Individual differences in auditory abilities

Gary R. Kidd, Charles S. Watson, and Brian Gygi

Citation: The Journal of the Acoustical Society of America **122**, 418 (2007); doi: 10.1121/1.2743154
View online: https://doi.org/10.1121/1.2743154
View Table of Contents: https://asa.scitation.org/toc/jas/122/1
Published by the Acoustical Society of America

Individual differences in auditory abilities

Gary R. Kidd, Charles S. Watson, and Brian Gygi
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 30 June 2006; revised 13 January 2007; accepted 7 April 2007)

Performance on 19 auditory discrimination and identification tasks was measured for 340 listeners with normal hearing. Test stimuli included single tones, sequences of tones, amplitude-modulated and rippled noise, temporal gaps, speech, and environmental sounds. Principal components analysis and structural equation modeling of the data support the existence of a general auditory ability and four specific auditory abilities. The specific abilities are (1) loudness and duration (overall energy) discrimination; (2) sensitivity to temporal envelope variation; (3) identification of highly familiar sounds (speech and nonspeech); and (4) discrimination of unfamiliar simple and complex spectral and temporal patterns. Examination of Scholastic Aptitude Test (SAT) scores for a large subset of the population revealed little or no association between general or specific auditory abilities and general intellectual ability. The findings provide a basis for research to further specify the nature of the auditory abilities. Of particular interest are results suggestive of a familiar sound recognition (FSR) ability, apparently specialized for sound recognition on the basis of limited or distorted information. This FSR ability is independent of normal variation in both spectral-temporal acuity and of general intellectual ability. © *2007 Acoustical Society of America.*

[DOI: 10.1121/1.2743154]

PACS number(s): 43.66.Fe, 43.66.Mk, 43.71.Es [MSS] Pages: 418–435

I. INTRODUCTION

The study of individual differences has a long history in psychology, most notably in the area of mental or cognitive abilities (e.g., Cattell, 1885; Galton, 1869, 1883; Binet, 1903; Thurstone, 1938; Carroll, 1993). Rather than simply documenting the range of abilities on various tasks, a primary goal of the subfield of differential psychology has been to identify the independent dimensions on which individuals vary. The identification of such dimensions, often termed *abilities*, has been a crucial component in the development of theories of intelligence and cognition. The nature of these abilities, and the extent to which they are utilized in different cognitive tasks, has been investigated with a wide range of tests (e.g., Sternberg, 1977; Hunt, 1978; Carroll, 1993). In some cases, experimental manipulations have both validated and further elucidated the nature and significance of specific abilities (e.g., Snow and Lohman, 1988).

The present study applies the individual differences approach to the study of auditory abilities. The goal was to identify the number and nature of distinct auditory abilities that underlie performance on a broad range of auditory discrimination and recognition tasks. The potential usefulness of such a determination is illustrated by the hypothesis that problems with speech recognition are a consequence of limited temporal (or spectral) processing ability. Rather than test this hypothesis by comparing speech recognition performance to performance on some single test of temporal resolution, a more valid approach would ask whether speech tests in general correlate strongly with a variety of different temporal and spectral processing measures. An extension of this approach would determine the relative contributions of each of several independent auditory abilities (assuming that such abilities do exist) to performance on speech recognition tasks. That strategy is employed in the research reported in this article.

"Auditory ability" is an abstract concept, comparable to general intelligence or physical strength. It is often a practical convenience to characterize individuals as more or less intelligent or strong, despite the uncertain referents of those concepts or *intervening variables* (MacCorquodale and Meehl, 1948). The validity of intervening variables depends on two properties, the first of which is the existence of a group of different measures (behavioral or psychophysical tests in the case of auditory abilities), all of which are significantly correlated with each other. The second property, which can elevate an intervening variable's status from a statistical inference to something more substantial, is a demonstrated neurophysiological correlate of those psychophysical measures, such as the rate of neural transmission, length of the cochlea, or volume of neural tissue in one of the auditory areas of the CNS. If multiple psychophysical measures of an ability are reliably demonstrated, there is a strong likelihood that a neural correlate will eventually be discovered. The work reported here addresses the first of these properties by using principal components analysis and structural equation modeling to identify distinct auditory abilities that underlie performance on a wide range of auditory tests.

Despite a number of earlier efforts, there is only modest agreement about the number or nature of auditory abilities. An exception to this generalization is found in one basic ability, auditory sensitivity as reflected in the audiogram. If persons with clinically significant hearing loss are included, this is clearly a primary dimension on which individuals differ and which predicts performance on many tasks, including speech recognition. This observation was made early in auditory research and has been reflected in the development of

0001-4966/2007/122(1)/418/18/$23.00

the Articulation Index (French and Steinberg, 1947) and in more recent derivatives of that index (Pavlovic, 1984; Humes and Riker, 1992). Other basic auditory abilities have also been discussed, particularly ones that appear essential for tasks demanding resolution of complex stimuli, notably temporal acuity (e.g., Hirsh, 1959; Green, 1971; Watson, 2004) and spectral resolving power (e.g., Feth and O'Malley, 1977; Moore and Glasberg, 1986).

Auditory sensitivity is thus regarded as the major dimension on which listeners differ, reflecting the severe consequences of variation in this ability, and the frequency with which it is diminished, particularly in old age. Loss or developmental impairment of other auditory abilities, such as temporal or spectral resolving power, has also been postulated as the cause of disorders such as delayed language development or dyslexia (e.g., Tallal *et al.*, 1993). No specific measures of these abilities have, however, achieved the canonical status of the pure-tone audiogram as the measure of sensitivity.

The identification of basic auditory abilities using a purely psychophysical methodology requires the testing of a large number of people on a large number of auditory tests. The size requirements are dictated by well-established principles relating test reliability, intertest correlations, and the expected range of performance (see, for example, Gorsuch, 1983). The set of auditory tests must be sufficiently large that it can be reasonably assumed to exceed the number of discrete auditory abilities and include multiple tests of each ability.

Application of factor analysis, or principal components analysis, is a systematic method by which to identify subsets of similar tasks that provide measures of a common underlying ability. These techniques, originally developed by Spearman (1904), have been widely employed in the search for primary mental abilities by Thurstone (1947) and many more recent investigators. A limitation of this approach is that the discrete abilities that can be identified depend on the selection of measures included in the test battery. This requirement has made it difficult to compare the results of earlier factor analytic studies of auditory abilities, because of only moderate overlap among the tests that were employed. This is illustrated by the range of tests included in the Seashore battery (Seashore *et al.*, 1939, 1960), which includes many tests of rhythm and of the perception of musical segments that are not found in any of the other major studies, but it has only a few measures of basic auditory temporal or spectral discrimination abilities. Nevertheless, there has been sufficient commonality among the tests used in some of the earlier studies to suggest the existence of a few discrete primary auditory abilities, as described in the following section.

A. Previous factor analytic studies of auditory abilities

Factor analytic studies of auditory abilities conducted prior to the mid-1980s have been reviewed in an earlier article (Johnson, Watson, and Jensen, 1987). Ten such studies used widely differing collections of auditory tests, methodologies, and numbers and types of subjects, ranging from high school sophomores (Elliott *et al.*, 1966) to prison inmates (Stankov and Horn, 1980). While the variety of tests employed makes it difficult to compare across the studies, most of the studies supported the existence of multiple auditory abilities. These included the abilities to discriminate differences in pitch, intensity, and duration, plus the higher-level ability of auditory memory. One of the earliest and most complete of these studies was conducted by Karlin (1942), a student of Thurstone. Karlin found that the ability to recognize speech was surprisingly independent of the listeners' spectral and temporal discrimination abilities.

More recently, in a remarkable survey of factor analytic studies of human cognitive abilities, Carroll (1993) devoted one of 11 ability-specific chapters to "Abilities in the Domain of Auditory Reception." Carroll observed that, "The domain of individual differences in auditory receptive abilities has received relatively little attention in the factor analytic literature." Nevertheless, he was able to find 38 published datasets that provide some basis for establishing the structure of the domain of auditory abilities. Carroll considered only one of these sets to provide a reasonable statistical account of individual differences in speech processing (Hanley, 1956). Carroll's list of candidate factors included "hearing acuity" (auditory sensitivity measured with either tones or speech), "speech sound discrimination," "spectral, temporal and intensive discrimination," "speech perception under distortion," "cognitive relations" (among tonal patterns), "musicality," and several others.

Watson *et al.* (1976) observed large ranges of performance in tonal pattern discrimination studies, spurring an interest in individual differences. In an effort to learn more about individual listeners who showed exceptionally good or poor ability to detect changes in these spectral-temporal patterns, the Test of Basic Auditory Capabilities (TBAC) was developed (Watson *et al.*, 1982a, b). This battery has since been used in a number of investigations (e.g., Christopherson and Humes, 1992; B. Watson and Miller, 1993; Drennan and Watson, 2001; Surprenant and Watson, 2001; Jakobson *et al.*, 2003). The original TBAC included three single-tone discrimination tests, three tests of temporal pattern discrimination, and two speech tests. (These are subtests 1 through 8 of the current test battery and are described below.) In the first use of the TBAC, 127 adults with normal hearing were tested in a free field. There was virtually no association between performance on the nonspeech subtests and the two speech tests, supporting Karlin's (1942) finding of a dissociation between the recognition of "social sounds" and measures of auditory acuity. Later, Surprenant and Watson (2001) conducted a replication of this study in which three additional speech tests (identification of sentences, words, and CVs) were added to the eight TBAC subtests. Analysis of the data from this extended battery showed three factors, one for speech recognition, a second for nonspeech discrimination tests with both simple and complex stimuli, and a third for temporal-order discrimination, supporting the earlier findings of independence between performance with speech and nonspeech stimuli.

Although spectral and temporal resolving power vary significantly among normal-hearing adults, the average spectral and temporal abilities of normal-hearing adults are clearly more than sufficient for the demands of speech recognition under quiet listening conditions. The lack of association between speech and nonspeech tests in these studies was not, however, due to a limited range of variation in speech recognition or in spectral-temporal acuity. Surprenant and Watson (2001) found a range of speech-to-noise ratios of about 7.0 dB for 50% correct identification of words in CID sentences (Davis and Silverman, 1970), and threshold measures for spectral and temporal discrimination differed by factors of from 3 to 10 times. Failure of measures of discrimination acuity to predict performance in the identification of nonsense syllables, words, or sentences pointed to some other ability as the source of the considerable range of speech recognition scores in noise. Another observation suggested that this hypothesized ability might be cognitive rather than auditory (more central than peripheral). Watson *et al.* (1996) found a modest correlation ($r=0.52$, $p<0.005$, $N=90$) between speech recognition by ear alone and that by eye alone (speechreading). While not a particularly strong association, this correlation is considerably larger than that between speech recognition and any single measure of spectral or temporal acuity obtained with nonspeech stimuli.

While this correlational evidence might appear to support arguments that speech recognition is "special" (i.e., explained by a speech-specific mechanism not utilized in the perception of nonspeech stimuli), the data are subject to other interpretations. In Surprenant and Watson (2001), as well as in virtually all prior factor analytic studies of auditory abilities, the speech tests differed from all of the nonspeech tests in at least two ways, in addition to the apparent differences between speech and most laboratory-generated stimuli. The first is that speech perception has been evaluated with recognition tests rather than with discrimination tests, placing quite different demands on the listeners. Discrimination tests generally reward the strategy of learning to focus on a dimension or property that is subject to change, while recognition tests encourage attention to larger patterns of spectral and temporal information. A second, more fundamental, difference is that the speech of one's native language is an extremely familiar stimulus, while the nonspeech stimuli used in these studies are novel laboratory-generated sounds. Most of a listener's knowledge about the nonspeech stimuli is gained during a relatively brief period of testing. However, listeners have extensive explicit and tacit knowledge about the structure of speech sounds, including the many constraints (linguistic and physical) that govern this class of sounds. An individual's ability to effectively use this knowledge is likely a major factor accounting for differences in speech recognition performance, especially in tests that require listeners to identify speech stimuli that have been degraded or masked by noise. Under these conditions, listeners who can use their knowledge to develop better attentional strategies (or better "guessing" strategies) might have an advantage over others with equally acute spectral and temporal resolution.

