

Effects of the Carrier Phrase on Word Recognition Performances by Younger and Older Listeners Using Two Stimulus Paradigms

Richard H. Wilson¹ Victoria A. Sanchez²

¹Department of Speech and Hearing Sciences, Arizona State University, Tempe, AZ

²Department of Otolaryngology-Head & Neck Surgery, University of South Florida, Tampa, FL

Address for correspondence Richard H. Wilson, Department of Speech and Hearing Sciences, Arizona State University, Tempe, AZ 85281 (e-mail: wilsonr1943@gmail.com).

J Am Acad Audiol 2020;31:412–441.

Abstract

Background In the 1950s, with monitored live voice testing, the vu meter time constant and the short durations and amplitude modulation characteristics of monosyllabic words necessitated the use of the carrier phrase amplitude to monitor (indirectly) the presentation level of the words. This practice continues with recorded materials. To relieve the carrier phrase of this function, first the influence that the carrier phrase has on word recognition performance needs clarification, which is the topic of this study.

Purpose Recordings of Northwestern University Auditory Test No. 6 by two female speakers were used to compare word recognition performances with and without the carrier phrases when the carrier phrase and test word were (1) in the same utterance stream with the words excised digitally from the carrier (VA-1 speaker) and (2) independent of one another (VA-2 speaker). The 50-msec segment of the vowel in the target word with the largest root mean square amplitude was used to equate the target word amplitudes.

Research Design A quasi-experimental, repeated measures design was used.

Study Sample Twenty-four young normal-hearing adults (YNH; $M = 23.5$ years; pure-tone average [PTA] = 1.3-dB HL) and 48 older hearing loss listeners (OHL; $M = 71.4$ years; PTA = 21.8-dB HL) participated in two, one-hour sessions.

Data Collection and Analyses Each listener had 16 listening conditions (2 speakers \times 2 carrier phrase conditions \times 4 presentation levels) with 100 randomized words, 50 different words by each speaker. Each word was presented 8 times (2 carrier phrase conditions \times 4 presentation levels [YNH, 0- to 24-dB SL; OHL, 6- to 30-dB SL]). The 200 recorded words for each condition were randomized as 8, 25-word tracks. In both test sessions, one practice track was followed by 16 tracks alternated between speakers and randomized by blocks of the four conditions. Central tendency and repeated measures analyses of variance statistics were used.

Results With the VA-1 speaker, the overall mean recognition performances were 6.0% (YNH) and 8.3% (OHL) significantly better with the carrier phrase than without the carrier phrase. These differences were in part attributed to the distortion of some words caused by the excision of the words from the carrier phrases. With the VA-2 speaker, recognition performances on the with and without carrier phrase conditions

Keywords

- ▶ auditory perception
- ▶ carrier phrase
- ▶ normal hearing for pure tones
- ▶ sensorineural hearing loss
- ▶ speech perception
- ▶ word recognition

by both listener groups were not significantly different, except for one condition (YNH listeners at 8-dB SL). The slopes of the mean functions were steeper for the YNH listeners (3.9%/dB to 4.8%/dB) than for the OHL listeners (2.4%/dB to 3.4%/dB) and were <1%/dB steeper for the VA-1 speaker than for the VA-2 speaker. Although the mean results were clear, the variability in performance differences between the two carrier phrase conditions for the individual participants and for the individual words was striking and was considered in detail.

Conclusion The current data indicate that word recognition performances with and without the carrier phrase (1) were different when the carrier phrase and target word were produced in the same utterance with poorer performances when the target words were excised from their respective carrier phrases (VA-1 speaker), and (2) were the same when the carrier phrase and target word were produced as independent utterances (VA-2 speaker).

Introduction

This study was designed to determine the influence that the carrier phrase has on the word recognition performance of recorded monosyllabic words by both young adults with normal hearing for pure tones (YNH) and older adults with sensorineural hearing loss (OHL). Why do we typically have a carrier phrase for testing word recognition abilities? Two articles help answer this question. First, in describing their “New Standard Articulation Lists,” Fletcher and Steinberg¹⁵ (page 810) used a random card selection algorithm to compile nonsense syllables of the *con-vow-con* type. (Note: *con-vow-con* refers to [CVC or CNC]—consonant-vowel nucleus-consonant syllable structure). To make the test paradigm similar to connected speech, a pool of *short introductory sentences* was compiled to each of which was appended one of the CVC syllables. Example introductory sentences included (page 813): *Listen carefully to ____, I am about to say ____, and Thirteen will be ____*. These introductory sentences likely were the genesis of the carrier phrase as we know it today. Second, in the classic article describing the development of the phonetic balance (PB)-50 word lists, Egan¹³ (page 977) stated that the *carrier sentence* (e.g., *You will write*) was necessary for the following three reasons: (a) to alert the listener of the forthcoming test item (word), (b) to agitate the particles of carbon in the microphone, and (c) to enable the speaker to modulate his voice so as to keep the level of his voice even from word to word. Egan goes on to say: “For most purposes, the carrier sentence and not the test item should be used to monitor voice level. Thus, no attempt should be made to compensate for the typical differences in the speech power used in pronouncing the different sounds in the test items. When only the carrier sentence is monitored, the test item should be spoken with the same general effort as the rest of the carrier sentence.”

Historically, the carrier phrase, which typically takes the form of *You will say ____* or *Say the word ____,* has been used with word recognition testing since the dawn of speech audiometry. In the mid-20th century, monitored live voice (MLV) was primarily used as the presentation mode of

speech materials with the volume unit (vu) meter (ASA³) providing a visual monitoring of the signal amplitude (a vintage vu meter is shown in Supplemental Figure S1, supplemental to the online version of this article). Before the evolution of digital signal processing in the 1980s, several studies considered the influence that the carrier phrase had on word recognition performance. For the most part, these early studies were fairly well conducted considering the analog instrumentation and techniques available at the time to manipulate the speech signals. MLV, which among other things has reliability issues, was used by some studies. When recorded materials were used, magnetic tape was the medium that for signal manipulation required mechanical cutting and splicing, which lacked the precision and versatility available with digital waveform editors. These predigital studies are summarized in the following text and detailed along with the role of the vu meter in speech audiometry in the Appendix.

Martin et al²² recorded the Central Institute for the Deaf (CID) W-22 words (Hirsh et al²³) both with and without a carrier phrase and presented them to YNH and OHL listeners at 30 dB, re: their speech recognition thresholds (SRT), i.e., 30-dB sensation level (SL). At that single, high presentation level, word recognition performance was not affected by the presence or absence of the carrier phrase, but when surveyed, 14 of the 30 OHL participants preferred listening with the carrier phrase. There were two major issues with this study. First, different words were used with the two carrier phrase conditions. Second, a single relatively high presentation level was used that produced maximum recognition performances for all conditions. Gladstone and Siegenthaler¹⁹ used three carrier phrase conditions and a no carrier phrase condition for studying 25 CID W-22 words they recorded. The materials were presented to YNH listeners at 5-dB SL, re: the SRT. With the carrier phrase conditions, mean recognition performances ranged from 47.2% to 56.4%, whereas with the no carrier phrase condition, performance was 40%. This finding is understandable as the carrier phrases at this low presentation level probably served to define the listening intervals for the target

words, which resulted in slightly better performances. In another study (Gelfand¹⁸), the CID W-22 lists were presented by MLV at 35-dB SL (re: the SRT) to 50 adult listeners with sensorineural hearing loss. Each listener received four lists, one with and one without a carrier phrase in each ear. Mean recognition performances were significantly better with the carrier phrase (73.9%) than without the carrier phrase (69.2%), which is about a two-word difference. In addition to the single, high presentation level, each word utterance was unique (MLV), which precluded precise control of the presentation levels of each utterance, especially utterances that did not include a carrier phrase.

Lynn and Brotman²⁷ stepped up the technical sophistication by studying the Hirsh recordings of the 27 CID W-22 words whose initial consonants were voiceless stops. The voiceless stop provided a silent interval in which the splice of magnetic tape could be made separating the carrier phrase from the target word without generating a click. The words were presented with and without a carrier phrase in speech spectrum noise (SSN) at a 0-dB signal-to-noise ratio (S/N, SNR) to 20 YNH listeners. Recognition performance was 10% better with the carrier phrase (37.4%) than without the carrier phrase (27.5%). Thus, in a difficult listening condition (0-dB S/N), the carrier phrase enhanced word recognition performance over presentation of the target words without the carrier phrase, which may be attributable to at least two reasons: (a) the carrier phrase helped define the listening interval in the degraded speech task and (b) possible informational cues in the “silent interval” between the end of the carrier phrase and the word onset may have been lost when the tape splice occurred. The importance in this study is that the comparisons were made with the same utterance of each target word. Coupled with earlier data, the authors suggested that the carrier phrase contained intelligibility cues that made a positive contribution to the recognition of the target words. Although a signal light was used with the words without a carrier phrase to define visually the listening interval, consideration must be given to the likelihood that the audible carrier phrase enhanced recognition performance in the adverse listening condition used in the experiment (0-dB S/N) by defining the listening interval in the same modality as the target word. In addition, the carrier phrase may have provided a normalization sample for the target word (e.g., Bladon et al⁶; Pisoni³⁷; Johnson²⁴) or perceptual attunement for the target word (Markham and Hazan³⁰).

More recently, Bonino⁵ evaluated recognition performance on the Auditec version of the PBK words (Haskins²²) presented with and without a carrier phrase. The without carrier phrase condition was constructed digitally by replacing the carrier phrase with a silence interval while maintaining the structure of the three 50-word PBK lists. Three masking conditions were studied (20-talker babble, 2-talker babble, and SSN) using different words for each condition and combinations of three SNRs. The general finding for the two babble conditions was that performance was better when the carrier phrase was used than when the words were presented without a carrier phrase. With the two babble conditions, recognition performance was 7.1-17.8%

better when the carrier phrase was used, but with SSN, performance was the same with and without the carrier phrase. Of the two babble conditions, the performance difference between the carrier phrase and no carrier phrase conditions was 10% larger with the 2-talker babble, which was a difference attributed to a larger perceptual (informational) masking component associated with the 2-talker babble (Carhart et al⁸; Pollack³⁹). The authors reasoned that the better performance with the carrier phrase was owing to an *auditory grouping cue* that the carrier phrase provided for the target word in the two babble conditions, which was a cue that provided little or no information to the target word in the energetic SSN masker that lacked an informational masking component.

To study the influence that the carrier phrase has on word recognition performance in a quiet listening condition, the current study purposed to improve on the limitations of the previous studies by using two paradigms of monosyllabic CNC words presented at multiple presentation levels to both YNH listeners and OHL listeners. One paradigm involved recording the carrier phrase and word as a single utterance, whereas the second paradigm involved recording the carrier phrase and word as independent utterances. The traditional measures of central tendency were examined as well as the recognition performances (a) by the individual listeners and (b) on the individual words.

Methods

From the literature, then, there are a variety of answers to the question of whether or not the carrier phrase influences word recognition performance. The current study compared word recognition performances on the Northwestern University Auditory Test No. 6 (NU-6; Tillman and Carhart⁴⁶) for the identical word utterances both with a carrier phrase and without a carrier phrase. Two speakers were studied, each with different carrier phrase and target word paradigms. The first speaker (VA-1) used the carrier phrase, *Say the word ___* (Causey et al⁹; Department of Veterans Affairs¹²). In this recording, each of the 200 NU-6 carrier phrase and target word utterances was unique. The second speaker (VA-2) recorded the NU-6 words as part of a pool of words used to evaluate several sets of word recognition materials spoken by the same speaker (McArdle and Wilson³³; Wilson et al⁴⁹). The carrier phrase, *You will cite*, was followed by three utterances of the target word, two of which were discarded. For the final version of the VA-2, NU-6 words, the same carrier phrase utterance followed by a 150-msec silent interval preceded each of the 200 words. Thus, with the VA-1 speaker, the carrier phrase and target word were one utterance (one acoustic stream), whereas with the VA-2 speaker, the carrier phrase was one utterance and the target word was a second utterance (two independent acoustic streams) that could be conjoined digitally.

Materials

Both versions of NU-6 investigated were spoken by professional female talkers (VA-1 and VA-2) and originally were recorded

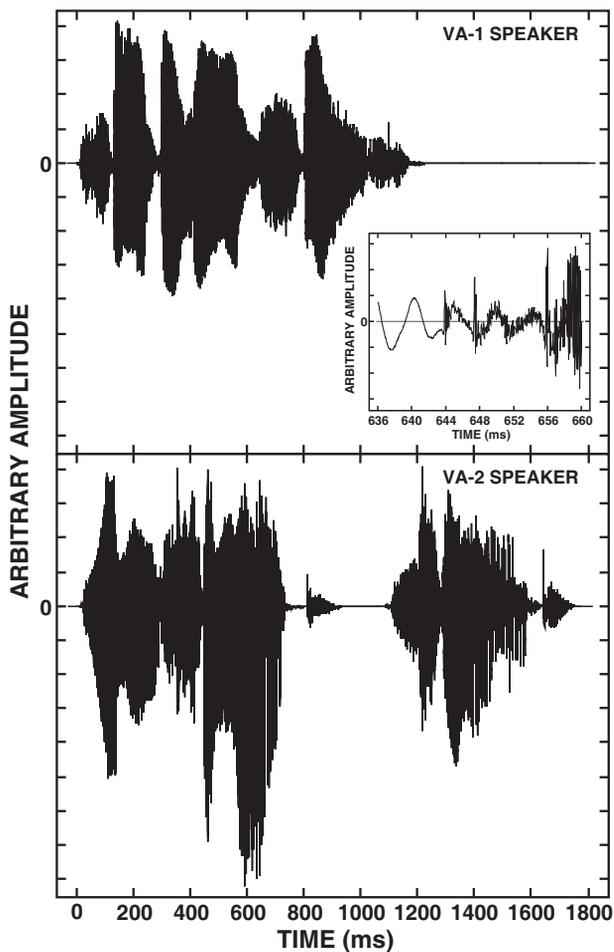


Fig. 1 Example waveforms of *Say the word said* (VA-1 speaker, upper panel) and *You will cite said* (VA-2 speaker, lower panel) are shown. The inset in the upper panel is the transitional segment of the /d/ in *word* transforming into the /s/ in *said*. A higher resolution version of the upper panel inset is illustrated in Supplemental Figure S2.

using the ANSI (2010) prescribed method of monitoring the level of the carrier phrase on a vu meter with the target word uttered in a natural manner following the carrier phrase. Because of a substantial amplitude asymmetry in the carrier phrase of the VA-2 recording (Robjohns⁴²), which can be seen in the lower panel of **Figure 1** around 600 msec, the overall levels of the VA-2 materials were out of necessity displaced further from the maximum digitization range of the waveform editor (Adobe Audition, CS6; Adobe Systems, Inc.¹) than were the overall levels of the VA-1 materials. This amplitude difference in the two materials was reflected by the 1000-Hz calibration tones that differed an average of 5.7 dB with the VA-2 recording lower. With consideration to headroom issues, the amplitude asymmetry in the VA-2 carrier phrase was the primary energy source that peaked the vu meter for those materials, which as can be seen in the lower panel of **Figure 1** produced artificially small root-mean-square (rms) amplitudes of the VA-2 target words in relation to the amplitude of the carrier phrase. Because the time constant of a vu meter precluded an accurate amplitude measure of the monosyllabic words (Wilson⁴⁷), the following procedure was used to equate

the amplitudes of the words in the VA-1 and VA-2 versions of NU-6. The intent of this calibration procedure was to specify the amplitude of the target word in a manner that was replicable. First, the 50-msec segment of the vowel in each word with the largest sustained rms amplitude was identified and quantified (rms), which usually, but not always, corresponded to the maximum envelope amplitude. For the VA-1 speaker, the mean 50-msec vowel segment amplitude for the 200 NU-6 words was -9.7 dB (re: maximum digitization range, standard deviation [SD] = 1.2 dB), and for the VA-2 speaker, the mean was -15.4 dB (SD = 0.4 dB), which interestingly and somewhat coincidentally is the same 5.7-dB difference observed with the two calibration tones. Because the amplitude asymmetry of the VA-2 carrier phrase prohibited amplification of VA-2 to the levels of the VA-1 version, the amplitude of each target word by both speakers was adjusted so that the 50-msec maximum vowel segment rms approximated -15.4 dB, re: maximum digitization range. The rms of the 1000-Hz calibration tone was -17.0 dB, re: maximum digitization range. For the VA-1 NU-6 materials, the level of each carrier phrase and target word utterance was adjusted, but for the VA-2 materials, the amplitudes of the target words and the common carrier phrase used with each word were adjusted independently. This recalibration of the materials had no overall effect on the relation between the calibration tone and the average level of the words with the VA-2 version but lowered the level of the VA-1 materials by an average of 5.7 dB with respect to the level of the calibration tone. In this manner, the calibration of the VA-1 materials was altered from the original standard.

The waveforms in **Figure 1** illustrate examples of the carrier phrases and target word *said* by both speakers. The waveform in the upper panel by the VA-1 speaker is *Say the word said*, which has a total duration of 1,189 msec, of which 655 msec is the carrier phrase and 531 msec the target word. Although the segment boundary between the carrier phrase and the target word in the upper panel at about 650 msec is obscure in the overall waveform, an increase in the temporal resolution of the waveform around the segment boundary reveals waveform details that are depicted in the **Figure 1** inset with an enhanced version provided in Supplemental Figure S2 (supplemental to the online version of this article). The 24-msec waveform segment in the insert clearly shows the transition from the /d/ in *word* at 642 msec to the /s/ in *said* at 658 msec. During this 16-msec coarticulation segment (Fowler^{16,17}; Redford et al⁴⁰), the periodic /d/ waveform has the random/s/ waveform superimposed on it. In this example, *said* was excised from the carrier phrase at 658 msec into the 1,189-msec waveform. A minority of the 200 NU-6 words by the VA-1 speaker were characterized by some degree of coarticulation with the carrier phrase. For the most part, the words coarticulated with the carrier phrase were easily separated from their carrier phrase while maintaining perceptual clarity. A few words (probably <10), however, were difficult to separate from their carrier phrases with maintained intelligibility, which is similar to the degradation of intelligibility that occurs with some words when they are excised from sentences or conversational speech (Pickett and Pollack³⁸). These words had intelligibility clarity,

except for some distortion of the initial consonant. The intelligibility of these few VA-1 words was expected to be reflected in poorer recognition performances when the words were spoken without the carrier phrase than when the words were spoken with their parent carrier phrase. The lower panel of **Figure 1** depicts the 1,772-msec waveform, *You will cite said*, produced by the VA-2 speaker. The 950-msec carrier phrase is separated from the 672-msec target word *said* by a 150-msec silent interval. The 150-msec silent interval between the /t/ in *cite* and the /s/ in *said* is apparent and by design void of any coarticulation. With the VA-2 speaker, the target words were independent of the carrier phrase.

