–19.

Raven, J. C., Raven, J., & Court, J. H. (1988). *The Mill Hill Vocabulary Scale.* London: H. K. Lewis & Co..

Read, T. C. R., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Spinger-Verlag.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

Salthouse, T. A., Wright, R., & Ellis, C. L. (1979). Adult age and the rate of an internal clock. *Journal of Gerontology, 34,* 53–57.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325–345.

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

Surwillo, W. W. (1964). Age and the perception of short intervals of time. *Journal of Gerontology, 19,* 322–324.

van Domburg, P. H. M. F., & ten Donkelaar, H. J. (1991). *The human substantia nigra and ventral tegmental area.* Berlin: Springer-Verlag.

Vanneste, S., & Pouthas, V. (1999). Timing in aging: The role of attention. *Experimental Aging Research, 1999,* 25, 49–67.

Wearden, J. H. (1991a). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation, 22,* 59–83.

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

Wearden, J. H. (1994). Prescriptions for models of biopsychological time. In M. Oaksford & G. D. A. Brown (Eds.), *Neurodynamics and psychology* (pp. 215–236). London: Academic Press.

Wearden, J. H. (1999). Beyond the fields we know: Exploring and developing scalar timing theory. *Behavioural Processes, 45,* 3–21.

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

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

Wearden, J. H., & Lejeune, H. (1993). Across the Great Divide: Animal psychology and time in humans. *Time and Society, 2,* 87–106.

Wearden, J. H., Rogers, P., & Thomas, R. (1997). Temporal bisection in humans with longer stimulus durations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 50B,* 79–94.

Wearden, J. H., Wearden, A. J., & Rabbitt, P. M. A. (1997). Age and IQ effects on stimulus and response timing. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 962–979.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised.* New York: The Psychological Corporation.

Welford, A. T. (1958). *Ageing and human skill.* New York: Oxford University Press.

Received January 31, 2000
Revision received August 14, 2000
Accepted November 20, 2000 ■