B. Expanded test battery for the present project

Despite the support provided by Karlin (1942) and Surprenant and Watson (2001) for the independence of the abilities to process speech and nonspeech sounds, there is ample evidence in the literature that temporal and spectral acuity are necessary to distinguish among the sounds of speech [e.g., the long line of research with vocoders, begun by Dudley (1939)]. Thus, persons with significantly worse than average spectral or temporal acuity might reasonably be expected to be correspondingly worse at recognizing speech. It is a common assertion among audiologists that, although amplification can return sensitivity to normal values, speech perception by the hearing impaired remains below that of nonimpaired persons because their auditory resolving power is inadequate. While this assertion may be true for persons with impaired hearing, it does not necessarily follow that the variation in resolving power among those with normal sensitivity will predict their speech recognition abilities.

An obvious problem with the earlier studies is the limited number of tests of spectral and temporal acuity, perhaps failing to include other discrimination tests that would predict speech recognition. Several spectral and temporal measures were therefore added to yield an expanded TBAC battery, including: (1) ripple-noise discrimination, a measure of spectral resolution as a function of the depth of troughs in the frequency spectrum (Yost *et al.*, 1978); (2) detection of amplitude modulation in Gaussian noise, measured at different modulation rates (a measure of temporal resolution first studied by Viemeister, 1979); and (3) detection and discrimination of temporal gaps, a test reported to be correlated with speech recognition (e.g., Glasberg and Moore, 1989; Snell *et al.*, 2002; Tyler *et al.*, 1982). The association between gap detection and speech recognition is perhaps the most widely cited finding in support of a relation between nonspeech and speech processing abilities. However, that association has been observed only with older or hearing-impaired listeners, and some studies have failed to find it even with these populations (e.g., Nelson, Nittrouer, and Norton, 1995; Strouse *et al.*, 1998). To some extent, the discrepancies may be due to differences in the details of the tests used to assess nonspeech and speech abilities (e.g., the duration and complexity of the stimulus in which a gap appears, the location of the gap, the type of speech stimuli and background noise). The present study assesses the relation between temporal processing and speech recognition in a large population of normal-hearing listeners, using several measures of temporal processing in addition to gap detection and gap discrimination, and several measures of speech recognition. The use of a large population ($N=340$) was expected to provide a wide range of performance on all measures and allows for a more definitive assessment of the relation among speech and nonspeech abilities.

An additional test was added to the battery to evaluate the hypothesis suggested earlier, that the difference between performance on speech and nonspeech tests is a consequence of differences in the familiarity of the stimuli used in these tests. For this purpose, a recognition test for familiar *nonspeech* sounds was added to the battery. The stimuli were environmental sounds produced by both animate and inanimate sources (e.g., dogs barking, doors slamming, cars starting). These sounds were presented in Gaussian noise, follow-

TABLE I. A summary of the 19 subtests of the extended version of the Test of Basic Auditory Capabilities (TBAC-E).

Subtest	Stimulus	Detect/identify	Manipulation
(1) Pitch discrimination	250-ms 75 dB SPL 1000 -Hz tone	ΔF	Frequency
(2) Single-tone intensity discrimination	250-ms 75 dB SPL 1000 -Hz tone	ΔI	Intensity
(3) Single-tone duration discrimination	100-ms 75 dB SPL 1000 -Hz tone	ΔT	Duration
(4) Pulse-train discrimination	Rhythmic sequence of six 20-ms 1000-Hz tones	ΔT	Relative duration of pauses between tones
(5) Embedded test-tone loudness	Sequence of nine (or eight) contiguous 40-ms tones (300–3000 Hz)	Presence/absence of middle tone	Duration of middle tone
(6) Temporal order for tones	Sequence of four contiguous tones (550–710 Hz)	Change in temporal order of middle two tones	Duration of middle two tones
(7) Temporal order for syllables	Sequence of four contiguous CV syllables	Change in temporal order of middle two syllables	Syllables duration
(8) Syllable identification	VC syllables in cafeteria noise	Syllable (3AFC)	Natural variation in cafeteria noise
(9–12) Sinusoidal amplitude modulation detection	500-ms noise sample with sinusoidal amplitude modulation (8, 20, 60, and 200 Hz)	Presence of AM	Modulation depth
(13) Ripple noise discrimination	500-ms noise sample with spectral "ripple"	Presence of ripple	Ripple depth
(14) Gap detection	750-ms noise sample	Silent gap in temporal center	Gap duration
(15) Gap duration discrimination	750-ms noise sample	Change in gap duration	Gap duration
(16) Nonsense syllable identification	Nonsense CVC syllable in noise	Syllable (4AFC)	S/N
(17) Word identification	One, two, and three-syllable words, in noise	Word (4AFC)	S/N
(18) Sentence identification	Meaningful sentences (4 to 10 words) in noise	Sentence (free recall)	S/N
(19) Environmental sound recognition	Familiar environmental sounds in noise	Familiar sounds (3AFC)	S/N

ing procedures similar to those used with the speech stimuli.

The new expanded test battery consists of 19 tests, as listed in Table I. Although there are certainly other tests or stimulus conditions that might have provided additional useful information, no earlier study has assessed performance on as wide a range of speech and nonspeech tests with a population of this size. The availability of SAT scores for a large portion of the subject population made it possible to assess the extent to which performance on these tests is determined by general intellectual abilities.[1]

II. METHOD

A. Subjects

The listeners were 340 adults (100 men, 240 women, ages 18–31, mean=22.4) recruited primarily from the Indiana University student population, but also from the nonstudent population in and around Bloomington. All tested within normal limits [<20 dB HL from 250 to 4000 Hz (ANSI, 1989)] on a hearing screening. All student listeners gave permission to access their academic grade point average and their scores (verbal and quantitative) on the Scholastic Aptitude Test (SAT). The listeners were paid for their participation.

B. Stimuli

Stimuli were delivered using a digital audio tape deck (Panasonic, SV3500). The stimuli were presented diotically over EAR 3A insert earphones. All stimuli were presented at 75 dB SPL, with the exception of the increased-level comparison tones in the intensity discrimination test.

This expanded version of the Test of Basic Auditory Capabilities (TBAC, Watson *et al.*, 1982a, b) includes the original eight subtests, plus 11 new subtests that were added to examine auditory abilities that may not have been addressed by the original battery. Trials in each of the subtests, except for speech identification, are structured in a modified 2AFC format in which a standard stimulus is followed by two test stimuli, one of which is different from the standard. The listeners use a computer keyboard to indicate their response selections. Trials are arranged in groups of six, and the level of difficulty is systematically increased from trial to trial, within each group, in logarithmic steps. For most tests, eight levels of difficulty are tested over 72 trials, presenting the six easiest levels in the first 36 trials, followed by an increase in difficulty of two log steps for trials 37–72. The actual numbers of easy and difficult groups of trials varies among the subtests, as described below. Standard stimuli in all subtests were presented at 75 dB SPL.

Details of the 19 subtests of the Test of Basic Auditory Capabilities, Expanded (TBAC-E)

Subtest 1: *Pitch discrimination*. The standard is a 1-kHz, 250-ms tone. The values of the frequency increments range from 2 to 256 Hz in equal logarithmic steps.

Subtest 2: *Single-tone intensity discrimination*. The standard is a 1-kHz, 250-ms tone. The increments range from 0.5 to 8.0 dB.

Subtest 3: *Single-tone duration discrimination*. The standard is a 1-kHz, 100-ms tone. Increments in duration range from 8 to 256 ms in equal logarithmic steps.

Subtest 4: *Pulse-train discrimination*. The standard stimulus consists of six 20-ms pulses of a 1-kHz tone. These pulses are arranged in three pairs, with 40 ms of silence between members of a pair and 120 ms between pairs. The temporal structure of the "different" sequence is varied by increasing the separation between members of each pair, with a corresponding decrease in the between-pair time (and, thus, a constant interval between the first tones in each of the successive pairs). Increases in within-pair separation ranged from 5- to 50-ms increments, in equal logarithmic steps. Thus, the first, third, and fifth tones are fixed in time, while the onsets of the second, fourth, and sixth tones are delayed by varying amounts. This is a minimal test of the ability to detect changes in the relative timing of events in a repeated temporal sequence (i.e., rhythm). However, it should be noted that, in principle, this test can be performed solely on the basis of changes in absolute durations.

Subtest 5: *Embedded test-tone loudness*. Subjects listen for one member of a sequence of nine tones with frequencies ranging from 300 to 3000 Hz. A different, randomly selected series of nine tones is presented on each trial. The task is to detect the presence of the fifth tone in the sequence. The tone is absent in the standard. The duration of all tones except the fifth, or target tone, is 40 ms. The test is made more difficult by reducing the duration of the target tone from 200 to 10 ms, in equal logarithmic steps. The name of this test derives from the listener's experience. Only the duration of the target tone is directly manipulated.

Subtest 6: *Temporal order for tones*. The task is to discriminate the order in which two equal-duration tones (550 and 710 Hz) are presented. The duration of the two tones is varied from 20 to 200 ms in equi-log steps. The tones are presented without a gap between them and are preceded and followed, without gaps, by 100-ms "leader" and "trailer" tones at 625 Hz.

Subtest 7: *Temporal order for syllables*. This is a speech analog to subtest 6, in which the listener is to discriminate the syllable sequence /ta/ka/ from /ka/ta/, when the two CV syllables are preceded and followed by the syllables /fa/ and /pa/. The listener's task is thus to discriminate /fa/-/ta/-/ka/-/pa/ from /fa/-/ka/-/ta/-/pa/, and subjects are told that only the middle syllables vary. The duration of the syllables was varied from 75 to 250 ms in five steps.

Subtest 8: *Syllable identification*. This is a subset of the Dubno and Levitt (1981) nonsense syllable test. Nonsense VC syllables are presented in cafeteria noise and a three-alternative forced-choice identification procedure is used. Syllables are presented in a carrier phrase (e.g., "You will mark *oothe* please"). The original test was modified by reducing the number of alternatives on each trial to the correct response plus the two most likely errors, as shown in confusion matrices provided to us by the authors.

Subtests 9–12: *Sinusoidal Amplitude Modulation (SAM) noise discrimination*. Fifty-four independent 500-ms samples of Gaussian noise are sinusoidally amplitude modulated at four different rates: 8, 20, 60, and 200 Hz. For each AM rate, eight different modulation depths are used, ranging from -18 to -32 dB for the slowest modulation rate, to -10 to -24 dB for the fastest rate. [Modulation depth is here expressed as $20\log(m)$, where m is a modulation index that ranges from 0.0 to 1.0.] The modulated and unmodulated stimuli are equated for total rms energy. The standard is always a 500-ms broadband noise with no amplitude modulation, and one of the two test stimuli is modulated. Each modulation rate is tested in separate blocks of 54 trials, in the order of increasing modulation rate. Each block consists of four 6-trial groups with the largest modulation depths, followed by five 6-trial groups with the smallest depths.

