Research Note

Phoneme and Word Scoring in Speech-in-Noise Audiometry

Curtis J. Billings,[a,b] Tina M. Penman,[a] Emily M. Ellis,[a]
Lucas S. Baltzell,[a] and Garnett P. McMillan[a]

Purpose: Understanding speech in background noise is difficult for many individuals; however, time constraints have limited its inclusion in the clinical audiology assessment battery. Phoneme scoring of words has been suggested as a method of reducing test time and variability. The purposes of this study were to establish a phoneme scoring rubric and use it in testing phoneme and word perception in noise in older individuals and individuals with hearing impairment.
Method: Words were presented to 3 participant groups at 80 dB in speech-shaped noise at 7 signal-to-noise ratios (−10 to 35 dB). Responses were scored for words and phonemes correct.

Results: It was not surprising to find that phoneme scores were up to about 30% better than word scores. Word scoring resulted in larger hearing loss effect sizes than phoneme scoring, whereas scoring method did not significantly modify age effect sizes. There were significant effects of hearing loss and some limited effects of age; age effect sizes of about 3 dB and hearing loss effect sizes of more than 10 dB were found.
Conclusion: Hearing loss is the major factor affecting word and phoneme recognition with a subtle contribution of age. Phoneme scoring may provide several advantages over word scoring. A set of recommended phoneme scoring guidelines is provided.

M any individuals experience perception-in-noise difficulties, particularly in challenging listening environments, such as in an automobile or a busy restaurant. However, speech perception testing in quiet is still the clinical norm. Although there are several important clinical reasons to include speech-in-noise testing (Boothroyd, 2008; Wilson, 2011), it is not routinely included in the clinical test battery, presumably because of time constraints. Several methods for reducing speech-in-noise test time have been proposed, including reducing the number of trials and using phoneme scoring rather than word scoring. Given that patient difficulties frequently include listening in background noise (Kochkin, 2002), we set out to establish phoneme scoring guidelines and to understand the effects and effect sizes of word and phoneme scoring on clinically relevant factors of age and hearing impairment given that phoneme scoring has been explored primarily in younger, normal-hearing individuals (e.g., Gelfand & Gelfand, 2012; McCreery et al., 2010).

Phoneme scoring has been suggested as an alternative to whole-word scoring because of several advantages, including an increased number of tokens in the same amount of test time, reduced variability in scores, information about specific phoneme perception errors, and reduced influence from top-down processing, such as effects of the listener's lexicon (Boothroyd, 2008; Gelfand & Gelfand, 2012; Markides, 1978; McCreery et al., 2010; Olsen, Van Tasell, & Speaks, 1997). There are drawbacks, however, including difficulty in comparing scores across clinics due to tester bias and issues with scoring accuracy as a result of dialectal differences between tester and testee (Boothroyd, 1968; Markides, 1978). A concrete set of phoneme scoring guidelines to be used in audiological testing would improve the consistency of a phoneme measure across testers and clinics; however, to our knowledge, such a measure is not available in the literature. Therefore, a goal of this study was to develop a formal set of phoneme scoring rules to help minimize tester bias and improve ease of testing. A further limitation of the literature is the limited population on which phoneme scoring has been tested. To justify clinical use, a better understanding of phoneme scoring effects across age and hearing loss is necessary.

Effects of age and hearing loss have been studied extensively over the last three decades. It is clear from the literature that as hearing thresholds decrease, whether due

to presbycusis or noise damage, the ability to understand speech in background noise also declines. This decrease in understanding occurs for syllables (Gelfand, Piper, Silman, 1986; Humes & Christopherson, 1991; Humes & Roberts, 1990), words (Souza & Turner, 1994; Stuart & Phillips, 1996; Studebaker, Sherbecoe, McDaniel, & Gray, 1997; Summers & Molis, 2004; Wiley et al., 1998; Wilson, 2011), and sentences (Bacon, Opie, & Montoya, 1998; Dubno, Dirks, & Morgan, 1984; Gordon-Salant & Fitzgibbons, 1995). Some of the perception-in-noise literature also demonstrates an effect of age such that older individuals perform worse than younger individuals (Divenyi & Haupt, 1997; Dubno et al., 1984; Gelfand et al., 1986; Stuart & Phillips, 1996; Studebaker et al., 1997; Wiley et al., 1998), although aging effect sizes were somewhat smaller than those associated with hearing loss. In contrast, other aging studies revealed no significant aging effect (Gordon-Salant & Fitzgibbons, 1995; Humes & Roberts, 1990; Souza & Turner, 1994). It may be that the magnitude of the age effect is small compared with the more robust effect of hearing loss and is therefore seen only in specific test conditions. Inclusion of older individuals and individuals with hearing loss in this study will help to assess phoneme scoring in broader clinical audiology populations.

