Word Recognition and the Articulation Index in Older Listeners with Probable Age-Related Auditory Neuropathy

George A. Gates*
M. Patrick Feeney*
Roger J. Higdon†

Abstract
This retrospective analysis of existing data was derived from 957 members of a population-based cohort who participated in a prior study of the prevalence of central auditory dysfunction. Word recognition scores (WRS) at three intensity levels were compared to predicted scores based on the Articulation Index (AI) and the Thornton-Raffin 95 percent critical differences. In 112 (11.7 percent) participants, one or more word recognition scores were significantly below the predicted score, which we consider a subtle sign of possible auditory neuropathy. In contrast, classic signs of retrocochlear dysfunction were found in only three people (0.3 percent) using rollover of the performance-intensity function for phonetically balanced word lists, in two (0.2 percent) people using the guideline of Yellin et al (1989), and in 54 people (5.6%) using a 20-point difference between the AI (x 100) and the WRS. Subtle signs of possible auditory neuropathy were more frequent than the classic signs. Comparing WRS at several high presentation levels to the AI is suggested as a method to screen for subtle neuropathy. Elderly listeners whose WRS fall below the Thornton-Raffin 95 percent critical difference based on AI should be considered for further testing for age-related auditory neuropathy.

Key Words: Aging, articulation index, auditory neuropathy, presbycusis, word recognition

Abbreviations: AI = Articulation Index; AI100 = Articulation Index x 100; tr-low = WRS falling outside the Thorton-Raffin low (5 percent) boundary; tr-high = WRS falling outside the Thorton-Raffin high (95 percent) boundary; WRS = Word Recognition Score

Sumario
Este análisis retrospectivo de datos ya conocidos se derivó de 957 miembros de una cohorte poblacional que participó en un estudio previo de prevalencia de disfunción auditiva central. Se compararon los puntajes de reconocimiento de palabras (WRS) en 3 niveles de intensidad con la predicción de puntajes basados en el Índice de Articulación (AI) y en las diferencias críticas al 95% de Thornton-Raffin. En 112 participantes (11.7%), uno o más de los puntajes de reconocimiento de palabras fueron significativamente mejores que el puntaje de predicción, lo que nosotros consideramos un signo sutil de posible
Age-related hearing loss (presbycusis) is a multifactorial disorder with peripheral (i.e., cochlear) and, often, central nervous system components (CHABA Working Group on Speech Understanding and Aging, 1988). The major peripheral pathologic changes are hair cell loss, strial atrophy, spiral ganglion cell degeneration, and combinations thereof (Schuknecht, 1964). Age-related central auditory dysfunction is less well defined and may include changes due to aging at any level of the central auditory pathways and associated areas.

The close relation of sensory (cochlear) function and auditory nerve function makes it difficult in the average clinical setting to assess the differential contribution of these important components. Indeed, because poor speech understanding may result from both sensory and neural dysfunction, the term, “sensorineural hearing loss,” is still used clinically to indicate this dilemma. This usage is justified because impaired sensory and neural function undoubtedly coexist in many cases.

The dividing line between the central and peripheral parts of the auditory system is ambiguous. Because the spiral ganglion cells reside in the cochlear spiral, these neural cell bodies are, technically, located in the periphery even though they are the first in a chain of neural connections. A peripheral auditory nerve lesion, for example, loss of ganglion cells, and a more central auditory nerve lesion, for example, a tumor in the internal auditory canal (i.e., a retrocochlear lesion) may have identical behavioral hearing test results. Although electrophysiologic audiometry may help make this distinction, it is seldom used in the evaluation of people with presbycusis.

Histologic evidence of primary auditory neuropathy (spiral ganglion atrophy) is evident in about 15 percent of human temporal bones from people with hearing loss (Schuknecht and Gacek, 1993). However, a corresponding number of adults do not display obvious clinical evidence of auditory neuropathy (Gates and Popelka, 1992). Is this discrepancy due to the subtle nature of age-related auditory neuropathy, or could it be the case that behavioral clinical methods in current use are insensitive to early or mild changes? Age-related atrophy of the spiral ganglion (neural presbycusis) is theoretically the most common type of auditory neuropathy when one accounts for its histologic prevalence (Schuknecht and Gacek, 1993) and the rising age of our population. Because of the adverse effect of neuropathy on speech understanding and hearing aid use, methods to screen for even mild cases of neuropathy should be sought.