Once the materials were set to the same relative amplitudes, three replications of the level adjusted files were made and attenuated (-8, -16, and -24 dB, respectively) using a batch processing routine in the waveform editor. The design of the study involved 3,200 stimulus words (200 words, 2 speakers, 2 carrier phrase conditions, and 4 presentation levels). From past experiences and for practical reasons, participant involvement was limited to two, one-hour sessions in each of which 25 practice items and 400 test items were administered. To accommodate the number of stimuli and the practical limitations, a “foursome” design paradigm evolved to produce 800 stimuli over 2 test sessions for each participant in the quartet. The 800 stimuli for each listener involved two, 50-word lists (1 list for each speaker with different words in each list), 2 carrier phrase conditions (with and without the carrier phrase), and 4 presentation levels. For each speaker and carrier phrase condition, the 50 words at the 4 presentation levels (200 stimuli) were randomized and concatenated into eight, 25-word tracks (the waveform of an example track is shown in Supplemental Figure S3, supplemental to the online version of this article). Each 25-word track contained randomized words at randomized presentation levels from one speaker in one of the carrier phrase conditions. In this manner, each word was presented to the listener eight times (four presentation levels \times two carrier phrase conditions). Two reasons tempered concerns about excessive exposure to the stimulus words, i.e., learning effects. First, the design involved a range of randomized low and high presentation levels and the order of the words was randomized. Second, as Miller et al³⁵ and later Thwing⁴⁵ demonstrated, even successive presentations of the same word at the same level result in only modest (2 dB or so) improvements in word recognition performance with the largest improvements made (a) between the 1st and 2nd presentations, and (b) at the higher presentation levels. The order of the four sets of eight tracks then was randomized for each participant and compiled in a counterbalanced scheme in which each of the four sets of tracks was ordered randomly before the subsequent sequence of four tracks was ordered randomly. The only restriction was the odd numbered participants started with the VA-1 materials and the even numbered participants started with the VA-2 materials, after which the speaker tracks were alternated. The 25-word practice lists were taken from the list of words/conditions used with another listener

in the foursome; thus, the practice words were never the test words for a specific listener. The tracks incorporated a 3.5-sec interstimulus interval and had 142-210 sec durations depending on the carrier condition and the speaker.

Participants

The 24 YNH listeners (19 females), who were recruited from the local university community, ranged in age from 20 to 32 years ($M = 23.5$ years; $SD = 2.8$ years). The YNH participants had pure-tone thresholds at the octave frequencies ≤ 20 -dB HL (ANSI²) with a 3-frequency, pure-tone average (PTA; 500, 1000, and 2000 Hz) of 1.3-dB HL ($SD = 4.8$ dB) in Session 1 and 1.8-dB HL ($SD = 4.9$ dB) in Session 2. The 48 OHL listeners (27 females) with sensorineural hearing loss ranged in age from 60 to 82 years and met the following inclusion criteria for the test ear: (a) 60-85 years of age, (b) English was their first language, (c) 500-Hz thresholds ≤ 35 -dB HL, (d) 1000-Hz thresholds ≤ 45 -dB HL, and (e) a PTA of < 42 -dB HL. The OHL participants ($M = 71.4$ years; $SD = 5.1$ years) were recruited from a clinical and research database and from study announcements posted in paper and electronic forms. The OHL listeners had mean PTAs for the test ear of 21.8-dB HL ($SD = 9.3$ dB) and 21.5-dB HL ($SD = 9.3$ dB) in Sessions 1 and 2, respectively. The mean pure-tone thresholds and SDs for the test ear of the two groups of listeners in Session 1 are shown in **Figure 2**. The individual demographic data are provided in Supplemental Table S1 (supplemental to the online version of this article), for each group including age, pure-tone thresholds, and PTAs. All participants were cognitively intact and scored higher than 23 on the Montreal Cognitive Assessment (Nasreddine et al³⁶). Educationally, 6.3% of the OHL listeners were high-school graduates, 35.4% had some college degrees, 22.9% had

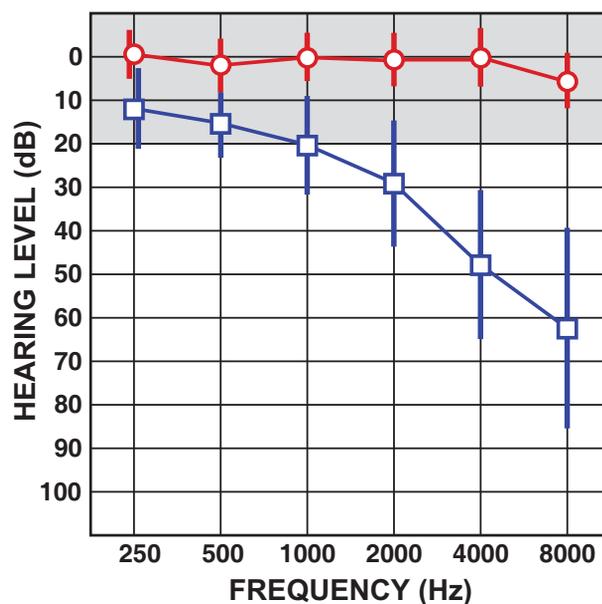


Fig. 2 The mean test ear pure-tone audiograms for the 24 YNH listeners and the 48 OHL listeners involved in the study. The vertical lines represent ± 1 SD. The individual participant thresholds for the test ear in both test sessions are listed in Table S1.

undergraduate degrees, and 35.4% had one or more graduate degrees. The University of South Florida Institutional Review Board approved all recruitment, screening, experimental, and compensation procedures before the start of any study activities. The participants were compensated for their time with a \$15 dollar gift certificate following each session.

Procedures

Two, 1-hour test sessions were conducted over a 1- to 98-day interval (YNH, $M = 8.5$ days, $SD = 5.5$ days; OHL, $M = 17.0$ days, $SD = 23.4$ days). In Session 1, the consent process was completed, demographic data were collected, and pure-tone thresholds were established with the automated method for testing auditory sensitivity (AMTAS™, Margolis et al^{28,29}) procedure using a tablet (Optiplex 780; Dell, Round Rock, TX) in conjunction with an audiometer (Audiostar Pro; Grason-Stadler, Eden Prairie, MN) and Sennheiser HD 200 earphones (Hanover, Germany). The test protocol then was explained, the protocol instructions given, and questions answered, followed by a practice list and 16 experimental lists. A break was provided following presentation of the eighth word list with other breaks provided as requested by the participant, which seldom happened. The protocols in Session 2 were identical with an AMTAS recheck of the 500-, 1000-, and 2000-Hz thresholds, followed by a review of the test protocol, instructions, a practice list, and the remaining 16 experimental lists. Both sessions were conducted in the same double-walled sound booth (IAC Acoustics, 120act, North Aurora, IL).

The *.wav files of speech materials were reproduced by a Windows Media Player on a computer (Optiplex 7010; Dell) and routed through an audiometer (Equinox, Model AC440; Interacoustics, Middelfart, Denmark) to a TDH-39P earphone encased in a supra-aural cushion. Four presentation levels were used with each listener group (0-, 8-, 16-, and 24-dB SL [YNH] and 6-, 14-, 22-, and 30-dB SL [OHL]) with the PTA at 500, 1000, and 2000 Hz from Session 1 used as the reference. The nontest ear was covered with a dummy earphone. The verbal responses of the listeners were recorded in a spreadsheet.

Results and Discussion

The mean data for the two carrier phrase conditions, the two speakers, and the two participant groups are presented with some discussion about the various relations among the variables including statistical analyses. Remember that for each participant, the same 50-word utterances at the same presentation levels were used for both carrier phrase conditions with different 50-word sets used for the two speakers, which made each word its own control. Because the mean data often obscure important underlying features in the individual participant data and in the individual word data, those two variables were examined in detail. Although not posed as primary questions in the current study, the NU-6 data also were evaluated with respect (a) to the recognition performances on the VA-1 and VA-2 versions of NU-6 with

the carrier phrase, (b) to the performances by the female and male OHL listeners, (c) to previous data on the VA-1 version with the carrier phrase, and (d) to the recognition performances on the VA-1 and VA-2 versions parsed into the traditional randomization A 25- and 50-word lists.

Mean Results and Statistical Analyses

The mean recognition performance functions for the two carrier phrase conditions, listener groups, the speakers are illustrated in the upper four panels of ►Figure 3 with a comparison of the data from the two speakers in the with carrier phrase condition depicted in the lower two panels. A numeric summary of the data in ►Figure 3 is provided in ►Table 1, including the mean overall recognition performances across the four presentation levels, the 50% points calculated from the polynomial equations used to describe the mean data, and the slopes of the mean functions at the 50% points calculated from the 1st derivatives of the polynomial equations. (Note: As Egan¹³ [page 961] indicated, word recognition performance is not “uniformly sensitive” over the range of possible performances with maximum sensitivity around the 50% point where performances are most variable, contrasted to the extremes of the function where performances are least variable because of floor and ceiling effects. For this reason, the 50% point is the ideal point on a function to evaluate differences between variables. Also, the slope of the function at the 50% point provides an accurate estimate of the traditional linear slope of a word recognition function between the 20% and 80% correct points [$m = \Delta y/\Delta x$]. Importantly, as Wilson and Margolis⁴⁸ demonstrated, the best estimate of the slope of a function is not the slope of the mean function but rather the mean of the slopes of the individual functions that comprise the mean function, which is steeper than the slope of the mean function. In addition, examination of the functions of the individual listeners/conditions provides measures of variability not observable with the mean functions.) The mean percent correct recognition data for the two carrier phrase conditions at each presentation level are listed in ►Table 2 along with the SDs for participants and for words along with the differences (Δ) between the mean recognition performances with carrier phrases minus the performance without carrier phrases and their associated level of significance. The individual listener data are listed in Supplemental Figures S2–S5 (supplemental to the online version of this article). A three-way analysis of variance (ANOVA) was performed on the arcsine-transformed data (Studebaker⁴⁴) in ►Table 2 from the YNH listener and the OHL listener groups to determine the effects of carrier phrase, presentation level, and speaker on recognition performance. (Note: For each ANOVA, assumptions were verified including assessing for outliers by examining performance data boxplots and verifying normality with the Shapiro-Wilk’s test. Sphericity was assessed by the Mauchly’s test, and, if significant, then the Greenhouse-Geisser result is reported, which produces a mixed number df .) All main effects from the three-way ANOVAs were significant for both groups of listeners: carrier phrase [YNH, $F_{(1, 23)} = 71.8$, $p < 0.0001$; OHL, $F_{(1, 46)} = 62.3$, $p < 0.0001$], presentation level [YNH_{(1, 23)} = 17.0, $p < 0.0001$; OHL, $F_{(1.3, 63.1)} = 438.1$, $p < 0.0001$],}

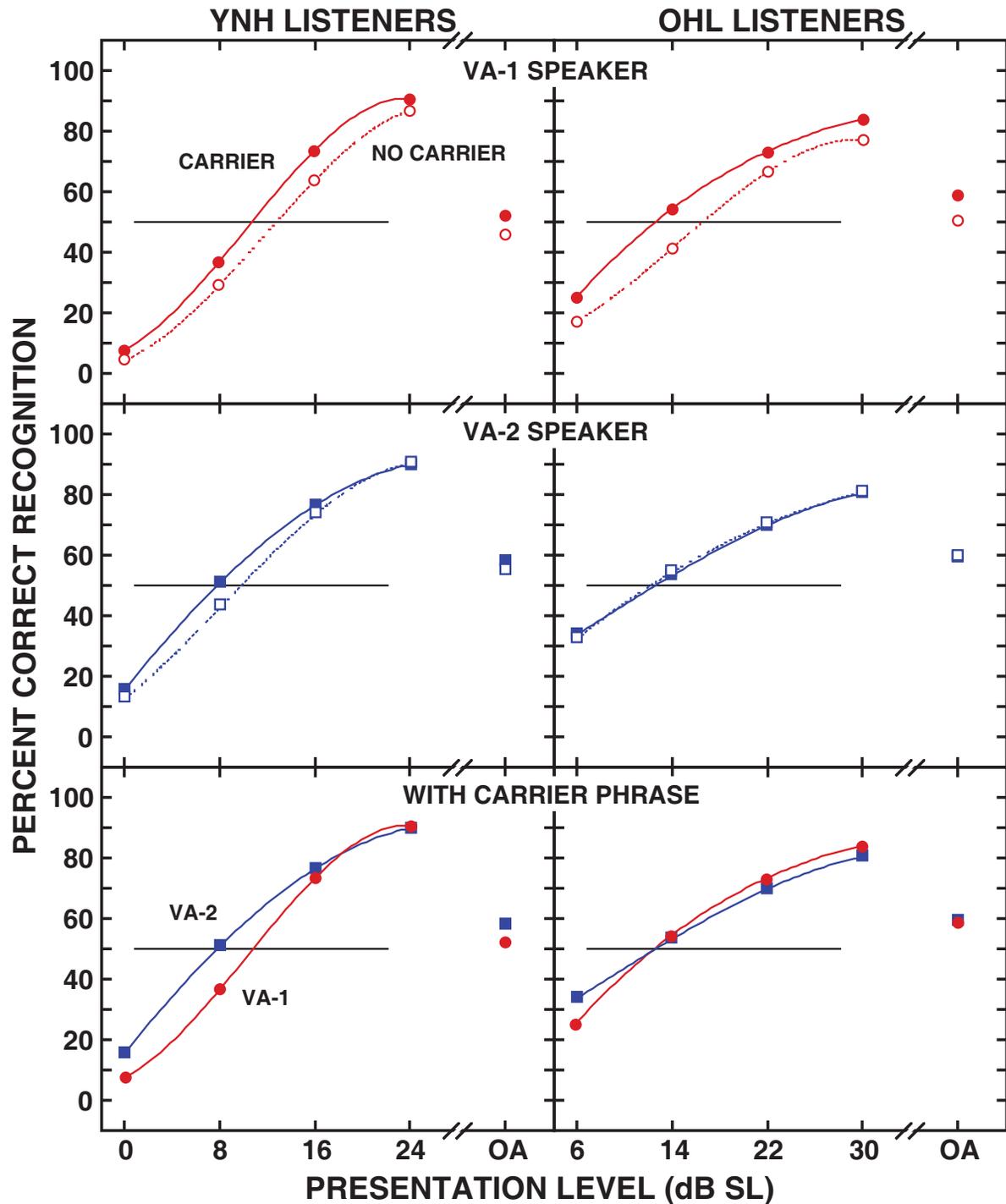


Fig. 3 The mean percent correct by 24 YNH listeners (left panels) and 48 OHL listeners (right panels) on NU-6 with the carrier phrase (filled symbols) and without the carrier phrase (open symbols) spoken by the VA-1 speaker (upper panels) and the VA-2 speaker (middle panels). The SDs for both the participant and word groupings are listed in ►Table 2. A comparison of the VA-1 and VA-2 versions with the carrier phrase is presented in the bottom panels. The overall (OA) recognition performances are depicted to the right in each panel. Third-degree polynomials are used to describe the data. The data for the individual listeners are listed in Supplemental Tables S2–S5 (supplemental to the online version of this article).

and speaker [YNH, $F_{(1, 23)} = 86.1$, $p < 0.0001$; OHL, $F_{(1, 46)} = 26.1$, $p < 0.0001$]. Although there were no significant three-way interactions, there were several significant two-way interactions. These significant main effects and interactions warranted further review to understand the interrelationships and to determine how recognition performance was influenced by the combinations of the conditions (speaker,

presentation level, and carrier phrase). The significant interactions and the main effects were further analyzed by *post hoc* examination with Bonferroni corrections made for multiple comparisons. Remnants of these analyses are incorporated into the results presented in this and subsequent sections.

The aforementioned interactions of the conditions are apparent from the data in ►Figure 3 that show recognition

Table 1 The mean overall percent correct at the 4 presentation levels of the 4 listening conditions are listed along with the mean sensation levels (dB) of the 50% correct points calculated from the polynomial equations and slopes of the functions (%/dB) at the 50% points calculated from the 1st derivatives of the polynomial equations used to describe the mean functions in ►Figure 3. These data are listed for both speakers, both carrier phrase (CP) conditions and both subject groups.

Group/Variable	VA-1 Speaker		VA-2 Speaker	
	With CP	Without CP	With CP	Without CP
YNH listeners				
Overall (% correct)	52.5	46.5	58.4	55.4
50% point (dB SL)	10.7	12.7	7.8	9.6
Slope @ 50% (%/dB)	4.8	4.4	3.9	4.1
OHL listeners				
Overall (% correct)	59.4	51.1	59.7	60.2
50% point (dB SL)	12.5	16.4	12.4	11.9
Slope @ 50% (%/dB)	3.2	3.4	2.4	2.6

Table 2 The mean percent correct recognition performances on the NU-6 words are listed for the two speakers (VA-1 and VA-2), the four presentation levels (dB SL), and the two carrier phrase conditions (with the carrier phrase and without the carrier phrase). The standard deviations for the subjects and for the words also are listed along with the overall mean performances (OA) at the four presentation levels. The difference in the mean recognition performances (Δ = the mean performance with carrier phrase minus the mean performance without carrier phrase) also are listed. The individual subject data are in Supplemental Tables S2-S5.

With Carrier Phrase					Without Carrier Phrase			
Speaker	dB SL	M	Subject SD	Word SD	M	Subject SD	Word SD	Δ
Young listeners with normal hearing								
VA-1								
	0	8.0	11.2	12.3	5.0	7.6	9.6	3.0*
	8	37.1	24.3	24.6	29.5	25.2	20.2	7.6**
	16	73.9	16.9	23.3	64.2	20.2	23.9	9.7**
	24	91.1	6.9	14.7	87.3	9.7	17.0	3.8*
	OA	52.5			46.5			6.0
VA-2								
	0	15.9	16.7	15.4	13.3	14.5	14.7	2.6
	8	51.1	24.9	26.1	43.5	23.7	25.9	7.6**
	16	76.7	16.0	21.8	74.1	15.4	22.1	2.6
	24	90.0	8.1	15.8	90.7	8.5	14.1	-0.7
	OA	58.4			55.4			3.0
Older listeners with sensorineural hearing loss								
VA-1								
	6	25.4	13.8	18.5	17.8	13.4	14.5	7.6**
	14	54.5	16.2	23.1	41.9	18.2	21.2	12.6**
	22	73.5	16.1	20.4	67.1	17.2	20.6	6.4**
	30	84.4	12.4	15.4	77.7	15.7	18.2	6.7**
	OA	59.4			51.1			8.3**
VA-2								
	6	34.3	16.7	20.9	33.2	14.8	21.6	1.1
	14	53.7	16.0	24.2	55.1	17.4	23.5	-1.4
	22	70.3	16.1	20.6	71.1	14.8	21.4	-0.8
	30	80.8	13.9	16.8	81.4	13.4	17.2	-0.6
	OA	59.7			60.2			-0.5

* $p < 0.003$; ** $p < 0.0001$.

performance varied with speaker, carrier phrase condition, and presentation level, but not always in a similar manner for each listener group at the four presentation levels. There were statistically significant two-way interactions between the speaker and carrier phrase [YNH, $F_{(1, 23)} = 17.0$, $p < 0.0001$; OHL, $F_{(1, 46)} = 204.5$, $p < 0.0001$], between the speaker and presentation level [YNH, $F_{(17, 41)} = 14.9$, $p < 0.0001$; OHL, $F_{(2.5, 112.9)} = 42.5$, $p < 0.0001$], and between the carrier phrase and presentation level, but only for the YNH group [$F_{(2.0, 45.5)} = 575.3$, $p < 0.001$]. Specifically, it can be seen in ►Figure 3, top panels, that for the VA-1 speaker throughout the ranges of presentation levels, recognition performances for both groups of listeners were better with the carrier phrase than without the carrier phrase. For the VA-1 speaker, the overall recognition performances for the with and without the carrier phrase conditions were 52.5% and 46.5%, respectively, for the YNH listeners, and 59.4% and 51.1%, respectively, for the OHL listeners. For the VA-1 speaker, at each presentation level for both groups of listeners, performance with the carrier phrase was significantly better than that without the carrier phrase ($p < 0.0125$; see Supplemental Table S6 (supplemental to the online version of this article) for detailed statistical results). The reduction in recognition performances was anticipated for the without carrier phrase condition as a few words had distorted initial consonants when separated from their carrier phrases, regardless of where on the waveform the separations were made. Because the OHL listeners have "internal" distortions owing to audibility and general degradation issues associated with aging, it is reasonable to think that these distortions in the excised words for the VA-1 version impacted the OHL listeners slightly more than the YNH listeners, which might explain the larger separation between the recognition performance functions for the OHL listeners.