The purpose of this study was to (a) establish phoneme scoring guidelines, (b) apply these guidelines in testing older and hearing-impaired groups, and (c) determine if scoring method (word vs. phoneme) modifies the effects of age and hearing loss. This information may aid the audiologist or researcher interested in speech-in-noise testing. In particular, the analysis will demonstrate how phoneme scoring may or may not modify results relative to word scoring. In addition, comparison of participant groups will provide effect sizes specific to age and hearing impairment effects.

Method

Participants

Three groups were tested: Group A consisted of 20 younger participants (mean age: 25.5 years; range: 18–35 years; 10 women, 10 men) with normal hearing (500, 1000, and 2000 Hz pure-tone average [PTA]: 7.3 dB HL), Group B consisted of 20 older participants (mean age: 68.5 years; range: 60–79 years; 10 women, 10 men) with normal to borderline normal hearing (PTA: 11.3 dB HL), and Group C consisted of 20 older participants (mean age: 71.9 years; range: 63–84 years; nine women, 11 men) with hearing impairment (PTA: 33.7 dB HL). Figure 1 shows the mean audiometric data for the three groups. Participants had 226-Hz tone tympanometric measures of peak pressure ranging from −80 to 45 daPa, ear canal volume ranging from 0.9 to 2.3 cm^3, and static admittance ranging from 0.2 to 2.3 mmho. Participants reported good overall health and had a normal score (as normed by Crum, Anthony, Bassett, & Folstein, 1993) on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975).

This study was part of a larger experiment and shared some participants with other published work (Billings, Penman, McMillan, & Ellis, 2015).

Signals and Noise

Signals consisted of words from the Northwestern University Auditory Test Number 6 (NU-6). The 200 words spoken by a woman with the accompanying carrier phrase ("Say the word…") were extracted from the Department of Veterans Affairs Speech Recognition and Identification Materials (Disc 4.0, Track #17-24, 2006). The background noise was continuous speech-shaped noise made on the basis of the spectrum of female-talker recordings from the 720 IEEE sentences (Institute of Electrical and Electronic Engineers, 1969). Both signals and noise were low-pass filtered at 4000 Hz (Butterworth filter cutoff at 4000 Hz transitioning to 5000 Hz, beyond which no frequency content was passed) to account for hearing differences between groups. Filtering is indicated in Figure 1 as gray shading. The root-mean-square levels of 10-s concatenated versions of the NU-6 words and speech-shaped noise were measured independently using a dBC weighting. Signal level was 80 dB SPL, and noise level was manipulated to result in seven signal-to-noise ratios (SNRs) (−10, −5, 0, 5, 15, 25, and 35 dB SNR) as measured in a 2-cc coupler. During the testing session, signal levels of 50, 60, and 70 dB were also tested; however, only the 80-dB SPL results are included herein. Thirteen NU-6 words were presented for each test condition, similar to the 10 words used by Boothroyd (2008) and Gelfand and Gelfand (2012). Words were presented sequentially using the 200-word list, and conditions were randomized such that the specific 13 words used for a given condition varied across participants. Signals and noise were presented through separate channels of a Grason-Stadler, Model 61 audiometer (Eden Prairie, MN) using an Etymotic, ER-2 insert earphone (Elk Grove Village, IL) to the right ear of the participant. All testing was done in a double-walled, sound-isolated booth meeting ambient noise level standards for audiometric rooms (American National Standards Institute, 1991).

Behavioral Testing and Scoring

Participants were asked to repeat the word following the carrier phrase "Say the word…" If the participant did not hear or understand the whole word, he or she was strongly encouraged to guess and repeat audible sounds. Two judges confirmed the responses in real time; if a disagreement occurred between the two judges, the differences were discussed or the participant was asked to repeat the response. Disagreements were rare for word scoring, occurring less than .01% of the time, whereas phoneme scoring resulted in disagreements about 1.7% of the time. A word score and phoneme score were generated for each condition. The word score was out of 13 possible, and the phoneme score was out of 39 possible phonemes (three possible phonemes in each word). Ten practice words were presented prior to testing to acquaint each participant with the

