Developmental Psychology
1994, Vol. 30, No. 6, 855–863

Copyright 1994 by the American Psychological Association, Inc.
0012-1649/94/$3.00

Developmental Increase in the Duration of Memory for Tone Pitch

Timothy A. Keller and Nelson Cowan

This research examined developmental change in the duration of memory for tone pitch. In Experiment 1, the persistence of memory for pitch was examined with a 2-tone comparison task in children 6–7 and 10–12 years old and in adults. Because pitch perception differences could contaminate the measure of memory, the frequency difference between tones was adjusted for each subject until a criterion level of performance was reached. In a subsequent test phase, the resulting frequency difference was maintained but the time between tones was varied. Performance deteriorated across the intertone interval more quickly in younger than in older subjects. Experiment 2 demonstrated that the developmental difference in pitch memory persistence is unlikely to be based on the development of strategic processing.

Research on the development of memory has focused on relatively complex stimuli (e.g., words, pictures, and stories) that the more advanced subjects can remember by rehearsing, elaborating on, and reorganizing (for a review, see Kail, 1990). This research has suggested that the development of memory in childhood consists primarily of increases in knowledge and the ability to use mnemonic strategies. However, the predominant emphasis on knowledge and strategies has caused researchers generally to skip over the more fundamental question of whether there is developmental change in the retention of even the simplest stimuli. This question is important because most information-processing models of memory assume that strategic control processes operate on temporarily active representations of the sensory features of incoming information. An adequate theoretical understanding of memory development, including strategic components, thus depends on knowledge about the stability or change with development in sensory representations. It has even been proposed, albeit without direct evidence, that the persistence of information in memory is one of several factors producing individual differences in the level of cognitive functioning (Jensen, 1993; Lehrl & Fischer, 1988).

Research with adults suggests that detailed sensory information about a stimulus (in audition as in other modalities) persists in memory for about 10 to 30 s (for reviews, see Cowan, 1984, 1988, 1993). Little is known about whether the duration of persistence changes with development. However, such a developmental change might be expected on the basis of neurophysiological considerations. For example, the temporal lobe areas of the cerebral cortex involved in audition show maturational changes in thickness, neuronal density, and histological structure at least through the age of 10 years (Rabinowicz, 1980). Evidence from recent electrophysiological studies in humans suggests that these maturing cortical areas are the primary neural bases of auditory sensory memory, including memory for pitch (e.g., Hari et al., 1984; Lü, Williamson, & Kaufman, 1992; Näätänen, 1992).

Although there is little behavioral research that would indicate directly whether the memory for the pitch of individual tones changes during childhood, some studies have suggested no change in properties of acoustic memory (e.g., Engle, Fidler, & Reynolds, 1981; Frank & Rabinovitch, 1974). However, they have not looked specifically for changes in the memory retention period. Among the few auditory studies that have examined the retention interval, the results are encouraging although inconclusive. Cowan and Kielbasa (1986) found that 4-year-old children lost memory for pairs of spoken nonsense syllables across a delay of 15 s filled with a silent, nonverbal distractor activity, whereas no memory loss was observed in adults in a comparable situation. Sipe and Engle (1986) found that memory for unattended syllables in a selective listening task was lost faster in poor readers than in good readers.

We know of no relevant studies of developmental changes in the *persistence of memory* for tone pitch. It is known that memory for complex tonal sequences improves with age (e.g., Trehub, Morrongiello, & Thorpe, 1985), but this improvement apparently involves an increasing understanding of musical structure. To examine the temporal properties of auditory memory per se, one must minimize structural considerations. In adults, for nearly a century, the persistence of auditory memory has been investigated with two-stimulus comparison tasks requiring judgments about the pitch of the second, comparison tone in relation to the first, standard tone, with a variable intertone interval (ITI; e.g., Angell & Harwood, 1899; Harris, 1952; Wickelgren, 1969). The use of a roving standard tone in such tasks prevents the build-up of a long-term memory for the standard stimulus (Bull & Cuddy, 1972; Harris, 1952). When the frequency difference between the standard and comparison tones is relatively small, the comparison judgment depends on a detailed memory of the pitch of the standard tone, because no differential categorization of the tones can be carried out.

Timothy A. Keller and Nelson Cowan, Department of Psychology, University of Missouri.

This work was supported by National Institutes of Health Grant R01 HD-21338. The article is a version of Timothy A. Keller's master's thesis. We thank Don Kausler, Martha Parnell, Scott Saults, Ken Sher, and Jim Thelen for helpful comments, and Caroline Elder and Prima Wagen for assistance in recruiting and running subjects.

Correspondence concerning this article should be addressed to Nelson Cowan, Department of Psychology, University of Missouri, 210 McAlester Hall, Columbia, Missouri 65211. Electronic mail may be sent to psycowan@MIZZOU1.missouri.edu.

Cowan (1984) showed that there is a great deal of convergence in the outcomes of such studies in adults, with a rapid loss of memory across the first few seconds of the ITI and a more gradual memory loss continuing for at least 10 s. We examined this sort of task developmentally in the present study.

There is a long history of research in which children have been asked to compare the pitches of pairs of tones. These tasks have shown a marked increase in ability with age. Simple tasks in which the comparison tone always follows the standard by less than 1 s show significant yearly increases in accuracy at least until adolescence (e.g., Bently, 1966; Kidd & Kidd, 1966). The frequency difference limen (the smallest difference in frequency required for a predefined threshold level of performance) decreases between early childhood and adulthood (Maxon & Hochberg, 1982). However, within these two-tone comparison tasks, sensory or perceptual development could lead to an increase in the strength or resolution of the representation of tone pitch, therefore allowing a more accurate comparison of tones. This is a factor that must be controlled before the persistence of memory for tone pitch can be measured in a developmental study.

Experiment 1

The first experiment examined the persistence of memory for tone pitch developmentally. Because children younger than the age of 6 years have difficulty in consistently applying categorical response labels to changes along the pitch dimension (Andrews & Madeira, 1977; Jeffery, 1958; Webster & Schlentrich, 1982), an age of 6–7 years (still well before cortical maturation is complete) was selected as the youngest subject group. The other two groups represent late preadolescents (10–12 years) and adults. A *same* or *not-the-same* response was used for the tone comparison task to minimize conceptual difficulties that young children may have in applying terms such as *higher, lower,* or *different* to changes in tone pitch. The task was made attractive to children through the incorporation of animation and synthesized speech into the feedback provided.