The effect of the carrier phrase was different for the VA-2 speaker. As can be seen in ►Figure 3 (middle panels), the carrier phrase differences were minimized with the overall recognition performances with and without carrier phrases, respectively, of 58.4% and 55.4% (YNH) and 59.7% and 60.2% (OHL). Collectively, for the two groups of listeners, with the exception of the 8-dB SL condition with the YNH listeners, the mean recognition performances with and without the carrier phrases were the same. As shown in the last column of ►Table 2, the difference (Δ) in performance with and without the carrier phrase for the OHL listeners was about 61%, which was confirmed by paired sample t -tests as not statistically significant. A different performance pattern across sensation levels was seen for the YNH listeners, with the carrier phrase providing an overall 3.0% benefit (58.4–55.4% in ►Table 2). With the YNH listeners, recognition performance was significantly 7.6% better (51.1–43.5%) with the carrier phrase than without the carrier phrase but only at 8-dB SL, which was confirmed with a paired sample t -test ($p < 0.0001$; see Supplemental Table S6 for all p values). Thus, unlike the OHL listeners, the YNH listeners were able to take advantage of the carrier phrase and obtain better recognition performance but only at a challenging presentation level that was less than 50% correct and greater than floor effects.

The differences between the carrier phrase conditions also were reflected at the 50% points on the mean functions in ►Figure 3, and the values at this point on the function are listed in ►Table 1. For the VA-1 speaker, 50% recognition performances with the carrier phrase were achieved at 2.0-dB (YNH, 12.7- to 10.7-dB SL) and 3.9-dB (OHL, 16.4- to 12.5-dB SL) lower presentation levels compared with the without the carrier phrase condition. For the VA-2 speaker, the 50% recognition performances for the two carrier phrase conditions differed by 1.8 dB (YNH, 9.6- to 7.8-dB SL) and –0.5 dB (OHL, 11.9- to 12.4-dB SL). Similarly, for both speakers, the slopes of the mean functions at the 50% point in ►Figure 3 were steeper for the YNH listeners (3.9–4.8%/dB) than for the OHL listeners (2.4–3.4%/dB) and slightly steeper (~1%/dB) for the VA-1 speaker than for the VA-2 speaker. The 50% points and slope values for both groups across the different talkers and conditions are discussed more in the following text including inferential statistics.

In randomized experiments like the current study involving two test sessions and multiple presentations of the target words, albeit at different presentation levels and over a range of time intervals, the expectation is that slightly better recognition performances are expected in the second session. This improved performance in the second session is attributable to the listeners becoming familiar with the test environment and with the listening and response tasks (Miller et al³⁵). The mean recognition performance functions obtained in each of the two test sessions are depicted in Supplemental Figure S4 (YNH) and Supplemental Figure S5 (OHL). At the 50% point on the recognition functions, both listener groups averaged 1.1-dB improvement in Session 2 compared with Session 1 with slightly larger improvement on the VA-2 materials (1.6 dB) than on the VA-1 materials (0.7 dB). In terms of overall percent correct for the various conditions, the performances in Session 2 averaged 2.7% (YNH) and 2.8% (OHL) better than in Session 1. These improvements amount to about a one-word improvement in Session 2.

Because previous studies of the effect that the carrier phrase has on word recognition performance involved only slivers of the variables incorporated into the current study, comparisons to the earlier studies are limited. The current data are for the most part in agreement with the Martin et al³² study in which no difference in recognition performance was found with and without the carrier phrase when presented at a high presentation level (30-dB SL), which is an easy listening condition. With YNH listeners, Gladstone and Siegenthaler¹⁹ reported 7–16% better recognition performances on 25, W-22 words presented at 5-dB SL (re: SRT) with the carrier phrase than without the carrier phrase, which is similar to the relations observed with the YNH listeners in the current study. On YNH listeners, Lynn and Brotman²⁷ reported for words with voiceless stop initial consonants 10% better performance in SSN with the carrier phrase (37.4%) than without the carrier phrase (27.5%), which is similar to the 7.6% difference observed in the current study at 8-dB SL with the VA-2 speaker. These latter studies suggest that in difficult listening conditions at which recognition performance is <50% but greater than floor effects, the with carrier phrase condition produces slightly

better performances from YNH listeners than does the without carrier phrase condition.

In summary and in review of the data in ►Tables 1 and 2, and in ►Figure 3, the preliminary conclusion with the VA-1 speaker data is that when the target words were excised from the carrier phrase, recognition performance decreased slightly, which is probably attributable to a reduction in and distortion of the associated initial consonant cues in the waveform caused by separating some of the target words from their carrier phrases. In comparison to the YNH listeners, this waveform distortion was amplified in the OHL listeners by audibility issues precipitated by their sensorineural hearing loss compounded by general auditory system degradation related to aging factors. Thus, with the VA-1 speaker paradigm, the results should be viewed not as the carrier phrase-enhancing word recognition performance but rather as a decrease in performance without the carrier phrase when the words are excised from the carrier phrase and presented alone. With the VA-2 speaker across both listening groups, mean performances on seven of the eight presentation levels studied were the same for both carrier phrase conditions, indicating the carrier phrase provided no appreciable enhancement in recognition performance for either group of listeners. Underlying the rather systematic mean recognition performance functions shown in ►Figure 3 are collections of recognition functions of the individual participants and of the individual words that are quite different from one another, reflecting variability in the data. These differences are minimal for YNH listeners but amplified for OHL listeners. This diversity of performances, which simply reflects the heterogenetic characteristic of word recognition performances, is considered for the individual participants and the individual CNC words in the following two sections.

Individual Listener Recognition Performances

The main objective of this study was to compare word recognition performance under two carrier phrase conditions (the target word with the carrier phrase and the target word without the carrier phrase). ►Figure 4 presents bivariate-plot comparisons of the percent correct data at each of the four presentation levels, for the two carrier phrase variables, and for the VA-1 speaker (red circles) and VA-2 speaker (blue squares) by the YNH listeners (upper two rows) and the OHL listeners (lower two rows). The data with the carrier phrase are on the ordinate and without the carrier phrase are on the abscissa. The diagonal line in each panel represents equal performances on the two carrier phrase conditions with datum points (a) above the diagonal line indicating better performance on the words presented with the carrier phrase and (b) below the diagonal line indicating better performance on the words presented without the carrier phrase. The numbers in parentheses are the percent of listeners whose recognition performances were above, on, and below the diagonal line. Several features can be noted in ►Figure 4. First, it is obvious from the distributions that most individual participant datum points are above the line of equality, indicating better recognition performance with the carrier phrase (61.5% of the 576

comparisons [$576 = 2 \text{ speakers} \times 4 \text{ presentation levels} \times 72 \text{ participants}$]) than without the carrier phrase (25.9%), the remaining 12.7% having equal performances.

Second, with the VA-1 speaker across the four presentation levels, overall better recognition performances were obtained with the carrier phrase than without the carrier phrase by an average of 65.6% of the YNH listeners (1st row of panels in ►Figure 4) and by 81.8% of the OHL listeners (3rd row of panels). By contrast, across the four presentation levels, only 10.4% of the YNH listeners and 12.0% of the OHL listeners had better performances without the carrier phrase. The remainder of the two participant groups had equal performances on the two carrier phrase conditions. Not only did more listeners in both participant groups have better performances with the carrier phrase than without the carrier phrase but also the mean performance differences of the datum points above the line of equality (i.e., the average percent above the line of equality with the carrier phrase) were substantially larger (YNH, 9.7%; OHL, 10.9%) than the mean performance differences of the datum points below the line of equality without the carrier phrase (YNH, -2.8%; OHL, -4.7%). Not surprising, as the group recognition performances previously discussed indicated statistically significant better performance with the carrier phrase for the VA-1 speaker, the data in ►Figure 4 emphasize that for both groups of listeners with the VA-1 speaker, there were substantially more and larger recognition performance differences with the carrier phrase than without the carrier phrase.

Third, with the VA-2 speaker, the two listener groups produced different performance relations. The results from the YNH listeners were only slightly different from those obtained with the VA-1 speaker. As can be seen in the 2nd row of panels in ►Figure 4, averaged across the four presentation levels, 54.2% of the YNH listeners had better performances with the carrier phrase, 27.1% had better performances without the carrier phrase, and 18.7% of the YNH listeners had equal performances. In concert with that distribution, the average performance advantage with the carrier phrase was 7.5%, whereas the average advantage without the carrier phrase was 3.8%. Although the target words with and without the carrier phrase were identical utterances produced independent of the carrier phrases, the presence of the carrier phrase slightly enhanced performances by the YNH listeners, especially at the 8-dB SL presentation level that had an average difference between performances on the two carrier phrase conditions of 7.6%, whereas the differences at the other three presentation levels were $\leq 2.6\%$. Perhaps, the carrier phrase enhanced recognition performance by defining the listening interval in the difficult listening portion of the recognition function above the area in which floor effects influenced performance. In sharp contrast to the YNH results with the VA-2 speaker, the OHL listeners (4th row of panels in ►Figure 4) produced recognition performances that were essentially the same for the two carrier phrase conditions. Overall, slightly more of the OHL group produced slightly better performances without the carrier

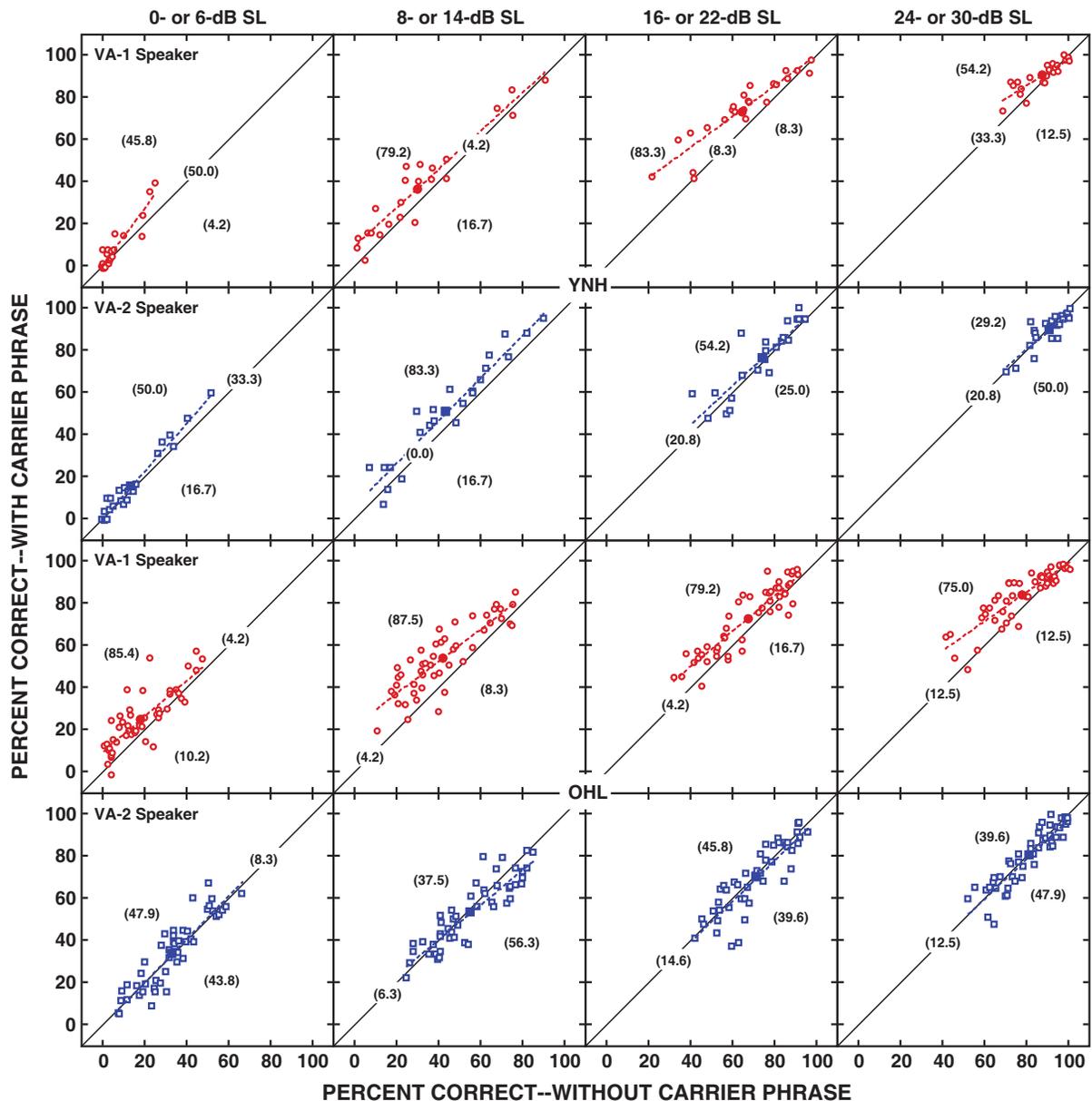


Fig. 4 Bivariate plots of the average recognition performances with the carrier phrase (ordinate) and without the carrier phrase (abscissa) obtained from the 24 YNH listeners (upper two rows of panels) and the 48 OHL listeners (lower two rows of panels) are depicted from the lowest to the highest presentation levels (upper abscissa labels). The data from the VA-1 speaker (red symbols) and from the VA-2 speaker (blue symbols) were jittered on both axes with a random additive algorithm from 20.8% to 0.8% in 0.2% steps. The numbers in parentheses are the percent of nonjittered datum points above, on, and below the line of equality. The dashed lines are the linear regressions used to describe the nonjittered data, with the larger filled symbols showing the means of the data in each panel.

phrase (46.9%) than with the carrier phrase (42.7%), which was reflected in the mean recognition performance functions (→Figure 3). In addition, the average recognition performance advantages (departure from the line of equality) with and without the carrier phrase were similar, 6.0% and 6.4%, respectively. As a group then, the OHL listeners did not receive any benefit from the carrier phrase in the VA-2 speaker paradigm.

Fourth, floor and ceiling effects on recognition performance are evidenced in the YNH listener data by the large percent of individuals (20.8-50.0%) with equal performances on the two carrier phrase conditions at the lowest and highest presentation levels. This effect was not observed

with the OHL listeners probably because many of the performance values did not approach the floor and ceiling limits owing to the relatively limited range of presentation levels used coupled with the gradual slope of the recognition function that is a characteristic of recognition performances by OHL listeners.

To this point, the mean data in →Figure 3, the statistical analyses, and the individual bivariate plots in →Figure 4 demonstrate better recognition performances throughout the range of presentation levels in three of the four listener/speaker conditions (YNH listeners, both speakers; OHL listeners, the VA-1 speaker) when the carrier phrase was used; only the OHL listeners with the VA-2 speaker produced

equivalent performances with and without the carrier phrase. Now, it is instructive (and challenging) to examine the individual recognition performance functions from the 72 individual listeners, samplings of which are illustrated in **Figures 5** and **6** (each of the four functions for each of the 72 listeners are listed in Supplemental Tables S7-S14 and shown in Supplemental Figures S6-S13, supplemental to the online version of this article). Recall that the four presentation levels for each listener group were selected to produce ideally two

points greater than 50% performance and two points less than 50%. Because of the questions being posed in this study, performances around 0% and 100% correct (floor and ceiling, respectively) were not of intended interest. The average recognition performance ranges over the 24-dB presentation-level continua were 82.7% and 75.8% (YNH) and 59.5% and 47.4% (OHL) for the VA-1 and VA-2 speakers, respectively. Thus, as anticipated, for the YNH listeners, the range of percent correct performances was 23-28% greater than the

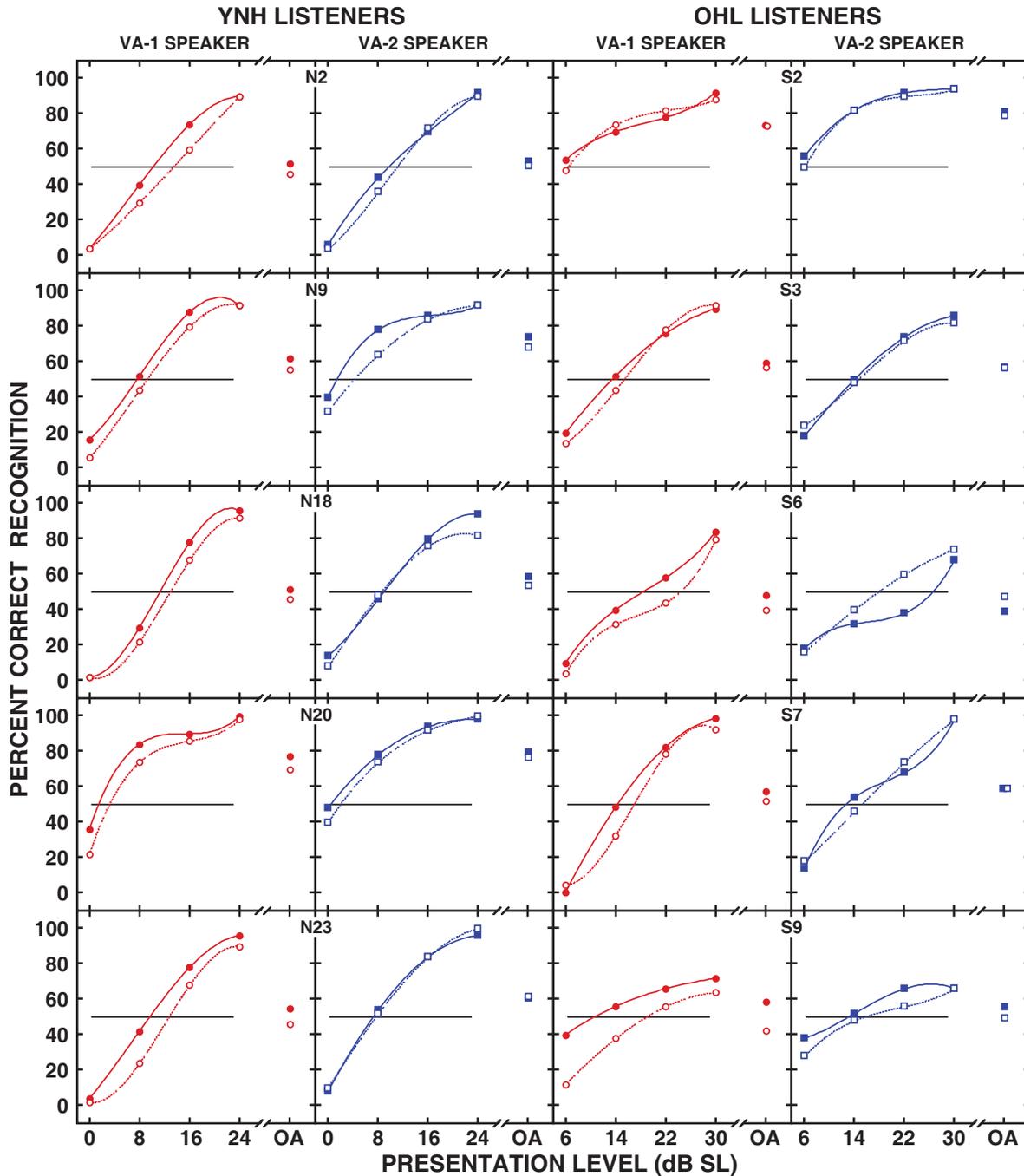


Fig. 5 The mean data at the four presentation levels from five representative YNH listeners (left columns; 0- to 24-dB SL) and 5 OHL listeners (right columns; 6- to 30-dB SL) are shown for the VA-1 speaker (red symbols, columns 1 and 3) and the VA-2 speaker (blue symbols, columns 2 and 4). The filled symbols represent the data with the carrier phrase and the open symbols represent the data without the carrier phrase. The overall means of the performances (OA) at the four presentation levels are also depicted. All of the YNH and OHL functions are included in Supplemental Figures S6-S13.