Figure 1. Mean thresholds for Group A (blue line), Group B (green line), and Group C (red line). Thresholds are displayed for the right ear only. To account for high-frequency threshold differences between Group A and Group B, all stimuli were low-pass filtered at 4000 Hz (depicted by gray shading). Error bars represent the standard errors of the mean.
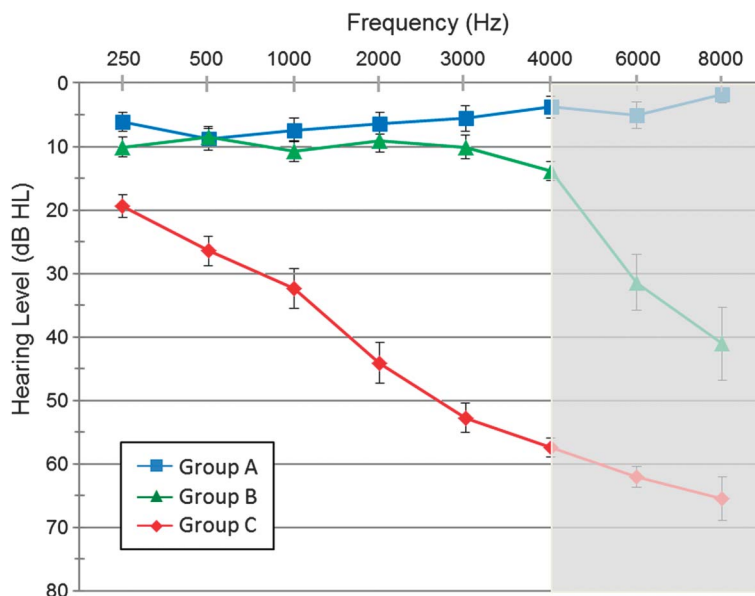
task. The first five practice words were presented at 80 dB SPL with a 45 dB SPL background noise, and the last five practice words were presented at 80 dB SPL with an 85 dB SPL background noise.

Method of Analysis

A two-stage approach was used to determine each subject's SNR50 and SNR70 measures (the 50% or 70% points, respectively, on an individual's psychometric function or the SNR value at which an individual achieves 50%/70% correct) and to evaluate the effects of age and hearing level on mean speech recognition. In the first stage, logistic regression (using SAS version 9.4) was used to model each participant's psychometric function (word or phoneme score as a function of SNR). SNR50 and SNR70 measures were deduced analytically using an "effective dose" methodology described in Collett (1991) and used previously (Billings et al., 2015). Extrapolation of the SNR50 or SNR70 outside of the range of stimulated SNRs was not allowed and resulted in missing values for some individuals in Group C (four individuals were missing an SNR50 for words and one for phonemes; seven individuals were missing an SNR70 for words and five for phonemes).

In the second stage, the mean word and phoneme SNR50 measure was modeled using a linear mixed model generalization of the multivariate analysis of variance. Mean SNR70 measures were modeled in the same way. Each model included response type (words and phonemes) and subject group (A, B, and C). Correlated random subject-level intercepts for words or phonemes were included in the model to account for the correlation among repeated measurements on each subject. Specific group comparisons focused on the isolation of hearing loss effects (Group B vs. Group C) and age effects (Group A vs. Group B). We also considered age and PTA as continuous covariates (as opposed to group assignments) in the fitted model. This resulted in considerably higher Akaike's information criterion (AIC) of 1,950.03 compared with the simpler model with categorical age/hearing impairment assignments (AIC = 1,814.72). The AIC is a commonly used measure of the goodness of fit of a statistical model. In accordance with this, we used the categorical age/hearing level model for inferences.

Results

The Appendix contains the phoneme scoring guidelines that were developed over the course of testing several pilot participants. Challenges associated with scoring consonant clusters, pronunciation variations (vowel coloring/dialectal differences), and two-syllable responses were addressed. These scoring guidelines were used for scoring all study participants. Figure 2a illustrates that increases in SNR resulted in an improvement in performance (mean percentage correct) across all conditions for both word (left) and phoneme (right) scores. All three groups were similar in this respect. However, Groups A and B appeared to perform similarly to each other, and Group C performed more poorly for all SNR conditions.