Auditory nerve dysfunction is traditionally characterized by poorer word recognition than expected by the pure-tone threshold levels. However, what defines the critical difference between the expected and

Abreviaturas: AI = Indice de Articulación; AI100 = Indice de Articulación x 100; tr-low = WRS fuera del límite bajo (5 por ciento) del Thornton-Raffin; tr-high = WRS fuera del límite alto (95 por ciento) del Thornton-Raffin; WRS = umbral de reconocimiento de palabras
the actual score is open to interpretation (Yellin et al, 1989). Several readily available behavioral methods have been used to suspect neural dysfunction: (1) rollover of the performance-intensity function of phonetically balanced word lists (PI-PB), (2) poorer word recognition scores (WRS) than predicted by the articulation index (AI) (Gates and Popelka, 1992), or (3) poorer WRS than predicted by the high-frequency pure-tone average (Yellin et al, 1989).

We compared the predictive utility of these methods by reexamining a large data set containing pure-tone and speech audiometric results from a population-based cohort of adult volunteers. These data were obtained from hearing testing done during the 18th Examination Cycle of the Framingham Heart Study Cohort and reported earlier (Gates et al, 1990; Cooper and Gates, 1991). The present analysis suggests a new clinical strategy to screen for age-related auditory neuropathy that uses the articulation index and the Thorton-Raffin critical differences (Thornton and Raffin, 1978) to determine whether observed WRS are poorer than expected.

METHODS

Rationale

The data in this report were initially gathered in a prospective manner for the purpose of determining the prevalence of central auditory dysfunction in a population-based sample of older people. Since that time, concepts of central auditory dysfunction have been evolving. The present report examines these data post hoc in light of the newer concepts of auditory neuropathy.

Terminology

The auditory nerve consists of the axons of Type I and Type II afferent cochlear neurons of the spiral ganglion plus efferent fibers of the olivocochlear bundle. Atrophy of the cochlear neurons in the spiral ganglion may be identified in human temporal bone sections (Schuknecht and Gacek, 1993). Retrocochlear dysfunction refers generally to test abnormalities associated with tumors of the auditory nerve, although any auditory nerve lesion is, technically, “retrocochlear.”

The archetypical retrocochlear clinical picture is that of a patient with very poor word recognition ability and near normal threshold sensitivity.

“Auditory neuropathy” was the term originally applied to people with present outer hair cell function as evidenced by otoacoustic emissions, and absent auditory brainstem responses (ABR) and absent acoustic reflexes (Starr et al, 1996). The term implies abnormal function, most often dyssynchrony, of the auditory nerve, which includes the synapse with the inner hair cell and, undoubtedly, in some cases, the inner hair cell as well. These patients usually have very poor word recognition and various levels of threshold loss. The histopathology of auditory neuropathy is uncertain, but in congenital and hereditary cases it may be associated with peripheral neuropathy as well.

We propose the term “age-related auditory neuropathy” to indicate spiral ganglion degeneration and possibly other neural changes that occur in older people as a synonym for neural presbycusis. The term implies a graduated rather than an all-or-none condition, such that patients might have dysfunction ranging from mild to severe. Implicit in this concept is the probable progressive nature of age-related neural degeneration. Our working theory is that word recognition at high presentation levels provides the best method to identify suboptimal word recognition performance in people with age-related auditory neuropathy.

Participants

A total of 1662 ambulatory Framingham cohort members completed an initial audiometric survey (pure-tone thresholds, tympanometry, and word-recognition testing in the better ear) (Gates et al, 1990). Of these 1259 (76 percent) were eligible for central testing, and 1026 had one or more central auditory tests. Eligibility criteria were normal tympanograms and present acoustic reflexes, no history of ear surgery or hearing loss since childhood, and less than a 21 dB difference between pure-tone threshold averages for the two ears. The first two criteria were used to exclude those with possible middle-ear disease, either past or present. The third and fourth were designed to minimize the effect of acquired conditions (e.g., unilateral sudden deafness, tumor, infection) other than aging.
The central auditory findings have been reported earlier (Cooper and Gates, 1991).