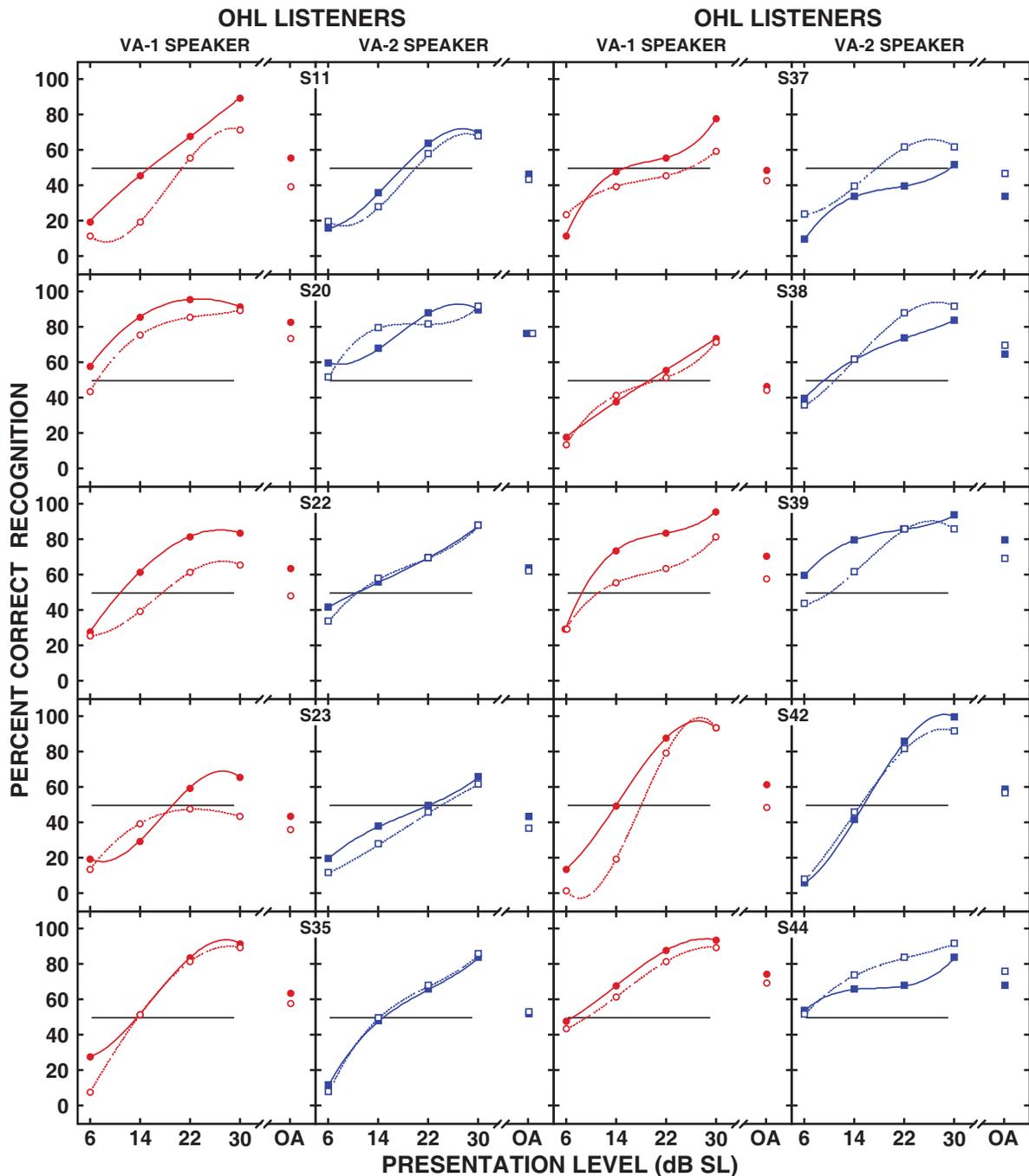


Fig. 6 The mean data at the four presentation levels from 10 representative OHL listeners are shown for the VA-1 speaker (red symbols, columns 1 and 3) and the VA-2 speaker (blue symbols, columns 2 and 4). The filled symbols represent the data with the carrier phrase and the open symbols represent the data without the carrier phrase. The overall means of the performances (OA) at the four presentation levels are also depicted. All of the individual OHL functions are included in Supplemental Figures S9-S13.

range of performances by the OHL listeners, which is a relation that was reflected in the slope differences observed between the performance functions for the two listener groups in ►Figure 3. The remainder of this section focuses (a) on the sensation levels (dB) at which 50% recognition performance occurred and on the slopes of the functions at the 50% point, calculated from the polynomial equations and the 1st derivatives of those equations, respectively, used to describe the four recognition performance functions for each

of the 72 listeners, and (b) on the variability of the functions of the individual listeners.

As mentioned earlier, the differences between the carrier phrase conditions were reflected at the 50% points and slopes on the mean recognition functions in ►Figure 3, which are listed in ►Table 1. It is also important to evaluate and appreciate the 50% points and slopes on the individual condition/participant functions because of the unique characteristics associated with the individual function

Table 3 The mean sensation levels (dB) at the 50% points and standard deviations (dB) for the speaker and listener groups calculated from the individual listener polynomial equations used to describe the individual data are listed along with the mean slopes of the functions (%/dB) at the 50% points calculated from the 1st derivatives of the individual listener polynomial equations. The column Δ s are the with carrier phrase minus without carrier phrase performance differences, whereas the row Δ s are the VA-1 speaker minus the VA-2 speaker performance differences. Statistically significant differences are denoted with asterisks (*). The individual subject data are listed in Supplemental Tables S2-S5.

Group/Speaker	50% Point (dB SL)			Slope @ 50% Point (%/dB)		
	With Carrier	Without Carrier	Δ	With Carrier	Without Carrier	Δ
YNH listeners						
VA-1 speaker						
<i>M</i>	10.5	12.5	-2.0*	6.0	5.7	0.3
<i>SD</i>	4.7	5.2		1.4	1.5	
VA-2 speaker						
<i>M</i>	8.0	9.2	-1.2**	5.4	5.3	0.1
<i>SD</i>	5.3	5.0		1.9	1.4	
Δ	2.5*	3.3*		0.6	0.4	
OHL listeners						
VA-1 speaker						
<i>M</i>	13.4	16.9	-3.5*	3.8	4.1	-0.3
<i>SD</i>	5.6	6.0		1.7	1.8	
VA-2 speaker						
<i>M</i>	13.7	12.7	1.0	3.0	3.6	-0.6
<i>SD</i>	7.5	6.1		1.6	1.5	
Δ	-0.3	4.3*		0.8	0.5	

* $p < 0.0001$

** $p < 0.004$

midpoints. The mean sensation levels (dB) (and SDs) and the slopes of the recognition functions (%/dB) at the 50% points derived from the polynomial equations and 1st derivatives of the individual listeners are listed in ►Table 3. As expected, the sensation levels (dB) of the mean 50% points in ►Table 3 computed from the individual listener data closely reflect the 50% points of the mean functions listed in ►Table 1, which were derived from the mean functions and not the individual listener functions. (Note: Throughout the article, measures of central tendency computed from the raw data and from various transformations of the raw data [e.g., polynomial equations] produce slightly different results, owing to the different transformations and algorithms involved in the respective processes.) The differences between the corresponding sensation levels of the 50% points (►Table 1 value minus ►Table 3 value) range 0.2 dB for the YNH listeners and the VA-1 speaker from 10.7- to 12.7-dB SL with and without a carrier phrase, respectively, to 10.5- and 12.5-dB SL. At the other extreme, with the OHL listeners and the VA-2 speaker with a carrier phrase, the range was -1.3 dB. These changes are 3–10%. By contrast, the slopes of the functions listed in ►Table 3 are somewhat steeper than the slopes of the mean functions listed in ►Table 1, with differences ranging from 0.6%/dB (OHL, both speakers, with a carrier phrase) to 1.5%/dB (YNH, VA-2, with a carrier phrase), which are 19-39% steeper slopes than the slopes of the mean functions.

Within ►Table 3, the recognition performance differences (Δ = performance with the carrier phrase minus performance without the carrier phrase) are listed for the various speaker and listener group combinations, with the level of significance noted with asterisks. These differences were evaluated with one-way ANOVAs with listener group as a between-subject variable. Performance differences among the 50% points (dB SL) between the listener groups were statistically significant for all carrier phrase and speaker conditions [VA-1, without carrier phrase $F_{(1, 69)} = 9.5$, $p < 0.003$; VA-1, with carrier phrase $F_{(1, 70)} = 5.0$, $p < 0.03$; VA-2, without carrier phrase $F_{(1, 70)} = 5.7$, $p < 0.02$; VA-2, with carrier phrase $F_{(1, 69)} = 11.3$, $p < 0.001$], with the 50% points for the YNH listeners at significantly lower levels than for the OHL listeners. Similarly, the differences among the function slopes between the groups were also statistically significant for all carrier phrase and speaker conditions [VA-1, without carrier phrase $F_{(1, 69)} = 13.6$, $p < 0.000$; VA-1, with carrier phrase $F_{(1, 70)} = 30.9$, $p < 0.0001$; VA-2, without carrier phrase $F_{(1, 70)} = 22.4$, $p < 0.0001$; VA-2, with carrier phrase $F_{(1, 67)} = 30.8$, $p < 0.0001$], with the slopes of the functions for the YNH listeners being significantly steeper than for the OHL listeners.

Within each listener group, another two-way ANOVA was performed to determine if the 50% points calculated from the polynomial equations differed with respect to the speaker and

carrier phrase conditions. All main effects were significant for both groups of listeners: speaker [YNH, $F_{(1, 23)} = 80.5$, $p < 0.0001$; OHL, $F_{(1, 45)} = 14.8$, $p < 0.0001$] and carrier phrase [YNH, $F_{(1, 23)} = 46.3$, $p < 0.0001$; OHL, $F_{(1, 45)} = 25.2$, $p < 0.0001$]. A significant speaker and carrier phrase two-way interaction was observed, but only for the OHL group [$F_{(1, 45)} = 30.6$, $p < 0.0001$]. The results of these two-way ANOVAs assessed the interrelationship of speaker and carrier phrase on the 50% points that was then followed with paired-sample t -tests with Bonferroni correction to determine each performance difference. The differences between the 50% points for the VA-1 speaker with and without the carrier phrase noted in **Table 3** were significant for both listener groups [YNH, $\Delta = -2.0$ dB, $t_{(24)} = 5.9$, $p < 0.0001$; OHL, $\Delta = -3.5$ dB, $t_{(46)} = 8.7$, $p < 0.0001$]; however, the difference was only significant for the VA-2 speaker with the YNH group [$\Delta = -1.2$ dB, $t_{(24)} = 3.2$, $p < 0.004$]. Furthermore, the without carrier phrase 50% points for the VA-1 and VA-2 versions were significantly different for both listener groups [YNH, $\Delta = 3.3$ dB, $t_{(24)} = 7.8$, $p < 0.0001$; OHL, $\Delta = 4.3$ dB, $t_{(46)} = 8.74$, $p < 0.0001$], but the with carrier phrase condition comparison was only significantly different for the YNH group [$\Delta = 2.5$ dB, $t_{(24)} = 5.7$, $p < 0.0001$]. The speaker differences for the with the carrier phrase condition are considered in more detail in a subsequent section.

In **Figures 5** and **6**, representative recognition functions from the individual YNH and OHL listeners are displayed for the VA-1 and VA-2 speakers in adjacent panels, with open symbols representing the without carrier phrase condition and the filled symbols representing the with carrier phrase condition. As a summary of the data in each panel, the overall mean recognition performances at the four presentation levels are depicted for the two carrier phrase conditions. Several of the datasets in **Figures 5** and **6** are considered in some detail to acquaint the reader with notable aspects of the functions from each of the listeners. Beyond these considerations, the reader is encouraged to study the results presented in the **Figures 5** and **6** and in Supplemental Figures S6-S13 to gain an appreciation of the variability inherent in the individual participant data with respect to the individual recognition functions and the relations between the functions of the two carrier phrase conditions. (Note: The mean percent correct recognition for each listener at each of the 16 variable conditions [2 speakers \times 2 carrier phrase conditions \times 4 presentation levels] are listed in Supplemental Figures S2-S5.)

Recognition performance functions for 5 of the 24 YNH listeners are depicted in the two leftmost columns of **Figure 5**. Consider first participant N2. With the VA-1 speaker, the recognition performance ranges were 86% with and without the carrier phrases, both of which had 90% maxima and 4% minima, with 50% points at 7.6- and 9.2-dB SL for the without carrier and with carrier conditions, respectively. With the VA-2 speaker, N2 had similar performances with 92% and 90% maxima, 6% and 4% minima, and 50% points at 9.6- and 11.0-dB SL, respectively, for the with and without carrier phrase conditions. The functions for N18 and N23 in **Figure 5** relationally correspond to the functions for N2 as did

the functions for most of the 24 YNH listeners. The results from participant N9 were similar to those for N2, N18, and N23 with a maximum of 92% for each of the four functions. There were, however, somewhat higher minima by N9 for the VA-1 speaker with carrier phrase condition (16%) and for both VA-2 carrier phrase conditions (40% and 32%) than were observed with the other YNH listeners. Participant N20 had near-maximum recognition performances on each of the four functions (98-100%) with minima performance ranges from 22% to 48% with the overall two without carrier phrase conditions being 8-14% lower performances with each speaker. The commonality that N20 has with N9 is the abbreviated VA-2 functions that barely drop less than 50%, which is a characteristic shared with only one other YNH listener, N19 in Supplemental Figure S7. For these three YNH listeners, consideration was given to the PTAs as the culprit in producing the relatively high recognition performances at the three highest presentation levels. If the response criteria by these listeners for the pure-tone thresholds were conservative, then the presentation levels of the speech materials would be at artificially high levels, producing inflated recognition performances. It is difficult, however, to substantiate or disprove this line of reasoning. For N9 and N19, the PTAs in the two test sessions were the same, whereas for N20, the PTA in the second session was 5-dB lower than in the first session. As discussed in the following text, there are potentially other reasons that should be considered to account for these abbreviated functions.

Consider next the OHL listener data in **Figures 5** and **6**. First, observe the general shapes of the recognition performance functions and the range of performances that they encompass. For example, S2 in **Figure 5** obviously has abbreviated recognition functions that with one exception is greater than 50% correct. Five other OHL listener datasets included in **Figures 5** and **6** had similar variants of these incomplete functions, including S9, S20, S22, S39, and S44. These 6 participants represent 40% of the 14 OHL listener results depicted in the figures; of the 48 OHL listeners, 18 (37.5%) had abbreviated functions. The most commonly occurring functions, which occurred in 30 of the 48 OHL listeners (62.5%), were more complete across the recognition performance range than the abbreviated functions, extending greater than and less than the 50% points. Examples of this group with complete functions included in **Figures 5** and **6** are S3, S6, S7, S11, S35, and S42 with the remaining three participants (S23, S37, and S38) having functions in **Figure 6** that are abbreviated but for the most part are distributed greater than and less than the 50% point. S23 and S37 had essentially equal performances on the materials spoken by the two speakers. By contrast, S38 had somewhat better performances on the materials spoken by the VA-2 speaker than on the materials spoken by the VA-1 speaker.

Second, with the OHL recognition performance functions in **Figures 5** and **6**, consider the relations among the responses on the two carrier phrase conditions. With the VA-1 speaker, all of the OHL listeners performed overall better with the carrier phrase than without the carrier phrase, the only interesting difference being the variability of the differences among the presentation levels for each

listener. With the VA-2 speaker, S6, S35, S37, S38, and S44 in **Figures 5 and 6** had better recognition performance without the carrier phrase than with the carrier phrase. In fact, with the VA-2 speaker, 27 of the 48 OHL listeners (56.3%) had overall better performances without the carrier phrase than with the carrier phrase. A different relation was observed for the VA-1 speaker, with all of the OHL listeners achieving better overall performances with the carrier phrase than without the carrier phrase. These various relations among the recognition performances on the experimental variables (speakers, presentation levels, and carrier phrases) highlight the importance of attention being given to the variables that combine to produce the iconic mean psychometric functions shown in **Figure 3**.