Figure 2b shows mean SNR50 and SNR70 measures for words and phonemes as a function of group. A significant difference between Groups B and C was observed

Figure 2. Psychometric functions (above) and mean SNR50 and SNR70 performance estimates (below) displaying word (left column) and phoneme (right column) results. Group differences are displayed as the parameter within each subplot and are generally much larger as a function of hearing impairment (Group B vs. Group C) than they are for age (Group A vs. Group B). Differences between Groups A and B are approximately 3 dB, and differences between Groups B and C are usually larger than 10 dB. All error bars represent the standard errors of the mean.
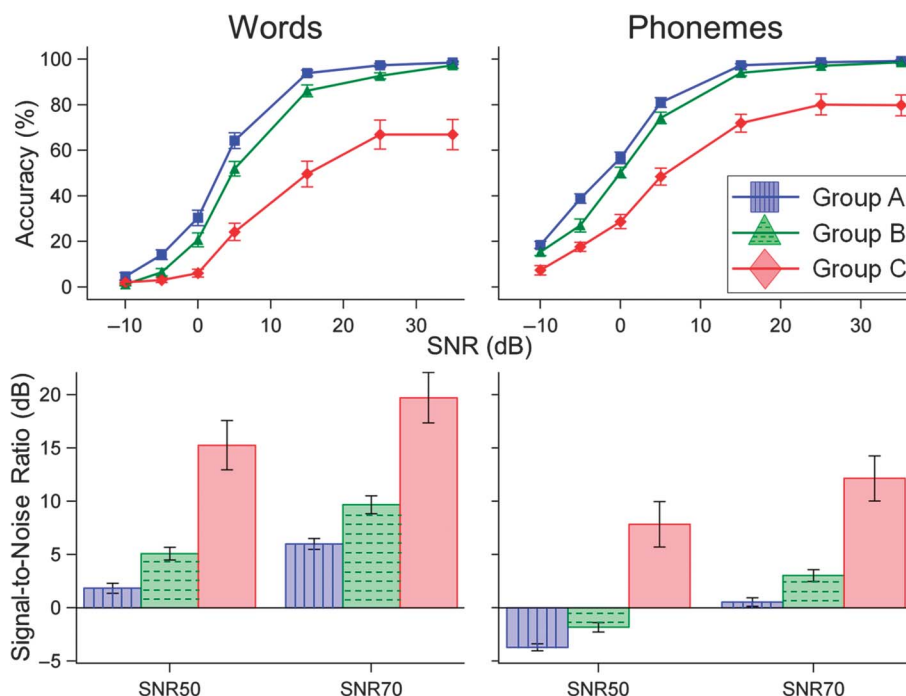
for words (SNR50: effect size = 12.3, SE = 1.75, t = 7.01, $p < .001$; SNR70: effect size = 11.8, SE = 1.65, t = 7.16, $p < .001$) and phonemes (SNR50: effect size = 9.7, SE = 1.73, t = 5.61, $p < .001$; SNR70: effect size = 9.12, SE = 1.62, t = 5.65, $p < .001$). Therefore, on average, Group C participants needed SNRs that ranged between 9 and 12 dB larger than Group B participants to perform similarly.

Differences between Groups A and B are also displayed in Figure 2. Significant differences were seen only for word SNR70 measures (SNR70: effect size = 3.7, SE = 1.5, t = 2.47, $p = .014$); other word (SNR50: effect size = 3.2, SE = 1.7, t = 1.9, $p = .058$) and phoneme (SNR50: effect size = 1.9, SE = 1.7, t = 1.12, $p = .266$; SNR70: effect size = 2.5, SE = 1.5, t = 1.66, $p = .098$) comparisons were not significant at a .05 α level. On average, the Group B participants needed an SNR that was 2 to 4 dB larger than Group A participants to perform similarly.

Figure 3 shows word and phoneme mean psychometric functions and word-to-phoneme correlations by group. Differences between word and phoneme scoring are evident in the figure with phoneme scores resulting in better performance than word scores at any given SNR but particularly at nonextreme SNRs at which floor and ceiling effects are present. Despite these differences in overall performance across word or phoneme scoring method, remarkable correlations between the measures are apparent (see Figure 3).

Pearson product–moment correlations were all significant ($p < .05$) with large correlation coefficients (Group A: SNR50 = 0.84, SNR70 = 0.84; Group B: SNR50 = 0.53, SNR70 = 0.75; Group C: SNR50 = 0.79, SNR70 = 0.91). The strength of these word-to-phoneme correlations suggests the possibility of using one measure in place of the other; however, it was important to test if different scoring methods resulted in different group effects. For this reason, scoring methods were compared as a function of group differences (e.g., Group B minus Group C for words vs. Group B minus Group C for phonemes). There was not a significant age effect difference between words and phonemes for SNR50 ($F(1, 297) = 1.86$, $p = .174$) or SNR70 ($F(1, 248) = 1.18$, $p = .278$). However, there were hearing loss effect differences between word and phoneme scores for SNR50 ($F(1, 297) = 6.23$, $p = .013$) and for SNR70 ($F(1, 248) = 4.57$, $p = .033$). This effect is demonstrated by larger word effect sizes (i.e., 12.3 and 11.8 dB) than phoneme effect sizes (i.e., 9.7 and 9.12 dB) when comparing Groups B and C above.

