A Multimodal Assessment of Sensory Thresholds in Aging

Joseph C. Stevens, L. Alberto Cruz, Lawrence E. Marks, and Stephen Lakatos

John B. Pierce Laboratory and Yale University, New Haven, Connecticut.

Young and elderly subjects yielded forced-choice detection thresholds in each of seven sensory tasks: (1) taste of sodium chloride, (2) smell of butanol, (3) cooling, (4) low-frequency vibration, (5) high-frequency vibration, (6) low-frequency hearing, and (7) high-frequency hearing. Average scores across these tasks nearly perfectly separated the 22 elderly from the 15 young subjects. For individual modalities, however, separation between the groups varied from complete (high-frequency touch) to negligible (low-frequency hearing). Scores on the Boston Picture Naming Test and especially the Wechsler Logical Memory Test correlated strongly with average threshold score (Pearson r = .80) and moderately with scores on individual modalities. This sensory-cognitive link is not caused, as might be supposed, by diminishing age-related capacity to handle the detection task, because the very same task resulted in negligible age effect (low-frequency hearing) and large effect (high-frequency hearing) in the same subjects.

A COMMON strategy in research on aging of the senses has been to compare cross-sectionally thresholds representative of the life span. In our laboratory this has included forced-choice detection thresholds for smell (Stevens, Cain, & Weinstein, 1987; Stevens & Cain, 1987; Stevens & Dadarwala, 1993), taste (Stevens, Cruz, Hoffman, & Patterson, 1995), and temperature (Stevens & Choo, in press), and spatial acuity thresholds for touch (Stevens, 1992; Stevens & Patterson, 1995; Stevens & Choo, 1996; Stevens, Foulke, & Patterson, 1996). Besides illuminating properties of sensory aging in the respective modalities, this work spurred the present study: to compare directly several modalities in the same group of individuals of various ages.

We are interested not only in various sense modalities (smell, taste, touch, hearing) but also in various submodalities of hearing and touch, that is, stimulus conditions within a given sense modality (low- vs high-frequency stimulation of touch and hearing in the present study). For simplicity we use the term "modality" to cover both kinds of sensory continua unless stated otherwise. The distinction between modalities and submodalities is sometimes fuzzy. For example, high and low frequencies of vibration are almost certainly detected by different types of receptors in the skin, suggesting that they could be considered as two different senses.

Practical and theoretical considerations dictated what, how, and whom to study. We decided on seven different detection thresholds: two for touch (low- and high-frequency vibration), two for hearing (low- and high-frequency tone), and one each for taste (NaCl), smell (butanol), and temperature (cooling). Visual thresholds, for example, had to be omitted because adequate dark adaptation would have required nearly 80 additional hours of subject time and expense; likewise, numerous other interesting and potentially useful stimulus conditions, such as intermediate (speech) frequencies in hearing, would have added considerably to an already heavy agenda of testing. We studied groups of 15 young and 22 elderly subjects only, and for the sake of individual reliability we tested each subject four different times on each of the seven threshold tasks, thereby yielding for analysis a data set of over 1,000 forced-choice thresholds.

We also tested these same persons for cognitive functioning on two well-known tests, the Boston Picture Naming and Wechsler Logical Memory tests, primarily as a subject screening device to control for senile dementia (none diagnosed). It turned out, however, that these tests provided useful information concerning the relation between cognitive and sensory capabilities in aging. The present study is primarily psychophysical and sensory in nature, and at the time of its design we were unaware of recent linking of cognitive and sensory factors in aging (Lindenberger & Baltes, 1994; Baltes & Lindenberger, 1997; Salthouse, Hancock, Meinz, & Hambrick, 1996). Fortunately, the screening tests chosen proved adequate to verify and strengthen these reports.

Past study of these sense modalities has revealed that weakening of absolute sensitivity with aging is common. Nevertheless, the various threshold tasks were selected because some modalities appear more vulnerable to aging than others. For example, high frequencies of vibratory touch and hearing have proved more vulnerable than low frequencies; smell and taste, more vulnerable than (although less thoroughly tested) temperature. The present study permits comparison of the same subjects' performance on the various threshold tasks. It also permits comparison of reliability coefficients characterizing the same subject's performance on a variety of tasks, thereby furnishing guidelines on the amount of testing needed for adequate evaluation of a given modality. Finally, it reveals the degree to which the thresholds of young and elderly persons overlap for different modalities. Thus, multimodal study of the same subjects can furnish kinds of information that are not available directly from separate unimodal studies.