Procedures

The audiometric methods have been reported previously (Gates et al, 1990). History and basic audiometric test data from that report are used herein for the participants included in this report. All testing was done by certified audiologists in a 1200 series Industrial Acoustics Company, Inc. sound room using a calibrated (ANSI S3.6-1969 [r-1973]) (American National Standards Institute, 1969) audiometer (Grason-Stadler 1715) and a properly maintained stereo cassette tape deck (Sony TC-FX22). The test battery used the recorded versions of the Central Institute for the Deaf (CID) W-22 phonetically balanced 50-item word lists produced by Auditec of St. Louis. Complete sets of WRS at three intensity levels for both ears (40 dB SL, 50 dB HL, and 90 dB HL) were obtained from 957 people, who are the subject of this report. These presentation levels were chosen to optimize the testing within the time constraints of the Framingham Heart Study. The test protocol did not allow for establishing formal comfortable and uncomfortable loudness levels.

The CID W-22 lists were selected because of their widespread use for word-recognition testing. Fifty items were used for each test. List 1A was used to test the better ear at 50 dB HL. The better ear was defined as the one with the numerically smaller PTAlow (0.5, 1.0, 2.0 kHz). If the PTAlow was equal for the two ears, the right ear was designated the better ear (the right ear was so designated in 52 percent of cases). Lists 2A and 4A were then used to test both ears at 40 dB SL with the better ear being tested first. Lists 1B, 2B, 4B, and the lists from Form C were then used, as necessary, to obtain measures at 50 dB HL in the poorer ear and 90 dB HL in both ears. Lists were presented to the test ear in quiet although speech noise was used to mask the contralateral ear when appropriate. All 957 participants had tests at 40 dB SL, 50 dB HL, and 90 dB HL. The data from the right ear of each subject were used for the present study. When converted to the appropriate SL for each subject by subtracting the three frequency pure-tone average from 50 for the 50 dB HL presentation or 90 for the 90 dB HL presentation, the range of presentation levels was from 12 dB SL to 90 dB SL with a mean (+ sd) of 47.7 + 18.5 dB SL.

Articulation Index

An AI was calculated for each word-recognition test for each subject at each of the three presentation levels using the method of Popelka and Mason (Popelka and Mason, 1987). This method was chosen because of its applicability to CID W-22 word lists and its scalability for various presentation levels. The AI varies from 0–1 and is based on nine audiometric frequency bands. In an earlier report of these data (Gates and Popelka, 1992), the AI x 100 (AI100)-WRS difference in the right ear at 50 dB HL was found to be normally distributed with about a 10 percent standard deviation. This indicates that no more than two percent of the cases could have a statistically “abnormal” AI100-WRS discrepancy. We now believe that cases of mild auditory neuropathy may be included in this distribution because normal, as a statistical concept, does not necessarily indicate normal in a physiologic sense.

Analyses

The present analysis was conducted to develop exploratory models of the relations between WRS and AI at multiple sensation levels (SL). The effective SL was determined by the pure-tone average threshold at 0.5, 1.0, and 2.0 kHz. As expected, there was a moderately strong correlation (r = 0.635) between the SL and the AI. The complex relations among these factors was investigated by an analysis of variance (ANOVA), which indicated that SL, AI, the interaction of SL and AI, and gender each had independent and significant effects on WRS (data not shown).

Because the AI calculation inherently adjusts for SL, we sought to develop an AI-based method to calculate an expected range of WRS for each presentation level. In addition, the highly right skewed WRS data necessitated exploring several models to determine the most appropriate representation of the central tendency of the data. STATA v 8.0. was used for all analyses (Statacorp., 2003). First, the binomial probability of a perfect word-recognition score (50 items correct) was estimated for each AI level while adjusting for repeated measures.
(i.e., three tests per person) using the General Linear Model (GLM) for binomial regression with repeated measures. Second, a nonlinear regression of WRS on AI using the logistic function:

\[ Y = \frac{b_1}{1 + (-b_2(x - b_3))} \]

was done without a repeated measures term. The predicted curves for each model were essentially identical. The nonlinear regression method, which gave the best visual fit of the WRS and SL, was used in Figure 1. The binomial regression model gave the best visual fit for the WRS and AI and was used in Figure 2 to describe the expected score for each AI level.