Subtest 13: *Ripple noise discrimination*. Seventy-two independent 500-ms digital samples of Gaussian noise are low-pass filtered at 3000 Hz. Sinusoidal ripples are created in the power spectrum by delaying the samples by 5 ms and adding them to the input. Before adding, the delayed signal is attenuated by an amount ranging from 0 dB (maximum ripple) to 14 dB (slight ripple), in seven 2-dB steps. The resulting eight "rippled" (i.e., spectrally modulated) noises are then high-pass filtered at 300 Hz and are equated for rms. The standard is always a 500-ms broadband noise with the same bandpass filtering as the "rippled" samples, but with a uniform power spectrum. One of the two comparison patterns is rippled. The test consists of 72 trials: six easier 6-trial groups (deep ripple: 0–10-dB attenuation) followed by six more difficult 6-trial groups (shallower ripple: 4–14-dB attenuation).

Subtest 14: *Gap detection*. Seventy-two independent 750-ms digital samples of Gaussian noise have gaps of silence of eight different durations at their temporal centers. The gap durations range from 0.5 to 64 ms, in log steps, while the total durations remained constant. The noises have 0.5-ms cosine ramps at the beginning and end of the silent intervals. The standard is always a 750-ms broadband noise that has no gap in it. One of the two comparison stimuli contains a gap. The test consists of 72 trials: six easier 6-trial groups (gaps ranging from 2 to 64 ms) followed by six more difficult 6-trial groups (gaps from 0.5 to 16 ms).

Subtest 15: *Gap-duration discrimination*. Seventy-two independent 750-ms digital samples of Gaussian noise have silent gaps of 40 ms placed at their temporal centers. Increments in gap duration range from 4 to 200 ms, in log steps. The total duration of the noise is kept constant while the gap duration increases, resulting in an increase in total duration with increases in gap duration. The noise bursts have a 0.5-ms cosine ramp at onset and offset. The standard is always a 750-ms broadband noise with a 40-ms silent gap at its temporal center. One of the comparison patterns contains a larger gap (with total noise duration remaining constant), while the other is identical to the standard. The test consists of 72 trials: six easier 6-trial groups (gap increments ranging from 12 to 200 ms) followed by six more difficult 6-trial groups (gap increments from 4 to 65 ms).

Subtests 16–18: *Speech tests*. Equal speech and noise levels for 0-dB speech-to-noise ratios were set by digitally recording at a peak VU-meter reading equal to that of the Gaussian noise samples. Various speech-to-noise ratios were achieved by attenuating the speech. Presentation levels were determined by setting a 1-kHz calibration tone, recorded with the same rms value as the Gaussian noise, to 75 dB SPL. Talkers were two male and two female students in the Department of Theatre and Drama at Indiana University, all of whom were native speakers of Standard American English (Midwestern dialect). Responses for all speech tests are recorded either on paper response sheets or by using a computer keyboard and monitor. In either case, the response alternatives are presented on each trial for the nonsense-syllable and word tests, while the sentence test is given in open-set format.

For all speech tests, the speech-to-noise (S/N) ratios were adjusted after the first 146 subjects had been tested and estimates of psychometric functions were examined. (Original and adjusted values are provided for each test below.) The new S/N values were selected to minimize testing at values that produce near-chance or near-perfect performance and to provide equal increments in percent correct between each step. New S/N values were interpolated from the psychometric functions, for values of percent correct from 40% (15 percentage points above chance) to 90% in steps of 12.5 percentage points. For the sentence identification test (which used an open-ended response protocol), S/N values were interpolated for percent-correct values from 20% to 90% in steps of 17.5 percentage points. The resulting S/N values for each subtest are given below.

Subtest 16: *Nonsense syllable identification*. A set of 100 nonsense CVC syllables uses all of the standard English consonants and vowels in random combinations. The stimuli are mixed with broadband Gaussian noise at five speech-to-noise ratios (initial values: −3, −7, −11, −15, −19 dB; adjusted values: −0.4, −5.1, −8.4, −11.5, −15.4 dB). A four-alternative forced-choice procedure is used. Three foils (incorrect CVCs) for each of the stimuli were generated by changing either the initial consonant, the vowel, or the final consonant (one for each foil). The response options are presented, and listeners are asked to indicate the sound that was presented on each trial. Listeners are given 3 s to indicate their response. The set of 100 stimuli is presented twice in separate blocks of trials using two different random orders. Stimuli are presented with decreasing S/N values within each group of five trials.

Subtest 17: *Word identification*. The set of 100 stimuli consisted of 60 one-syllable, 25 two-syllable, and 15 three-syllable words. The stimuli were mixed with broadband Gaussian noise at five speech-to-noise ratios (initial values: −3, −7, −11, −15, −19 dB; adjusted values: −2.9, −7.8, −11.1, −14.4, −18.3 dB). The method was the same as that for nonsense syllables, except for a slightly different strategy for generating foils. Three foils were generated by modifying the beginning, middle, or ending portion of each word (one for each foil). To accommodate the constraints of English words with different numbers of syllables, it was necessary for the modified portion to range from single consonants or vowels to full syllables.

Subtest 18: *Sentence identification*. A set of 40 sentences was composed for this test. The number of words per sentence ranged from four to ten. All sentences consisted of familiar words and concepts. Sentences were highly predictable so that listeners' abilities to use context would be evaluated by this test. The stimuli were mixed with broadband Gaussian noise at five speech-to-noise ratios (initial values: −4, −6, −8, −10, −12 dB; adjusted values: −6.3, −7.9, −9.0, −10.1, −11.7 dB). Listeners were allowed 6 s to write the words they could identify, in order, on a response sheet. All 40 sentences were presented twice, using two random orders in two blocks of trials. As with the other speech tests, stimuli were presented with decreasing S/N values within each group of five trials. The total number of correctly identified words (regardless of order) was recorded for each sentence.

Subtest 19: *Environmental sound identification*. Twenty-five digitally recorded sounds were selected from two high-quality sound effects collections (Hollywood Edge and Sound FX The General). All were 16-bit stereo sounds sampled at 44.1 kHz. The sounds were converted to monaural format and presented diotically. The particular sounds were selected from among several hundred options, based on criteria of familiarity, detectability, and pairwise discriminability. These 25 sounds and several others from the collections cited above have been extensively studied by our research group (Gygi, 2001; Gygi, Kidd, and Watson, 2004, in press). The sounds were all perfectly recognizable at high event-to-noise ratios (e.g., in excess of +20 dB). To simulate difficult listening conditions, samples were prepared by mixing each sound with Gaussian noise at eight different event-to-noise ratios (Ev/N). Equal sound and noise levels for 0 dB Ev/N were achieved by equating the rms values. The level of the environmental sound was adjusted to create other Ev/N values. The levels were selected so that the probability of identification would range from near chance to near perfect, which was achieved in seven 3-dB steps. The range for each event is centered on the Ev/N level corresponding to an average probability of detection of 0.66, as determined in a preliminary experiment with a group of four normal-hearing college students. Minimum Ev/Ns ranged from −31.6 to −16.9 across the set of sounds. A 3AFC format is employed, in which one sample is played to the listeners who then select the correct response from among three alternatives. The two incorrect alternatives on each trial are those sounds from the same 25-stimulus catalog that were determined in pilot studies to be most likely to be confused with the sound actually presented. Trials are presented in groups of six with increasing Ev/N levels within each group. The test consists of two 150-trial blocks, each with a different a random sequence of the 25 sounds. In the first block, the sounds are presented at the six highest Ev/N levels. In the second block, the sounds are presented at the six lowest Ev/N ratios.

C. Procedures

Participants were tested in a sound-treated room in groups of up to 12 listeners per session. Testing of each participant was conducted over four 90-min sessions on consecutive weekdays (one session per day). The sequences of tests and of stimuli within each test were the same for all participants. The rationale is that, in an individual differences study, the goal is to eliminate variance due to any factors other than differences among the subjects. Subtests 1–8 were

TABLE II. Split-half reliability coefficients for the 19 TBAC subtests.

TBAC-I Subtests	Reliability	Added subtests	Reliability
Embedded tone	0.723	SAM 8 Hz	0.786
Pitch	0.819	SAM 20 Hz	0.815
Pulse train	0.816	SAM 60 Hz	0.702
Duration	0.755	SAM 200 Hz	0.717
Loudness	0.878	Ripple	0.759
Temporal order	0.807	Gap discrim.	0.560
Syllable ID	0.501	Gap detect.	0.612
Syllable sequence	0.766	Nonwords	0.787
		Words	0.635
		Sentences	0.795
		Env. sounds	0.827

administered in the first session, subtests 9–15 in the second, subtests 16–18 in the third, and subtest 19 in the final session. All tests began with detailed recorded instructions, which included two practice trials. Subjects were encouraged to make their “best guess” on each trial, regardless of their confidence level.

III. RESULTS

A. Reliability of TBAC measures

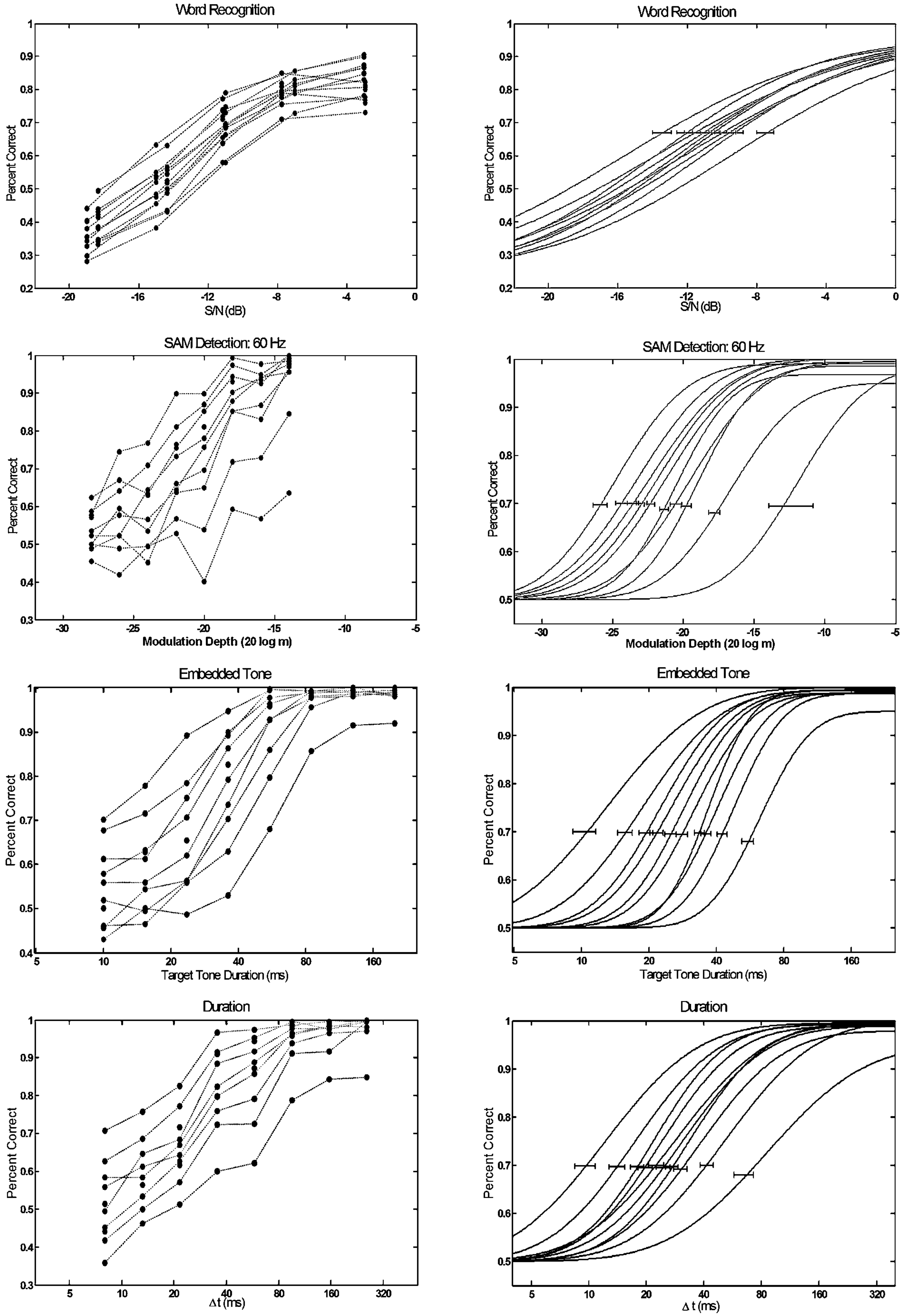
Christopherson and Humes (1992) examined the reliability of the original eight TBAC subtests. The tests were administered multiple times to the same listeners and were found to be reliable; Cronbach’s alpha values were above 0.7 for all but the Syllable Identification subtest, which had a value of 0.58. Performance on all subtests changed little over six repeated administrations of the TBAC. The new TBAC tests employed here were constructed following the same principles as the earlier tests and utilized stimuli and tasks that have been used extensively in previous research. The reliability of all subtests was estimated as part of this research project, using a split-half procedure, rather than the repeated-test procedure used by Christopherson and Humes. Following a resampling strategy (see Good, 2006), correlations for 1000 randomly selected pairs of split halves were computed for each subtest, and reliability was computed using the Spearman-Brown prediction formula applied to the mean correlation. The results are shown in Table II. Consistent with the findings of Christopherson and Humes, reliability coefficients for all of the first eight subtests were above 0.7, except for the Syllable Identification subtest. Reliability for the newer subtests were in the same range, with only three of the subtests falling below 0.7 (with the lowest coefficient at 0.56). Thus, reliability of most subtests is fairly high, and the less reliable subtests have more than adequate reliability for the present purposes. As with any test, performance on these tests may be influenced by factors other than the abilities they were designed to measure (such as general intelligence or motivation). However, these tests appear sufficiently stable and reliable for use in estimating the associations among them. The question of the nature and number of the abilities that account for performance on these tests is addressed in Sec. III C.