The biggest problem in using a two-tone pitch comparison task in a developmental study is that forgetting rates cannot be meaningfully compared when the initial levels of performance (i.e., performance at very short ITIs) differ among age groups. This can occur because of perceptual or conceptual differences between subjects of different ages. To examine forgetting rates free of these factors, we attempted to equate task difficulty across age groups by first determining the smallest frequency difference between the standard and comparison tones (the Δf estimate) for which each subject could attain a criterion level of performance. The Δf estimate is obtained by using a relatively short (2-s) ITI and a relatively high performance criterion (.841, or approximately 84% correct). This Δf is then used in tests in which the dependent measure is Δt, the ITI that produces a specific, predetermined criterion level of performance. This type of test was conducted both for the original, 84% correct performance level and for a second, lower performance level (.707, or approximately 71% correct).

Age differences in Δt that occurred equally at the 84% and 71% correct criterion levels would reflect the development of *perceptual* or conceptual factors in pitch comparison. However, if subjects of all ages are equated for their performance level at a 2-s ITI, as they should be (with performance at 84%) when the criterion Δf is used for each subject, then differences in the increase in Δt needed to lower performance to a 71% level can only be attributed to a developmental change in the persistence of memory for pitch.

An adaptive tracking procedure based on the work of Levitt (1971) was used to determine a criterion Δf and then to test for the critical Δt at two performance levels. In such a procedure, one looks for stimuli that produce a criterion level of performance. If the subject makes the correct perceptual judgment, the stimuli are made more difficult, and if the subject answers incorrectly, the stimuli are made easier. The adaptive tracking is considered finished when there have been seven *reversals* in tracking, or points at which the stimulus adjustments had to switch from decreases in the stimulus difference to increases, or vice versa.

The criterion 84% and 71% performance levels that we used emerge naturally from the adaptive tracking procedure (Levitt, 1971). If one makes the stimulus conditions harder every time the subject gets four trials in a row correct and makes the stimulus conditions easier every time the subject gets one trial wrong, then the stimuli theoretically should approach a level at which the subject can answer correctly 84% of the time. If, on the other hand, one makes the stimulus conditions harder every time the subject gets two trials in a row correct and makes the conditions easier every time the subject gets one trial wrong, then the stimuli theoretically should approach a level at which the subject can answer correctly 71% of the time. We refer to the test procedures used to estimate the Δt levels required for these two performance criteria as separate adaptive *tracks* (Levitt, 1971). Within the Δt test session, trials within the 84% and 71% tracks were randomly interleaved.

Method

Subjects

The data reported here are from 24 subjects per age group. The mean ages of the subjects whose data were included were 20 years 0 months (*SD* = 24.65 months), 10 years 7 months (*SD* = 8.01 months), and 6 years 10 months (*SD* = 7.24 months) for the three groups. Data from additional subjects were discarded because the subjects dropped out before completing the experiment (five 6- to 7-year-olds and one 10- to 12-year-old); failed to reach a criterion level of performance for 80 Hz, the largest Δf value permitted (two 6- to 7-year-olds); produced a Δt adaptive tracking protocol that leveled off at 1 s, the shortest duration permitted (which occurred in the 84% track only, for three 6- to 7-year-olds, two 10- to 12-year-olds, and three adults); or produced a Δt adaptive tracking protocol that leveled off at 30 s, the longest duration permitted (which occurred in the 71% track only, for two 6- to 7-year-olds and five adults).

Adult subjects were recruited through sign-up sheets placed at various locations at the University of Missouri—Columbia campus and were paid $5 per hr for their participation. Potential subjects in the two younger age groups were randomly selected from records provided by the local school system. Children were then recruited through letters and telephone contact with their parents. For their participation, the 6- to 7-year-olds received a book at the end of each session, as well as a bookmark on which stickers earned during the course of the experiment were placed. Ten- to 12-year-olds were given their choice of a book or $5 at the end of the session. None of the subjects had a known hearing impairment or a history of recurrent ear infections, and all were free of colds at the time of testing.

Apparatus and Stimuli

Subjects were run one at a time in a sound-attenuated chamber. The experiment was controlled by an Apple Macintosh II microcomputer located outside the chamber. The auditory stimuli were created with a sound-editing program and were digitally stored on the hard disk of the microcomputer. They consisted of 200-ms-long tones (sinusoidal waves) with 20-ms rise and fall times. The tones were output through the computer's digital-to-analog converter under control of the program that ran the experiment and were presented through TDH-39 earphones at 80 dB(A), measured with a GenRad Model 1565-B sound-level meter equipped with a 9-A-type earphone coupler.

Seven sets of tones were created, with each set corresponding to one of seven Δf values (80, 40, 20, 10, 5, 2.5, or 1.25 Hz). Each of these tone sets contained nine separate tones ordered from lowest to highest in frequency, so that the difference between adjacent tones in a set corresponded to the Δf value for that set, and the center tone in each set was always 405 Hz. In all of the procedures described later, all but the highest and lowest frequency tones in a set could be used as the first or standard tone, allowing the second or comparison tone on not-the-same trials to be either higher or lower than the standard with equal probability.

Visual displays were presented on a 19-in. (48.3 cm) AppleColor high-resolution RGB monitor as black images on a white background. Subjects initiated each trial and entered responses by manipulating a computer mouse to select the appropriate "button" displayed on the screen. These buttons each were approximately 3.5 × 4 cm, and three of them were displayed across the bottom of the computer screen at equal intervals. The middle button was labeled with the word *go* and was used to initiate each trial. The left button was labeled with the word *same* and contained, as an illustration of the concept, images of two identical circles above the label. The right button was labeled *not the same* and contained images of a circle and a triangle above the label.