As previously mentioned, of the 48 OHL listeners, 18 (37.5%) had abbreviated functions, in which almost all datum points were at or greater than 50% correct and 30 (62.5%) had more complete functions most of which were balanced about the 50% point (see Supplemental Figures S9-S13). The recognition functions grouped and evaluated in this manner are

illustrated in **Figure 7**, in which the mean recognition performance functions for the 18 OHL listeners with abbreviated functions are shown on the left panels with the more complete functions from the 30 OHL listeners depicted in the right panels. The relations between the functions for the two carrier phrase conditions are maintained with each of the subgroups, i.e., with VA-1 the functions for the two carrier conditions are separated, whereas with VA-2, the functions are intertwined. The functions for the two groups of OHL listeners are displaced by about 5-8 dB on the x-axis. With over a third of the functions for the OHL listeners abbreviated, it is easy to understand the influential impact these data had on the mean group functions depicted in **Figure 3**. Obviously, with each of the 18 listeners with abbreviated functions, had an additional lower presentation level been used, the recognition function would have been defined more completely. As suggested earlier in this article and in an earlier article (Wilson⁵²), it is possible, but difficult to prove, that the PTAs of these participants were inflated, owing to unique listener response criteria to the pure tones

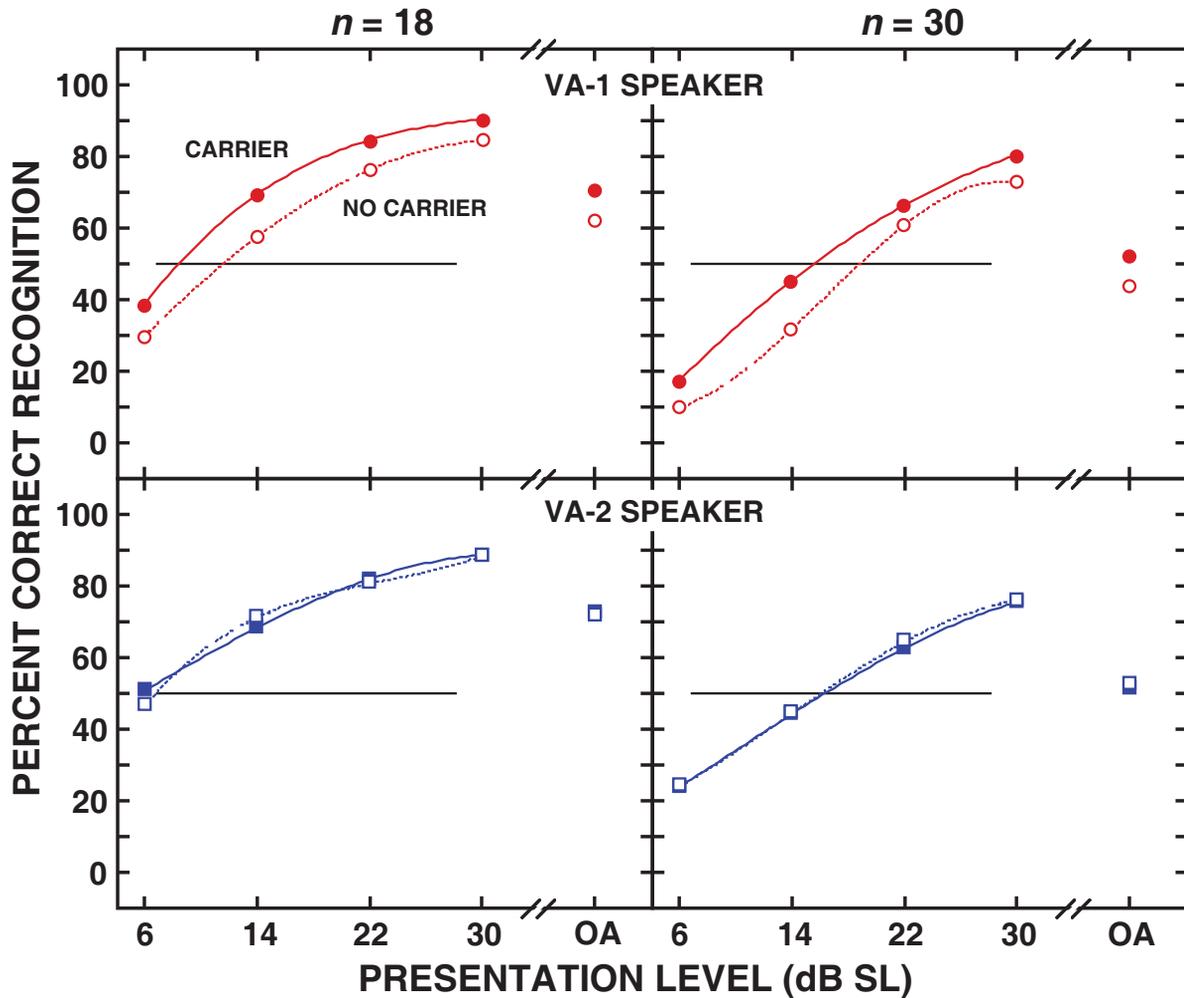


Fig. 7 The mean percent correct by 48 OHL listeners on the 200 NU-6 words with the carrier phrase (filled symbols) and without the carrier phrase (open symbols) spoken by the VA-1 speaker (upper panels) and the VA-2 speaker (lower panels). The functions from 18 of the OHL listeners (left panels) were abbreviated with the majority of points above 50% correct, whereas the functions from the remaining 30 OHL listeners were characterized by two datum points above and below 50% correct. The overall recognition performances (OA) are depicted to the right in each panel.

that elevated the presentation levels of the speech signals. The fact that many of these listeners had equal or near equal recognition performances at the two highest presentation levels adds support to this line of reasoning. With some OHL listeners, the inclusion of the 2000-Hz threshold in the PTA might be pushing the PTA artificially high in terms of the application of the sensation level to speech materials. The concept of the sensation level as applied to YNH listeners may not have the same implications when applied to some OHL listeners. Finally, another contributing factor might be that in comparison to the other listeners, these OHL participants with abbreviated functions were able to understand the words at lower presentation levels than the other

listeners, which could be related to possible loudness function differences among the OHL listeners (Knight and Margolis²⁵). Clinically, this phenomenon of abbreviated functions with some OHL listeners would be unnoticed because the effect is at the lower presentation levels of the recognition function as opposed to the higher levels of the function at which word recognition performance is typically evaluated in the clinic.

Recognition Performances on Individual Words

As with the recognition functions for the individual participants, the functions for the individual words exhibit a volatility that is lost in the averaging process that produced the

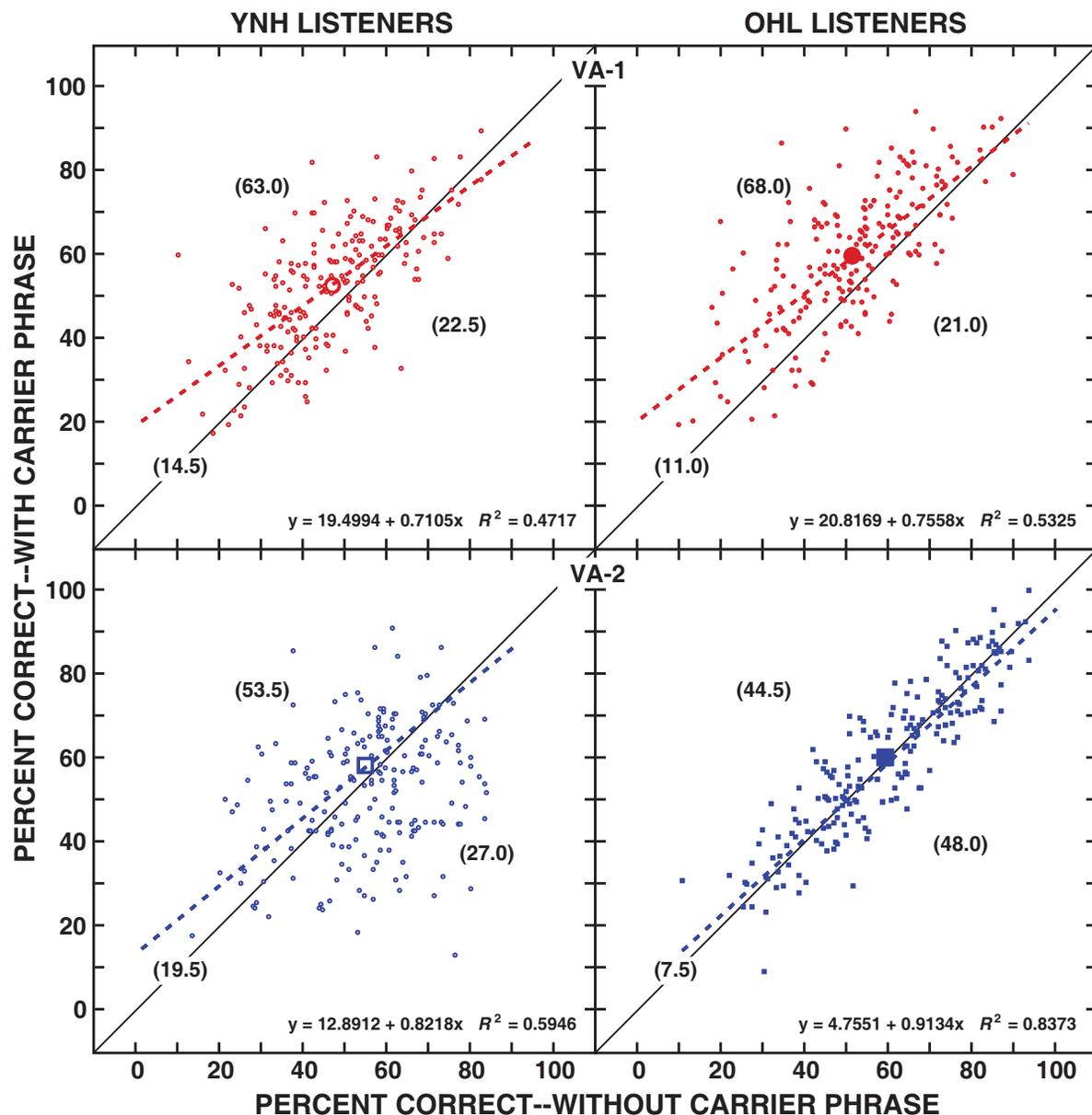


Fig. 8 Bivariate plots of the average percent correct recognition at the four presentation levels on each of the 200 NU-6 words spoken with a carrier phrase (ordinate) and without a carrier phrase (abscissa) by the VA-1 (upper panels) and VA-2 (lower panels) speakers are depicted for the YNH listeners (left panels) and the OHL listeners (right panels). The data were jittered on both axes with a random additive algorithm from -0.5% to 0.5% in 0.05% steps. The numbers in parentheses are the percent of nonjittered datum points above, on, and below the line of equality. The dashed lines are the linear regressions used to describe the nonjittered data with the four larger symbols showing the means of the 200 words.

systematic mean recognition functions depicted in ► **Figure 3** (the percent correct recognition for each of the 200 words are listed in Supplemental Tables S15–S18, Supplemental to the online version of this article). ► **Figure 8** contains bivariate plots of the overall recognition performances with the carrier phrase (ordinate) and without the carrier phrase (abscissa) for each of the 200 NU-6 words spoken by the VA-1 speaker (upper panels) and the VA-2 speaker (lower panels). The data for the YNH listeners are in the left panels and for the OHL listeners are in the right panels. The datum points are the average recognition performances at the 4 presentation levels for each of the 200 words. First, with the VA-1 speaker, 63% (YNH) and 68% (OHL) of the words had performances that were better with the carrier phrase than without the carrier phrase. Likewise, for the YNH and OHL listeners, respectively, 22.5% and 21.0% of the words had better performances without the carrier phrases and 14.5% and 11.0% had equal performances. Datum points above the line of equality had an average departure of 12–14%, whereas the average departure of the points below the line of equality about half, -5% to -7%. The slopes of the regressions (0.71%/ [YNH] and 0.76%/ [OHL]) indicate a strong relation between the two carrier phrase conditions. Finally, for the VA-1 words, the R^2 values for the linear regressions used to describe the data in ► **Figure 8** were similar for the two listener groups, 0.47 (YNH) and 0.53 (OHL), both of which reflect the variability in the data. Second, with the VA-2 speaker, the bivariate plots are a bit more interesting. It is obvious from the distributions of datum points in the lower panels of ► **Figure 8** that the YNH distribution is somewhat different from the OHL distribution, but similar to the distributions observed with the

VA-1 speaker. With the VA-2 speaker, the distributions of datum points are more evenly distributed greater than and less than the lines of equality with 53.5% (YNH) and 44.5% (OHL) greater than the line and 27.0% (YNH) and 48.0% (OHL) less than the line. In addition, with the OHL listeners, the average departures from the lines of equality were smaller than were observed with the VA-1 speaker (YNH listeners, 10.4% and -9.5%; OHL listeners, 6.1% and -6.6%). This latter relation is reflected in the R^2 values of the linear regressions that were higher for the OHL listeners (0.84) than for the YNH listeners (0.59), which is a reflection of less variability for the former and a more homogeneous relation between the two carrier phrase conditions for the OHL group than in the YNH group. The slopes of the VA-2 regressions (0.82%/ [YNH] and 0.91%/ [OHL]) are slightly steeper than the slopes for the VA-1 speaker. The relations observed among the variables in the bivariate plots in ► **Figure 8** provide unique insights into the response components associated with the individual words that underlie the mean functions depicted earlier in ► **Figure 3**.

In terms of the recognition performances on the individual words, the parameters used in ► **Figure 8** enabled a detailed examination of the two carrier phrase conditions both within the listener groups and within the speaker groups. Missing in ► **Figure 8** was the presentation level variable that was needed to make comparisons of recognition performances on the individual words between the listener groups and between the speaker groups. The 50% point on the recognition function of each word met this requirement. As the carrier phrase conditions were shown in ► **Figure 3** to differ between the two speaker conditions, only

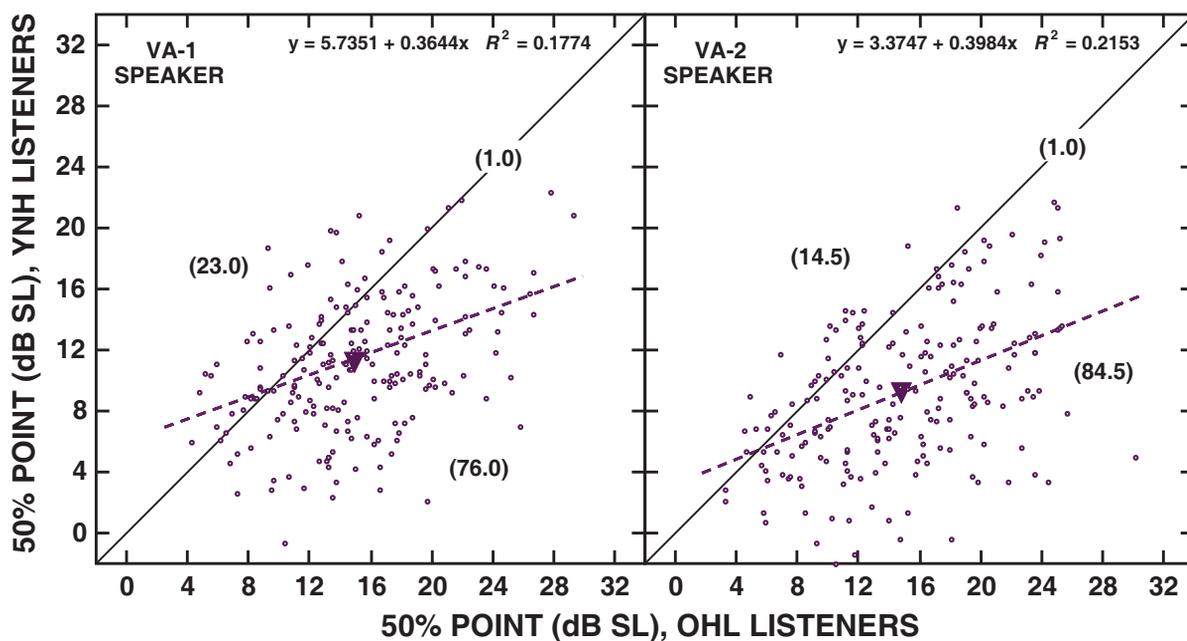


Fig. 9 Bivariate plots of the 50% points (dB SL) of each of the 200 NU-6 words spoken with the carrier phrase by the VA-1 speaker (left panel) and by the VA-2 speaker (right panel). The data were calculated with the Spearman-Kärber equation and were from 24 YNH listeners (ordinate) and 48 OHL listeners (abscissa). The numbers in parentheses are the percent of the 200 words above, on, and below the line of equality. Points below the line indicate better recognition performances by the YNH listeners. The regression equations are listed in each panel and the means are depicted with the larger symbols.

the with carrier phrase condition was involved in the following analyses. The Spearman-Kärber equation, which incorporates the presentation level as a variable, was used to determine the location of the 50% recognition point (dB SL) on the recognition function of each of the 200 NU-6 words for the VA-1 speaker and for the VA-2 speaker. ► **Figure 9** shows bivariate plots of the 50% points for the YNH listeners (ordinate) versus the OHL listeners (abscissa) for the VA-1 speaker (left panel) and VA-2 speaker (right panel). Again, the

numbers in parentheses in each panel are the percent of the 200 words that are above, on, and below the line of equality. Points below the line indicate better recognition performances by the YNH listeners. For both speakers, the vast majority of better performances were obtained by the YNH listeners (76.0% and 84.5% for VA-1 and VA-2, respectively) than by the OHL listeners (23.0% and 14.5%). For the VA-1 and VA-2 speakers, respectively, the average departures from the lines of equality were 2.8 dB and 2.0 dB (YNH) and -5.8 dB

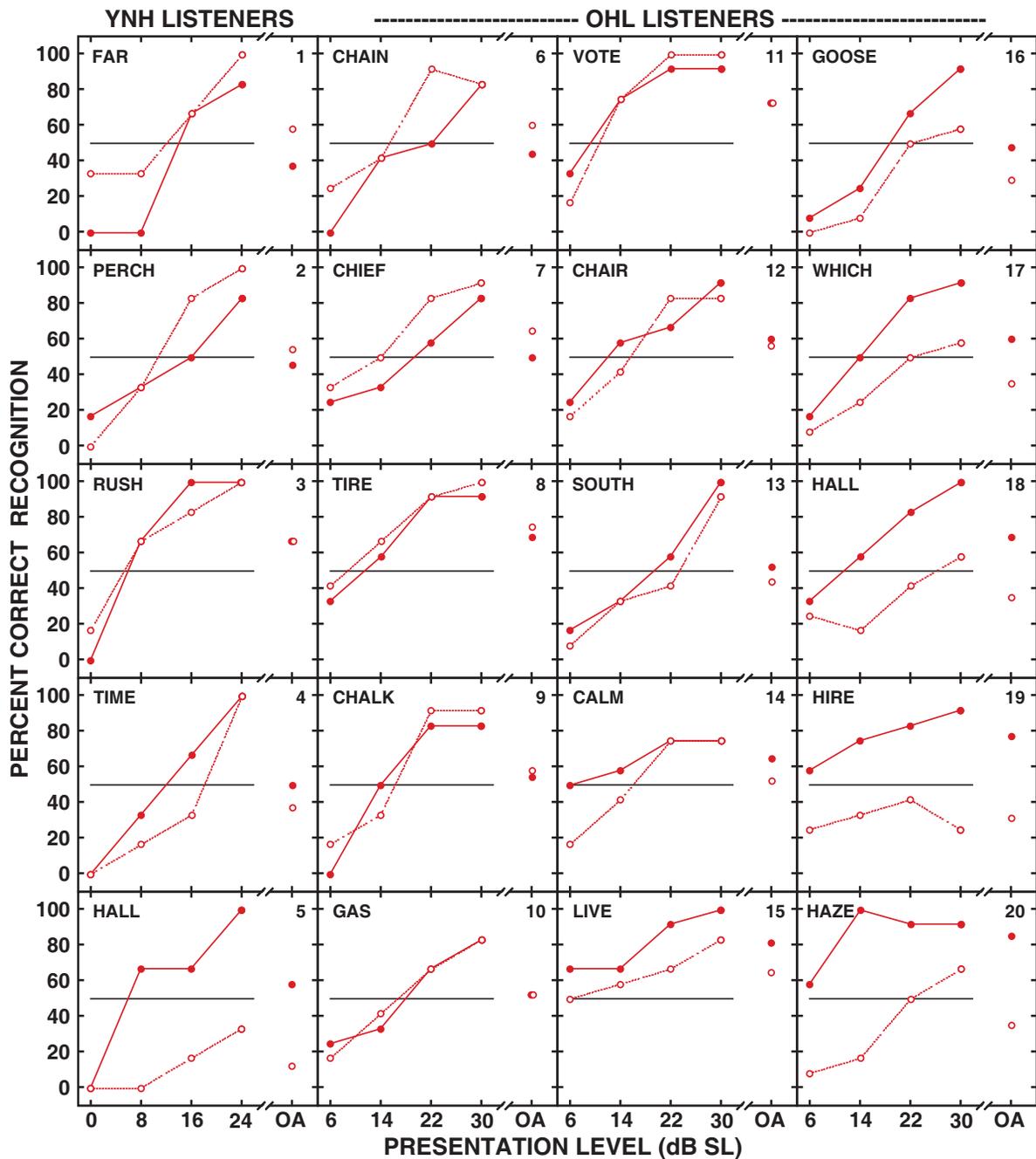


Fig. 10 The psychometric functions of representative individual stimulus words spoken by the VA-1 speaker without the carrier phrase (open circles) and with the carrier phrase (filled circles) that were obtained from the YNH listeners (panels 1-5) and OHL listeners (panels 6-20). The mean overall performance at the four presentation levels for each word (OA) was used to order the sequencing of the words. The functions are sequenced to represent the range of differences observed between the two carrier phrase conditions. The individual data for each word spoken by the VA-1 speaker are listed in Supplemental Tables S15 and S17.