Earlier studies (Cain & Gent, 1991; Rabin & Cain, 1986; Stevens, Cain, & Burke, 1988; Stevens & Dadarwala, 1993;
Subjects (6 male and 16 female) over 65 years (mean = 2.74) and another of 22 older persons, selected more or less randomly, and compared cross-sectionally on a variety of short detection threshold tests. Nevertheless, a number of general conclusions and hypotheses emerged from the multimodality strategy, and the study provides a kind of prologue to a more full-scale epidemiological survey of sensory decline with aging.

One should note that this kind of study is no substitute for thorough examination of each modality by itself, with appropriate sampling over the adult life span and adequate attention to the main stimulus parameters relevant to each. The unique feature of the present study is its attention to covariation among the several modalities examined in the same subject groups. Despite this caveat, it may be noted that the two subject groups gave data that are in good agreement with the body of knowledge of sensory aging.

Method

Overall Plan and Subjects

The first of five sessions was a 1-hr interview/screening. Potential subjects were recruited from advertisements posted at nearby university buildings and at two senior centers in the New Haven community. Subjects were read a brief questionnaire regarding medical history and administered two cognitive tests: the Boston Naming Test (last 30 items), which tests ability to name line drawings of objects (Kaplan, Goodglass, & Weintraub, 1983) and the Wechsler Logical Memory Test, revised version (1987), which tests memory for two stories, immediately and 30 min after hearing them read aloud; the subject’s score was the number of items (of 49 possible) recalled in their retelling of the two stories.

Although potential subjects were paid for this screening session, not all were selected for further study. With certain restrictions, those selected were done so on a first-come, first-served basis to provide two groups, one consisting of 15 younger subjects (6 male and 9 female) between ages 18 and 27 (mean = 22.1, SD = 2.74) and another of 22 older subjects (6 male and 16 female) over 65 years (mean = 76.7, SD = 5.58). (No significant sex-related or sex by age-related difference emerged in any dependent variable of this study, perhaps owing to the small number of subjects.) Excluded from further testing were persons who reported being smokers at any time during the previous 10 years, persons diagnosed as diabetic, persons giving evidence of dementia (no one did), persons who use a hearing aid, and persons presenting physical trauma in the region of the skin tested (none).

Thereafter, each subject was tested in four 2-hr sessions on four different days, not necessarily consecutive. Payment was $100 upon completion of all five sessions. In each of the 2-hr sessions a subject was tested on all seven modalities, with intervening rest periods. For logistic reasons the modalities were tested in two blocks: (1) taste, smell, and temperature (modalities balanced for order), and (2) hearing and touch (sense modalities counterbalanced for order and, within a sense modality, high and low frequency balanced); blocks 1 and 2 were also balanced for order. In this way, potential order effects were avoided insofar as circumstances permitted. The taste, smell, and cooling testing was done in an ordinary laboratory test chamber. The auditory and vibrotactile testing was done in an audiometric test chamber (IAC).

Psychophysical Method

All thresholds were measured by an adaptive, two-interval, forced-choice procedure. On each trial the subject confronted sequentially a signal stimulus and a blank stimulus (order randomized) and had to decide which was the signal, if necessary by guessing. The level of the signal was determined by the “two correct down, one wrong up” rule (Wetherill & Levitt, 1965). Following two correct consecutive choices at a given signal level, the level was decreased by a specified amount on the next trial, and after one wrong choice, the level was increased by that amount on the next trial. Testing began at a specified signal level, but the recorded track started after the first wrong choice and was continued until a fixed number of reversals in the direction of the track (from down to up or from up to down) had occurred. The reversals numbered 9 in the case of the smell and taste and 12 in the case of temperature, touch, and hearing. The sampling of tastes and odors took longer than the other stimuli, which was offset by shorter tracks (fewer trials). For every task, threshold was defined as the average of the last six reversals. Given the two-down, one-up rule, the threshold corresponds to 71% correct (Wetherill & Levitt, 1965).

Following a suggestion by Wetherill and Levitt (1965), the step by which the signal level changed from one trial to the next was relatively gross early in the track and relatively fine later on. The rationale was to locate quickly the neighborhood of the threshold (through the first six reversals for touch, hearing, and temperature and through the first three for smell and taste) and thereafter to refine its location over the last six reversals. For each modality a detailed description of the stimuli including starting levels, initial step size, and final step size follows.