B. Range of abilities

The range of performance on each of the 19 subtests for 338 subjects is summarized in Table III. Two subjects were excluded from the analysis because of extremely low performance on one or more subtests, suggesting either misunderstanding of the instructions or lack of motivation. The entries are estimated threshold values for the group of subjects in each decile, with decile assignment based on subjects’ overall percent correct scores for each subtest. Thresholds are based on estimates from psychometric functions fitted to data for all subjects within a decile, as shown in Fig. 1. This method was chosen in preference to fitting the psychometric functions to data from individual subjects because of the

TABLE III. Threshold (70% correct) estimates for each decile for the 19 subtests of the TBAC-E.

Subtest	1	2	3	4	5	6	7	8	9	10
(1) Pitch (ΔF, Hz)	36.27	15.45	13.99	12.55	10.96	8.78	7.88	6.59	4.95	3.08
(2) Intensity (ΔI, dB)	2.37	1.16	0.84	0.70	0.65	0.55	0.48	0.45	0.34	0.29
(3) Duration (ΔT, ms)	67.20	41.34	30.58	27.47	24.71	22.54	19.67	18.04	14.11	9.68
(4) Pulse train (ΔT, ms)	22.83	14.92	11.53	10.38	9.69	7.94	7.26	6.21	4.93	3.22
(5) Embeded tone (T, ms)	55.95	42.64	35.89	33.54	28.16	25.15	21.73	19.49	15.75	10.63
(6) Temporal order tones (T, ms)	119.97	77.73	64.74	56.97	51.22	45.99	41.65	35.73	28.13	21.96
(7) Temporal order syllables (T, ms)	256.11	159.45	135.64	114.15	105.42	100.10	86.94	87.59	80.89	73.70
(8) Syllable ID: VC $[P(c)]$	0.61	0.69	0.71	0.73	0.74	0.75	0.77	0.78	0.80	0.84
(9) SAM 8 Hz (mod. depth, dB)	−16.34	−20.52	−22.60	−23.48	−25.02	−25.90	−26.48	−27.31	−28.25	−29.73
(10) SAM 20 Hz (dB)	−14.05	−16.52	−20.84	−22.60	−23.94	−24.53	−26.08	−26.81	−27.61	−30.87
(11) SAM 60 Hz (dB)	−12.95	−17.70	−19.82	−20.60	−21.38	−22.38	−22.94	−23.55	−24.37	−25.88
(12) SAM 200 Hz (dB)	−7.08	−14.30	−16.32	−16.97	−17.34	−17.94	−18.52	−19.13	−20.12	−22.04
(13) Ripple noise (dB)	0.64	−1.75	−3.58	−4.70	−5.06	−5.90	−7.43	−8.08	−8.65	−11.00
(14) Gap detection (T, ms)	5.10	3.06	2.59	2.34	2.00	1.83	1.54	1.40	1.09	0.71
(15) Gap discrimination (ΔT, ms)	67.68	48.91	46.56	41.73	37.04	31.39	30.93	28.24	24.36	14.61
(16) Syllable ID (CVC) (S/N)	−4.49	−5.77	−6.53	−7.08	−7.62	−8.03	−8.45	−8.82	−9.41	−9.35
(17) Word ID (S/N)	−7.37	−9.23	−9.56	−9.96	−10.46	−10.73	−11.04	−11.54	−11.99	−13.39
(18) Sentence ID (S/N)	−7.07	−7.70	−7.88	−8.05	−8.25	−8.34	−8.39	−8.57	−8.84	−9.22
(19) Environmental sounds ID (S/N)	−10.73	−11.98	−12.42	−12.91	−12.94	−13.15	−13.53	−13.72	−14.17	−14.92



Word Recognition

Word Recognition

SAM Detection: 60 Hz

SAM Detection: 60 Hz

Embedded Tone

Embedded Tone

Duration

Duration

FIG. 1. Performance of each decile group on 18 of the 19 TBAC-E subtests. (The required data were not available for the Syllable Identification subtest.) Both percent-correct data and fitted functions are shown for the first four subtests. Panels on the left show percent correct at each stimulus level. Panels on the right show fitted functions and 90% confidence intervals for the threshold estimates derived from the fitted functions. Fitted functions and confidence intervals are shown for the remaining subtests.

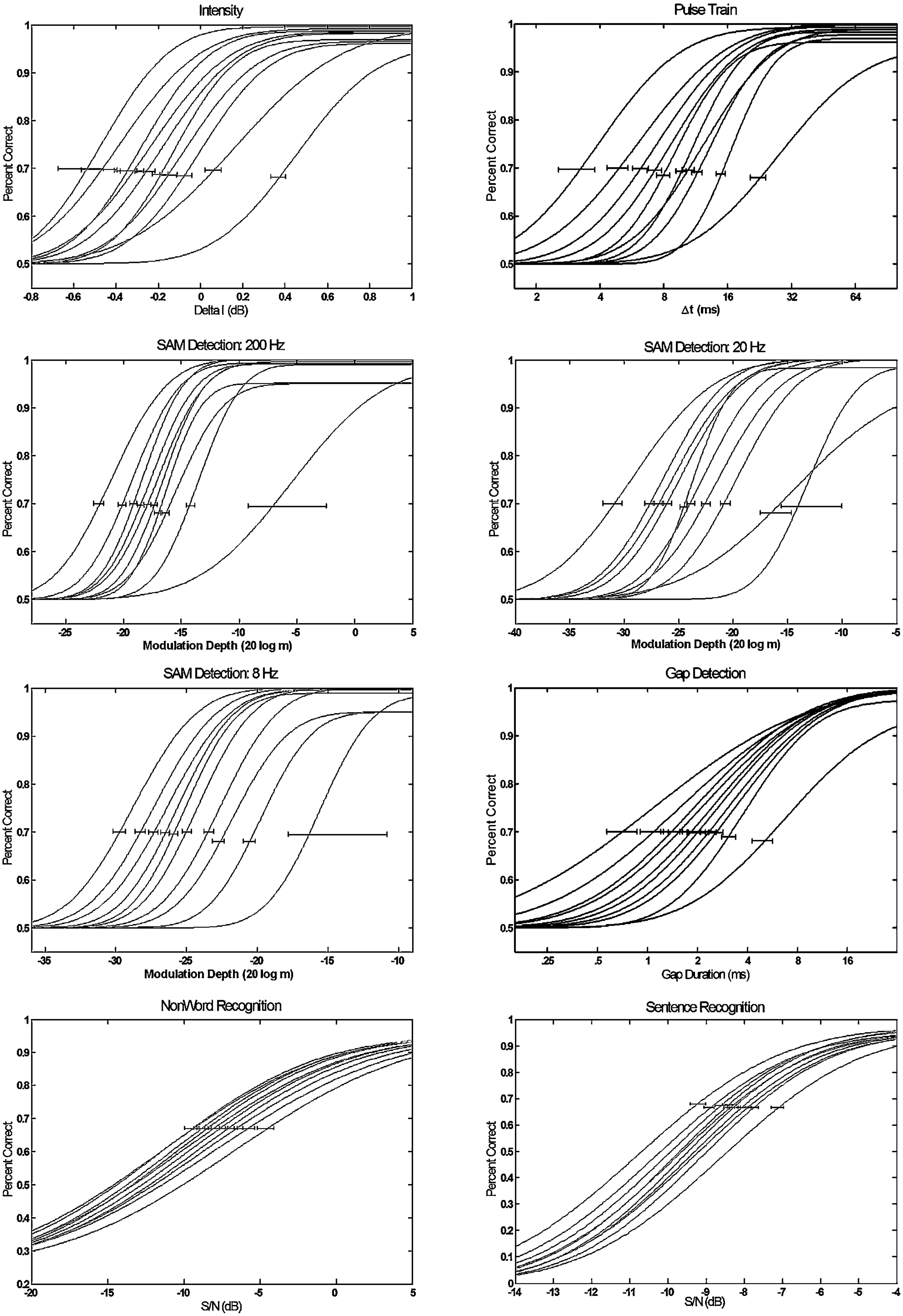


FIG. 1. (*Continued*).