Instructions and auditory feedback were presented over the earphones by using commercially available speech synthesis software (MacinTalk, Apple Computer, Inc., Cupertino, CA) under control of the program that ran the experiment. This synthesized speech was synchronized with an animated visual display of a talking dog by means of commercially available animation software (HyperAnimator, Bright Star Technologies, Inc., Bellevue, WA), also under control of the program that ran the experiment. In addition, for the Δt test, a horizontally oriented bar approximately 1-cm high and variable in width was centered above the dog's head. The width of this bar corresponded to the length of the delay interval for a trial and varied from 2 cm for a 1-s delay to 19 cm for a 30-s delay. Initially the bar was empty, but it was gradually filled in black from left to right to indicate for the subject the proportion of the ITI that had elapsed. The purpose of this bar was to provide information that would minimize the extent to which subjects' attention might wander from the task during long ITIs.

Procedure

Each subject completed a single session approximately 1 hr in length. The session began with a series of training phases designed to ensure that all of the subjects comprehended the basic tone comparison task. During the three training phases, all auditory stimuli were drawn from the 40-Hz Δf tone set. Phases 1 and 2 incorporated simultaneous auditory and visual examples of same and different stimulus pairs, to facilitate application of the same versus not-the-same concept to the domain of pitch. Adult subjects began the experiment with Phase 3 of the training procedure. Training was followed by a Δf determination procedure and then, finally, the task of primary interest, a Δt testing procedure.

Training. The first phase of training was intended to familiarize the subjects with the use of the computer mouse, the response buttons displayed on the monitor, and the concepts of same versus not the same that they represented. At the beginning of this phase, the animated dog introduced itself and asked for the subject's name, which the experimenter entered phonetically through a keyboard, permitting the dog to refer to the subject by name throughout the experiment. The dog told the subject that they were going to play a game in which it (the dog) would say when to push one of the buttons and then would say if the right one was pushed. The experimenter repeated what the dog had said to ensure that the subject could understand the synthesized speech, and then demonstrated how to move the mouse to select a button on the screen. Even the youngest age group had little difficulty understanding the instructions or manipulating the mouse.

On each trial, the dog said either "Please push the button with two shapes that *are* the same" or "Please push the button with two shapes that are *not* the same." The experimenter repeated the instruction when necessary. When the instruction was to push the same button, the appropriate response resulted in the dog lifting its ear and the presentation of a tone drawn randomly from the 40-Hz Δf set, followed by a 2-s ITI and then presentation of the same tone again. In addition, a visually displayed black circle was presented in synchrony with the onset of each tone in an area of the screen next to the dog's ear. When the instruction was to push the not-the-same button, the appropriate response had a similar result, but in this case the second tone was either 40 Hz higher or 40 Hz lower than the first, and the visual image synchronized with the second tone was a triangle instead of a circle. An inappropriate response to either instruction resulted in the dog frowning, saying, "Oops, you pushed the wrong button," and then repeating the instruction.

The trials for Phase 1 of training were divided into blocks of four, consisting of two same and two not-the-same instructions, with the order randomly determined within each block. For one of the not-the-same responses, the second tone was higher than the first, and for the other, the second tone was lower. The criterion for moving on to Phase 2 was a correct response to each of the four trials in a block. When an incorrect response was made to any trial within a block, a new randomly ordered block of trials was begun. Throughout this and all subsequent phases of the experiment, a pair of hands was displayed at the top of the screen. After each correct response, fingers of the hands were shown to indicate the cumulative number of correct responses made by the subject. When the hands displayed 10 fingers, the dog whistled, praised the subject, and told the subject to put a sticker on his or her bookmark. The number of fingers displayed was then reset to 0.

The aim of Phase 2 of training was to ensure that subjects could use the concepts of same and not the same and to begin to associate them to tone pairs. The stimuli were as in Phase 1, but the task was to make a same or not-the-same judgment. Although the instructions emphasized attention to the auditory stimuli, correct performance could be achieved on the basis of the visual stimuli alone. When the child initiated the trial, a tone pair accompanied by shapes was presented. The subject used the buttons to label each pair of dual-modality stimuli as the same or not the same, and the dog then told the subject whether or not the response was correct. The criterion for moving on again was correct performance for all trials in a block of four.

The purpose of the third and last phase of training was to ensure that subjects could apply the concepts of same and not the same specifically to the auditory stimuli. The trials were identical to those of Phase 2 with the exception that the visual stimuli (circles and triangles) were not displayed. The criterion for the completion of this phase was two consecutive correct blocks of four trials.

The Δf pretesting. Determination of the individual Δf level to be used in subsequent testing was accomplished by means of an adaptive tracking procedure that adjusted the Δf for not-the-same trials (in which the two tones differed) on the basis of performance on previous not-the-same trials. The ITI between standard and comparison tones was held constant at 2 s throughout this procedure. Same trials (in which the tones were identical) were included to keep subjects honest and to prevent them from developing a strong expectation that the tones will differ, but these trials were not used by the tracking algorithm. Feedback was provided after every trial.

If the subject got four consecutive not-the-same trials correct at a given Δf level, the Δf value on the next such trial was decreased to half of the previous Δf. If the subject answered incorrectly on a not-the-same trial, the Δf value on the next such trial was increased to twice the previous Δf. This four-down, one-up rule allows the adaptive staircase procedure to track the Δf level required by the subject for a performance level of 84% correct (see Levitt, 1971). For same trials, the tone presented as both standard and comparison always was drawn from the Δf set used for the most recent not-the-same trial. To ensure that a sufficient number of not-the-same trials were presented within a reasonable length of time, we set their probability at .67 and the probability of same trials at .33.

Testing began with a Δf value of 20 Hz. The computer program kept track of the number of reversal points (the Δf values at which a transition between increases and decreases in Δf occurred). Testing continued until seven reversals had occurred. The first reversal point was discarded, and the geometric average of the remaining six reversal points was taken as the estimate of the subject's 84% correct Δf threshold.

Because the maximum Δf value used was 80 Hz, the procedure could not accurately track the Δf threshold for subjects who missed a not-the-same trial at this level. When this situation occurred, the Δf value remained at 80 Hz. If the subject missed another not-the-same trial within the next block of four, the subject's data were excluded from all analyses. The decision to exclude these few subjects (as mentioned in the *Subjects* section earlier) was based on the assumption that nondiscrimination with a Δf level this large was likely to indicate inattentiveness or conceptual difficulty with the task. Provision was also made for excluding subjects whose track decreased below the minimum Δf level of 1.25 Hz, but this situation never actually occurred during the experiment.