Next, 95 percent critical differences were computed for each of the 2871 test instances using the Thorton and Raffin critical differences (Thornton and Raffin, 1978). The binomial probability derived from the GLM model (above) of a perfect word-recognition score for each test instance was compared to the actual score, and each actual score was then designated as outside or inside the Thorton-Raffin critical difference boundary assigned to the respective AI. Those outside the critical difference boundary were designated as \( tr\)-high for those above the boundary and \( tr\)-low for those below it. Participants were noted as having none, one, two, or three \( tr\)-low WRSs.

In addition, three dummy outcome variables were created. The first was applied to all instances in which the AI x 100 (AI100) was 20 percentage points greater than the WRS at any SL, and was termed “AI100-WRS discrepancy” to indicate one definition of retrocochlear dysfunction (Gates and Popelka, 1992). The second variable was applied to the participants whose WRS at the highest level was 20 percentage points lower than either of the two WRS scores at lower intensity levels, and was termed “rollover” to indicate abnormal decline of the PI-PB function with increasing presentation level (Jerger and Jerger, 1971). The third variable was calculated to determine cases with WRS below expected based on the PTA\(_{\text{high}}\) (average of 1, 2, and 4 kHz) using the method of Yellin et al 1989.

**RESULTS**

Of the 957 participants, 559 were women and 398 were men. The mean age was 72.1 + 5.6 years (range 63–92) and did not differ by gender. Only right-ear data were used. The overall mean WRS was 89.1 + 12.5 percent with a range from 0–100. Seven hundred and forty-nine of the 957 people (78.3 percent) had a normal WRS of 90 percent or greater at 40 dB SL, and 867 of the 957 (90.1 percent) had a normal WRS at SLs of 35 dB and greater. The relations of WRS and SL are displayed by gender in Table 1.

The relation of the WRS to presentation level in dB SL is displayed in Figure 1 using the nonlinear regression method. The mean presentation level was 47.3 + 18.8 dB above the participants’ pure-tone thresholds. Note
that the predicted score reached 90 percent at 45 dB SL and continued to increase to 98 percent at 90 dB SL.

In Figure 2, the relation of the predicted WRS to the AI100 is displayed with the Thornton-Raffin 95 percent bounds. There were 165 out of 2871 (5.6 percent) WRS tests that were designated tr-low in 112 participants (11.7 percent) and 610 of 2781 (21.3 percent) WRS tests that were designated tr-high (in 424 participants [44.0%]). Fifty-one men and 61 women were in the tr-low group. This gender difference was not significant (p = 0.85). However, significantly more women than men (285 vs. 139) were in the tr-high group (p < 0.001). There was an age effect on the Thornton-Raffin distribution with the tr-low outliers being significantly older (77.6 ± 7.3 years) than the tr-high outlier cases (70.3 ± 4.6 years, p < 0.0001) and the cases within the Thornton-Raffin bounds (72.3 ± 5.5 years, p < 0.0001).

Of the tr-low cases, 25 people had only one tr-low test at a sensation level below 40 dB SL, which could be attributed to low audibility as well as possible neuropathy. The remaining 87 tr-low people (9.1 percent of participants) had one or more tr-low tests at SL of 40 dB or higher. Thus, these 87 people demonstrated poorer WRS than could be accounted for by either audibility or the AI. For convenience we label these people as auditory neuropathy suspects (ANS). The ANS group included all the “rollover” and AI100-WRS discrepancy cases (see below). Within the ANS group, 12 participants were tr-low on all three tests, 29 were tr-low in two tests, and 46 had one WRS that was tr-low.

Rollover of the PI-PB function of 20 percentage points or more occurred in only three participants, all of whom were women and were tr-low at one or more SL.

A 20 percentage point or greater discrepancy in the AI100-WRS scores occurred in 76 AI-WRS comparisons from 54 of the 957 participants (5.6 percent) (28 women and 26 men; the gender difference was not significant [p = 0.923]). All three rollover cases also had AI100-WRS discrepancy, and all the rollover and AI100-WRS discrepancy cases were below the Thornton-Raffin 95 percent confidence levels for one or more AI levels (i.e., tr-low).