In summary, results from word and phoneme scoring used in this experiment indicated a large and significant effect of hearing loss on speech perception in noise. A smaller effect of age was significant or near significant when word scoring was used. It was not surprising that phoneme scoring led to better overall SNR50s and SNR70s but only modified the hearing loss effect significantly (no change on

Figure 3. Word and phoneme psychometric functions (left) and word–phoneme correlations (right) for all three groups for signals presented at 80 dB SPL (error bars represent standard errors of the mean). In general, the overall shape of the function is similar across scoring methods with phoneme scores being 10%–20% better than word scores. Word–phoneme correlations reveal strong relationships between scoring methods.
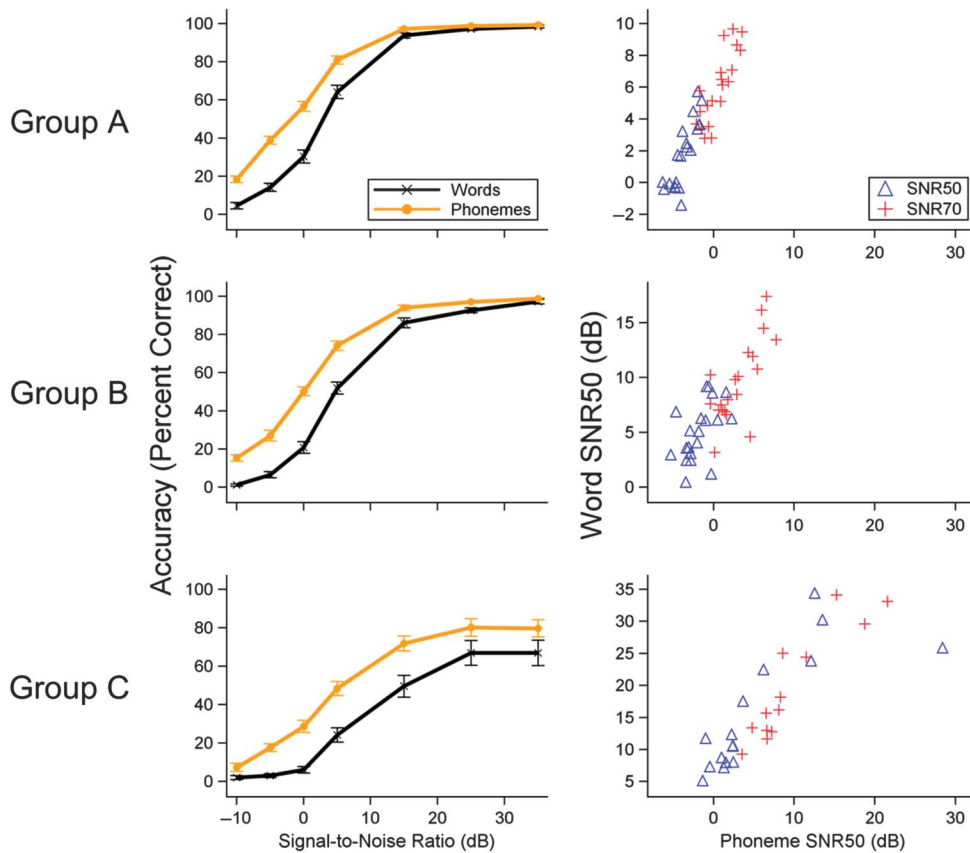
the age effect) such that word scoring resulted in a larger hearing loss effect than phoneme scoring.

Discussion

Considerable research and attention has been devoted to listening in adverse conditions generally and speech perception in background noise specifically. Contributing factors to performance include both those associated with degradation of the signal and/or source (e.g., characteristics of the speaker, interfering signals, environmental modifications) and those associated with limitations of the listener (e.g., peripheral auditory deficits, cognitive impairment, linguistic considerations; for a review, see Mattys, Davis, Bradlow, & Scott, 2012). Phoneme scoring may reduce test time and allow for additional trials (effectively reducing variability) or the acquisition of additional information about a patient's ability to understand speech in complex listening environments. This information could improve both hearing-related diagnosis and management of difficulties in background noise. Two motivating factors suggest the importance of gaining additional insight into an individual's ability to understand speech

in background noise: (a) Listening in noise is the primary complaint of hearing aid users (Kochkin, 2002) and one of the most difficult listening situations for many populations, such as those who are older and those with hearing loss; if these complaints are to be addressed effectively, the case can be made that clinical speech-in-noise testing must be completed (Wilson, 2011). (b) The idea of hidden hearing loss is emerging in the literature (Kujawa & Liberman, 2009), and it is likely that some of the functional difficulties associated with these subclinical deficits are present in clinical populations and may best be seen in more complex listening tasks, such as understanding speech in background noise.