From the subject's point of view, the task was the same as in Phase 3 of training, except that subjects were told, "Now when the tones are not the same, they will sometimes be a little harder to tell apart." As in the training phases, feedback was provided after each response, and stickers were awarded to subjects in the two younger age groups after every 10 correct responses.

The Δt testing. Each subject's estimated Δf value was rounded to the nearest of the seven possible Δf levels, and this level was used throughout the Δt testing procedure. As explained earlier, two separate adaptive tracks were used during this phase of the experiment (presented in an interleaved manner) to provide estimates of Δt at two separate points on the underlying psychometric function (84% and 71%). Every time the tracking algorithm required an increase in the ITI, it was multiplied by 1.5, and every time the tracking algorithm required a decrease in the ITI, it was divided by 1.5. The starting points for the two adaptive tracks were 2 s (for the 84% track) and 5 s (for the 71% track).

If the Δf levels that were obtained in the last phase remain valid, there should be no age difference in Δt for the 84% track; each age group should yield a Δt near 2 s. On the other hand, if memory for pitch is lost more quickly in younger subjects, then they should yield a Δt in the 71% correct track that is shorter than the Δt of the older subjects in that track.

The two adaptive tracks were pseudorandomly interleaved during the course of the session, with the constraint that the probability of a Track 1 trial was set at .667 and the probability of a Track 2 trial was set at .333. Because Track 1 requires more correct responses before Δt is increased, it requires that more trials be presented before the same number of reversals have occurred as on Track 2, and the higher probability of Track 1 trials was a convenient way to make the most efficient use of the subject's time. Within each track, the probability of same stimuli on a given trial was set at .333, but these trials had no effect on the Δt determination procedure. The first of seven reversals on each track was discarded and the Δt estimate for each track was calculated as the geometric mean of the next six reversals.

Because an adaptive tracking procedure provides a valid estimate of a threshold only if performance is a monotonic function of the variable being manipulated, it was necessary to set a lower limit on Δt to prevent backward masking effects (e.g., Massaro, 1972) from degrading performance at short ITIs. A Δt value of 1 s was chosen as this lower limit. If the subject missed a not-the-same trial on a particular track at a Δt of 1 s, the Δt value remained at 1 s until a not-the-same trial was answered correctly, and then 1 s was recorded as the reversal point. However, data from subjects for whom the ITI failed to increase above 1 s within the next three blocks of trials (12 and 6 not-the-same trials for the two tracks, respectively) were discarded. Similarly, an upper limit on Δt of 30 s was also set. (See the *Subjects* section for a breakdown of subjects' data discarded for various reasons.)

Results and Discussion

The Δf Test

The mean estimates of Δf at the 84% correct threshold with a 2-s ITI were 36.04 Hz (SD = 15.33), 25.55 Hz (SD = 16.19), and 16.99 Hz (SD = 11.03) for the 6- to 7-year-olds, 10- to 12-year-olds, and adults, respectively. A one-way analysis of variance (ANOVA) on these data verified the significant main effect of age, $F(2, 69) = 10.519, p < .0001$. Follow-up pairwise comparisons (Tukey honestly significant difference [HSD]) indicated that the mean estimate for the 6- to 7-year-olds differed reliably from that of the 10- to 12-year-olds ($p < .05$) and from that of the adults ($p < .01$). The two older groups did not differ significantly from one another. This effect is consistent with previous developmental studies of pitch discrimination (e.g., Maxon & Hochberg, 1982).

It is clear that these Δf values are higher than the minimal sensory thresholds for comparable tones, in fact by almost an order of magnitude (e.g., Green, 1976, pp. 260–263). That can be attributed to a number of factors. For example, we did not use highly practiced subjects screened for their ability, and we used a roving-standard procedure, which does not permit the subject to build up a stable long-term memory representation of the standard tone. The Δf values were in a range comparable with what has been used in other memory studies with similar conditions (e.g., Massaro, 1970). Moreover, the absolute magnitude of Δf is of no concern for the present aim, which is to examine memory loss over time rather than sensory ability. It is important only that the Δt at one performance level be equated across age groups (attained by individually adjusting Δf), so that the difference among groups in the longer Δt, producing a lower, 71% performance level, can be attributed to the rate of forgetting rather than sensory ability.

The Δt Test

The mean Δt estimates based on the geometric average of each subject's six reversals on Track 1 (the 84% correct track) were similar for the three age groups: 2.42 s (SD = 1.17), 3.05 s (SD = 1.81), and 2.73 s (SD = 1.60) for the 6- to 7-year-olds,

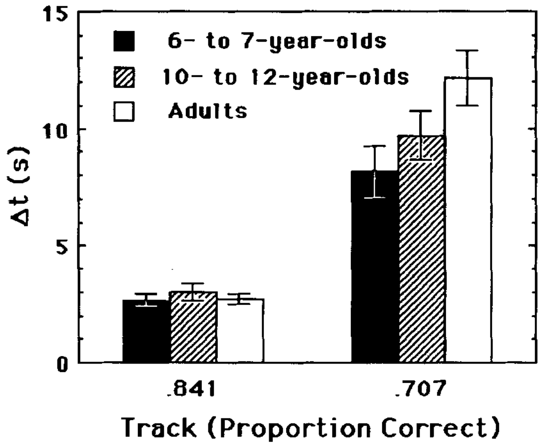

Figure 1. Mean Δt test performance by age for Track 1 (.841 proportion correct) and Track 2 (.707 proportion correct) for Experiment 1. Error bars represent the standard error of the mean. Δt = intertone interval that produces a specific, predetermined criterion level of performance.

10- to 12-year-olds, and adults, respectively. This suggests that the adjustments of Δf worked as planned. The fact that no subject reached the 30-s limit for Δt in this track (as indicated earlier in the *Subjects* section) provides assurance that the similarity among age groups cannot be attributed to performance being near a ceiling level of performance; no such ceiling was reached.