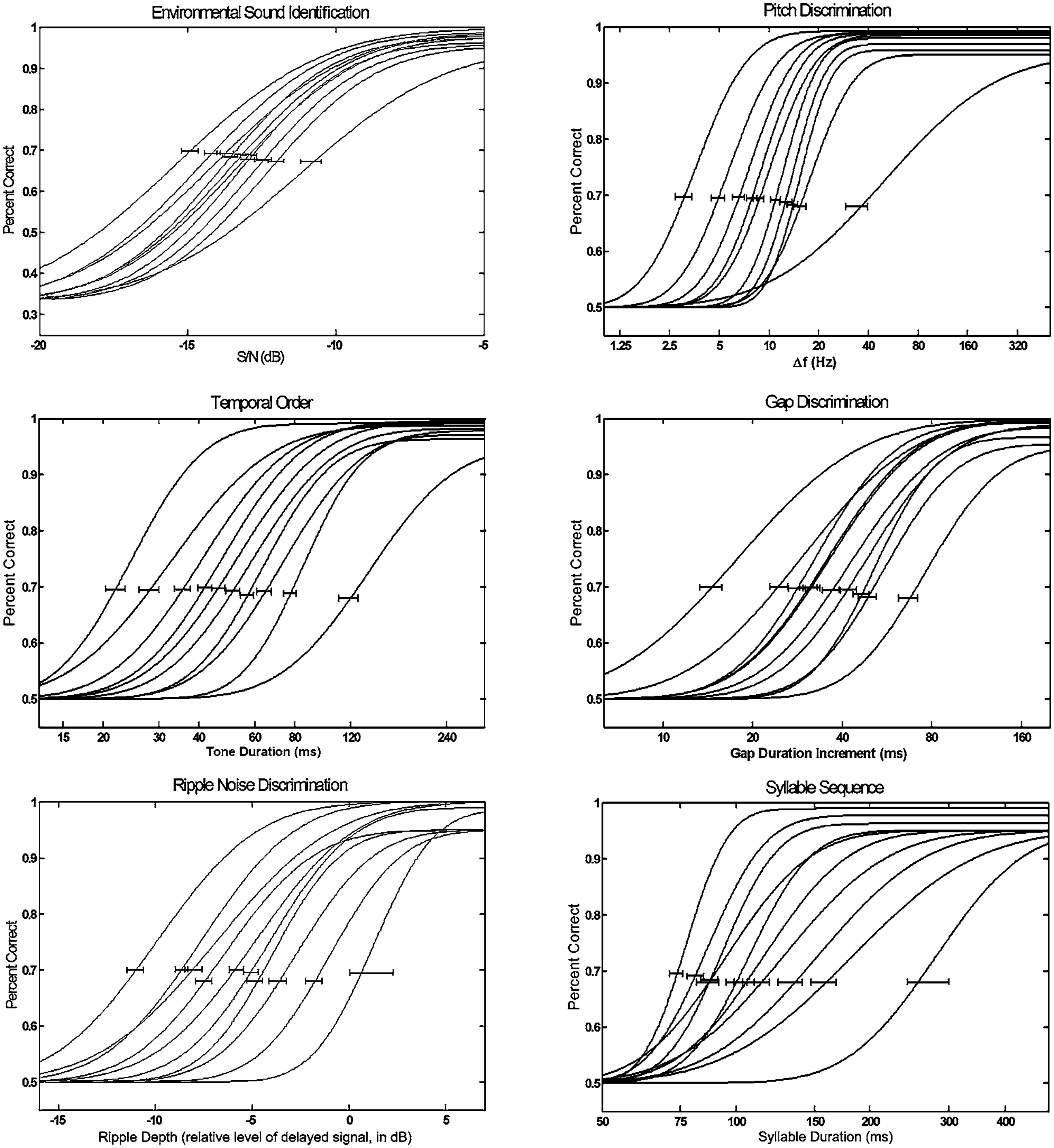


FIG. 1. (*Continued*).

limited number of trials per subject. Thus, the range of threshold values in Table III does not represent the full range of thresholds for individual listeners. The underlying distributions were approximately normal, with significant deviations from normality occurring primarily at the low end of the distribution. Because equal numbers of subjects were included in each decile (by definition), deciles near the tails of the distribution include a greater range of performance than those near the middle. Although exceptional performance was observed at both ends of the distributions, all distributions were unimodal, with no evidence of separate populations of listeners with exceptionally low or high acuity on any measure.

Thresholds were estimated by fitting a sigmoid function (cumulative Gaussian) to the data for each decile using a maximum-likelihood method developed by Wichmann and Hill (2001a, b). This method uses a parametric bootstrap technique (a Monte Carlo resampling method with maximum-likelihood parameter estimates) that allows for the estimation of confidence intervals for threshold estimates. In this application, the lower limit of the function was constrained to be at chance (which varied with the task), and lambda (a lapse parameter that sets the upper limit of performance) was constrained to be within the range of 0.0 to 0.5. The threshold was defined as the stimulus level corresponding to 70% correct (for cases in which lambda=0.0) or slightly less for nonzero values of lambda. This represents a type of correction for inattention in that the actual percent-correct value at which the threshold is estimated is a constant proportional distance between chance and the upper limit of the fitted function (defined by lambda).

Perhaps the most remarkable aspect of the data in Table III is the broad range of performance on many of the tests. The ranges of threshold values are close to those reported by Surprenant and Watson (2001) for the first seven TBAC subtests, with substantial differences occurring for only two of the subtests. The best performance on the Embedded Tone subtest (i.e., in the highest deciles) was considerably better than in the earlier study, as was the worst performance (first decile) on the Duration subtest. In general, performance of the top 10% on each subtest approaches that found with highly trained listeners using more traditional psychophysical procedures, while the worst 10% performed quite poorly. For example, the top group was able to detect a 0.3% change in frequency (or 5.3 cents), while the bottom group required a 3.6% change in frequency (or 61.7 cents). Thresholds in terms of duration, or of changes in duration, in temporal tasks increased from 3.5 to 7.0 times across the deciles, and thresholds for modulation depth and ripple depth ranged from 12 to 16 dB. The range of threshold S/N values for recognition of familiar sounds (speech and environmental sounds) may seem small in comparison (2 to 6 dB) to those of the unfamiliar laboratory test sounds. But, it should be considered that because of the steep psychometric functions for these stimuli, relatively small changes in S/N translate into large changes in percent-correct identification. For example, at a S/N of −10 dB, the worst subjects could understand only about 18% of the words in sentences, while the best subjects could understand about 52%. Others (Surprenant and Watson, 2001; Bronkhorst and Plomp, 1992) have found a larger range for sentence recognition thresholds of roughly 6 to 7 dB. This represents a substantial range in the ability of normal-hearing listeners to understand speech in noise.

Fitted psychometric functions for each decile are shown in Fig. 1 for 18 TBAC subtests.[2] [No slope or threshold estimates could be derived for the Syllable Identification subtest (subtest 8) because of the stimulus values used.] Raw data and fitted functions are shown in separate plots for four of the subtests.[3] These four subtests were chosen because they are representative of the different patterns of results in the full set of tests and because they represent each of the four factors identified in the factor analysis described below. The 90% confidence intervals indicated for each of the threshold estimates show consistent accuracy, with the exception of the first decile for the 8-Hz SAM subtest. The reduced accuracy results from the restricted range of above-chance performance for this decile, suggesting that even the easiest stimulus levels were too challenging for the worst listeners. This was an issue for all SAM subtests, the Ripple subtest, and the Syllable Identification subtest (subtest 8). Despite the reduced accuracy, these are valid threshold estimates. They indicate a low level of performance at the same stimulus levels that result in near-perfect performance for most listeners. Such a large difference in performance at these levels may indicate that, for these tasks, the worst subjects are not simply less sensitive to the to-be-detected stimulus dimension; they may be using a different listening strategy that is inefficient or inappropriate for the task.

The families of psychometric functions are quite orderly, with thresholds decreasing consistently across the deciles. The largest change in threshold typically occurs between the first and second deciles. Slopes of functions for adjacent deciles are generally similar across the deciles, and any changes in slope tend to be gradual. For 13 of the subtests, slopes tended to increase slightly with increasing thresholds, while the opposite was true for the remaining cases. Thus, there was no consistent tendency for slopes to change systematically with the ability of the listeners.

C. Identification of basic auditory abilities

The considerable variance on most of the subtests suggests a broad range of abilities. It also provides an opportunity to examine the covariance as a means of identifying the number of distinct abilities that underlie performance on this collection of tests. As a first step in an attempt to identify these underlying abilities, a principal components analysis was conducted using arcsine-transformed overall percent-correct scores for each of the 338 subjects and all 19 subtests. This analysis produced four orthogonal factors with eigenvalues greater than 1.0, which together accounted for approximately 50% of the total variance. Attempts to extract additional factors produced uninterpretable factors with eigenvalues less than 1.0. The factor loadings (varimax rotated) for each of the subtests are shown in Fig. 2.[4] In this figure, the tests are grouped according to the factor on which they have the highest loading. Within each factor, the tests are ordered according to the strength of the loading on that factor. The cross loadings of each test with the other three factors are also shown. Examination of the cross loadings reveals that, within each factor, there are some tests with high primary loadings and low cross loadings, while other tests have one or more relatively high cross loadings. Thus, some tests appear to be close to pure measures of each of the underlying factors, while others are influenced by more than one underlying factor or ability.

The factor names were chosen to reflect the tests with the highest loadings on each factor and also the variety of tests that have their primary loading on that factor. The four factors can be interpreted as four independent abilities that underlie performance on this collection of auditory tasks. Although these will be referred to as auditory abilities, this is not meant to imply that these abilities are exclusively in the auditory domain. The same abilities may also be important for the performance of tests with stimuli in other modalities, especially when the test requires the discrimination of temporal structure or the identification of stimuli presented in noise. The nature of the four factors or abilities is discussed in more detail below.

1. Loudness-duration factor

The strongest loadings for intensity and duration discrimination on this factor suggest a basis in the sensitivity to

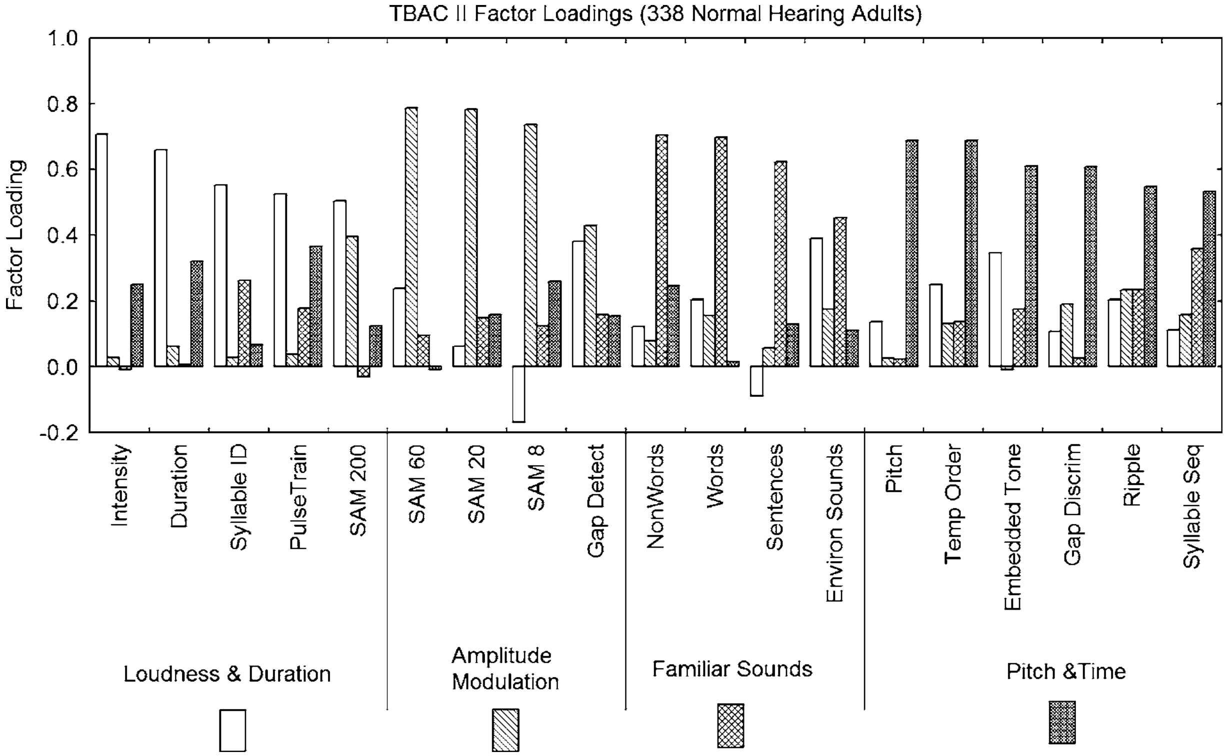


FIG. 2. Factor loadings for each of the 19 TBAC-E subtests. Subtests are grouped according to the factor on which they have the highest loading and ordered within each factor according to the magnitude of the primary loading on that factor.