Although somewhat dependent on the particular modality, the psychophysical tracks usually took from 10 to 30 min and consisted of 25 to 50 forced-choice trials each.
Stimuli

Taste.—A series of 20 concentrations of NaCl was constructed by successive 0.25 log dilutions starting from a mother solution of 1 mol/liter of Baker-grade NaCl in deionized water (DHOH). These were stored under refrigeration, warmed to room temperature for use in testing, and replaced after 3 weeks of use. On a given trial the subject faced two 30-ml plastic medicine cups, one containing about 5 ml of an NaCl (signal) solution, the other the same quantity of DHOH (blank). On the first trial, the signal concentration was the tenth dilution step (young subjects) or the eighth step (elderly subjects), chosen because of the approximately threefold age-related difference found earlier for NaCl (Stevens, 1996; Bartoshuk, Rifkin, Marks, & Bars, 1986). For taste and smell only, the tracks began at different levels for young and elderly subjects. The reason for this was the shorter length of these tracks and the concern to avoid “starting bias” (Simpson, 1989) that can result from starting the track too far from the subject’s threshold. The gross step size (first part of track) was two dilution steps (0.5 log concentration), the fine step size was one dilution step (0.25 log concentration). The sampling method was sip and spit. The subject rinsed with DHOH thoroughly at the start of the session and after each stimulus; all stimuli and rinses were expectorated.

Smell.—A series of 15 concentrations of n-butanol was constructed by successive ternary dilutions with DHOH and stored in polypropylene “shampoo” bottles (250-ml), starting from a stock solution of 4% (v/v), which corresponds to concentration in air (headspace) of 3,100 ppm as determined by gas chromatography (see Stevens & Dadarwala, 1993). The gross step size (first part of track) was two dilution steps (ninefold concentration change), the fine step size (later part) was one dilution step (threefold change).

Temperature (cool).—The stimuli were delivered to the right hand (thenar eminence) via a 2.2 × 2.2 cm Peltier heater/cooler transducer. This device was under the control of a computer that monitored a sensor at the interface between the module surface and the skin. A track began with 30-s acclimation to a baseline temperature of 33 °C. A trial consisted of a sequence of a signal (a 3-sec cooling pulse, including rise time) and a blank (3-sec) in random order, separated by an interval of 2 s. A brief auditory beep marked the start and the end of both stimulus intervals. The stimulus was reckoned as the downward change in skin temperature in degrees Celsius from the 33 °C baseline, or Δ °C, reported here without regard to sign (as calibrated via the sensor on the skin). Overall shape of the pulses was approximately the same regardless of pulse magnitude, with uniformly rapid rise and fall times, always far faster than 0.05 °C/s, which Kenshalo, Holmes, and Wood (1968) reported to be the rate of change faster than which warming and cooling thresholds are independent of rate.

The starting pulse was always 1°, gross step size (first part of track) was 0.2°, fine step size 0.1° (latter part). About one-third of the way into the experiment, it became apparent that these step sizes were too gross to accommodate the exquisite sensitivity of some young subjects, whose responses in the latter part tended to alternate regularly between ΔC = 0 and ΔC = 0.1. Thereafter the exact up-down tracking rule gave way to a new rule, whereby step size was narrowed progressively on successive correct responding and widened on successive incorrect responding. As in the other tracks, the “two correct down, one wrong up” pattern was in force, and the threshold was computed as the average of the last six reversal levels as usual.

Touch (low- and high-frequency vibrotactile signals).—The stimuli were 250-ms bursts of 20-Hz and 250-Hz sinusoids delivered to the thenar eminence of the right hand. Sinusoidal signals were filtered, linearly shaped in rise and decay (both 50 ms), and attenuated using Coulbourn Instruments acoustic modules, then amplified and delivered to an accelerometer-calibrated, minishaker-driven piston (Brüel and Kjaer) having a 1-cm² circular contact surface. This surface projected about 5 mm upward through a hole (diameter about 2 mm larger than that of the piston) in a rigid table top over which the subject placed the hand, thereby pressing the piston into the skin. The shaker was mounted on a triple-beam balance, to which was added a constant mass (40 g) so that the force of contact with the hand was held constant. The table top served as a fixed (nonvibrating) “surround.” During all threshold determinations, the subject listened through earphones to a white noise loud enough to mask the audibility of the vibrotactile signals.

Starting levels were 44 dB (20 Hz) and 36 dB (250 Hz) re 1 µm, measured peak to peak. The gross step size was 4 dB (first part of the track), the fine step size, 1 dB (second part).

For the young subjects, the experimental sessions were fully automated. Visual information on a computer screen signaled concurrently the occurrence of each observation interval. After stimulus presentation, subjects pressed one of two foot pedals corresponding to “stimulus occurred in the first interval” and “stimulus occurred in the second interval.” Pilot testing of elderly subjects disclosed that they had considerable difficulty coordinating the tasks (i.e., looking at the monitor, attending to concurrent tactile stimulation, pressing pedals to respond) and that, as a result, they made frequent response errors even when they could clearly perceive the signal. In order to lighten the task demands for the elderly subjects, we placed an experimenter beside them in the test chamber who pressed the appropriate response pedal in accordance with the subject’s verbal
response (e.g., “I felt the first [second] one”). This procedure permitted subjects to focus attention on the signal and thereby reduced the likelihood that threshold differences between young and elderly subjects could result in some significant way from nonsensory factors.