For Track 2 (the 71% correct track), the mean Δt estimates differed by age. They were 8.26 s ($SD = 5.39$), 9.69 s ($SD = 5.41$), and 12.12 s ($SD = 5.91$) for the 6- to 7-year-olds, 10- to 12-year-olds, and adults, respectively. These group means are shown in Figure 1. The figure demonstrates that the permissible difference in ITI between the two levels of performance was largest for the adults, somewhat smaller for the 10- to 12-year-olds, and smallest for the 6- to 7-year-olds.

To confirm this description of the data, we submitted the data to a 3 (age) × 2 (track) mixed ANOVA. The results of this analysis of course revealed a large main effect of track, $F(1, 69) = 130.244$; the mean for Track 1 was 2.75 s, and the mean for Track 2 was 9.99 s. Of more theoretical importance, though, there also was a significant Age × Track interaction, $F(2, 69) = 3.414$, $p < .05$. Separate tests of the simple main effect of age for each track indicated that there was a reliable difference among the age group means for the 71% correct track, $F(2, 69) = 3.200$, $p < .05$, but not for the 84% correct track, $F(2, 69) = 0.697$, $p > .5$. Follow-up Tukey HSD pairwise comparisons revealed that the simple main effect of age on the 71% correct Δt estimates resulted from a significant difference between the 6- to 7-year-olds and the adults ($p < .05$). (The intermediate, 10- to 12-year-olds' mean Δt estimate did not differ reliably from either of the other groups.) This pattern of findings suggests that there is a gradual developmental increase in the persistence of memory for pitch.

One potential objection to this conclusion is that, instead, there could have been a developmental change in the response bias. If younger subjects had a stronger bias toward responding same, then they would stop responding to differences in tones at shorter ITIs than would older subjects. However, the pattern of responding on same or catch trials indicates that a response bias interpretation would be incorrect. Older subjects not only yielded higher estimates of Δt on Track 2, but they also correctly responded to same trials in that track slightly more often than the younger subjects. Whereas the 6- to 7-year-olds were correct on .70 of those trials ($SD = .22$), and 10- to 12-year-olds were correct on a similar .69 of them ($SD = .22$), adults were correct on .75 of them ($SD = .15$). Including only those catch trials that occurred at the subject's final Δt, there was virtually no difference between the youngest group (.81) and the oldest group (.82).

In summary, the present experiment has disentangled the effects of perceptual sensitivity and forgetting on developmental change of performance in a two-tone pitch discrimination task. The finding that the frequency difference (Δf) required by subjects to achieve a predefined criterion level of tone comparison performance decreases with age is consistent with previous developmental investigations of frequency discrimination ability (Maxon & Hochberg, 1982). In the test phase of this experiment, results of the prior Δf determination pretest were used to construct a situation at which subjects of all three age groups reached a standard performance level (84%) with a standard ITI (2 s). It was shown further that the ITI that permitted a particular lower level of performance (71%) nevertheless differed considerably across age groups (about 8, 10, and 12 s in the three groups). This difference cannot be attributed to the sensory differences between groups given the matched performance levels at a 2-s ITI, whereas it can be attributed to differential forgetting of pitch during the ITI among the three groups.

Experiment 2

Although the first experiment demonstrates that pitch memory persists longer in older subjects, it does not specify the nature of the change in the memory system. One can distinguish between *structural* changes, or changes in the automatic persistence of memory, and *strategic* changes, or changes in the voluntarily applied mnemonic strategies that might serve to prolong the memory for pitch. Given that many more of the previous studies of memory development have focused on strategic changes (see Kail, 1990), the more interesting finding would be structural change. However, a strategic contribution is theoretically possible. For example, older subjects conceivably may find some way to rehearse the first tone in a pair during the ITI better than the younger subjects can, although previous research offers little evidence that such fine details of pitch can in fact be rehearsed (e.g., Deutsch, 1970; Massaro, 1970, Wickelgren, 1969).

The contribution of strategic processes to pitch discrimination in the two-tone comparison task used in Experiment 1 can be estimated with further research in adult subjects. In one of the two conditions of the present experiment, adults were permitted to rehearse the first tone in a pair during the ITI, as in Experiment 1. However, in the other condition, they were required to engage in a demanding musical imagery task during the ITI, which should drastically impair any ability that they had to rehearse tone pitch. The task to be performed during the

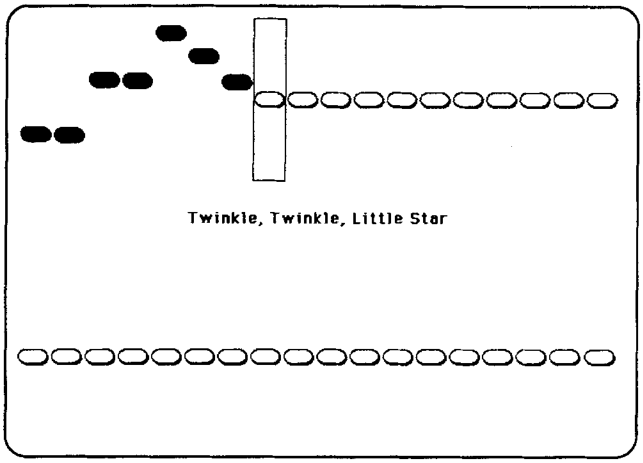

Figure 2. Illustration of the response screen used for the musical imagery task in Experiment 2. In this example of an ongoing protocol, the subject already has recorded seven notes, the sixth one incorrectly.

ITI always was silent, which is necessary to avoid interference with a modality-specific, auditory sensory memory (Cowan, 1984).

If the developmental difference in memory occurs because of tone pitch rehearsal in older subjects, then adults' performance in the musical imagery task should be lower than in the rehearsal condition, and more similar to performance of the youngest group of subjects in Experiment 1. On the other hand, if there is little contribution of tone rehearsal to the developmental change that we have observed, then there should be little or no effect of the nature of the silent task carried out during the ITI.

The musical imagery task involved mentally imagining the melody of a familiar song and indicating its contour. Weber and Brown (1986) previously have shown that subjects produce similar tonal contours for songs that are actually heard and songs that are imagined on the basis of long-term memory representations.