changes in overall energy as an explanation of the commonality of these subtests. The remaining three subtests that have their highest loadings on this factor have their highest cross loading on the factor to which they seem logically to belong: Syllable ID on the Familiar Sound Factor, Pulse Train on the Pitch-Time Factor, and SAM 200 on the Amplitude Modulation Factor. However, the Pulse Train subtest does require subjects to detect changes in duration, and the 200-Hz SAM test involves rapid (i.e., short-duration) changes in intensity. Note that the duration of the intensity dips for the 200-Hz SAM test are comparable to threshold gap durations in the Gap Detection subtest, which has similar loadings on the Loudness-Duration and Amplitude Modulation factors. (A reviewer suggested that the 200-Hz SAM stimuli might have stronger pitch strength when deeply modulated, which might in turn be perceived as greater loudness, which seems a plausible argument.) The high cross loadings indicate that these tests require a second ability, almost to the same degree as the primary Loudness-Duration ability. The role of this ability in the Syllable Identification test ("you will mark *ooze*, please") is less obvious than in the other tests with high loadings on this factor, but the loadings also suggest that the use of duration and intensity cues may be especially important in discriminating among the alternatives when nonsense VCs are presented in this context (cafeteria noise).

2. Amplitude modulation factor

This factor includes the SAM noise discrimination tests with modulation rates of 8, 20, and 60 Hz, as well as the gap detection subtest. The modulation rates of subtests in this factor are considerably slower than the 200-Hz rate for the SAM test that loaded most strongly on the Loudness-Duration Factor. This is consistent with the change in perceptual quality as the modulation rate increases. At the slower rates, the change in amplitude has a time-varying or "fluttering" quality, while at the fastest rate, the modulation is heard as an overall "roughness" or "buzzing" quality. Inclusion of the gap detection test in this factor seems understandable as it is yet another measure of the ability to follow brief changes in level. As with the 200-Hz SAM test, the amplitude fluctuation is brief (roughly 1 to 5 ms), and both tests have relatively high (and similar) loadings on the Loudness-Duration and Amplitude Modulation Factors. It appears that both tests utilize the two underlying abilities (envelope following and duration/intensity discrimination) in roughly equal amounts.

3. Familiar sounds factor

In previous studies with the TBAC, this factor has been designated a speech factor (e.g., Surprenant and Watson, 2001). However, earlier studies did not include any familiar sounds other than speech. While the strongest loading for the environmental sounds test is on this factor, unlike the nonsense-word, word, and sentence tests, it has a relatively strong cross loading on another factor (Loudness-Duration). This suggests the existence of a familiar sound recognition ability that is not specific to speech. Implications of a single factor with high loadings for tests of both speech and other familiar sounds are discussed in a later section.

4. Pitch and time factor

The six subtests that load most strongly on this factor require two types of pattern processing: spectral (the pitch and ripple-noise discrimination tests) and spectral-temporal (temporal order for tones, embedded test tone, gap discrimination, and syllable sequence). The syllable sequence subtest might have been expected to have its highest loading on the familiar sounds factor. However, it is perceptually quite different from the other speech tests in that it consists of sequences of natural speech tokens (/fa/ /ta/ /ka/ /pa/) that are made difficult to discriminate by deleting multiple cycles of the fundamental pitch from the vowels. As the components become brief, they lose their similarity to natural speech, and the test becomes similar to the nonspeech temporal-order test in which temporal-order judgments are based on changes in holistic pattern properties rather than an identification of the order of the components. Because all of the spectral-temporal subtests in this factor consist of brief patterns that are made more difficult by shortening durations that change the overall sound of the patterns, it may be that these tests are performed by attending to holistic properties that affect the overall sound quality. This interpretation is consistent with earlier reports indicating that subjects can *detect* changes in the temporal order of brief tones based on holistic properties, even when they are unable to identify the order of the tones (Hirsh, 1959; Watson, 2004).

Although the four factors described above do not capture the wide range of auditory abilities included in Carroll's (1993) analysis (primarily because of the narrower range of auditory tests included), the variety of tests within this narrower domain provides considerable information about the nature of the underlying abilities. The types of tests and stimuli were similar to those included in two of Carroll's factors: (1) Spectral, Temporal, and Intensive Discrimination of Nonspeech Sounds and (2) Speech Recognition under Distortion. Three of the present factors can be seen as a finer differentiation of the first factor, suggesting different abilities underlying different types of temporal and spectral tests. The Familiar Sounds factor supports Karlin's (1942) finding of a separate factor for understanding speech under difficult listening conditions (Carroll's "Speech Recognition under Distortion"). The inclusion of an environmental sound test in the present study and its major loading on the factor that otherwise contains only speech tests suggests that this factor may be related to a more general ability.

5. General auditory and intellectual abilities

Before considering the nature of these four factors and the underlying abilities they may represent in more detail, the possible role of general auditory or intellectual abilities in the performance of these tests should be considered. The influence of a general ability, and the degree to which this ability may be auditory or intellectual, will be evaluated through comparison of different models of the latent variable structure that underlies the measured variables, using structural equation modeling.[5] The role of intellectual abilities will be evaluated by examining the correlations between SAT and latent variable scores.

The collection of models evaluated for this analysis is illustrated in Fig. 3. The simplest, model 1, has a single latent variable and represents the hypothesis that a single general ability underlies all measured variables. Model 2 is based on the principal components analysis shown in Fig. 2. This model includes four latent variables (corresponding to the four factors), each associated with a different set of measured variables. Each measured variable is associated with the latent variable that corresponds to the factor on which it had the highest loading in the principal components analysis. For simplicity, no cross loadings from the principal components analysis are represented in this model. The third model is a combination of the first two, in which there are four independent specific auditory factors and a general factor associated with all of the measured variables (and independent of the four specific factors). Model 4 is a version of the second model in which the four auditory factors are allowed to correlate with each other. (Model 1 can also be seen as a special case of model 4, in which all correlations among the factors are set to unity.) Finally, model 5 is the most general model in that it includes all of the components of the other four models.

A summary of the goodness of fit for all five models is given in Table IV. Models 1 and 2 are not particularly good fits to the data. Root-mean-square-error of approximation (RMSEA) values less than 0.05 and comparative-fit index (CFI) values of 0.95 or higher are generally taken as evidence of a good fit (see Bollen, 1989; Raykov and Marcoulides, 2000). Although a four-factor solution was considerably better than a single-factor solution in the principal components analysis, there is no advantage for the simplified four-factor model over a single-factor model in this context. This is largely due to the absence of cross loadings in this model and to a different standard in evaluating the goodness of fit. Also, recall that the four-factor principal components solution was far from a perfect fit, accounting for roughly half of the variance.

Models 3 and 4 both provide fairly close fits to the data. However, the model with a general factor provides a significantly better fit to the data than the correlated-factors model ($\chi^2=64$, $df=13$, $p<0.001$). This provides further support for the existence of four independent factors, but it also indicates the existence of a general factor accounting for additional variance. Model 5, which allows correlations among the four auditory factors, results in a small, but statistically significant ($\chi^2=19$, $df=6$, $p<0.01$), improvement in the fit measures. However, given the similarity of the fit measures (and the power of the significance test with such a large N), there is no compelling argument for the more complex and difficult-to-interpret model 5.

The standardized factor loadings for model 3 are shown in Fig. 4. All loadings for associations with the general factor are significant. Nearly all loadings for associations between measured variables and the four specific factors are also significant, the exceptions being those for SAM at 200 Hz, Gap Discrimination, Ripple Noise Discrimination, and Syllable Sequence. The loadings for the four specific auditory factors are generally consistent with those from the principal components analysis, supporting the interpretation given earlier.

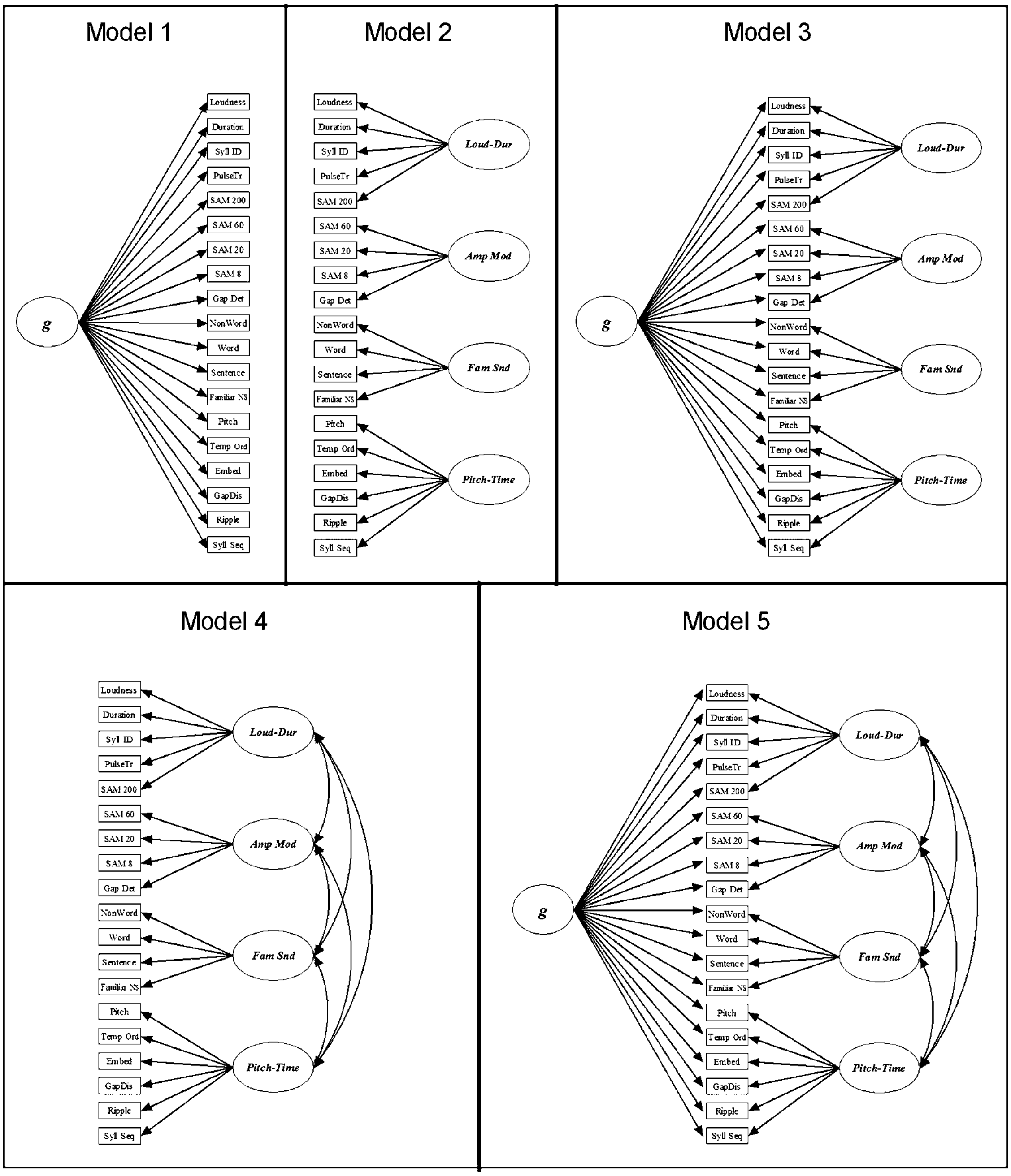


FIG. 3. Five models of auditory abilities. Model 1 represents a single general auditory ability accounting for performance on all tasks. Model 2 represents four independent auditory abilities associated with specific subsets of tasks (based on the principal components analysis) and no general ability. Model 3 includes a general ability and four independent abilities. Model 4 is a version of model 2 that allows correlations among the four auditory abilities. Model 5 is a version of model 3 that allows correlations among the four auditory abilities.