**Hearing (low- and high-frequency acoustic signals).**—The auditory stimuli were 200-Hz and 3,000-Hz sinusoidal signals generated by Coulbourn Instruments acoustic modules. Signals were 250 ms long, with 50-ms rise and decay ramps, and after attenuation were fed to calibrated headphones (Telephonics model TDH-39P). Starting level was 58 dB SPL for 200 Hz and 64 dB SPL at 3,000 Hz. The gross step size was 4 dB (first part of track); fine step size, 0.5 dB (second part). As in the touch experiment, sessions were automated for young subjects, but the elderly subjects were accompanied by an experimenter, who recorded their responses in the manner described above.

**RESULTS**

In evaluating aging across various sense modalities, one faces the complication that stimulus measures are not directly comparable. This is obvious when comparing different modalities, e.g., smell with hearing. There is no a priori way of telling whether an age-related fivefold concentration deficit in olfaction is more or less serious than a fivefold sound pressure deficit in hearing (7 dB). Less obvious but nonetheless true, there is no a priori way of telling whether a 7-dB hearing loss is more or less serious than a 7-dB vibrotactile loss. A common metric does not necessarily imply comparability of differences. Yet comparison of these various modalities in the same subjects is the objective of the present study. Such a comparison (titled “Analysis Across Modalities” below) was achieved by (1) pairwise correlations between modalities and (2) analysis of variance performed after z-score transformations of the thresholds. At the same time, it is also useful first to describe each modality separately using its own appropriate stimulus metric (titled “Analysis Within Modalities” below), to which we first turn.

**Analysis Within Modalities**

Figure 1 displays all individual thresholds for each modality (panels 2–8) and (in panel 1) the individual scores on the Boston Naming Test and the Wechsler Logical Memory Test, plotted as a function of subject’s age. The horizontal ticks at the center of the panels mark each young’s (left) and each elder’s (right) threshold averaged across the four sessions. Memory scores based on immediate and delayed (30-min) recall were highly intercorrelated (Pearson r = .954), and for simplicity’s sake the averages of the two are plotted in Figure 1. In accordance with long-standing practice, all thresholds are plotted in logarithms or equivalent logarithmic transform (decibels, dilution steps) of the metric appropriate for the modality: molar concentration for taste, concentration parts per million (ppm) in air for smell, Δ °C (disregarding sign) for temperature (cooling), sound pressure for hearing (high and low frequencies), and peak-to-peak amplitude in micrometers for touch (high and low frequencies). Each of panels 2–8 displays a separate regression line for the thresholds obtained in each session (1 to 4). Each panel also gives the equation of the (average) regression line and the Pearson r relating threshold to age.

Regression lines in the absence of middle-aged subjects presupposes that the aging effect, as assessed cross-sectionally, is linear over the age span from youth to advanced age. More detailed study of smell (Cain et al., 1995), taste (Bar-toshuk et al., 1986), hearing (Bunch, 1929), low-frequency touch (Gescheider, Bolanowski, Hall, Hoffman, & Verrillo, 1994), and current measures of warming and cooling thresholds at multiple body sites (Stevens & Choo, in press) are consistent with a linear regression. Gescheider et al. (1994) have described aging of high-frequency vibration with two line segments, one for persons younger than 65–70 years, another, steeper, for those older so there is room for refinement. Regression is used mainly to show that there are real age-related differences in absolute sensitivity; alternative statistical analysis (t tests comparing young and elderly groups) leads to the same conclusion (see Figure 2). When comparing studies of a given modality, in which age distribution is often quite diverse, regression serves as a more convenient yardstick than, for example t tests.

Four features of panels 2–8 in Figure 1 are noteworthy.

1. The slopes of all seven regression lines are positive, indicating an average rise of threshold (decline in sensitivity) with age. Except for panel 7 (low-frequency hearing), these slopes are also statistically significant, i.e., there was a reliable age effect in six of the seven modalities. In these six the mean differences between the two age groups as a whole were also significant by individual unpaired t tests, but the t test for low-frequency hearing was not (cf. Figure 2).