In the current experiment, we implemented word and phoneme scoring of NU-6 words presented in background noise. As expected, an improvement in SNR resulted in an improvement in performance for all three groups. The psychometric function appeared to asymptote at 100% around 15 dB SNR for normal-hearing participants (Group A and Group B), whereas the asymptote for participants with hearing impairment (Group C) appeared to occur at higher SNRs, generally around 25 dB SNR, and topped out at 80% (see Figure 2). It is important to consider

the relevance of the SNRs tested in this study as they relate to real-world listening environments. Table 1 shows the performance of the three groups at several SNRs representative of everyday listening situations (i.e., home, store and hospital, cocktail party, and transportation; Pearsons, Bennett, & Fidell, 1977; Plomp, 1977; Markides, 1986). This table illustrates the extreme difficulty that older individuals experience at SNRs that are typical for very difficult listening situations, with accuracy remaining below 50% for words and phonemes in most signal-in-noise listening situations.

A significant and large effect of hearing loss was found (Group B vs. Group C). On average, a 9 to 12 dB better SNR was needed by older individuals with hearing impairment for their performance to be similar to more normally hearing older individuals. In contrast, the results of this study demonstrated that age effects were limited to word scoring when using an SNR70 measure. This result may be helpful when using SNR50 measures to compare a younger group with an older group that may or may not have hearing loss. Any differences in SNR50 measures would be likely due to hearing loss rather than age. For SNR70 measures, however, these data demonstrate that age is a small but significant contributor. On average, older, normally hearing individuals needed SNRs 2 to 4 dB better to perform similarly to their younger counterparts in background noise (see Figure 3). This suggests that if a 2- to 4-dB SNR improvement can be provided to more normally hearing older listeners through various methods (e.g., communication strategies, FM systems, directional microphones), performance may be comparable to younger listeners. It is interesting that recent data demonstrate that the just-noticeable difference in SNR is about 3 dB on average (McShefferty, Whitmer, & Akeroyd, 2015), indicating the importance of small changes in SNR.

It is apparent from this study that for NU-6 words, the effect size of hearing loss is at least two times larger than the effect size of age, consistent with other findings in the literature. Gordon-Salant and Fitzgibbons (1995) demonstrated a difference of 6.53 dB SNR when comparing younger, normal-hearing with younger, hearing-impaired groups and of 5.23 dB when comparing older, normal-hearing with older, hearing-impaired groups. In contrast, the difference was 0.00 dB when comparing normally

hearing individuals from different age groups and 1.3 dB when comparing individuals with hearing impairment from different age groups; however, the effect sizes of the current study were about twice the size. Souza and Turner (1994) presented speech-shaped and high-pass noise to various participant groups and found that the younger, normal-hearing group achieved an approximate mean of 82% accuracy; the younger, hearing-impaired group was approximately 60%; and the older, hearing-impaired group was approximately 66%. This resulted in a difference of 22% when comparing the normal-hearing and hearing-impaired groups and only a 6% difference when comparing the younger and older groups.

Phoneme scoring guidelines were established as part of this study (see Appendix), which may be useful in the future clinically or in research to facilitate phoneme scoring. There are potentially multiple advantages to phoneme scoring, including decreased test time due to increased number of tokens per word, information about specific confusions or difficult-to-perceive sounds that could be used clinically, and reduced influence from top-down processes. In fact, the modulation of top-down influences may have modified the scoring effect (i.e., hearing loss effect sizes were larger for word scoring than phoneme scoring) such that when word scoring was used, older, normally hearing individuals were able to take greater advantage of context cues than their counterparts with hearing impairment, possibly relating to the limited cognitive resources caused by the presence of hearing loss.

A clear advantage of phoneme scoring in this study was less missing data. The vertical shift of the phoneme functions relative to the word functions (see Figure 3) resulted in three additional SNR50 data points and two additional SNR70 data points. In clinical cases in which individuals with hearing impairment are tested, phoneme scoring will allow for more complete psychometric functions or fewer missing SNR50/SNR70 values. Another potential advantage of phoneme scoring is the increased number of tokens compared with word scoring, resulting in increased test–retest reliability and decreased variability (Boothroyd, 2008; Gelfand & Gelfand, 2012). The decreased variability can be seen in Figure 3; in general, the spread of data as shown by error bars reveals phoneme results to be a more precise measurement relative to word results.