Method

Subjects

Fifteen adult subjects were recruited and were paid $5 per hr as in Experiment 1. Of these, 3 were eliminated because their performance on an adaptive track designed to hold performance at 71% correct exceeded the maximum delay of 30 s.

Apparatus and Stimuli

The Δf and Δt testing procedures included the apparatus and stimuli used in Experiment 1. However, for the musical imagery condition within the Δt test, a second computer (an Apple Macintosh SE equipped with a MicroTouch touch screen) was used for the imagery task. It was positioned to the right of the monitor for the tone comparison task, at a comfortable viewing distance and within easy reach of the subject.

Figure 2 shows the appearance of the response screen for the musical imagery task during a typical trial. On each trial, 1 of 24 possible familiar song titles (e.g., Jingle Bells, Yankee Doodle) was continuously displayed in the center of the computer screen. The task was to represent the pitch contour of the song by vertically moving, one by one, oval buttons (5 × 8 mm) within a horizontally arranged series. The active response area began at the top leftmost button, moved rightward as the subject responded, and then continued at the bottom left and again moved rightward. As shown in the figure, this active area always was marked with a vertical bar (45 × 8 mm), and the response buttons turned from white to black as they were used, so as to mark the subject's place.

The response was not recorded until the subject removed his or her finger from the screen, which made it possible for the subject to visually guide the sliding indicator to the desired spatial location within the response area. Responses were recorded as the number of vertical pixels between the bottom of the response area and the bottom of the sliding indicator and were converted to a value on a scale from 0.00, representing the lowest possible position, to 10.00, representing the highest. Subjects could change a previous response by using an arrow key on the keyboard attached to the computer to move the active response area back one position at a time.

Procedure

The training and Δf pretesting procedures were identical to those used in Experiment 1, except that the animated talking dog was eliminated to make the most efficient use of the subjects' time. Feedback was instead provided by visual display of the words *correct* or *wrong* after each trial.

Following the Δf pretest, subjects were shown the 24 song titles to be used for the musical imagery task. It was confirmed that all subjects knew all of the songs, although in a few cases the experimenter had to hum the melody to refresh the subject's memory. The experimenter then demonstrated the musical imagery task with a randomly chosen title from the list, by humming the song aloud while simultaneously indicating the relative pitch of each note on the touch screen. When the subject understood the task, the touch screen demonstration was repeated with the experimenter silent, and subjects were told that during the experiment it was important that they rely on their auditory images but maintain silence, rather than overtly humming the notes. The subject was then given 12 practice trials with different titles drawn randomly without replacement from the list, and the subject was allowed as much time as needed to represent the first 20 notes of the song. Subjects were told that if they came to a point in the song at which they did not know any more, they could start back at the beginning of the song. After each of these practice trials, the experimenter discussed with the subject any discrepancies between the response contour the subject produced and the true pattern of pitch changes commonly used in singing the song.

After the 12 practice trials with the musical imagery task alone, subjects were given 12 trials of practice in interpolating the musical imagery task between the two tones of the tone comparison task. During these practice trials, the delay interval remained constant at 10 s, the tones were drawn from the subject's individually determined Δf set, and the song titles were randomly drawn without replacement from the remaining 12 titles not used in the previous practice phase.

On each trial, the subject was to examine the song title displayed in the center of the right monitor so that the tune would be readily accessible. Then the subject was to click on the *begin* button displayed on the left monitor to initiate presentation of the standard tone. After termination of the standard tone, the words "perform musical imagery task" appeared in the center of the left monitor. Subjects were to begin the imagery task at once and continue to perform it until the comparison tone was presented, whereupon they were immediately to stop performing the musical imagery task and to make a same or different judgment about the pair of tones. Feedback about the tone comparison judgment was then displayed, and the subject was prompted to press the return key to display the next song title on the right-hand monitor. No feedback was given regarding performance on the musical imagery task, but subjects were told that the accuracy of their responses would be scored.

After the musical imagery practice, the procedure for the Δt test phase was as in Experiment 1, except that no 84% correct track was used and there were two 71% tracks randomly interleaved with equal probability, one for the rehearsal condition and one for the musical imagery condition. On half of the trials, the message displayed on termination of the standard tone read "Rehearse Tone 1." Subjects were told that when this message appeared, they were to concentrate their attention on remembering the pitch of the first tone during the delay interval, using whatever strategy worked best for them. On the other half of the trials, the "perform musical imagery task" message appeared instead. The cue revealing the intertone task was placed after the presentation of the standard tone to ensure that subjects would give equal attention to the standard tone across conditions. As in Experiment 1, Δt testing continued until at least seven reversals in Δt level had occurred on each track and Δt was calculated as the geometric mean of the second through seventh reversals.

Scoring of Musical Imagery Performance

Two objective measures of performance on the musical imagery task were used, one related to the speed of responding and the other to the accuracy of the contour produced. The speed measure was simply the number of responses made on a trial divided by the delay interval for the trial. To score the accuracy of subjects' musical imagery responses, we compared the relative pattern of changes in the response contours with the true pattern of relative pitch changes commonly accepted for the songs. The method of coding the true pattern for each song and the response pattern produced by the subject was taken from Parsons (1975), with each successive note represented as higher, lower, or the same as the preceding note. Thus, the position of the first note of each contour was not scored (i.e., subjects can start an imagined melody on any pitch). In addition, information about the musical interval between successive notes was not considered in this method of scoring. Only the pattern of relative changes or repetitions of notes was scored. For example, Figure 2 shows the first seven responses of a subject to the song title "Twinkle, Twinkle Little Star." The true pattern of notes for this song would be, "unscored first note, repetition, higher, repetition, higher, repetition, lower." The contour depicted in Figure 2 would be coded as "unscored first note, repetition, higher, repetition, higher, lower, lower."

Because subjects could not be expected to place the bar at exactly the same vertical position as the preceding response when representing a repetition in pitch, a difference in vertical position between successive responses of ±2.5 mm was coded as a repetition. A subject's score on a trial was the proportion of responses beyond the first that matched the true pattern of notes. Thus, for the example in Figure 2, the score would be .833 (five correct responses out of six total responses, excluding the first). The mean rate of responding and mean proportion of correct responses were calculated across trials for each subject. Because both of these measures were ratios that were indirectly dependent on the delay interval, they were arcsine transformed before statistical analysis.