The loadings for the general factor are greater than those for the specific factors for 11 of the 19 subtests. This is most pronounced for the subtests of the Pitch and Time factor (with the exception of the Pitch subtest) and, to a lesser extent, for subtests of the Loudness-Duration factor (with the exception of the Loudness subtest). Loadings for the specific factors are generally greater than or equal to those for the general factor for subtests in the Amplitude Modulation factor and the Familiar Sounds factor, although the general factor plays a greater role in Gap Detection and Environmental Sound Identification.

Although all measured variables are based on tests designed to assess auditory abilities, it is important to determine the extent to which nonauditory intellectual abilities contribute to performance. The extent to which any of the latent variables in model 3 reflect intellectual abilities was

TABLE IV. Goodness-of-fit statistics for five models of auditory abilities.

Model	*df*	$\chi2$	RMSEA	CFI
(1) One general factor	152	556	0.089	0.907
(2) Four auditory factors	152	614	0.095	0.887
(3) One general and four independent factors	133	189	0.035	0.983
(4) Four correlated factors	146	253	0.047	0.969
(5) One general and four correlated factors	127	170	0.032	0.986

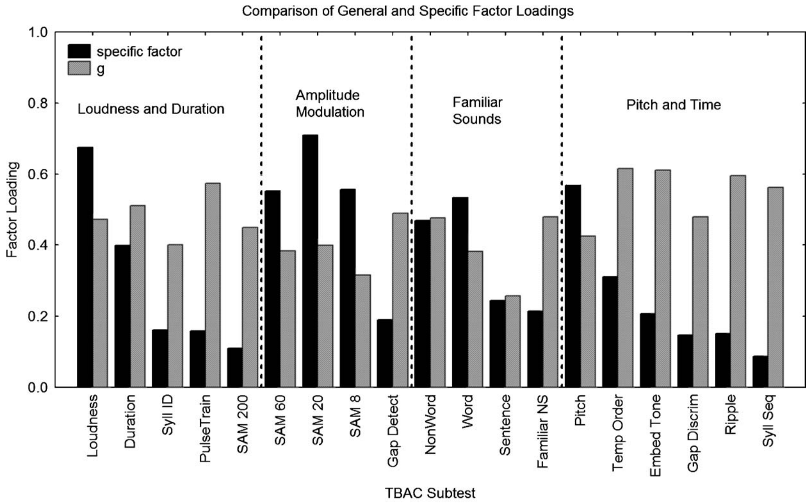


FIG. 4. Factor loadings for the latent variables in model 3. This shows the degree to which performance on each subtest is influenced by the general auditory ability (gray bars) and the specific auditory ability (black bars) with which it is associated in the model. Each section of the figure shows the loadings for a different specific factor.

evaluated by examining the correlations between the latent variable scores and SAT scores for each of the latent variables.

Math and verbal SAT scores were available for approximately 70% of the subjects (235 of 338 subjects).[6] The correlations between these scores and the latent variable scores from model 3 are shown in Table V. All correlations with latent variables are relatively low, indicating that variation in general intellectual abilities within this population accounts for an extremely small portion of the variance in performance. Some correlations are significant, however, with the highest occurring between SAT scores and the latent variable scores associated with the "general" ability. The data also indicate that verbal abilities play a minor role in familiar sound recognition and that both math and verbal abilities have a small influence on performance with the Pitch/Time tests. Overall, the magnitude of the correlations indicates that the abilities associated with these latent variables are closely associated with the ability to discriminate or identify different types of auditory stimuli, rather than with more general intellectual abilities associated with nonauditory or amodal aspects of the tests. However, the possible involvement of some other type of general nonauditory ability (e.g., temporal processing, attention, memory), independent of intellectual abilities but not specific to auditory stimuli, cannot be ruled out.

TABLE V. Correlations between latent variable scores and SAT scores (verbal and quantitative).

	Loud dur	Amp mod	Fam snd	Pitch time	*g*
SAT-V	–0.02	0.03	0.18[a]	0.14[b]	0.19[a]
SAT-M	–0.03	0.03	0.08	0.15[b]	0.22[c]

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

IV. DISCUSSION

A. Familiar sound recognition ability

The data suggest the existence of an FSR ability, but additional evidence will be required to more precisely specify the nature of that ability. This is closely related to an ability identified by Carroll (1993), based largely on the work of Stankov and Horn (1980), which was labeled "Speech Perception Under Distortion." The factor loadings suggest that the perceptions of familiar speech and nonspeech sounds rely on the FSR ability to different degrees. However, there is clearly a common ability that partially accounts for the recognition of both types of familiar sounds under difficult listening conditions. This may be an ability to make use of limited information that does not fully specify the stimulus event. Such an ability might be based on one or more of the following:

1. Efficiency in the location or activation of stored information (memory) about familiar sounds. Rapid access to items in the lexicon and in an equivalent storage system for environmental sounds could facilitate the consideration of alternatives in a recognition task.
2. A problem-solving or guessing strategy that effectively fills in the missing pieces of an incomplete signal (i.e., an ability to identify the "whole" based on fragments). This might reflect a more effective use of stimulus knowledge to generate hypotheses and rule out alternatives with regard to the missing information. Evidence of such an ability in both the auditory and visual modalities was reported by Watson *et al.* (1996).
3. Effective dynamic focusing of attention on the most informative spectral-temporal locations within a familiar sound. The locations may be those where distinguishing information is present or where the background noise is least disruptive. Such attentional strategies would require

the use of information about the properties of the familiar sounds to be identified and of the background or interfering sounds.

These three skills are not mutually exclusive, and the FSR ability may reflect a combination of them. That is, efficient access to stored information about familiar sounds can facilitate guessing or stimulus-completion strategies or help allocate attention. It should also be noted that these abilities are not necessarily purely auditory. The memory retrieval, stimulus-completion, or attentional strategies that make up the FSR ability may facilitate the recognition of any familiar stimulus (or perhaps any familiar temporal pattern) regardless of its sensory modality. That FSR is independent of the other abilities identified in this study does not mean that spectral and temporal resolving power are unimportant for the recognition of speech and other familiar sounds. The findings indicate only that individual differences in spectral and temporal resolving power among young normal-hearing listeners do not account for individual differences in the ability to recognize speech or environmental sounds under difficult listening conditions. A more severe degradation of spectral acuity can certainly impair the ability to understand speech, as is evident in the effects of processing speech through a vocoder with only a few channels (Dudley, 1939; Shannon *et al.*, 1995; Friesen *et al.*, 2001). The identification of an FSR ability, independent of spectral-temporal resolving power does, however, strongly suggest that cognitive approaches to improving speech understanding in hearing-impaired individuals should be explored. Although no hearing-impaired individuals were tested in this study, a corresponding range of the FSR ability seems certain to exist in the hearing-impaired population and thus to account for differences in speech recognition abilities among listeners with similar hearing loss.

B. Spectral and temporal processing abilities

The remaining three specific factors identified in this study reflect a variety of spectral or temporal processing abilities. These three factors represent a further differentiation of a factor identified by Carroll (1993), which he described as encompassing spectral and temporal processing for nonspeech complex stimulus patterns. This division of Carroll's single factor into three is a consequence of a wider range of relevant tests than in any of the earlier investigations considered by Carroll. There was some evidence in the datasets analyzed by Carroll for a distinction between spectral abilities (e.g., frequency and timbre discrimination) and temporal abilities (e.g., intensity, duration, and rhythm discrimination), but the data did not provide clear support for the existence of these two independent factors.

Rather than revealing a distinction between spectral and temporal abilities, the present study suggests the presence of three distinct types of temporal processing abilities. The *Loudness-Duration* factor reflects the ability to detect changes in overall energy (and duration). This ability has an obvious element of temporal processing, suggestive of the energy detection model (the "leaky integrator" model) developed within signal detection theory (Green and Swets, 1966; Jeffress, 1967). The *Pitch-and-Time* factor suggests a combined spectral-temporal pattern processing ability that is closely associated with spectrum and timbre discrimination. A range of tests with both spectral (i.e., frequency or timbre) and temporal (i.e., duration or timing) aspects is included. This factor may largely reflect the ability to discriminate holistic changes in the sound quality of a spectral-temporal pattern, whether due to spectral changes, temporal changes, or both. Finally, the *Amplitude Modulation factor* (consisting of SAM detection at the slower rates and gap detection) suggests an envelope-following ability that is distinct from an energy detection ability.

Despite the relatively large number of auditory tests (19) employed in the present study, compared to earlier investigations, a still wider range appears necessary to achieve a satisfactory definition of each of the four auditory abilities. The current findings provide a basis for the selection of additional tests that utilize the hypothesized abilities to different degrees. For example, the inclusion of other modulation-based tasks (e.g., modulation masking, modulation detection interference, modulation rate discrimination) might help clarify the nature of the ability that underlies the Amplitude Modulation factor, perhaps clarifying the role of modulation filters (Dau *et al.*, 1997a, b; Langner *et al.*, 2002) in these abilities. The inclusion of a rhythm discrimination test that required the discrimination of relative timing (e.g., by using time-transposed rhythmic patterns) plus other temporal discrimination tasks with different types of temporal structure over a range of time frames might help clarify the relations among the different temporal abilities observed in the present study. The inclusion of tasks that require the *discrimination* of familiar sounds with different types of stimulus changes would help to clarify the nature of the FSR ability and its relation to other auditory abilities.

C. The possibility of an "auditory *g*"

The latent variable analysis revealed the existence of a *general ability* in addition to the four independent auditory abilities. This represents common variance among all auditory tests that was not accounted for by the four factors. The influence of the general factor on performance was significant for all tests and was greater than that of the specific factors for a majority of the tests (11 of 19). However, both specific and general factors had a significant influence on all but four of the tests. The influence of specific factors was somewhat greater than that of the general factor for tests associated with the Amplitude Modulation and Familiar Sounds factors, while the opposite was true for the Pitch and Time factor and the Loudness-Duration factor.

Although a small portion of this general ability is related to math and verbal abilities, as measured by SAT scores, this ability is largely independent of intellectual abilities and appears to reflect a general facility with auditory stimuli. While many cognitive abilities, such as working memory or controlled attention, have been shown to be correlated with SAT scores (Daneman and Hannon, 2001; Engle *et al.*, 1999), as has general intelligence (Frey and Detterman, 2004; Beaujean *et al.*, 2006), there may be other amodal abilities that influence performance on auditory tests. For example, performance may be influenced by a general ability to process patterns in time or by the ability to detect subtle changes in complex multidimensional stimuli. It would be necessary to include a wider variety of nonauditory tests (e.g., visual, tactile, cognitive) to determine the extent to which the general ability identified in this study is specific to the auditory modality.