Table 1. Phoneme and word recognition accuracy (percentage correct) for each group at specific signal-to-noise ratios (SNRs) with representative listening environments.

Groups	−8 dB SNR (transportation)	−4 dB SNR	0 dB SNR (cocktail party)	+4 dB SNR (store/hospital)	+8 dB SNR (home)
Phonemes					
Group A	23	48	68	80	87
Group B	16	38	58	73	82
Group C	10	21	35	46	56
Words					
Group A	5	17	38	61	77
Group B	2	9	24	46	64
Group C	2	5	11	20	30

Although correlations between word and phoneme scoring method were very strong, differences between the measures were apparent. Figure 3 shows that the psychometric function asymptotes for the older, hearing-impaired group occurred at higher accuracies when phonemes were scored (about 80%) than when word scoring was used (about 70%). In contrast, asymptotes for the more normally hearing groups were similar for words and phonemes (i.e., near 100% for both groups). Given these observations, it was not surprising that differences between phoneme and word scoring were only significant when hearing impairment effects were considered (i.e., older, normal-hearing vs. older, hearing-impaired groups).

Conclusions

Phoneme scoring may be applied to current NU-6 word lists using the scoring guidelines provided (see Appendix), and phoneme scoring may be advantageous clinically and in research when compared with word scoring. Advantages of phoneme scoring include shorter test times or an increased number of tokens resulting in more precise measurement. It is interesting that phoneme scoring modified hearing loss effect sizes but not aging effect sizes. Older individuals with hearing impairment needed about a 10-dB better SNR to perform similarly to more normally hearing older individuals, and older, normally hearing individuals needed about a 3-dB better SNR to perform similarly to younger, normal-hearing individuals.

Acknowledgments

References

American National Standards Institute. (1991). *Maximum permissible ambient noise levels for audiometric test rooms* (ANSI S3.1-1991). New York, NY: Author.

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech, Language, and Hearing Research, 41,* 549–563.

Billings, C. J., Penman, T. M., McMillan, G. P., & Ellis, E. (2015). Electrophysiology and perception of speech in noise in older listeners: Effects of hearing impairment and age. *Ear and Hearing, 36,* 710–722.

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal of Audiology, 2,* 3–10.

Boothroyd, A. (2008). The performance/intensity function: An underused resource. *Ear and Hearing, 29,* 479–491.

Collett, D. (1991). *Modelling binary data.* London, England: Chapman & Hall.

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and education level. *JAMA: The Journal of the American Medical Association, 269,* 2386–2391.

Divenyi, P. L., & Haupt, K. M. (1997). Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects. *Ear and Hearing, 18,* 42–61.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America, 76,* 87–96.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Gelfand, S. A., & Gelfand, J. T. (2012). Psychometric functions for shortened administrations of a speech recognition approach using tri-word presentations and phonemic scoring. *Journal of Speech, Language, and Hearing Research, 55,* 879–891.

Gelfand, S. A., Piper, N., & Silman, S. (1986). Consonant recognition in quiet and in noise with aging among normal hearing listeners. *The Journal of the Acoustical Society of America, 80,* 1589–1598.

Gordon-Salant, S., & Fitzgibbons, P. J. (1995). Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index. *Journal of Speech and Hearing Research, 38,* 706–713.

Humes, L. E., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. *Journal of Speech and Hearing Research, 34,* 686–693.

Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research, 33,* 726–735.

Institute of Electrical and Electronic Engineers. (1969). *IEEE recommended practice for speech quality measures.* New York, NY: Author.

Kochkin, S. (2002). MarkeTrak VI: Consumers rate improvements sought in hearing instruments. *The Hearing Review, 9,* 18–22.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *JNeurosci: The Journal of Neuroscience, 29,* 14077–14085.

Markides, A. (1978). Whole-word scoring versus phoneme scoring in speech audiometry. *British Journal of Audiology, 12,* 40–46.

Markides, A. (1986). Speech levels and speech-to-noise ratios. *British Journal of Audiology, 20,* 115–120.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27,* 953–978.

McCreery, R., Ito, R., Spratford, M., Lewis, D., Hoover, B., & Stelmachowicz, P. G. (2010). Performance-intensity functions for normal-hearing adults and children using CASPA. *Ear and Hearing, 31,* 95–101.

McShefferty, D., Whitmer, M. W., & Akeroyd, M. A. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends in Hearing, 19,* 1–9.

Olsen, W. O., Van Tasell, D. J., & Speaks, C. E. (1997). Phoneme and word recognition for words in isolation and in sentences. *Ear and Hearing, 18,* 175–188.