Results and Discussion

The Δf Test

The mean Δf estimate for the 12 adult subjects was 12.91 Hz (SD = 5.41). As expected, the mean proportion correct on different trials was near the theoretically expected level of 84% (M = 83.4%, SD = 1.0). For same trials, performance was comparable, though slightly higher (M = 89.3%, SD = 11.7).

The Δt Test

The mean Δt estimate in the rehearsal condition was 13.759 s (SD = 7.551). For the musical imagery condition, the corresponding mean was very similar, 13.614 s (SD = 7.199). The difference between conditions did not approach significance. It is clear that there was no effect of the interpolated task, and both conditions yielded Δt estimates comparable with the adults in Experiment 1.

It theoretically would be possible for memory to be affected by musical imagery and for subjects still to reach similar Δt levels in both conditions, if there were a difference in response bias between conditions. Such a difference would show up empirically as a difference in the performance on catch trials in the two conditions. Though there was a higher proportion correct on catch trials in the rehearsal condition (M = .737, SD = .189) than in the musical imagery condition (M = .666, SD = .124), this difference did not approach significance. To investigate it further, we computed a nonparametric measure of sensitivity to the frequency difference taking into account performance on both same and different trials, A′ (Aaronson & Watts, 1987; Grier, 1971), from all trials at the subject's final Δt value. The difference in A′ between the musical imagery condition (M = .802, SD = .147) and the rehearsal condition (M = .844, SD = .115) did not approach significance. Thus, there was little, if any, effect of the nature of the silent strategy that subjects might use during the ITI on memory for pitch.

Musical Imagery Performance

Postexperiment interviews with the subjects indicated that they were in unanimous agreement that the musical imagery task effectively prevented them from rehearsing or thinking about the standard tone. All of the subjects maintained that the imagery task was very attention demanding. Three of the 12 subjects admitted to attempting initially to rehearse the standard tone during the imagery task, but all 3 claimed to have abandoned this attempt early in the course of the session because the perceived benefit did not outweigh the additional attentional demands. One additional subject claimed to have attempted a strategy of using the pitch of the standard tone as the starting pitch for the imagined tune, but he indicated that he also abandoned this in favor of a passive retention strategy because it did not seem to benefit his performance.

The objective measures of performance on the musical imagery task appear to confirm subjects' reports of the difficulty of the task. The mean rate of responding was 0.49 responses per second (SD = 0.22), much slower than subjects would be able to hum the tunes, and the mean proportion of correct responses was .620 (SD = .15), for which chance performance would be .33. The correlation between the rate of responding and the proportion of correct responses was not significant (r = .197, $p >$ .5, two-tailed).

If some subjects attempted to rehearse the first tone during the imagery task, then there should be a trade-off between performance in the imagery task and performance in the tone comparison task. No support for this possibility was found in the data, however. The correlation between subjects' rate of responding in the imagery task and Δt estimates was −.068 ($p >$ .8, two-tailed), and that between subjects' proportion of correct responses in the imagery task and Δt estimates was .127 ($p >$ .6, two-tailed). In contrast, the individual Δt estimates in the imagery and rehearsal conditions were positively related to one another ($r = .632$, $p < .02$, two-tailed), as they should be if both tap a common system of sensory memory storage.

In summary, the results of Experiment 2 suggest that rehearsal of the standard tone during the ITI is not an important determinant of the duration of auditory memory for pitch as measured by the present procedure. Therefore, it is unlikely that the development of strategic processing can account for the developmental difference observed in Experiment 1.

General Discussion

The primary goal of this study was to examine whether the rate at which information about the pitch of tones is lost from memory changes with development. Although some previous reports have indicated that children perform more poorly than adults at making comparisons between stimuli on the basis of auditory modality-specific pitch information (e.g., Maxon & Hochberg, 1982), the potential separate contributions of auditory sensory discrimination, pitch memory persistence, and pitch retention strategies were not addressed in those studies. The two experiments of the present study, taken together, indicate that there is a developmental increase in the persistence of memory for pitch between the ages of 6 to 7 years and adulthood, distinct from whatever other developmental factors exist. This is a rare finding in light of the literature on memory development in childhood (e.g., Kail, 1990), which has focused on the development of knowledge and strategies with little emphasis on possible developmental changes in basic memory structures.

The experimental logic used in this study was one in which the sensory retention factor was isolated from other factors. In a two-tone discrimination task, it is not initially clear what the separate contributions of sensory discrimination, memory retention, and strategic memory maintenance may be. In the present Experiment 1, we did not attempt to assess all factors but set up a situation in which effects of memory loss over time could be isolated from other effects. The frequency difference between tones was individually adjusted until a standard, 84% level of performance was reached using a 2-s ITI. Despite this careful matching, we found that the increase in ITI needed to lower performance to another standard level, 71% correct, was shorter in younger than in older subjects as shown in Figure 1. This difference can only be attributed to differential mean rates of forgetting of the tones across time in the different age groups. A developmental difference in sensory discrimination alone could not account for it, because the same frequency pairs were used in both the 84% and the 71% correct tracks of the Δt test procedure.

The results of Experiment 1 cannot distinguish between sensory memory persistence and active memory retention strategies. However, in Experiment 2 we addressed this issue by setting up a condition in which adult subjects were to imagine a melody between the two tones to be compared, greatly curtailing or eliminating any opportunity to rehearse the first tone in a pair. Given that this task did not noticeably lower performance (as measured by the Δt estimate for 71% correct) in comparison with a condition in which subjects were free to rehearse the first tone in a pair, the data suggest that strategic processing plays at most a minor role in performance. This result is, of course, in dramatic contrast to the retention of semantic categorical information such as letter names, for which a rehearsal-blocking task has a severe effect on performance (e.g., Peterson & Peterson, 1959). It is in contrast also to the actual presentation of interpolated tones during the ITI of a tone comparison task, which drastically reduces tone comparison performance regardless of whether the subject is required to attend to the interpolated stimuli or not (e.g., Deutsch, 1970; Massaro, 1970; Pechmann & Mohr, 1992).