2. For any given modality, there are large differences among subjects and even within the four thresholds of a particular subject, so that the overall scatter of thresholds is marked. Detailed inspection shows that the means of the four thresholds are generally less variable than the unaveraged thresholds, also a prominent feature of earlier studies (Cain & Gent, 1991; Rabin & Cain, 1986; Stevens et al., 1988; Stevens & Dadarwala, 1993; Stevens et al., 1995; Stevens & Cruz, 1996; Stevens & Choo, 1996). Brief threshold tests of the kind used here and in aging research in general, based as they are on only about 40 trials, tend to exaggerate individual differences by adding within-subject variability to between-subject variability. In doing so they also tend to obscure the effect of aging. Gauging the sensitivity of individuals with single brief threshold tests has been compared to assessing baseball players’ hitting abilities with their batting average in a single game: such averages will exaggerate individual differences and obscure real differences in ability. Longer-term averages are obviously more reliable.

3. Even based on four tests, the average thresholds of younger and older subjects overlap (examine the horizontal ticks in panels 2–8), but the degree of overlap depends greatly on the modality, varying from virtually total (low-frequency hearing), to little (taste, smell), to none (high-frequency touch). Plainly, aging has a more profound effect on some modalities than others. More is said later on this matter in analysis across modalities.
Figure 1. Individual test scores on the Boston Picture Naming Test and the Wechsler Logical Memory Test (panel 1) and individual thresholds (circles) for seven modalities (panels 2–8) as a function of subject age. The horizontal ticks at the center of the panels mark the locations of each subject’s mean (of four) thresholds. The lines shown in the figure are regression lines. The equations shown in panels 2–8 describe the average of the four regression lines in each panel, one each for each of the four sequential threshold tests. Also shown are correlation coefficients (Pearson r) with associated significance.

4. Despite the variability, the four regression lines fitted to the four sequentially obtained thresholds are fairly similar to one another. This implies that a single threshold from a large group of subjects may provide sufficient information to determine the magnitude of the aging effect for a given modality. Most studies of aging have in fact relied on single brief thresholds from relatively large groups of subjects. These can adequately reveal the overall magnitude of the aging effect, but conclusions drawn from them about individual subjects have sometimes been misleading (see Stevens et al., 1995). When it comes to the evaluation of the individual aging person, more information may be needed than can be supplied from only 40 or so trials.

How much information is provided by repeated testing? In addressing this question one may note that the total variance associated with a set of measurements has two sources: that associated with sampling a given individual’s threshold, or the within-subject variance, and that among the mean thresholds across the individuals in the subject population, or the between-subject variance. That is,

\[ \sigma^2_{\text{total}} = \sigma^2_{\text{within}} + \sigma^2_{\text{between}} \]

Further, the ability to discern a difference between the mean thresholds for groups of young and old individuals will depend on the standard errors associated with the mean thresholds within each group,

\[ \frac{\sigma^2_{\text{within}}}{n} + \frac{\sigma^2_{\text{between}}}{N} \]

where \( n \) is the number of repeated tests and \( N \) is the number of individuals. In principle, the reliability of the threshold measures should always improve with repeated testing, but this will be so only to the extent that the within-subject variance is substantial relative to the between-subject variance.
One empirical measure is to compute how the value of Student's $t$ comparing young and elderly subjects increases as a function of the number of threshold tests averaged to give a composite threshold. Figure 2 plots $t$ computed for $n = 1$ (mean of all single tests), 2 (mean of all pairs of tests), 3 (mean of all trios of tests), and 4 (mean of all four tests). For all modalities, $t$ increases with number of tests averaged to give a composite threshold; that is, discriminability of the two age groups improves with more extensive testing of the same subjects. The increase results from a shrinking of the variability of the averaged thresholds, not from any increase in the mean difference between the young and the elderly subjects. For all modalities, $t$ tends to level off with increasing number of tests, suggesting diminishing returns upon further testing. It is also clear that more is gained by repeated testing in some modalities (smell, taste, touch) than others (hearing, temperature). An interesting case is high-frequency hearing, for which repeated testing appears to be largely a waste of time. This finding doubtless relates to the high test–retest reliability of the high-frequency test, a subject to which we turn next.

**Test–retest correlations.**—For each of the seven modalities, Pearson correlation coefficients, $r$, were computed for each of six possible pairings of the four tests conducted on four different days. These were always positive in sign and nearly always statistically significant. Table 1 (column 2) lists the median (almost identical to the mean) of these six coefficients for each modality. The test–retest correlations are moderately high, except for low-frequency vibration and audition. Judging from the overlap in Figure 1 between young and elderly subjects' thresholds, these are the two modalities that show the smallest age effect. A large (small) aging effect means a large (small) range of threshold scores, and the larger the range, the larger the coefficient of correlation to be expected.

That these test–retest correlations vary from one modality to another (for high-frequency hearing they are especially large) suggests that the number of tests necessary to discriminate between younger and older subjects may also vary with the modality. The lower the value of $r$, the more tests may be necessary.

