

Effects of Open-Set and Closed-Set Task Demands on Spoken Word Recognition

Cynthia G. Clopper*
David B. Pisoni†
Adam T. Tierney‡

Abstract

Closed-set tests of spoken word recognition are frequently used in clinical settings to assess the speech discrimination skills of hearing-impaired listeners, particularly children. Speech scientists have reported robust effects of lexical competition and talker variability in open-set tasks but not closed-set tasks, suggesting that closed-set tests of spoken word recognition may not be valid assessments of speech recognition skills. The goal of the current study was to explore some of the task demands that might account for this fundamental difference between open-set and closed-set tasks. In a series of four experiments, we manipulated the number and nature of the response alternatives. Results revealed that as more highly confusable foils were added to the response alternatives, lexical competition and talker variability effects emerged in closed-set tests of spoken word recognition. These results demonstrate a close coupling between task demands and lexical competition effects in lexical access and spoken word recognition processes.

Key Words: Lexical competition, spoken word recognition, talker variability, task demands

Sumario

Las pruebas de reconocimiento de la palabra hablada con listas cerradas se utilizan frecuentemente en el contexto clínico para evaluar las habilidades de discriminación del lenguaje de sujetos hipoacúsicos, particularmente de niños. Los científicos del lenguaje han reportado efectos importantes de la competencia del lenguaje y de la variabilidad del hablante en tareas con listas abiertas de palabras, no así en listas cerradas, sugiriendo que las pruebas de reconocimiento de la palabra hablada con listas cerradas pueden no evaluar válidamente las habilidades de reconocimiento del lenguaje. La meta del estudio actual fue explorar algunas de las tareas que pueden relacionarse con esta diferencia fundamental entre el uso de lista cerradas y abiertas de palabras. En una serie de cuatro experimentos manipulamos el número y la naturaleza de las alternativas de respuesta. Los resultados revelaron que conforme se adicionan elementos de confusión a las alternativas de respuesta, los efectos de la competencia léxica y de la variabilidad del hablante emergen en las pruebas de reconocimiento de la palabra hablada con listas cerradas. Estos resultados demuestran un acoplamiento cercano entre los efectos de las demandas de las tareas y la competencia léxica, en los procesos de acceso léxico y de reconocimiento de la palabra hablada.

Palabras Clave: Competencia léxica, reconocimiento de la palabra hablada, variabilidad del hablante, demandas de la tarea

*Department of Linguistics, Northwestern University; †Department of Psychological and Brain Sciences, Indiana University; ‡Department of Cognitive Science, University of California, San Diego

Spoken word recognition tasks have been used for more than a half-century in settings as diverse as the testing of military radio equipment (Miller, 1946), studies of speech intelligibility (Horii et al, 1970), and clinical tests of the auditory capabilities of hearing-impaired individuals (Owens et al, 1981). Most of the early speech intelligibility tests described by Miller (1946) used open-set tests of word or syllable recognition. However, many researchers began to use closed-set speech intelligibility tests because they were faster and easier to administer and score (Black, 1957). The underlying assumption of the new closed-set multiple choice intelligibility tests was that the basic process of word recognition would be fundamentally the same, regardless of the response format of the test. Speech and hearing scientists thought that the only difference between open-set and closed-set tasks was chance performance ($1/L$ in open-set tests, where L is the size of the mental lexicon; $1/N$ in closed-set tests, where N is the number of response alternatives provided by the experimenter). Indeed, few researchers have explicitly examined the differences in task demands between open-set and closed-set tests.

The closed-set testing format has remained popular in clinical settings because it can be simply and rapidly administered and scored (Black, 1957); the results obtained from closed-set tests are reliable even with a small number of trials (Gelfand, 1998, 2003); and participants do not show learning over the course of the task (House et al, 1965). In addition, closed-set tests are commonly used with patients who do not perform well on open-set tests and with children because overall performance is typically better on closed-set tests than open-set tests given the more constrained nature of the task. Some of the more recent closed-set tests also enable clinicians and researchers to pinpoint perceptual difficulties that patients may have with specific features or phonemes (e.g., Owens et al, 1981; Foster and Haggard, 1987). Closed-set tests are so desirable in clinical settings that Schultz and Schubert (1969) converted a popular open-set clinical test of spoken word recognition, the W-22 (Hirsh et al, 1952), into a closed-set test in order to take advantage of the closed-set format.