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). *Speech levels in various noise environments.* Washington, DC: U.S. Environmental Protection Agency.

Plomp, R. (1977). Acoustical aspects of cocktail parties. *Acoustica, 38,* 186–191.

Souza, P. E., & Turner, C. W. (1994). Masking of speech in young and elderly listeners with hearing loss. *Journal of Speech and Hearing Research, 37,* 655–661.

Stuart, A., & Phillips, D. P. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear and Hearing, 17,* 478–489.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., & Gray, G. A. (1997). Age-related changes in monosyllabic word recognition performance when audibility is held constant. *Journal of the American Academy of Audiology, 8,* 150–162.

Summers, V., & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. *Journal of Speech, Language, and Hearing Research, 47,* 245–256.

Wiley, T. L., Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Klein, R., & Klein, B. E. K. (1998). Aging and word recognition in competing message. *Journal of the American Academy of Audiology, 9,* 191–198.

Wilson, R. H. (2011). Clinical experience with the Words-in-Noise Test in 3430 veterans: Comparisons with pure-tone thresholds and word recognition in quiet. *Journal of the American Academy of Audiology, 22,* 405–423.

Appendix

Phoneme Scoring Guidelines

Northwestern University Auditory Test Number 6 (NU-6) words consist of three syllables in a consonant-vowel-consonant structure. Given that targets and responses can consist of combinations of consonants and vowels (e.g., clusters, dipthongs, etc.), the three syllables are referred to linguistically as *onset* (initial grouping of consonants), *nucleus* (middle vowel portion), and *coda* (final grouping of consonants). For phoneme scoring, each element (i.e., onset, nucleus, and coda) is awarded 1 point if correct with a few scoring exceptions as follows:

1) Consonants (forming consonant clusters) are grouped with the onset or the coda rather than with the nucleus, with the exception of the [r]-colored schwa, which is treated as a single vowel sound [ɚ].

Examples:	Target	Response	
	search [sɚtʃ]	teach [titʃ]	(1 point for coda)
	shirt [ʃɚt]	short [ʃort]	(1 point for coda)
	shawl [ʃɑl]	sharp [ʃɑrp]	(2 points for onset and nucleus)

2) Vowel coloring (e.g., by laterals and glides) should be disregarded if possible. Try to determine the status of the vowel before it is influenced by the following consonant.

Examples:	Target	Response	
	goal [goʊl]	goat [goʊt]	(2 points for onset and nucleus)
	shall [ʃæl]	shout [ʃaʊt][a]	(1 point for onset)

3) In the case of two-syllable responses, the stressed vowel of the response (i.e., response nucleus) will be compared with the vowel in the target (i.e., target nucleus). Any phoneme(s) before or after the response nucleus are considered onset or coda, respectively, and will be compared with the target onset or coda.

Examples:	Target	Response	
	pole [poʊl]	poet [poʊɪt]	(2 points for onset and nucleus)
	hall [hɑl]	appall [əpɑl]	(2 points for nucleus and coda)

4) Due to dialectical differences,[b] certain phonetic distinctions are collapsed in certain conditioning environments. This collapsing is such that the variances in pronunciation shown below are treated as allophones of single phonemes.

• No distinction is made between the low back rounded [ɔ] and unrounded [ɑ] vowel in any environment.

Examples:	Target	Response	
	wall [wɔl]	far [fɑr]	(1 point for nucleus)
	jar [jɑr]	jaw [jɔ]	(2 points for onset and nucleus)

• No distinction is made between the high front tense [i] and lax [ɪ] vowel in the environment before a velar nasal consonant [ŋ], before the consonant cluster C[+ nasal]C[+ velar], or in other environments where they may be considered allophones.

Examples:	Target	Response	
	sing [sɪŋ]	king [kɪŋ]	(2 points for nucleus and coda)
	think [θɪnk]	seem [sim]	(1 point for nucleus)[c]
	link [link]	thin [θɪn]	(1 point for nucleus)

• No distinction is made between the mid front lax vowel [ɛ] and the mid front diphthong [eɪ] in the environment before a voiced velar consonant ([ŋ] or [g])

Examples:	Target	Response	
	beg [bɛg]	leg [leɪg]	(2 points for nucleus and coda)
	length [lɛŋθ]	strength [streɪŋθ]	(2 points for nucleus and coda)

[a]Diphthongs should be treated as single vowel sounds and cannot be broken into two component parts. [b]Dialectical differences between the NU-6 speaker and the listener must also be considered. [c][i] and [ɪ] are considered allophones of the same phoneme.