**Analysis Across Modalities**

**Pairwise comparison of variables.**—How strongly does threshold (mean of four for each subject) for one modality correlate with threshold for each modality and with the two tests of cognitive ability (picture recognition and story memory)? Table 2 lists the Pearson correlation coefficients $r$ for 36 pairings of these variables.

Of immediate interest is the positive correlation of all of the modalities with one another and the negative correlation of the two cognitive test scores with all of the modalities. Of these 14 negative correlations, 13 are individually statistically significant. Low scores on a cognitive test, especially the memory test, tend to be associated with high thresholds. Even the low-frequency auditory thresholds, which showed no significant relation to age, nevertheless correlated significantly (though weakly) with the cognitive scores.

These moderate negative correlations say nothing about a causal relation between cognition and sensory function—in either direction. Indeed, if cognitive weakness were the cause of threshold elevation, one would expect all the modalities to be affected approximately equally by aging. But this was not the case, most evident in the contrast between low-frequency hearing (no significant aging effect on thresholds) and high-frequency touch (complete experimental separation of the thresholds of older from younger subjects). The correlations between cognitive and sensory functions serve as reminders, however, that threshold tasks can be unwisely designed to place too heavy reliance on cognitive dexterity and thereby erroneously confuse sensory and cognitive effects of aging.
provide an answer to the interesting question as to whether action between age category and modality \( F(6,210) = 10.218, p < .0001 \) also emerged, showing that age exerts its effect nonuniformly across the various senses and even across different stimulus conditions within a given sense. For example, age affects temperature less than smell and taste, and low-frequency hearing and touch less than high-frequency hearing and touch. These modality-specific effects can be examined in Figure 3, which compares young and elderly mean z score on each modality, and in Figure 4, which displays the regression lines relating z scores to age. Here it is seen that the slopes range from nearly zero (.008 for low-frequency hearing) to relatively steep (.031 for high-frequency touch). Clearly, no single function could adequately characterize all modalities.

One view of aging supposes that some individuals age physiologically faster than others. If this physiological aging characterizes the whole sensory world of a subject, then he or she ought to show relatively uniform sensitivity losses across modalities. To what, if any, extent is this true? To address this question the subjects in each age category were divided into three performance ranges—lower, middle, and upper—according to whether their thresholds, averaged across all seven conditions, fell into the lower, middle, or upper one-third of their age group. Figure 5 plots the average z scores for each of the three groups of young and elderly subjects on each of the seven modalities. If the same subjects account for the aging effect across modalities, then we would expect the same histogram profile for each modality. However, no such common shape is to be strictly observed throughout, implying that individuals do not sort without qualification into sensory "losers" and "retainers." A first-order answer to the question is, however, possible by the analysis of z scores presented under the next heading.

A final word concerns what happens to the intermodality coefficients in Table 1 when age is kept constant by mean of first-order partial correlation, as was performed on the intramodality coefficients. The answer is that they are reduced nearly to zero (median \( r = .107, \text{n.s.} \)). In other words, the correlations in Table 2 are attributable almost exclusively to age.

Analysis of z scores.—As pointed out earlier, it is empirically meaningless to compare directly the raw threshold magnitudes obtained across modalities. To achieve a more meaningful comparison, for each modality the averages of the subjects' four thresholds were transformed into z scores with a mean of zero. The z scores for each modality enable us to characterize the relative distributions of young and elderly subjects within the total distribution for each modality. One can then compare these distributions across modalities in more exact quantitative ways than was otherwise possible.

Analysis of variance of the z scores gave a significant main effect \( P(1,35) = 127.339, p < .0001 \) for age category, confirming that elderly subjects had higher thresholds across the modalities than did young subjects. A significant interaction between age category and modality \( F(6,210) = 10.218, p < .0001 \) also emerged, showing that age exerts its effect nonuniformly across the various senses and even

Table 2. Pearson Correlation Coefficients, \( r \), for 21 Pairings of the Seven Modalities With Each Other and With Scores on the Two Cognitive Tests, the Boston Naming Test and the Wechsler Logical Memory Test