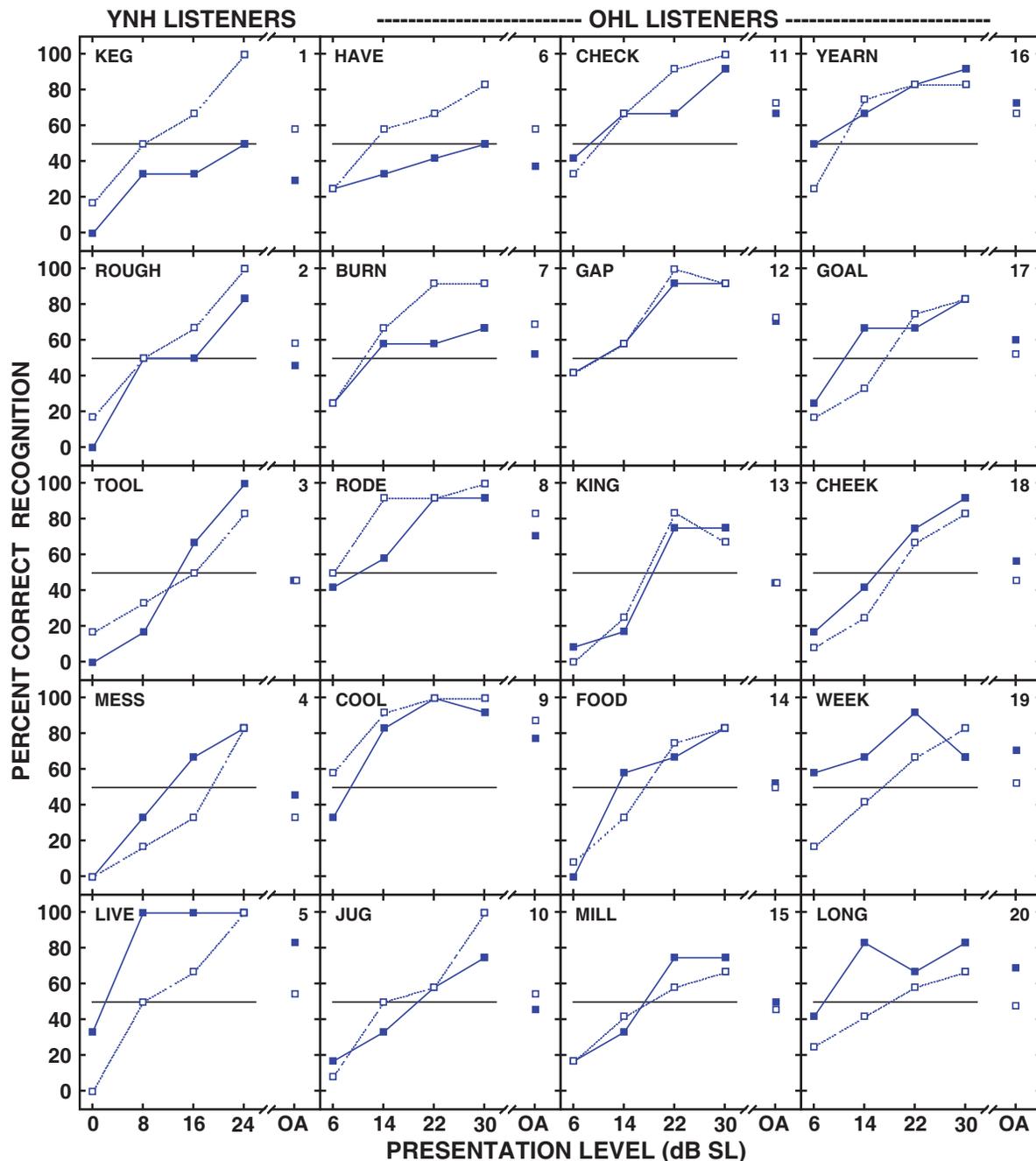


Fig. 11 The psychometric functions of representative individual stimulus words spoken by the VA-2 speaker without the carrier phrase (open squares) and with the carrier phrase (filled squares) that were obtained from the YNH listeners (panels 1-5) and OHL listeners (panels 6-20). The mean overall performance at the four presentation levels for each word (OA) was used to order the sequencing of the words. The functions are sequenced to represent the range of differences observed between the two carrier phrase conditions. The individual data for each word spoken by the VA-2 speaker are listed in Supplemental Tables S16 and S18.

and -7.0 dB (OHL). Thus, in terms of the number of words and the magnitudes of the 50% performance differences, the YNH listeners were better than the OHL listeners. There were, however, 46 words (VA-1) and 29 words (VA-2) on which the OHL listeners exhibited marginally better recognition performances than did the YNH listeners. The data in **Figure 9** cover the gamut of the relations between the with carrier phrase and without carrier phrase conditions, which when

each of which is averaged form the systematic mean recognition functions depicted in **Figure 3**.

The following two figures present the recognition performance functions at the four presentation levels and the overall performances for representative individual words spoken by the VA-1 speaker (**Figure 10**) and by the VA-2 speaker (**Figure 11**). In these figures, the words without the carrier phrase are depicted with open symbols and the words with the carrier phrase are shown with filled symbols. As with the

previous figures, the words without the carrier phrase are shown with open symbols and the words with the carrier phrase are shown with filled symbols. For each listener group, the sequence of datasets in ► **Figures 10** and **11** progresses vertically according to the overall recognition performance difference ($\Delta = \% \text{ correct with the carrier phrase} - \% \text{ correct without the carrier phrase}$) from the negative to the positive extremes. Examination of the differences between recognition performances obtained in the two carrier phrase conditions provides further insight into variability among the individual words. The percent correct differences (the recognition performances with carrier phrase minus the performances without carrier phrase) for each word are listed by the mean overall performance difference between the two carrier phrase conditions in Supplemental Table S19 (YNH) and Supplemental Table S20 (OHL) (Supplemental to the online version of this article).

For the VA-1 speaker, the overall range of performance differences between the two carrier phrase conditions for both groups of listeners was 66.7%, the extremes of which included HALL ($\Delta = 45.8\%$) to FAR ($\Delta = -20.8\%$) for the YNH listeners and HAZE ($\Delta = 50.0\%$) to CHAIN ($\Delta = -16.7\%$) for the OHL listeners. The mean overall performance differences were 6.0% (YNH) and 8.3% (OHL). The distributions of the performance differences deserve mention. Using 61 word to define the range of equal performance differences between the two carrier phrase conditions, 43.5% and 27.0% of the words had equal performances by the YNH and OHL listeners, respectively. Better performances were obtained with the carrier phrase by the YNH and OHL listeners on 46.5% and 61.0% of the words, respectively, contrasted to better performances without the carrier phrase on 10.0% and 12.0% of the words. As listed earlier, with the Δ values and as can be visualized in ► **Figure 10**, with most words, the differential relation is positive, reflecting better recognition performance with the carrier phrase than without the carrier phrase. In addition to the diversity of relations observed between the two carrier phrase conditions for the VA-1 speaker, it is noteworthy that the largest performance differences between the carrier phrase conditions were observed with words whose initial phoneme was /h/ (panels 5, 18, 19, and 20 in ► **Figure 10**). In fact, with the VA-1 speaker, words starting with /h/ had the largest performance differences with 5 of the 6 words with the largest differences (YNH, Supplemental Table S19) and with 6 of the 10 words with the largest differences (OHL, Supplemental Table S20). Words with the initial phoneme /h/ were not among the words spoken by the VA-2 speaker with the largest performance differences between carrier phrase conditions. These observations suggest that excising words with the initial phoneme /h/, which is a phoneme whose onset often is difficult to define, somehow systematically disrupts intelligibility of the word by eliminating cues in the ending of the carrier phrase, in the /h/ phoneme, or in both.

The distributions of recognition performance differences between the two carrier phrase conditions were somewhat different with the VA-2 speaker than with the VA-1 speaker. With the VA-2 speaker (► **Figure 11**), the overall range of performance differences between the two carrier phrase

conditions (a) for the YNH listeners was 58.3%, which included LIVE ($\Delta = 29.2\%$) to KEG ($\Delta = -29.2\%$), and (b) for the OHL listeners was 41.7%, which included LONG ($\Delta = 20.8\%$) to HAVE ($\Delta = -20.8\%$). The mean overall performance differences were 3.0% (YNH) and -0.5% (OHL). Again, using ± 1 word as the range to define equal performance differences between the two carrier phrase conditions, 43.5% and 32.5% of the words had equal performances by the YNH and OHL listeners, respectively, with better performances obtained with the carrier phrase by the YNH and OHL listeners on 38% and 31.5% of the words, respectively, contrasted to better performances without the carrier phrase on 18.5% and 36.0% of the words.

The main purpose of the data in the ► **Figures 10** and **11** is to illustrate the variability of the recognition performances within and between the carrier phrase conditions that was observed with the individual words. Again, the diversity of word recognition functions and the relations between the word functions with and without the carrier phrase can be observed. The first observation from the data in the two figures is that almost all of the functions for the individual words are systematic, i.e., as the presentation level increases or decreases, there is a corresponding increase or decrease in recognition performance. Second, as discussed earlier, in the y-axis domain, the positional relations between the functions for the two carrier phrase conditions cover a range of positive and negative values. Third, there is a lack of homogeneity regarding the ranges of the response functions of different words and sometimes within the two carrier phrase conditions for a given word, especially with the OHL listeners. For some words, the recognition functions encompass a good portion of the response range between 0% and 100%, whereas with other words, the response range is abbreviated across the presentation levels. The fact that there are both relatively complete and abbreviated response functions at the word level suggests that as has been long known (Stevens et al⁴³), the uniform calibration procedure as used in this study was not sensitive to critical auditory cues required for the intelligibility of some words and/or equal intelligibility of a word is not necessarily correlated with the presentation level of that word. To paraphrase Davis¹¹, from an amplitude prospective, all words are not created equal, i.e., equal presentation levels of words do not produce equal intelligibilities. Regardless of the amplitude calibration technique, although the amplitude of a word is a major contributor to the intelligibility of that word, other unquantified cues in the word waveform also make contributions to the overall intelligibility of the word.

In the nonexistent perfect world of word recognition, under reasonable conditions, increases in the presentation level produce increases in recognition performance that is a systematic relation. From time to time, this axiom is violated by a collection of reasonable suspects, ranging from participant inattention/distraction to insufficient data to tester scoring errors, etc., which collectively produce results contaminated with data noise that in turn produces nonsystematic or irregular recognition performance functions. To quantify the nonsystematics in the individual word functions, a simple algorithm was used to determine the instances in which

recognition performance decreased as the presentation level incremented across the four levels. If the percent correct on the subsequent higher presentation level was lower than the value of one word (16.7% for the YNH and 8.3% for the OHL), then the recognition function was considered irregular. Across the four listening conditions (2 speakers and 2 carrier phrase conditions), 1–5% of the word functions from the YNH listeners were irregular and 15–20% of the word functions from the OHL listeners were irregular. Examples of irregular word functions are shown in **Figure 10** (panels 6 and 19) and **Figure 11** (panels 12, 19, and 20). A more extensive, additional set of these irregular functions from 5 YNH listeners and 15 OHL listeners are illustrated in Supplemental Figure S14 (supplemental to the online version of this article).

Comparison of the VA-1 and VA-2 Speaker Versions

The speaker variable (VA-1 and VA-2) was one of the three significant main effects observed with the initial three-way ANOVA for both the YNH and OHL listeners ($p < 0.0001$). It was of interest to explore this finding further, especially because both sets of materials were calibrated using the same 50-msec vowel segment rms procedure. A secondary interest was to establish reference normative data for the VA-2 version of NU-6 with the carrier phrase in quiet as the only previous data available were obtained on 24 YNH listeners in four levels of SSN (Wilson et al⁴⁹). This comparison of the mean recognition data is depicted in the bottom two panels of **Figure 3** for the VA-1 (red circles) and VA-2 (blue squares) and listed in **Table 1**. At the 50% points, the YNH listeners (left panel) performed significantly 2.9-dB lower (better) on the VA-2 version (7.8-dB SL) than on the VA-1 version (10.7-dB SL) [$t_{(24)} = 5.7$, $p < 0.0001$], whereas the OHL listeners performed 0.1-dB better on the VA-2 than on VA-1, which was not a significant difference. The performance differences between speakers also were evaluated at each of the four presentation levels listed in **Table 2** with paired sample t -tests with Bonferroni correction. For the YNH listeners, recognition performances at 16- and 24-dB SL were not significantly different, but were significantly different at 0- and 8-dB SL by 7.0% and 14.0%, respectively [0-dB SL, $t_{(23)} = -5.6$, $p = 0.000$; 8-dB SL, $t_{(23)} = -7.9$, $p = 0.000$]. With the OHL listeners, performances were significantly different only at 6-dB SL by 8.9% [$t_{(46)} = -5.3$, $p < 0.0001$]. Thus, with the 200 NU-6 words, the functions for the VA-1 and VA-2 speakers are similar and intertwine with only slightly better consistent performances with the VA-2 version at lower presentation levels. This relation by both groups at the lower presentation levels is interpreted as an indication that the VA-1 speaker is slightly more difficult to understand than the VA-2 speaker, but only when the listening conditions are more difficult (degraded), in this case at the lower presentation levels.

As was mentioned earlier, the data in **Figure 3** indicate an orderly, systematic relation between the mean functions for the two speakers by both the YNH and OHL listener groups. Examination of the relations between the recognition performance functions for the two speakers with each individual listener, however, again revealed substantial

inter-subject variability. Representative VA-1 and VA-2 functions with the carrier phrase are depicted in **Figure 12** for 5 YNH listeners (1st column) and 10 OHL listeners (2nd and 3rd columns). For both groups of listeners, the data are ordered by the overall smallest to largest recognition performance differences between the functions for the two speakers (VA-1 minus VA-2), which are noted by the Δ in each panel. (Note: The functions in this format with the carrier phrase are depicted for each of the 72 listeners in Supplemental Figures S15–S17, supplemental to the online version of this article.) With the 24 YNH listeners, 22 listeners (91.7%) had better overall performances on the VA-2 version of NU-6 than on the VA-1 version. Participant N17 had the largest difference between the functions for the two speakers ($\Delta = -16.0\%$) with better performance on the VA-2 version than on the VA-1 version. At the other end of the performance continuum, participants N4 and N8 had overall 4.0% and 2.5%, respectively, better performances on the VA-1 version than on the VA-2 version. With the 48 OHL listeners, 25 listeners (52.1%) had better overall performances on the VA-2 version than on the VA-1 version, 22 listeners (45.8%) had better overall performances on the VA-1 version than on the VA-2 version, and 1 listener (S22) had equal overall performances on the two speaker versions. Most of the performance differences between speakers were $< 10\%$. There were, however, five OHL listeners who had overall performance differences (VA-1 minus VA-2 performances) $> \pm 10\%$; specifically, S35 and S37 had differences of 11.5% and 14.5% dB, respectively, whereas S14, S36, and S38 had differences of -11.0% , -14.0% , and -18.5% , respectively. These relations demonstrate again how different listeners respond to different speakers differently. Between these two extremes with the OHL listeners, a variety of relations exist between the two recognition functions of the various listeners. The point here is again the noticeable variability among listeners in their understanding of words spoken by different speakers. The VA-2 speaker was definitely better understood than the VA-1 speaker by some individual listeners, whereas the VA-1 speaker was better understood than the VA-2 speaker by other listeners, with some listeners exhibiting equal performances with both speakers. Furthermore, recognition performances by an individual on materials from two speakers were the same at some levels and different at other levels, all being part of the variability that was observed. As one would expect, the range of differences illustrated in **Figure 13** is substantially larger for the OHL listeners with sensorineural hearing loss than for the YNH listeners.