However, the change from open-set to closed-set tests is not neutral with respect to

theories of spoken word recognition (Luce and Pisoni, 1998). In addition to the difference in the level of chance performance, closed-set tasks are fundamentally different from open-set tasks in terms of their information processing demands, particularly with respect to the level of competition between potential responses. In closed-set tests, competition between alternatives is limited to the response set provided by the experimenter. In open-set tests of spoken word recognition, however, lexical competition exists at the level of the entire mental lexicon. Although many speech scientists believe that spoken words are recognized by identification of their component segments, recent theoretical work has questioned this fundamental assumption and proposed alternatives to this traditional segmental view of speech perception and spoken word recognition.

A number of recent models of spoken word recognition assume that spoken words are recognized relationally as a function of their phonetic similarity to other words in the listener's lexicon through processes of activation and inhibition (e.g., Cohort Theory [Marslen-Wilson, 1984], Shortlist [Norris, 1994], and the Neighborhood Activation Model [Luce and Pisoni, 1998]). The relative nature of lexical access and spoken word recognition was first described by Hood and Poole (1980), who found that some words were easier to recognize than other words, regardless of phonetic content or frequency of occurrence (see also Treisman, 1978). A reanalysis of the properties of Hood and Poole's (1980) "easy" and "hard" words by Pisoni et al (1985) confirmed that average word frequency did not differ between the "easy" and "hard" words. However, the two groups of words did differ significantly in terms of the number and nature of other words that were phonetically similar to them. The "hard" words in Hood and Poole's (1980) study were phonetically similar to many high-frequency words, whereas the "easy" words were phonetically similar to fewer words overall that were also lower in frequency.

In the Neighborhood Activation Model, the lexical neighborhood of a given word is defined as all of the words differing from that word by one phoneme by substitution, deletion, or insertion (Luce and Pisoni, 1998). Lexical activation and competition reflect the contribution of possible word candidates

from these lexical neighborhoods. In several experiments, Luce and Pisoni (1998) found that words with few lexical neighbors were recognized better and were processed more quickly in lexical decision and naming tasks than words with many neighbors. These words are therefore considered “easy” words because they reside in “sparse” neighborhoods. Words with many high-frequency neighbors were recognized more poorly and were processed more slowly in lexical decision and naming tasks than words with few neighbors. These words are therefore considered “hard” words because they come from “dense” neighborhoods. Since word-frequency effects were not found in the naming task, Luce and Pisoni (1998) concluded that lexical competition effects are distinct and independent from well-known word frequency effects. Lexical competition effects have also been reported in priming tasks (Goldinger et al, 1989), as well as in word recognition tasks with hearing-impaired populations (Kirk et al, 1997; Dirks et al, 2001) and second language learners (Takayanagi et al, 2002). If closed-set tests of spoken word recognition assess the same basic processes as open-set tests, we should expect to see evidence of lexical competition effects in closed-set tests of spoken word recognition.

Exemplar-based models of the mental lexicon make another set of predictions about the factors that affect lexical access and spoken word recognition. In an exemplar-based model of the lexicon, the listener stores every utterance that he or she encounters in memory, although some decay in representational strength over time is typically assumed (Murphy, 2002). Recognition is facilitated when similar exemplars have previously been encountered either because the talker is familiar or the word is very frequent. Goldinger (1996) found evidence to support an exemplar approach to the lexicon in a series of spoken word recognition and recall tasks that demonstrated that target items that matched stored representations in terms of linguistic and talker-specific information were more easily accessed than nonmatching targets.

A number of other studies have also reported evidence to support the claim that talker-specific information is stored in memory and then used to facilitate later speech processing. In an early study, Craik and Kirsner (1974) found that performance

in a recognition memory task was better when the words were spoken by the same talker than by a different talker. Palmeri et al (1993) replicated this finding and found that subjects were able to explicitly identify whether an item had originally been presented in the same voice or a different voice. Finally, Nygaard et al (1994) trained listeners to identify ten different talkers by name over several days. In a subsequent word recognition test, they found that the listeners performed better when the words were produced by the familiar talkers they were exposed to during training than when the words were produced by new talkers. These studies all suggest that detailed information about a talker’s voice is not discarded or lost as a result of perceptual analysis but is encoded and retained in memory and used for processing spoken words.