<table>
<thead>
<tr>
<th>Memory</th>
<th>Naming</th>
<th>Taste</th>
<th>Smell</th>
<th>Cool</th>
<th>Touch, low</th>
<th>Touch, high</th>
<th>Hearing, low</th>
<th>Hearing, high</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naming</td>
<td>.57</td>
<td>-.74</td>
<td>-.59</td>
<td>-.45</td>
<td>-.49</td>
<td>-.78</td>
<td>-.38</td>
<td>-.60</td>
</tr>
<tr>
<td>Taste</td>
<td>-.74</td>
<td>.54</td>
<td>-.38</td>
<td>-.27*</td>
<td>-.52</td>
<td>-.54</td>
<td>.35</td>
<td>-.41</td>
</tr>
<tr>
<td>Smell</td>
<td>-.59</td>
<td>-.38</td>
<td>.56</td>
<td>.47</td>
<td>.46</td>
<td>.68</td>
<td>.35</td>
<td>.56</td>
</tr>
<tr>
<td>Cool</td>
<td>-.45</td>
<td>-.27*</td>
<td>.47</td>
<td>.42</td>
<td>.28*</td>
<td>.57</td>
<td>.22*</td>
<td>.26*</td>
</tr>
<tr>
<td>Touch, low</td>
<td>-.49</td>
<td>-.52</td>
<td>.47</td>
<td>.46</td>
<td>.28*</td>
<td>.57</td>
<td>.22*</td>
<td>.26*</td>
</tr>
<tr>
<td>Touch, high</td>
<td>-.78</td>
<td>-.54</td>
<td>.66</td>
<td>.68</td>
<td>.39</td>
<td>.57</td>
<td>.41</td>
<td>.59</td>
</tr>
<tr>
<td>Hearing, low</td>
<td>-.38</td>
<td>-.35</td>
<td>.35</td>
<td>.27*</td>
<td>.22*</td>
<td>.41</td>
<td>.26*</td>
<td>.19*</td>
</tr>
<tr>
<td>Hearing, high</td>
<td>-.60</td>
<td>-.41</td>
<td>.56</td>
<td>.50</td>
<td>.31</td>
<td>.61</td>
<td>.59</td>
<td>.19*</td>
</tr>
</tbody>
</table>

*Not statistically significant at \( p < .05 \).
DISCUSSION

This study provides some new insight into the nature of aging in the sensory domain, and confirms earlier insights as well. Of course, the sensory domain is vast territory. The choices of modalities and conditions were somewhat arbitrary, but nonetheless illustrative, as was the choice of the dependent variable (detection threshold). Full-scale assessment of sensory functions needs to include such classical measures as magnitude estimation, differential sensitivity, and spatial and temporal resolution. Of these, detection by forced choice has a near unique advantage of providing an objective measure applicable to all of the major sense modalities.

Aging has a pervasive but differential effect on the senses.—The outcome makes clear that aging takes a toll on many modalities; of the seven modalities selected for study, only low-frequency hearing escaped a statistically significant age effect. The mean $z$ scores on all seven threshold tests all but perfectly separated the sensory performance of our young and elderly subjects; the chances of an elderly person totally escaping aging’s impact are slim indeed. Low-frequency hearing appears to be nearly impervious to aging’s impact; thus it can serve as a comparison to gauge aging in other modality conditions [the approach taken by Bartoshuk et al. (1986) in a study of age-related taste changes]. Moreover, thresholds in one modality correlate significantly with thresholds in any other modality, and in our study age accounts for most of the variance.

It is, of course, a corollary that aging takes a greater toll on some modalities than on others. This finding challenges a rather common, unsupported assumption that aging causes a uniform decline of all physiological functions. Of course, the modalities for study were selected because earlier experiments on individual modalities suggested differential effects of aging. The difference between low- and high-frequency hearing has long been known to the audiologist (Bunch, 1929). The difference between low- and high-frequency vibrotaction also is well established (Verrillo, 1980; Kenshalo, 1987; Gescheider et al., 1994). Age-related differences in smell (for summary, see Stevens & Dadarwala, 1993) and taste (for summary, see Stevens et al.,
1995) have been documented repeatedly. The preponderance of evidence on temperature sensitivity from several studies reveals an aging effect for both warming and cooling of various body sites, including the thenar eminence (for a review, see Stevens & Choo, in press).

Patterns of aging.—An issue raised but incompletely answered is the degree to which decline of all the various modalities goes hand in hand with aging. Although significant pairwise correlations appear to be the rule, they do not prove a common aging factor. Some evidence for a common factor did emerge by comparing the profile of high-, medium-, and low-threshold performers overall with their profiles for the separate modalities. It seems likely that this factor may also account for the pervasive correlation of the sensory tests with the cognitive scores seen in Table 2. But it is also apparent that there are specific differences among some of the modalities that must be accounted for in other ways. When it comes to questions of this sort, principal factors analysis may eventually provide answers. Such might be a goal of a full-scale epidemiological study of sensory aging with number of subjects adequate to warrant more sophisticated statistical analysis.