D. Summary and conclusions

The following conclusions are based on modeling of the data from 338 normal-hearing participants on 19 auditory discrimination and identification tests and associations with verbal and quantitative SAT scores.

(1) The best-fitting model of these data is one in which there is one general auditory ability and four independent specific abilities.
(2) The four specific auditory abilities are
    (a) loudness-duration discrimination
    (b) amplitude modulation detection
    (c) familiar sound recognition (speech and environmental sounds)
    (d) spectral and temporal pattern discrimination
(3) Intellectual abilities, as measured by SAT scores, had very weak associations with the general and specific abilities, accounting for less than 5% of the variance in each case.
(4) The independence of the four specific abilities suggests that individual differences in speech recognition under difficult listening conditions are not due to differences in spectral or temporal acuity, but to differences in a *familiar sound recognition (FSR) ability*.
(5) A more precise specification of the nature of the abilities identified in this study will require further individual-differences research.

Although no hearing-impaired listeners participated in this study, it is assumed that clinical populations vary along the same auditory-ability dimensions as the young normal-hearing listeners tested in the present study. Thus, differences in the ability to understand speech under difficult listening conditions among hearing-impaired listeners with similar hearing loss are likely due, at least in part, to differences in the FSR ability.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC00250 from the National Institute on Deafness and other Communicative Disorders. The authors would like to thank Patricia Knapp for her help in preparing the manuscript for publication. The associate editor and two reviewers also made significant contributions to this article. We are especially grateful for the constructive criticisms made by a third reviewer.

---

[1] Although SAT scores may not be ideal measures of general intellectual abilities, two recent studies (Frey and Detterman, 2004; Beaujean *et al.*, 2006) have shown that the relation between SAT scores and IQ measures is quite robust and clearly sufficient for the current purpose. These scores were available, while the cost of individually administered IQ tests would have been prohibitive.

[2] Because two sets of S/N values were used for the speech subtests (as described in the Method section), deciles were computed separately for each of the two groups of listeners and then combined to form one set of deciles. The data plotted in Fig. 1 for the Word subtest thus include 10 points (representing two sets of five S/N values) for each decile.

[3] At the most difficult stimulus values tested, performance becomes nearly independent of the stimuli. At these levels, some listeners will perform above chance and others below (for finite numbers of trials). When listeners are grouped by overall performance, as in the present case, those in the lowest deciles will tend to be those with below-chance responses, yielding the appearance of below-chance asymptotes for those deciles. This appearance is, of course, merely an artifact of this method of grouping the listeners.

[4] A different version of this figure, based on an earlier analysis, appears as part of an overview of individual differences research in Watson and Kidd (2002).

[5] Although structural equation modeling (SEM) has become popular in many other fields, there are only a handful of uses of this approach in articles published in this journal. In brief, SEM is a family of statistical concepts that include (and interrelate) multiple regression analysis, factor analysis, (causal) path analysis, and linear regression analysis (see Bentler, 1980, 1986; Bollen, 1989). Its use in the present project is to compare several alternative models that might account for the performance of the 338 subjects on 19 subtests in this project, in terms of a variety of hypothesized auditory and cognitive abilities. LISREL version 8.7 from Scientific Software International was used for all SEM analyses.

[6] Some subjects were not students, some did not take the SAT, and some records were not available at the Indiana University Office of Records for unknown reasons.

ANSI (**1989**). ANSI S3.6-1989: Specifications for Audiometers. (American National Standards Institute, New York).

Beaujean, A. A., Firmin, M. W., Knoop, A. J., Michonski, J. D., Berry, T. P., and Lowrie, R. E. (**2006**). "Validation of the Frey and Detterman (2004) IQ prediction equations using the Reynolds Intellectual Assessment Scales," Personality & Individual Differences **41**, 353–357.

Bentler, P. M. (**1980**). "Multivariate analysis with latent variables: Causal modeling," Annu. Rev. Psychol. **31**, 419–456.

Bentler, P. M. (**1986**). "Structural equation modeling and Psychometrika: An historical perspective on growth and achievements," Psychometrika **31**, 35–51.

Binet, A. (**1903**). *The Experimental Study of Intelligence* (Schleicher, Paris).

Bollen, K. A. (**1989**). *Structural Equations with Latent Variables* (Wiley, New York).

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Carroll, J. B. (**1993**). *Human Cognitive Abilities: A Survey of Factor-analytic Studies* (Cambridge University Press, New York).

Cattell, J. M. (**1885**). "On the time required for recognizing and naming letters and words, pictures and colors," in *James McKeen Cattell: Man of Science, Vol. 1: Psychological Research*, edited by A. T. Poffenberger (Science, Lancaster, PA, **1947**), pp. 13–25.

Christopherson, L. A., and Humes, L. E. (**1992**). "Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC)," J. Speech Hear. Res. **35**, 929–935.

Daneman, M., and Hannon, B. (**2001**). "Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT," J. Exp. Psychol. Gen. **130**, 208–223.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Davis, H., and Silverman, S. R. (**1970**). *Hearing and Deafness* (Holt, Rinehart & Winston, New York).

Drennan, W., and Watson, C. S. (**2001**). "Sources of variation in profile analysis. I. Individual differences and extended training," J. Acoust. Soc. Am. **110**, 2491–2503.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.

Elliott, D. N., Riach, W. D., Sheposh, J. P., and Trahiotis, C. (**1966**). "Discrimination performance of high school sophomores on a battery of auditory tests," Acta Oto-Laryngol., Suppl. **216**, 1–59.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. (**1999**). "Working memory, short-term memory, and general fluid intelligence: A latent-variable approach," J. Exp. Psychol. Gen. **128**, 309–331.

Feth, L. L., and O'Malley, H. (**1977**). "Two-tone auditory spectral resolution," J. Acoust. Soc. Am. **62**, 940–947.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Frey, M. C., and Detterman, D. K. (**2004**). "Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability," Psychol. Sci. **15**, 373–378.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1167.

Galton, F. (**1869**). *Hereditary Genius: An Inquiry into its Laws and Consequences* (Macmillan, London).

Galton, F. (**1883**). *Inquiries into Human Faculty and its Development* (Macmillan, London).

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Good, P. I. (**2006**). *Resampling Methods: A Practical Guide to Data Analysis*, 3rd ed. (Birkhauser, Boston).

Gorsuch, R. L. (**1983**). *Factor Analysis* (Erlbaum, Hillsdale, NJ).

Green, D. M. (**1971**). "Temporal auditory acuity," Psychol. Rev. **78**, 540–551.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory* (Wiley, New York).

Gygi, B. (**2001**). "Factors in the identification of environmental sounds," Unpublished doctoral dissertation, Indiana University, Bloomington.

Gygi, B., Kidd, G. R., and Watson, C. S. (**2004**). "Spectral-temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–1265.

Gygi, B., Kidd, G. R., and Watson, C. S. (**2007**). "Similarity and categorization of environmental sounds," Percept. Psychophys. (in press).

Hanley, C. N. (**1956**). "Factorial analysis of speech perception," J. Speech Hear Disord. **21**, 76–87.

Hirsh, I. J. (**1959**). "Auditory perception of temporal order," J. Acoust. Soc. Am. **31**, 759–767.

Humes, L. E., and Riker, S. (**1992**). "Evaluation of two clinical versions of the articulation index," Ear Hear. **13**, 406–409.

Hunt, E. (**1978**). "Mechanics of verbal ability," Psychol. Rev. **85**, 109–130.

Jakobson, L. S., Cuddy, L. L., and Kilgour, A. R. (**2003**). "Time-tagging: A key to musicians' superior memory," Music Percept. **20**, 307–313.

Jeffress, L. A. (**1967**). "Stimulus-oriented approach to detection reexamined," J. Acoust. Soc. Am. **41**, 480–488.

Johnson, D. M., Watson, C. S., and Jensen, J. K. (**1987**). "Individual differences in auditory capabilities. I," J. Acoust. Soc. Am. **81**, 427–438.

Karlin, J. E. (**1942**). "A factorial study of auditory function," Psychometrika **7**, 251–279.

Langner, G., Albert, M., and Briede, T. (**2002**). "Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (Chinchilla laniger)," Hear. Res. **168**, 110–130.

MacCorquodale, K., and Meehl, P. E. (**1948**). "On a distinction between hypothetical constructs and intervening variables," Psychol. Rev. **55**, 95–107.

Moore, B. C. J., and Glasberg, B. R. (**1986**). "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 251–308.

Nelson, P. B., Nittrouer, S., and Norton, S. J. (**1995**). "Say-stay identification and psychoacoustic performance of hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 1830–1838.

Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.

Raykov, T., and Marcoulides, G. A. (**2000**). *A First Course in Structural Equation Modeling* (Erlbaum, Mahwah, NJ).

Seashore, C. E., Lewis, D., and Saetveit, J. (**1939**, **1960**). *Seashore Measures of Musical Talents* (The Psychological Corporation, New York).

Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). "Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity," J. Acoust. Soc. Am. **112**, 720–727.

Snow, R. E. and Lohman, D. F. (**1988**). "Implications of cognitive psychology for educational measurement," in *Educational Measurement*, 3rd ed., edited by R. L. Linn (American Council on Education & Macmillan, New York), pp. 263–331.

Spearman, C. (**1904**). "General intelligence objectively determined and measured," Am. J. Psychol. **15**, 201–293.

Stankov, L., and Horn, J. L. (**1980**). "Human abilities revealed through auditory tests," J. Educ. Psychol. **72**, 21–44.

Sternberg, R. J. (**1977**). *Intelligence, Information Processing, and Analogical Reasoning: The Componential Analysis of Human Abilities* (Erlbaum, Hillsdale, NJ).

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.

Tallal, P., Miller, S., and Fitch, H. A. (**1993**). "Neurobiological basis of speech: A case for the preeminence of temporal processing," in *Temporal Information Processing in the Nervous System: Special Reference to Dyslexia and Dysphasia, Annals of the New York Academy of Sciences, Vol. 682* edited by P. Tallal, A. M. Galaburda, R. R. Llinas, and C. v. Euler (NYAS, New York), pp. 27–47.

Thurstone, L. L. (**1938**). "Primary mental abilities," Psychometric Monographs, No. 1.

Thurstone, L. L. (**1947**), *Multiple Factor Analysis* (University of Chicago Press, Chicago).

Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing impaired listeners," J. Acoust. Soc. Am. **72**, 740–752.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Watson, B. U. and Miller, T. K. (**1993**). "Auditory perception, phonological processing and reading ability/disability," J. Speech Hear. Res. **36**, 850–863.

Watson, C. S. (**2004**). "Temporal acuity and the identification of temporal order: Related, but distinct, auditory abilities," Seminars in Hearing **25**, 219–227.

Watson, C. S., and Kidd, G. R. (**2002**). "On the lack of association between basic auditory abilities, speech processing, and other cognitive skills," Seminars in Hearing **23**, 83–93.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.

Watson, C. S., Qiu, W. W., Chamberlain, M., and Li, X. (**1996**). "Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**, 1153–1162.

Watson, C. S., Jensen, J. K., Foyle, D. C., Leek, M. R., and Goldgar, D. E. (**1982a**). "Performance of 146 normal adult listeners on a battery of auditory discrimination tasks," J. Acoust. Soc. Am. **71**, S73.

Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., and Jensen, J. K. (**1982b**). "An auditory discrimination test battery," J. Acoust. Soc. Am. **71**, S73.

Wichmann, F. A., and Hill, N. J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Wichmann, F. A., and Hill, N. J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Yost, W. A., Hill, R., and Perez-Falcon, T. (**1978**). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. **64**, 1166–1173.