Only two participants had a maximum word-recognition score below the lower boundary of Yellin et al (Yellin et al, 1989). Both of these cases had rollover of the PI-PB function and were also tr-low at their highest SL.

DISCUSSION

Speech perception is a critical element in geriatric hearing evaluation. Older patients’ nearly universal complaint is difficulty understanding speech. Normal peripheral speech perception depends on a combination of audibility, cochlear function, and neural integrity. An emerging goal of clinical auditory research is to be able to measure the contribution of each of these elements in order to select patients for future cell-specific therapy. We believe the present report to be a small step in that direction. Moreover, given the adverse effect of auditory nerve dysfunction on amplification success (Chmiel and Jerger, 1996), this approach has potential usefulness at present.

We recommend incorporating AI calculations and speech testing at multiple intensity levels for evaluating people with the complaint of difficulty understanding speech. Adding the AI permits one to evaluate whether the observed WRSs are within the range of expected results. We consider that
a WRS obtained at SL of 40 and above that lies below the Thornton-Raffin lower bound for the AI is a subtle sign of possible auditory neuropathy; objective assessment with otoacoustic emissions and auditory brainstem audiometry could be considered in these cases.

Speech testing at 40 SL only is not sufficient because the scores continue to improve with increasing presentation levels. As demonstrated herein, the maximum WRS may occur from 45 to 90 dB SL. Adding a test at just below loudness discomfort level (Beattie and Zipp, 1990) or 90 dB SL appears to stress the system to reveal more cases of possible neuropathy than are suspected by current guidelines. However, further research will be necessary to validate that hypothesis.

The present findings suggest that current clinical methods for suspecting retrocochlear neural dysfunction are very conservative and relatively insensitive to mild age-related auditory neuropathy. The archetype function—the PI-PB rollover—was rarely abnormal, and only a few participants showed a 20-point difference in the AI100-WRS function, but many more showed WRS that were lower than predicted using the Thornton-Raffin estimates (1978). Note that all participants had a present acoustic reflex at 1 kHz.

The Thornton-Raffin data were obtained from the clinical records of 4120 male veterans in their 50s and 60s who had 50-item CID W22 tests “at 40 dB SL whenever possible” (1978, 509). It is likely that there were unrecognized cases of auditory neuropathy present in that sample. If this is indeed true, the present method would underestimate the true prevalence of auditory neuropathy. Of interest, we found no differences in the Thornton-Raffin estimates between genders. There were an appreciable number of presentations—all with normal WRS—in which the actual WRS was higher than the Thornton-Raffin upper bound as predicted by the AI100. In most cases, this was due to the higher presentation levels used in the present study. Although the AI100 is an index of speech information transfer, it is recognized that a 100 percent WRS score can be achieved with an AI100 substantially lower than 100 percent (Halpin et al, 1996).

Current methods to diagnose primary auditory neuropathy depend on the presence of otoacoustic emissions and abnormal auditory brainstem responses. Emission testing is often abnormal in the elderly because of hair cell loss or strial dysfunction. Electrocochleography could be considered to further evaluate neuropathy in the presence of cochlear dysfunction (Arnesen, 1982).

Three-quarters of the tr-low cases had a three-frequency (0.5, 1.0, 2.0 kHz) pure-tone threshold average of 25 dB or poorer. Therefore, we would suspect that many of these cases had multifactorial presbycusis, that is, a combination of hair cell loss, strial dysfunction, and spiral ganglion atrophy. The present data do not provide a method to address the issue of primary versus secondary neural atrophy.

**SUMMARY**

A method for suspecting the presence of age-related auditory neuropathy is presented that is based on the Popelka-Mason method for determining an Articulation Index in conjunction with the Thornton-Raffin critical differences. Participants with WRS below the predicted levels who are considering amplification may wish to undergo further testing to determine the presence of auditory neuropathy. Future research into speech testing at high levels will be necessary to document our findings and confirm our hypothesis.

**Acknowledgment.** Gerald Popelka, Ph.D., made helpful suggestions for this report. Aimee Verrall assisted in data management and manuscript preparation.
REFERENCES