It should be noted here that the results of this experiment do not necessarily suggest that tone rehearsal is never a beneficial strategy for maintaining pitch information. For example, Pechmann and Mohr (1992) found that, for tone comparisons involving stimuli differing by a semitone, adults without musical training performed somewhat worse when required to attend to visual or verbal stimuli presented between the tones than when allowed to ignore those stimuli. However, subjects in that study performed at 96% correct with a 6-s unfilled ITI. We are proposing only that subjects do not benefit noticeably in the present procedure from the rehearsal of more fine-grained pitch information during a silent ITI.

Taken together, the results suggest that the persistence of memory for pitch changes with age. According to the previous literature (e.g., Cowan, 1984, 1988), this memory for pitch is just one instance of auditory sensory memory, in principle similar to memory for other acoustic properties of sounds such as timbre, loudness, and duration. It may be that auditory sensory memory for many or all of these properties changes with development. Although the neural basis of a developmental increase in the persistence of memory for pitch still is entirely unclear, it is plausible that reverberatory neural circuits in or surrounding the sensory cortical areas (Hebb, 1949; Lopes da Silva, 1991) produce responses to stimulation that last longer in more mature brains.

The potential theoretical impact of findings such as these are readily apparent. Although we used a two-tone comparison procedure, the underlying theoretical construct under investigation was a rapidly decaying auditory memory representation that presumably could be used in a variety of tasks. The duration of auditory memory might, for example, constrain a child's ability to remember and copy musical notes; and if similar results are obtained for speech stimuli, it might even act as one constraint on the ability to retain and carry out verbal instructions.

A further understanding of the developmental course of structural components of memory should have important implications for information-processing models and theories of memory development, in part because the efficiency of strategies depends on the characteristics of the structures on which those strategies operate. The present study clarifies the developmental course of one type of structural component, auditory memory, with a psychometric procedure that successfully overcame the difficult, confounding factor of group differences in nonmemorial processes. Application of a similar experimental logic to other types of memory could provide a clearer overall picture of memory development.

References

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A′ and B″ to below-chance performance. *Psychological Bulletin, 102,* 439–442.

Andrews, M. L., & Madeira, S. S. (1977). The assessment of pitch discrimination ability in young children. *Journal of Speech and Hearing Disorders, 42,* 279–286.

Angell, F., & Harwood, H. (1899). Experiments on discrimination of clangs for different intervals of time. *American Journal of Psychology, 11,* 67–79.

Bently, A. (1966). *Musical ability in children and its measurement.* New York: October House.

Bull, A. R., & Cuddy, L. L. (1972). Recognition memory for pitch of fixed and roving stimulus tones. *Perception & Psychophysics, 11,* 105–109.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191.

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory and Cognition, 21,* 162–167.

Cowan, N., & Kielbasa, L. (1986). Temporal properties of memory for speech in preschool children. *Memory and Cognition, 14,* 382–390.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science, 168,* 1604–1605.

Engle, R. W., Fidler, D. S., & Reynolds, L. H. (1981). Does echoic memory develop? *Journal of Experimental Child Psychology, 32,* 459–473.

Frank, H. S., & Rabinovitch, M. S. (1974). Auditory short-term memory: Developmental changes in precategorical acoustic storage. *Child Development, 45,* 522–526.

Green, D. M. (1976). *An introduction to hearing.* Hillsdale, NJ: Erlbaum.

Grier, J. B. (1971). Nonparametric indices for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.

Hari, R., Hamalainen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., Alho, K., Näätänen, R., & Sams, M. (1984). Responses of the primary auditory cortex to pitch changes: Neuromagnetic recordings in man. *Neuroscience Letters, 50,* 127–132.

Harris, J. D. (1952). The decline of pitch discrimination with time. *Journal of Experimental Psychology, 43,* 96–99.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Jeffery, E. (1958). Variables in early discrimination learning: II. Mode of response and stimulus difference in the discrimination of tonal frequencies. *Child Development, 29,* 531–537.

Jensen, A. R. (1993). Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science, 2,* 53–56.

Kail, R. (1990). *The development of memory in children.* New York: Freeman.

Kidd, A. H., & Kidd, R. M. (1966). Auditory perception in children. In A. H. Kidd & J. L. Rivoire (Eds.), *Perceptual development in children* (pp. 113–142). New York: International Universities Press.

Lehrl, S., & Fischer, B. (1988). The basic parameters of human information processing: Their role in the determination of intelligence. *Personality and Individual Differences, 9,* 883–896.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467–477.

Lopes da Silva, F. (1991). Neural mechanisms underlying brain waves: From neural membranes to networks. *Electroencephalography and Clinical Neurophysiology, 79,* 81–93.

Lü, Z.-L., Williamson, S. J., & Kaufman, L. (1992). Physiological measures predict behavioral lifetime of human auditory sensory memory. *Science, 258,* 1668–1670.

Massaro, D. W. (1970). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology, 83,* 32–39.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 79,* 124–145.

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: Sensitivity and discrimination. *Ear and Hearing, 3,* 301–308.

Näätänen, R. (1992). *Attention and brain function.* Hillsdale, NJ: Erlbaum.

Parsons, D. (1975). *The directory of tunes and musical themes.* Cambridge, England: Spencer Brown.

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory and Cognition, 20,* 314–320.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 61,* 12–21.

Rabinowicz, T. (1980). The differentiate maturation of human cerebral cortex. In F. Falkner & J. M. Tanner (Eds.), *Human growth 3: Neurobiology and nutrition* (pp. 97–123). New York: Plenum Press.

Sipe, S., & Engle, R. W. (1986). Echoic memory processes in good and poor readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 402–412.

Trehub, S. E., Morrongiello, B. A., & Thorpe, L. A. (1985). Children's perception of familiar melodies: The role of intervals, contour, and key. *Psychomusicology, 5,* 39–48.

Weber, R. J., & Brown, S. (1986). Musical imagery. *Music Perception, 3,* 411–426.

Webster, P. R., & Schlentrich, K. (1982). Discrimination of pitch direction by preschool children with verbal and nonverbal tasks. *Journal of Research in Music Education, 30,* 151–161.

Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology, 6,* 13–61.

Received May 28, 1992
Revision received May 6, 1994
Accepted May 6, 1994 ■