Similarly, talker variability has been shown to affect spoken word recognition performance in a number of open-set tests. In one early study, Creelman (1957) found that word recognition performance decreased as the number of different talkers in a block of trials increased. More recently, Mullennix et al (1989) replicated this result for both accuracy and response latency measures. This talker variability effect has also been found for clinical populations; Kirk et al (1997) found that adult participants with moderate hearing loss performed better on an open-set word recognition test when they were exposed to speech produced by only one talker than when the lists included samples from multiple talkers. When multiple talkers are presented in a single experimental block, the variability in the talker specific information leads to additional competition between potential response candidates. Like the lexical competition effects described above, we should therefore expect to observe talker variability effects in closed-set tasks that assess the same underlying processes as open-set tasks.

In a recent study, however, Sommers et al (1997) explicitly examined the effects of lexical competition and talker variability in both open-set and closed-set word recognition tasks. The participants included normal-hearing adults in quiet listening conditions, normal-hearing adults listening at two difficult signal-to-noise ratios, and adult cochlear implant users in quiet listening conditions. The stimuli were presented in single- and multiple-talker blocks and

included both lexically “easy” and lexically “hard” words. Each listener completed an open-set and a six-alternative closed-set word recognition test. The response alternatives in the closed-set task were taken from the Modified Rhyme Test (MRT; House et al, 1965), which consists of six CVC words differing in either the first or the last consonant.

Sommers et al (1997) replicated the earlier lexical competition and talker-variability effects in the open-set condition of their study. Performance was better on the “easy” words than the “hard” words and in the single-talker blocks than the multiple-talker blocks. These effects were robust across all of the listener groups, except the normal-hearing participants in the quiet listening condition in which ceiling effects were observed. In the closed-set test conditions, however, Sommers et al (1997) found no significant effects of lexical competition or talker variability for any of the four listener groups. These null results suggest that performance in closed-set tests of word recognition may be fundamentally different than performance in open-set tests, because lexical competition and talker variability are not significant variables in closed-set performance. That is, closed-set tests of spoken word recognition may lack validity because they do not truly assess the word recognition processes that they were designed to measure (Bilger, 1984; Walden, 1984). The goal of the present set of experiments was to examine word recognition processes in more detail to determine what factors contribute to the difference in performance between open-set and closed-set tests of speech perception.

One explanation for the difference in performance between open-set and closed-set tasks is that recognizing words in the closed-set tasks is simply easier than recognizing those same words in open-set tests. Overall word recognition performance may be better in closed-set than open-set tasks because the increased task demands in the open-set condition lead the listener to adopt different processing strategies. In the present study, we manipulated task difficulty in the closed-set task in three different ways. First, we inserted a delay of one second between the presentation of the auditory stimulus and the presentation of the response alternatives because several researchers have recently suggested that word frequency and

talker variability effects are affected by task demands such as memory load. For example, McLennan and Luce (2005) argued that talker-specific information is processed relatively slowly and that talker variability effects are attenuated when processing is rapid and/or easy. They used a long-term repetition priming paradigm and a shadowing task with two sets of nonwords. One set, the “easy” set, were very un-wordlike, while the other set, the “hard” set, were very wordlike. One group of participants responded immediately after the presentation of the stimulus item, and the second group responded after a 150 msec delay. Differential priming effects between the easy and hard words were found only in the delayed task, suggesting that the short response delay affected the listeners’ processing strategy.

The failure to find lexical competition and talker variability effects using closed-set word recognition tests may, therefore, be due to the relative ease of the closed-set tests and the speed with which such tests can be completed. In general, spoken words can be recognized with only partial acoustic-phonetic information (Grosjean, 1980). A listener does not necessarily need to access or contact the lexicon to perform the closed-set speech intelligibility task but, instead, can use a more general pattern-matching strategy. We might therefore expect to find lexical competition and talker variability effects in the closed-set task if the processing demands were made more difficult by requiring the listener to store the stimulus word in memory for a short period of time before responding. To test this prediction, we required the listeners to encode and maintain the stimulus item in memory for one second before making their response in the closed-set condition.

Second, we manipulated the number of response alternatives in the closed-set task because an increase in the number of response alternatives has also been shown to increase the difficulty of word recognition tasks and to interact with word frequency and talker variability effects. In their seminal paper on multimodal speech perception, Sumbly and Pollack (1954) examined the effects of response set size on spoken word recognition performance. They found that performance decreased as set size increased from two alternatives to 64 or more alternatives. These results are also consistent with traditional assumptions about the

relationship between chance performance and task difficulty. In particular, as the number of response alternatives increases, the level of chance performance decreases, and the task becomes increasingly more difficult. We might therefore predict that simply increasing the size of the response alternative set might increase lexical competition effects by promoting more competition between potential responses.