Individual versus average aging effects.—The results confirm earlier reports that the customary brief tests can exaggerate individual differences in sensitivity (Cain & Gent, 1991; Rabin & Cain, 1986; Stevens & Dadarwala, 1993; Stevens et al., 1995; Stevens & Cruz, 1996). For most modalities, brief tests provide rather unreliable estimates of a subject’s real sensitivity. An antidote is repeated threshold tests. The greater the number of tests averaged, the smaller the overall variability among the subjects of a particular age group, and the greater is the ability to separate the sensitivities of young and elderly persons. The outcome also demonstrates, however, that discriminating sensitivities may require more or less testing, depending on modality. Figure 2 provides guidelines: for example, relatively little is gained by repeated measures of hearing, much by repeated measures of touch. Even though the various modalities were assessed by a uniform procedure, it discriminated more economically in some modalities than in others.

Cognitive/sensory correlation.—A provocative finding of this study is the pervasive correlation between scores on two well-known cognitive tests, both picture naming and memory for stories, and the various threshold tasks. In Figure 6 is plotted for each subject the Wechsler memory score against the \( z \) score averaged across all seven modalities. The correlation is strong (Pearson \( r = .80 \)); the corresponding correlation for the Boston naming test was 0.59 (\( p < .0001 \)). This high correlation is attributable in large measure to age, for when the variance contributed by subjects’ chronological age is removed by partial correlation, the coefficient drops from .8 to .35 (\( p < .05 \)).

The correlations for individual modalities (shown in Table 2) are moderate but seem to pertain to all modalities, even low-frequency hearing (for which the aging effect obtained actually fell short of statistical significance). Of course, the correlations are highest for modalities showing the clearest aging effect (taste, smell, high-frequency hearing, and touch), but the correlations across the sensory domain suggest a common linkage.

One hypothetical linkage might be that the elderly are not cognitively up to the task. They might fail to attend as consistently (and therefore guess more often than young subjects), or they might fail to remember in which of the two-alternative forced-choice intervals the stimulus occurred, even though they are perfectly capable of sensing the signal. Although a degree of cognitive shortcoming in the task itself cannot be totally ruled out, it can be dismissed as the fundamental explanation of the aging effects. The reason is that the various threshold tasks had a common methodological design and therefore put about constant cognitive demand on the subject. If cognitive shortcoming were the determiner of threshold, then aging should affect low-frequency hearing and touch as much as it does high-frequency hearing and touch. But if anything is clear in the outcome it is that aging affects the high frequencies much more than the lows, as students of hearing and vibration well know.

The explanation for the correlation seems more indirect than by way of task demands. According to this view, a person who shows the marks of age in one way is likely to show the marks of age in another. Hence, a person whose memory is weaker by dint of age is, on the average, more likely to have weaker hearing as well. Correlations of this kind prove nothing about direct causation: weaker memory does not cause weaker hearing, or vice versa. Nevertheless, they are linked in some way, and so too apparently with the other senses studied.

Other recent studies have also uncovered a strong link between sensory and cognitive functioning in aging. Lindenberger and Baltes (1994) assessed visual acuity (of the “corrected” type) and audiometric thresholds in addition to five cognitive variables (speed, reasoning, memory, knowledge, and fluency) in 156 persons between 70 and 103
years. Although their sensory measures were hardly strictly comparable to forced-choice detection thresholds, they nonetheless showed quite clearly that sensory functioning is “a strong late-life predictor of individual differences in intellectual functioning.” Salthouse et al. (1996) went on to study visual acuity and cognitive functioning in 204 somewhat younger persons (18–80 years) but with similar outcome: corrected visual acuity shared a large proportion of the age-related variance in measures of memory, learning, and concept identification. Finally, Baltes and Lindenberger (1997) extended their earlier study to 687 persons between 25 and 103 years, with comparable links among visual acuity, audiometric thresholds, and intelligence. Like us, these authors favor the idea that some “common factor” contributes to age-related differences over a broad spectrum of both cognitive and sensory functioning.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health (NIH) grants AG 10295 and AG 04287 to Joseph C. Stevens. Stephen Lakatos was supported by National Research Service Fellowship Award F32 DC00174 from NIH. A preliminary unpublished summary of the present findings was presented to the NIH Workshop on Changes in Multimodal Sensory Systems with Aging, in Bethesda, Maryland, in September, 1996, convened by Dr. Judith A. Finkelstein. The authors thank Kenneth K. Choo for extensive technical assistance and Dr. Suzanne Corkin for valuable suggestions in the planning stages of this research.

Address correspondence to Dr. Joseph C. Stevens, John B. Pierce Laboratory, 290 Congress Avenue, New Haven, CT 06519. E-mail: jstevens@jbpiercce.org

REFERENCES


Received May 8, 1997
Accepted September 9, 1997