Finally, with speaker differences, **Figure 13** depicts bivariate plots of the 50% recognition points for the words spoken by the VA-1 speaker (ordinate) versus for the words spoken by the VA-2 speaker (abscissa) for the YNH listeners (left panel) and for the OHL listeners (right panel). With the YNH listeners, 60.0% of the words were better understood when spoken by the VA-2 speaker than when spoken by the VA-1 speaker (31.5%), with 8.5% of the words equally intelligible. The average departures from the line of equality were greater than 5.3 dB and less than -4.0 dB. In contrast with the OHL listeners, both the VA-1 and VA-2 words were

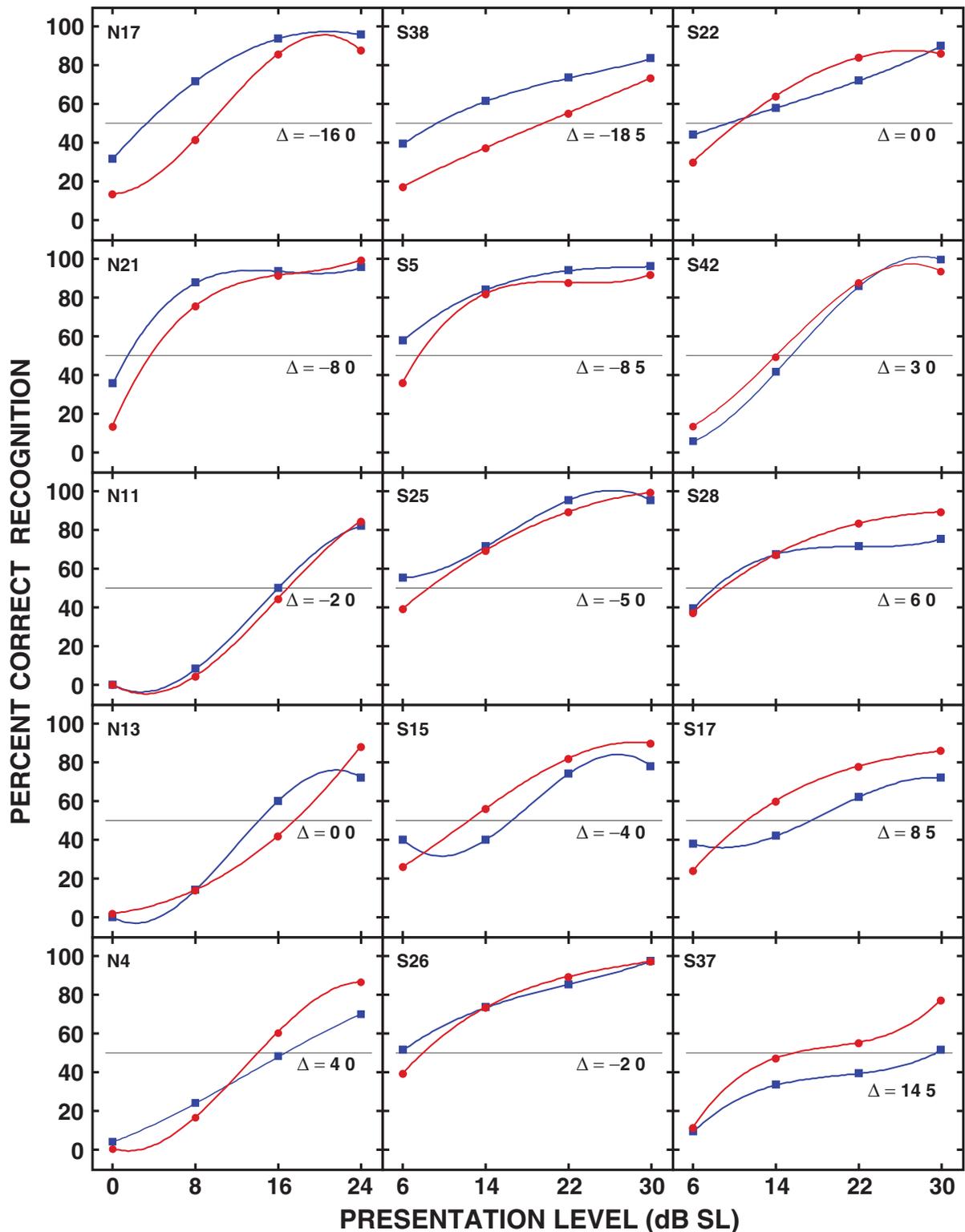


Fig. 12 The mean psychometric functions for the with carrier phrase conditions from representative YNH listeners (left column) and OHL listeners (center and right columns) for the VA-1 speaker (red circles) and the VA-2 speaker (blue squares). The overall recognition performance difference (VA-1 performance minus VA-2 performance) was determined for each listener (Supplemental Figures S15-S17 present the functions for all 72 listeners in this format). The Δ in each panel gives the overall percent correct difference between the two functions. The data for each subject group are arranged from the smallest to the largest difference between functions.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

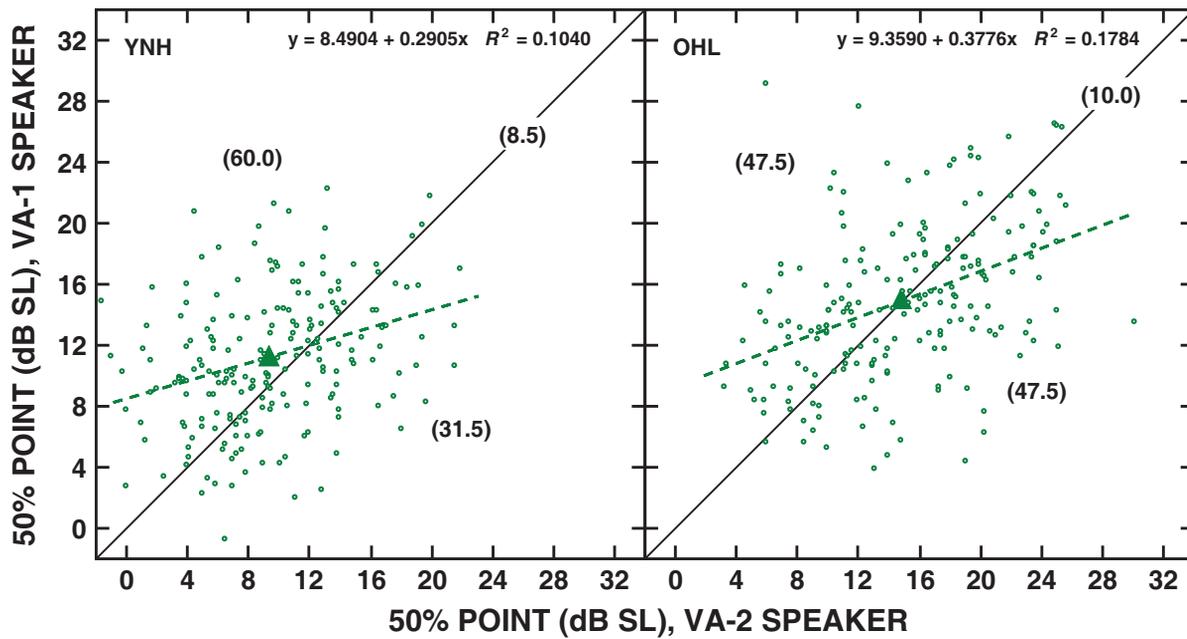


Fig. 13 Bivariate plots of the 50% points (dB SL) of each of the 200 NU-6 words spoken with the carrier phrase by the VA-1 speaker (ordinate) and by the VA-2 speaker (abscissa). The data were calculated with the Spearman-Kärber equation and were from 24 YNH and 48 OHL listeners. The numbers in parentheses are the percent of the 200 words above, on, and below the line of equality. The datum points were jittered using additive algorithms of -0.6 dB to 0.6 dB in 0.1 -dB increments (YNH) and -0.3 dB to 0.3 dB in 0.1 -dB increments (OHL). The regressions were based on nonjittered data. Data above the line of equality indicate better performances on the words spoken by the VA-2 speaker. The regression equations are listed in each panel, and the means are depicted with the larger symbols.

equally understood (47.5%) with departures from the line of equality of greater than 4.7 dB and less than -4.6 dB. Thus, the data for the individual words indicate substantial variability among the 50% points and differences in recognition performances in both directions by both listener groups on NU-6 spoken by two speakers. The distributions of the data in **Figure 13** and the regressions are yet another way to demonstrate that some words by both speakers were equally intelligible, both easy and hard to understand, whereas other words were easy to understand by one speaker and difficult to understand by the other speaker and vice versa.

Recognition Performances by the Female and Male OHL Listeners

The OHL listener group consisted of 27 females ($M = 70.6$ years, $SD = 5.9$ years) and 21 males ($M = 72.7$ years, $SD = 3.9$ years). The PTAs for the two test sessions were 22.8- and 22.3-dB HL for the females and 21.0- and 20.6-dB HL for the males. The mean audiograms shown in Supplemental Figure S18 (supplemental to the online version of this article) are very similar through the low and mid frequencies but separate at the higher frequencies. The mean 4000-Hz threshold was 19.3 dB poorer for the males (59.3-dB HL) than for the females (40.0-dB HL) and the mean 8000-Hz threshold was 8.9 dB poorer for the males, 71.6- and 62.7-dB HL. A one-way ANOVA indicated there were no significant differences between the males and females with regard to age, low- through mid-frequency pure-tone thresholds, or PTAs. Only at 4000 Hz was the mean male threshold significantly poorer than the female threshold [$F_{(1, 46)} = 18.1$, $p < 0.0001$]. The question was as follows: Did other gender

differences in the OHL listeners produce gender word recognition performance differences?

The recognition performance functions for the female and male OHL listeners are depicted in Supplemental Figure S19 (supplemental to the online version of this article), with the data from the VA-1 speaker in the upper panels and from the VA-2 speaker in the lower panels; the data without the carrier phrase are shown in the left panels, and the data with the carrier phrase are given in the right panels. For the VA-1 speaker, the 50% points without the carrier phrase were 15.4-dB SL (females) and 17.6-dB SL (males); with the carrier phrase, the 50% points were 11.5- and 14.1-dB for the respective groups of listeners. The slopes of the functions for the VA-1 version at the 50% point ranged from 2.7%/dB to 3.5%/dB. For the VA-2 speaker, the 50% points without the carrier phrase were 10.2-dB SL (females) and 14.6-dB SL (males). With the carrier phrase, the performances were similar, 10.9- and 14.3-dB SL, respectively, for the two groups. The slopes of the functions at the 50% points were more gradual with the VA-2 speaker than with the VA-1 speaker, ranging from 2.4%/dB to 3.0%/dB. Although the recognition performances by the female listeners were systematically 23 dB (6-8%) better than the performances by the male listeners at every presentation level for both speakers and both carrier phrase conditions, it is interesting that a mixed four-way repeated measures ANOVA indicated there were no significant four- or three-way interactions between gender and speaker, carrier phrase, or presentation level. The lack of statistical findings here indicates that gender, when considered with all of the other conditions, did not significantly influence recognition performance. In all probability, the lack of significance differences was influenced

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

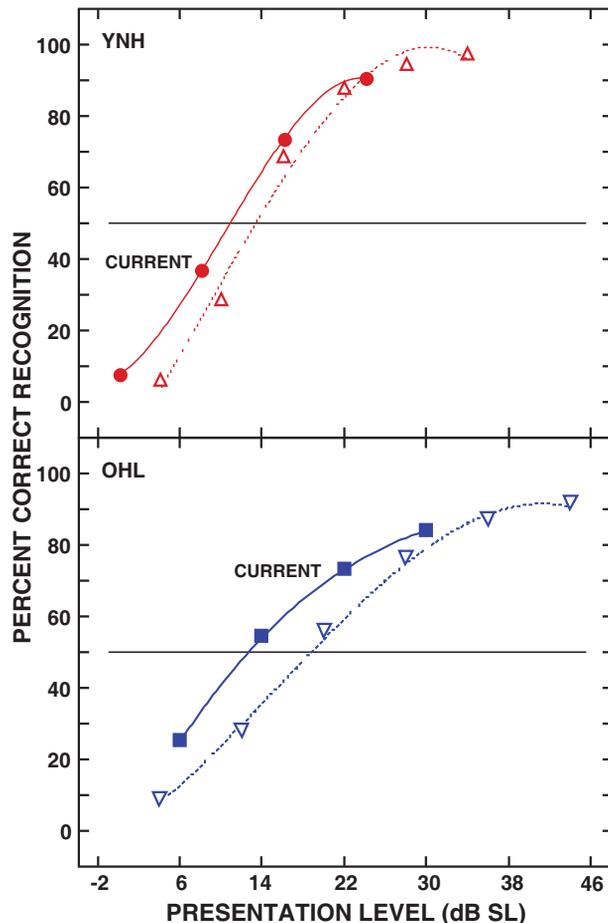


Fig. 14 Comparisons of the recognition performances on the 200 NU-6 words spoken with a carrier phrase by the VA-1 speaker are shown from the current study (filled symbols) and from an earlier study (open symbols, Wilson⁵²) by YNH listeners (upper panel) and OHL listeners (lower panel). Note that the Wilson data were shifted to a 5.7-dB higher presentation level because of the calibration technique used in the current study. The data are described with third-degree polynomials.

by the inherent variability in the data. The SDs were consistently $\sim 15\%$ for all conditions. That being said, perhaps the slightly poorer performances by the male listeners can in part be attributed to the difference in high-frequency sensitivity noted in the audiograms that may be indicative of other auditory functions, even at the cochlear level, that operate less efficiently as the system ages and suffers the impacts of the variety of environmental insults. It is also conceivable that there are probably other auditory functions that are not yet realized that contribute to the slight recognition differences that were observed. The understanding of speech involves a multifarious series of transformations from a speech signal that is a complex acoustic waveform to the perceptual interpretation of the signal as a meaningful unit (a word). To borrow from astrophysics, these other auditory functions and transformations, like the so-called auditory processing, until they are defined and understood should be considered *dark hearing*, i.e., something we know is present but defies and awaits definition, interpretation, and understanding.

Comparisons with Previous Data

Finally, it is informative to compare the current VA-1 speaker data to similar data from the same speaker obtained in an earlier study (Wilson⁵²), in which similar groups of YNH and OHL listeners were evaluated using a slightly different stimulus paradigm involving random presentation levels with the carrier phrase. The 5.7-dB calibration correction described earlier was applied to the previous data for comparison with the current data in **Figure 14**. As can be seen in the figure, the recognition performance functions for the two studies are similar with the data from the current study displaced at the 50% points to the lower presentation levels by 2.4 dB (YNH) and 6.0 dB (OHL). The slopes of the functions at the 50% points were the same for the YNH listeners (4.8%/dB) but slightly different for the OHL listeners, with the function for the current data being slightly steeper (3.2%/dB) than the function from the earlier study (2.9%/dB). The differences between the functions from the two studies are reasonable considering the different procedural differences involved in the two studies (50 words/participant at 4 presentation levels [filled symbols] in the current study and 100 words/participant at 6 presentation levels [open symbols] in the earlier study) and the different groups of participants, especially the OHL groups that were 56% female in the current study and 100% male in the earlier study. In addition, with the OHL listeners, the pure-tone threshold differences between the two groups may have contributed to the word recognition performance differences. The mean PTA in the current study (21.8-dB HL) was about 5-dB lower than the mean PTA in the 2019 study (26.7-dB HL); at 4000 Hz, the threshold difference increased to 16.4 dB, with only a modest difference at 8000 Hz (3.6 dB). It is interesting that the 16.4-dB difference at 4000 Hz is almost identical to the 16.5-dB difference observed between the female and male listeners in the current study.

25-Word and 50-Word Lists of NU-6, Randomization A

The origin of word recognition testing can be traced to the development of the PB-50 word lists by Egan¹³ who indicated that 50 monosyllabic words were necessary to achieve some semblance of phonetic balance (PB) along with other inclusion criteria including equal range of difficulty, common usage, and composition representative of everyday speech. In the ensuing years, although 50 words have become the standard for word recognition testing, the *de facto* standard has become lists of 25 words, typically the 1st and 2nd halves of the traditional (Randomization A) 50-word lists. This unofficial transition to shorter word lists was prompted by audiologists wanting to “save time” (Elpern¹⁴; Grubb²⁰) and studies that (a) question the need for phonetic balance (e.g., Campanelli⁷; Martin et al³¹), (b) demonstrated equal performances on 25- and 50-word lists at high presentation levels (Elpern¹⁴; Resnick⁴¹; Beattie et al⁴), and (c) demonstrated the relative homogeneity with respect to audibility of common, recorded monosyllabic words (Elpern¹⁴). Considering that brief history, recognition performances on the half lists (words 1-25 and 26-50) of NU-6, Randomization A by the VA-1 and VA-2 speakers with the carrier phrases were available in the current dataset and were evaluated along

with their 50-word parent lists, the results of which are shown in Supplemental Figures S20–S22 (supplemental to the online version of this article).

Earlier in this article when the VA-1 and VA-2 versions of NU-6 with the carrier phrase were compared, recognition performances at the 50% points were (a) 2.9 dB better on the VA-2 version by the YNH listeners and (b) essentially the same ($\Delta = 0.1$ dB) for the two versions by the OHL listeners (see **Figure 3**, bottom panels). Importantly, for clinical audiology, for each group of listeners, performances on the two NU-6 versions were essentially the same at the highest presentation levels. Supplemental Figures S20 and S21 present comparisons of the recognition performances on the VA-1 (red circles) and VA-2 (blue squares) versions of the 25- and 50-word lists of NU-6, Randomization A with the carrier phrase. In the figures, the data for the words 1–25 are in the left column of panels, for words 26–50 in the middle columns, and for words 1–50 in the right columns. Again, polynomial equations were used to describe the data from which the sensation level (dB) at the 50% point of each function was calculated. For the YNH listeners (Supplemental Figure S20), overall for each 25-word list, the VA-2 speaker produced recognition performances at the 50% point that were better than the performances produced by the VA-1 speaker. The differences (VA-1 performance minus VA-2 performance) ranged from 0.3 dB (List 3A) to 5.6 dB (List 1A), with the average difference reflecting the overall 2.9-dB difference observed in **Figure 3**. The slopes for the 25-word functions ranged from 4.2%/dB (List 3B) to 5.3%/dB (List 2A) for the VA-1 speaker and 3.6%/dB (List 4A) to 4.6%/dB (List 3B) for the VA-2 speaker. For the OHL listeners (Supplemental Figure S21), most of the 25-word functions for the two speakers were intertwined with differences at the 50% point, ranging from -2.1 dB (List 1A) to 5.8 dB (List 4A), the average of which was close to zero. The slopes for the 25-word functions ranged from 3.1%/dB (List 4B) to 3.7%/dB (List 2B) for the VA-1 speaker and 2.0%/dB (List 4A) to 2.9%/dB (List 3A) for the VA-2 speaker. For both groups of listeners, the performance functions for the respective NU-6, 50-word lists are depicted in the right columns of Supplemental Figures S20 and S21. Again, it should be noted for both groups of listeners, recognition performances at the highest presentation levels were essentially the same on both the 25- and 50-word lists, with most differences between recognition functions apparent at the more difficult listening conditions (lower presentation levels), which should be considered in the realm of degraded speech. Finally, in Supplemental Figure S22, the relations between the recognition performances on the first 25 words (black symbols) and the second 25 words (burnt orange symbols) of each NU-6, Randomization A list by both speakers are shown in Supplemental Figure S22. Again, the half-list recognition differences are minimal and demonstrate that 25-word lists of common monosyllabic words like those used in NU-6 can be used with confidence in the clinical setting, especially at presentation levels ≥ 30 -dB SL.

Summary and Conclusions

The basic question of this study was as follows: How do NU-6 word recognition performances compare with the carrier

phrase preceding the target word versus without the carrier phrase preceding the target word. The results are clear. When the target words were excised from the carrier phrase as with the VA-1 speaker materials, recognition performance decreased in comparison to when the target words were produced as a continuous acoustic stream following the accompanying carrier phrase. The effect was small, overall 6% (YNH) to 8% (OHL), but apparent at each presentation level. With VA-1, the performances at the 50% points on the mean functions with the carrier phrase and without the carrier phrase were, respectively, 10.7- and 12.7- dB SL (YNH) and 12.5- and 16.4- dB SL (OHL). The differences between carrier phrase conditions were significant. When the carrier phrase and target words were produced as separate acoustic streams as with the VA-2 speaker materials, almost identical mean recognition performances ($\pm 2\%$) were observed at seven of the eight presentation levels with only the YNH listeners at 8-dB SL demonstrating significantly 7.6% better performance with the carrier phrase. There was an occasional OHL listener, however, who performed noticeably better with the carrier phrase. For both speakers, the slopes of the functions at the 50% points were 3.9%/dB to 4.8%/dB (YNH) and 2.4%/dB to 3.4%/dB (OHL) with the VA-1 functions $\sim 1\%$ /dB steeper than the VA-2 functions. The current findings apply only to listening in quiet. In degraded speech listening conditions such as multitalker babble or other masking agents, the carrier phrase may cue the listener when to listen for the target word.