Several closed-set tests have also revealed talker-variability effects when the number of response alternatives was relatively large. For example, Verbrugge et al (1976) found significant talker variability effects in a vowel-identification test using hVd stimuli and a 15-alternative forced-choice task. They found that the participants' accuracy decreased as the number of different talkers increased. More recently, Nyang et al (2003) observed talker-variability effects in a 12-alternative forced-choice test of spoken word recognition using stimulus materials produced by native and non-native English speakers. Taken together, the findings from these two studies suggest that we might also expect to find talker variability effects in a closed-set task when more response alternatives are presented to the listeners. To test this hypothesis, we examined performance in both six-alternative and 12-alternative closed-set tasks.

Finally, we manipulated the phonetic confusability of the response alternatives because the nature of the response alternatives has been found to affect overall task difficulty in spoken word recognition. Pollack et al (1959) reported that the most important factor in determining closed-set word recognition performance in noise was the degree of confusability between the target and the foils. Typically, speech intelligibility tests are designed to maximize the similarity between the response alternatives and the target in order to accurately assess speech discrimination performance. For example, Black (1957) selected foils for the target words in his multiple-choice intelligibility test from incorrect responses to the same set of targets when they were presented in an open-set condition in noise. That is, he used response foils that he knew would be confusable with his target words based on data collected in a similar task with human participants.

A more common approach to the selection of response alternatives is to define an

objective measure of similarity, such as a single phoneme or feature substitution, and select foils accordingly. For example, the response alternative sets in Fairbanks' (1958) Rhyme Test differed only in the initial consonant. House et al (1965) elaborated on Fairbanks' (1958) Rhyme Test and constructed sets of CVC words in which all of the words in each set contained either the same initial consonant and vowel or the same vowel and final consonant in their Modified Rhyme Test. In the Minimal Auditory Capabilities (MAC) battery, Owens et al (1981) used a similar design for determining phoneme discrimination abilities for English CVC words differing either in initial consonant, final consonant, or vowel. Finally, Foster and Haggard (1987) constructed lists of targets and foils that differed only on individual featural dimensions based on minimal pairs. Despite this range of possibilities, however, no one has explicitly examined the role of phonetic similarity in determining closed-set word recognition performance. We predicted that the nature of the similarity between the response alternatives would affect competition between response alternatives, leading to an emergence of the lexical competition and talker variability effects in tasks where the response alternatives were highly confusable.

The primary goal of the current set of experiments was to explore the role of task demands in eliciting talker variability and lexical competition effects in closed-set tests of spoken word recognition. We manipulated the timing of the presentation of the response alternatives (Experiment 1), the number of response alternatives (Experiment 2), and the degree of phonetic confusability between the foils and the target (Experiments 3 and 4). In Experiment 1, we replicated previous research using relatively confusable foils in a six-alternative closed-set task. In Experiment 2, we increased the number of phonetically confusable response alternatives to 12 to examine the effects of set size and competition on performance. Experiments 3 and 4 were designed to examine the independent contributions of response set size and phonetic similarity. We used 12 phonetically dissimilar response alternatives in Experiment 3 and six highly confusable response alternatives in Experiment 4. We therefore had two six-alternative experiments (Experiments 1 and 4), two 12-alternative

experiments (Experiments 2 and 3), two highly confusable sets of response alternatives (Experiments 2 and 4), and two less confusable sets of response alternatives (Experiments 1 and 3). In all four experiments, we also manipulated the relative timing of the presentation of the response alternatives and the presentation of the auditory signal to explore effects of memory load on encoding and recognition. Participants were asked to identify isolated spoken words in noise under three response format conditions: (1) open-set; (2) closed-set (before), in which the response alternatives were presented before the auditory signal; and (3) closed-set (after), in which the response alternatives were presented after a short delay following the auditory signal. The results of these studies demonstrate that spoken words are not recognized in an invariant manner across different experimental conditions. The context of the experiment and the specific task demands differentially affect the amount of lexical competition and overall level of performance that are observed.