The individual participant functions and the individual word functions exhibited a volatility that were masked in the systematic mean recognition functions. All of the individual YNH and OHL participant functions were systematic in that increased presentation level produced a corresponding increase in recognition performance. For the VA-1 and VA-2 speakers, the range of recognition performances over the different 24-dB presentation level ranges were, respectively, 82.7% and 75.8% (YNH) and 59.5% and 47.4% (OHL), which were reflected in the slopes of the mean functions previously mentioned. Comparing recognition performances on NU-6 produced by the two speakers, 23 of the 24 YNH listeners (95.8%) performed overall better on the VA-2 version than on the VA-1 version, with one listener having equal performances. By contrast, 36 of the 48 OHL listeners (75.0%) had better overall performances on the VA-2 version, with 11 listeners (22.9%) being better on the VA-1 version and 1 listener having equal performance. These relations are a good example, especially with the OHL listeners, of different listeners responding to different speakers differently. Perhaps, the most interesting aspect of the individual participant functions occurred with 18 of the OHL listeners who had abbreviated recognition performance functions that mostly were exhibited at performances $> 50\%$. The remaining 30 OHL listeners had more complete functions that were fairly balanced with 2 data points greater than and 2 data points less than 50%. The reasons for these abbreviated functions, which had been observed in an earlier study (Wilson⁵²), ranged from issues with the use of the PTA as the reference for the sensation level to different loudness experiences for these two groups of listeners to factors yet

to be identified. This is yet another reason to retire the concept of the sensation level as used with word recognition testing in favor of fixed presentation levels like normal conversational levels (50-dB HL) plus a second level 20 dB or so higher that collectively would provide two or more points on the recognition performance function. The use of hearing level as the presentation level reference with word recognition testing is not a new concept (Wilson and Strouse⁵¹) being advocated in the 1st edition of the *Audiology Primer* in 1982 (Wilson et al⁵⁰) and more recently by Halpin and Rauch.²¹ After all, we do not listen to speech in terms of sensation level, but rather, we listen in fixed levels such as hearing level and sound-pressure level. The individual word recognition functions were more variable than the individual participant functions, which to some degree probably is related to the difference in sample sizes for the two variables. Some of the word functions covered the 0-100% range of performances, whereas other word functions had substantially shorter performances ranges. With the VA-1 speaker, most words exhibited better recognition performances with the carrier phrase (YNH = 63.0%; OHL = 68.0%) than without the carrier phrase (YNH = 22.5%; OHL = 21.0%), whereas with the VA-2 speaker, the words with the carrier phrase (YNH = 53.5%; OHL = 44.5%) were more equivalent than the words without the carrier phrase (YNH = 27.0%; OHL = 48.0%), especially with the OHL listeners.

Other relations gleaned from the dataset included the following: (1) a comparison of recognition performances on the VA-1 and VA-2 versions of NU-6 demonstrated for both listeners groups that at the higher presentation levels similar performances were achieved, whereas the VA-2 version was slightly better (easier) at the presentation levels at which performances were <50%; (2) there were 27 females and 21 males in the OHL group with the same (a) mean ages, (b) mean pure-tone thresholds except at 4000 Hz at which the mean male threshold was 16.5 dB lower, and (c) recognition performances that were systematically, but nonsignificantly, 6-8% better by the female listeners at all presentation levels; and (3) a comparison with data from a previous study (Wilson⁵²) involving the VA-1 speaker demonstrated performances to be 2.4-dB (YNH) and 6.0-dB (OHL) better in the current study, which was reasonable agreement considering the different procedures and calibrations used in the two studies.

Abbreviations

ANOVA	analyses of variance
CID	Central Institute for the Deaf
CVC	consonant-vowel-consonant
Δ	difference
<i>M</i>	mean
<i>m</i>	slope
MLV	monitored live voice
NU-6	Northwestern University Auditory Test No. 6
OA	overall
OHL	older adults with sensorineural hearing loss
PB	phonetic balance
PTA	pure-tone average

rms	root-mean-square
<i>SD</i>	standard deviation
SN, SNR	signal-to-noise ratio
SL	sensation level
SRT	speech-recognition thresholds
SSN	speech spectrum noise
VA-1	VA speaker 1
VA-2	VA speaker 2
YNH	young adults with normal hearing for pure tones

Notes

Portions of this work were presented at the annual conference of the American Auditory Society in Scottsdale, AZ, March 1, 2019.

Funding

This work was supported by the Rehabilitation Research and Development Service, Department of Veterans Affairs through the Auditory and Vestibular Dysfunction Research Enhancement Award Program (REAP) at the VA Medical Center, Mountain Home, TN. Additional support was provided by the Arizona State University Foundation. Portions of this work, including participant compensation, were supported by the National Institute of Mental Health (NIMH) and the Summer Research Institute at the Florida Mental Health Institute.

Acknowledgments

A substantial portion of this work was performed while the senior author was affiliated with the VA Medical Center, Mountain Home, TN. The authors give special recognition for significant contributions to the study, including help with participant recruitment, data collection, data entry, and study management: Morgan Oktela, Emily Bethune, Marsha-Gaye Allen, Payton Brown, Michelle Arnold, and Celia Escabi. Portions of this work were collected and presented in part of audiology doctoral projects at the University of South Florida.

References

- 1 Adobe Systems, Inc. Adobe Audition CS6. San Jose, CA; 2012
- 2 American National Standards Institute. Specifications for Audiometers (ANSI S3.6-2010). New York, NY: American National Standards Institute; 2010
- 3 American Standards Association. American Recommended Practice for Volume Measurements of Electrical Speech and Program Waves (C16.5-1954). New York, NY: American Standards Association; 1954
- 4 Beattie RC, Svihovec SA, Edgerton BJ. Comparison of speech detection and spondee thresholds for half versus full-list intelligibility scores with MLV and taped presentations of NU-6. *J Am Audiol Soc* 1978;3(06):267-272
- 5 Bonino AY, Leibold LJ, Buss E. Release from perceptual masking for children and adults: benefit of a carrier phrase. *Ear Hear* 2013;34(01):3-14
- 6 Bladon RAW, Henton CG, Pickering JB. Towards an auditory theory of speaker normalization. *Lang Commun* 1984;4(01):59-69
- 7 Campanelli PA. A measure of intra-list stability of four PAL word lists. *J Aud Res* 1962;2(01):50-55

- 8 Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. *J Acoust Soc Am* 1969;45(03):694–703
- 9 Causey GD, Hood LJ, Hermanson C, Bowling LS. The Maryland CNC test: normative studies. *Audiology* 1984;23(06):552–568
- 10 Chinn HA, Gannett DK, Morris RM. A new standard volume indicator and reference level. *Bell Labs Tech J* 1940;19(01):94–137
- 11 Davis H. Tests of hearing. In: Davis H, ed. *Hearing and Deafness: A Guide for Laymen*. New York, NY: Rinehart Books; 1947: 125–160
- 12 Department of Veterans Affairs. *Speech Recognition and Identification Materials*. Disc 4.0. Mountain Home, TN: VA Medical Center; 2006
- 13 Egan JP. Articulation testing methods. *The Laryngoscope* 1948;58(09):955–991
- 14 Elpern B. The relative stability of half list and full list discrimination tests. *The Laryngoscope* 1961;71(01):30–36
- 15 Fletcher H, Steinberg JC. Articulation testing methods. *Bell Sys Tech J* 1929;8(04):806–854
- 16 Fowler CA. Production and perception of coarticulation among stressed and unstressed vowels. *J Speech Lang Hear Res* 1981;24(01):127–139
- 17 Fowler CA. Parsing coarticulated speech in perception: effects of coarticulation resistance. *J Phonet* 2005;33(02):199–213
- 18 Gelfand SA. Use of the carrier phrase in live voice speech discrimination testing. *J Aud Res* 1975;15(02):107–110
- 19 Gladstone VS, Siegenthaler BM. Carrier phrase and speech intelligibility test score. *J Aud Res* 1971;11(01):101–103
- 20 Grubb PA. Some considerations in the use of half-list speech discrimination tests. *J Speech Hear Res* 1963;6(03):294–297
- 21 Halpin C, Rauch SD. Improvement in word recognition score with level is associated with hearing aid ownership among patients with hearing loss. *Audiol Neurotol* 2012;17(03):139–147
- 22 Haskins H. A phonetically balanced test of speech discrimination for children. Unpublished Master's Thesis. Evanston, IL: Northwestern University; 1949
- 23 Hirsh IJ, Davis H, Silverman SR, Reynolds EG, Eldert E, Benson RW. Development of materials for speech audiometry. *J Speech Hear Disord* 1952;17(03):321–337
- 24 Johnson K. Speaker normalization in speech perception. In: Pisoni DB, Remez RE, eds. *The Handbook of Speech Perception*. Malden, MA: Blackwell Publishing; 2005:363–389
- 25 Knight K, Margolis RH. Magnitude estimation of loudness II: loudness perception in presbycusis listeners. *J Speech Hear Res* 1984;27(01):28–32
- 26 Lobdell BE, Allen JB. A model of the VU (volume-unit) meter, with speech applications. *J Acoust Soc Am* 2007;121(01): 279–285
- 27 Lynn JM, Brotman SR. Perceptual significance of the CID W-22 carrier phrase. *Ear Hear* 1981;2(03):95–99
- 28 Margolis RH, Glasberg BR, Creeke S, Moore BC. AMTAS(r): automated method for testing auditory sensitivity: validation studies. *Int J Audiol* 2010;49(03):185–194
- 29 Margolis RH, Saly GL, Le C, Laurence J. Qualind(tm): a method for assessing the accuracy of automated tests. *J Am Acad Audiol* 2007; 18(01):78–89
- 30 Markham D, Hazan V. The effect of talker-and listener-related factors on intelligibility for a real-word, open-set perception test. *J Speech Lang Hear Res* 2004;47(04):725–737
- 31 Martin FN, Champlin CA, Perez DD. The question of phonetic balance in word recognition testing. *J Am Acad Audiol* 2000;11(09):509–513
- 32 Martin FN, Hawkins RR, Bailey HAT. The nonessentiality of the carrier phrase in phonetically balanced (PB) word testing. *J Aud Res* 1962;2(04):319–322
- 33 McArdle R, Wilson RH. Predicting word-recognition performance in noise by young listeners with normal hearing using acoustic, phonetic, and lexical variables. *J Am Acad Audiol* 2008;19(06): 507–518
- 34 McKnight JG. Some Questions and Answers on the Standard Volume Indicator (“vu meter”). 2006. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.193.1938&rep=rep1&type=pdf>. Accessed December 18, 2018
- 35 Miller GA, Heis GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 1951;41(05):329–335
- 36 Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment (MoCA): a brief cognitive screening tool for detection of mild cognitive impairment. *J Am Geriatr Soc* 2005; 53(04):695–699
- 37 Pisoni DB. Some thoughts on “normalization” in speech perception. In: Johnson K, Mullennix JW, eds. *Talker Variability in Speech Processing*. San Diego, CA: Academic Press; 1997:9–32
- 38 Pickett JM, Pollack I. Intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpt. *Lang Speech* 1963;6(03):151–164
- 39 Pollack I. Auditory informational masking. *J Acoust Soc Am* 1975; 57(01):S5
- 40 Redford MA, Kallay JE, Bogdanov SV, Vatikiotis-Tateson E. Leveraging audiovisual speech perception to measure anticipatory coarticulation. *J Acoust Soc Am* 2018;144(04):2447–2461
- 41 Resnick DM. Reliability of the twenty-five word phonetically balanced lists. *J Aud Res* 1962;2(01):5–12
- 42 Robjohns H. Why do waveforms sometimes look lop-sided? *Sound on Sound*. 2013. <http://www.soundonsound.com/sos/may13/articles/qanda-0513-1.htm>. Accessed May 13, 2015
- 43 Stevens SS, Egan JP, Miller GA. Methods of measuring speech spectra. *J Acoust Soc Am* 1947;19(05):771–780
- 44 Studebaker GA. A rationalized arcsine transform. *J Speech Hear Res* 1985;28(03):455–462
- 45 Thwing EJ. Effect of repetition on articulation scores for PB words. *J Acoust Soc Am* 1956;28(02):302–303
- 46 Tillman TW, Carhart R. An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6. Brooks Air Force Base, TX: USAF School of Aerospace Medicine Technical Report; 1966
- 47 Wilson RH. Amplitude (vu and rms) and temporal (msec) measures of two Northwestern University Auditory Test No. 6 recordings. *J Am Acad Audiol* 2015;26(04):346–354
- 48 Wilson RH, Margolis RH. Measurement of the auditory thresholds for speech stimuli. In: Konkle D, Rintelmann W, eds. *Chapter IV in Principles of Speech Audiometry*. Baltimore, MD: University Park Press; 1983:79–126
- 49 Wilson RH, McArdle R, Roberts H. A comparison of recognition performances in speech-spectrum noise by listeners with normal hearing on PB-50, CID W-22, NU-6, W-1 spondaic words, and monosyllabic digits spoken by the same speaker. *J Am Acad Audiol* 2008;19(06):496–506
- 50 Wilson RH, Shanks JE, Akin FW, Johnson E, Murnane O, Noe C, Smith S, Bratt G. *The Audiology Primer for Students and Health Care Professionals*. 3rd ed. Mountain Home, TN: VA Medical Center; 2009. <https://www.etsu.edu/crhs/aslp/audiology/documents/audiologyprimer.pdf>. Accessed May 16, 2019
- 51 Wilson RH, Strouse AL. Auditory measures with speech signals. In: Musiek FE, Rintelmann WF, eds. *Chapter 2 in Contemporary Perspectives in Hearing Assessment*. Boston, MA: Allyn & Bacon; 1999:21–66
- 52 Wilson RH. A comparison of word-recognition performances on the auditec and VA recorded versions of Northwestern University Auditory Test No. 6 by young listeners with normal hearing and by older listeners with sensorineural hearing loss using a randomized presentation level paradigm. *J Am Acad Audiol* 2019;30(05):370–395

Appendix

The vu meter

In speech audiometry, whether antiquated, nonstandardized MLV, or standardized recordings are used to present speech stimuli, the vu meter (see Supplemental Figure S1) is necessary to monitor within certain limits the amplitude of the speech signal that is fed to the output attenuator. (Note: McKnight³⁴ provides vu meter details in an unpublished article originally written in 1971; Lobdell and Allen²⁶ described a digital version of the vu meter.) The output level of an audiometer is the sum of the signal amplitude registered on the vu meter and the level indicated on the output attenuator. Thus, if the signal through the vu meter was -3 vu (or dB), then the output to the transducer would be 3 dB less than the level indicated on the output attenuator. When the vu meter is 0 vu, then the level of the signal to the transducer is the value indicated on the output attenuator. The vu meter was developed originally to monitor the level of signals involved in the broadcast and transmission line industries (Chinn et al¹⁰), a use that continues to this day. The vu meter was never intended as a precise measurement instrument. Two aspects of the vu meter need mention. First, the vu meter is in effect a mechanical averager that has a time constant of 300 msec ($\pm 10\%$), which basically means when the signal level is constant, it takes the monitor needle ~ 300 msec to move full-scale from the resting point to 0 vu. Second, speech is basically an amplitude-modulated signal with monosyllabic words 500-600 msec in duration (Wilson⁴⁷) and no sustained amplitude longer than 100 msec. By contrast, the carrier phrases are 600-1000 msec with a somewhat more constant airstream that makes carrier phrases easier and a little more accurate than monosyllabic words to monitor on a vu meter. These are among the reasons that the ANSI standard for audiometers (2010) specifies that when recording monosyllabic words, a carrier phrase should be used to monitor the level of the signal on a monitoring meter with the target word spoken in a natural manner following the carrier phrase. In effect, the level or amplitude of the target word is not specified, only estimated. A more detailed discussion of the amplitude calibration issues associated with the speech materials used in audiology is provided in a previous article (Wilson⁵²) that includes reference to earlier comments on the calibration of speech signals by Davis¹¹ and Stevens et al.⁴³

Predigital Literature

Martin et al³² investigated the “nonessentiality” of the carrier phrase in clinical speech audiometry. First, a “local” audio tape recording of the CID W-22 Lists 1A, 2A, and 3E (Hirsh et al²³) was made with the carrier phrase, *Say the word*, preceding each word. With List 3E, the authors state that *only the PB word was actually recorded on the tape* (page 319). Second, the List 4A words were recorded without the use of a carrier phrase and the speaker *tried to use equal stress on each word* (page 320). List 1A was always presented first as

a practice list with Lists 2A, 3E, and 4A randomized as the experimental lists presented at the 30-dB SL (re: the SRT). The 75 participants ($M = 43.4$ years) included 15 with normal hearing, 30 with conductive hearing losses, and 30 with sensorineural hearing losses. Each listener in the latter group had a PTA at 500, 1000, and 2000 Hz of ≥ 28 -dB HL, probably with reference to the 1951 ASA standard for pure-tone thresholds, which were about 10 dB higher than the current standard. Although no measures of central tendency were reported, ANOVA found no significant differences among the three lists, the conclusion being that the carrier phrase had no effect on recognition performance at 30-dB SL. This is not an unexpected finding, given that the recognition performances in all probability were at or near maximum. There was, however, one interesting observation. Following data collection, the participants were surveyed regarding their preference for the carrier phrase. Of the 45 participants in the normal hearing and conductive hearing loss groups, 41 preferred not to have the carrier phrase, which was a significant preference. By contrast, 16 of the 30 listeners with sensorineural hearing loss preferred having the carrier phrase, which statistically indicated no significant preference. In this study, it is unfortunate that the stimuli both with and without the carrier phrase were not identical and that for all conditions the single, high presentation level produced maximum or near-maximum performances.

The effects of three carrier phrases (*Say the word* ____, *You will say* ____, and *Point to the* ____) and a no carrier phrase condition were studied by Gladstone and Siegenthaler (1971) using recordings they made of the first 25 words in List 3B of the CID W-22s presented at 5-dB SL (re: the SRT). Thirty-two YNH listeners were studied. When the carrier phrases were used, recognition performances ranged from 47.2% to 56.4%, whereas performance on the no carrier phrase condition was 40.0%. The conclusion was that clinically the carrier phrase should continue to be used even though a clinically unrealistic low presentation level was examined. Another interpretation is that when audibility is low, e.g., 5-dB SL, the carrier phrase alerts the listener and defines the listening interval, which is a cue that slightly enhances word recognition performance.

The effect on word recognition performance was studied by Gelfand¹⁸ using the CID W-22 lists presented by MLV with and without a carrier phrase. Two 50-word lists, one with and one without the carrier phrase, were presented at 35-dB SL (re: the SRT) to each ear of 50 male participants aged 21-66 years with sensorineural hearing loss. All conditions were counterbalanced, and only one speaker was involved. Recognition performance was significantly ($p < 0.01$) 4.7% better with the carrier phrase ($M = 73.9\%$; $SD = 17.0\%$) than without the carrier phrase ($M = 69.2\%$; $SD = 17.3\%$). Although there was a significant difference between the two carrier conditions, 4.7% is only equivalent to a little more than a 2-word difference between conditions. As with the Martin et al study, Gelfand did not use the identical stimuli for the two carrier phrase conditions.

Lynn and Brotman²⁷ examined on young adults with normal hearing the effect that the carrier phrase, “You will say,” which was used in the Hirsh recording of the CID W-22s, had on the 27 target words with a voiceless stop initial consonant (/p/, /t/, and /k/). In the first experiment, the target words were removed from the carrier phrases and based only on the carrier phrase the participants were able to determine the initial consonant of the phantom target word most of the time. This observation suggested that in some conditions, the carrier phrase can contribute intelligibility cues to understanding the target word. In the second experiment, the recognition performances of the 27, W-22 words with a voiceless stop initial consonants were investigated in SSN

with and without a carrier phrase. Both the speech and noise were presented at 70-dB SPL. A light was used to alert the participant when the target word was presented in the without a carrier phrase condition. Recognition performance was significantly 10% better when the carrier phrase was used (37.4%) than when the target words alone were used (27.5%). These findings led to the conclusion that under these conditions in which the listening task was degraded (i.e., on the lower part of the psychometric function in noise), recognition performance was enhanced when the target words were presented with carrier phrases, even when the listening interval for the target words without the carrier phrase was defined visually.

Erratum: This article has been corrected as per the Erratum (DOI: 10.1055/s-0041-1740141) published online on November 29, 2021.