EXPERIMENT 1

Methods

Listeners

Sixty-eight participants (24 male, 44 female) from the Indiana University community were recruited to serve as listeners in Experiment 1. Data from eight participants were discarded prior to the analysis for the following reasons: four reported a history of a hearing or speech disorder at the time of testing; one was substantially older than the other participants; one did not complete the entire experiment; and the data from two participants were discarded due to experimenter error. The remaining 60 listeners were all 18- to 25-year-old monolingual native speakers of American English with no reported hearing or speech disorders at the time of testing. The listeners were randomly assigned to one of two listener groups: the single talker group ($N = 30$) or the multiple talker group ($N = 30$). The listeners

received partial course credit in an introductory psychology course for their participation in this experiment.

Stimulus Materials

A set of 132 CVC English words was selected for use in this study from the Modified Rhyme Test (MRT; House et al, 1965) and the Phonetically Balanced (PB; Egan, 1948) word lists. The words were divided into two groups based on measures of lexical competition (lexically "easy" vs. lexically "hard" words) with 66 words in each group. Mean log frequency was equated across the two groups of words, based on the lexical frequency data provided in Kucera and Francis (1967). Similarly, mean word familiarity, as judged by Indiana University undergraduates (Nusbaum et al, 1984), was also equated across the two groups of words. The lexically "hard" words had a significantly higher mean lexical density (defined as the number of words that differ from the target by a single phoneme substitution, deletion, or insertion) than the "easy" words. In addition, the lexical neighbors of the "hard" words had a significantly higher mean log frequency than the lexical neighbors of the "easy" words. As in Sommers et al (1997), the "easy" and "hard" sets of test words were defined based on these differences in mean density and mean neighborhood log frequency. Table 1 shows a summary of the means for each set of words on the four lexical measures, as well as significance values for the t-tests run to compare the means.

Five male talkers were selected from a total of 20 talkers (10 males and 10 females) who were recorded reading the MRT and PB word lists for the PB/MRT Word Multi-Talker Speech Database (Speech Research Laboratory, Indiana University). Goh (2005) conducted a similarity judgment task on the talkers included in the PB/MRT corpus with normal-hearing adult listeners. Based on his results, we selected five male talkers for the current experiment who were all highly discriminable from one another.

Each of the 132 words was spoken by each of the five talkers, for a total of 660 tokens. Each token was stored in an individual sound file in .wav format. For the present study, the tokens were degraded using a bit-flipping procedure written in

Table 1. Lexical Properties of the “Easy” and “Hard” Words

	Easy	Hard	
Mean Log Frequency	1.91	2.02	$p = .41$
Mean Familiarity	6.81	6.70	$p = .27$
Mean Neighborhood Density	16.03	24.86	$p < .001$
Mean Neighborhood Log Frequency	1.83	2.17	$p < .001$

Note: P-values indicate the results of independent sample t-tests.

Matlab (Mathworks, Inc.). In this procedure, degradation is introduced into the signal by flipping the sign of a randomly selected proportion of the bits in the signal. The higher the percentage of bits that are flipped, the more degraded the signal is. Based on pilot studies (see Clopper and Pisoni, 2001), 10% degradation was selected in order to produce responses below ceiling level in the closed-set conditions and above floor level in the open-set condition.

For the closed-set conditions, a six-alternative forced-choice task was designed. Each of the five response foils for each target word differed from the target by the substitution of a single phoneme. The foils were selected such that they were rated by undergraduates as having a familiarity rating of greater than 6.0 on a 7-point scale (Nusbaum et al, 1984). In addition, the goal was to have two foils higher in frequency, two foils lower in frequency, and one foil with approximately the same frequency as the target. Finally, two foils differed from the target with respect to the initial consonant, two with respect to the final consonant, and one with respect to the vowel. This general procedure for selecting foils could not be followed in all cases due to the constraints of the English language. Therefore, some sets of foils did not match the criteria with respect to minimum familiarity, frequency distributions, or phoneme substitution location, but in all cases the foils differed from the targets by the substitution of a single phoneme, and the familiarity of all foils was greater than 5.0 based on scores obtained from the Hoosier Mental Lexicon database (Nusbaum et al, 1984).

Procedure

The 132 stimulus words were presented one time to each listener for recognition. The stimulus words were randomly assigned

without replacement to one of three experimental blocks for each listener, for a total of 44 trials per block. Each block contained 22 “easy” words and 22 “hard” words. The experimental blocks differed in terms of response format: one block was open-set and two blocks were closed-set. The presentation order of the experimental blocks was counterbalanced across the listeners.

The listeners in the single talker group heard all 132 words produced by a single talker. All five talkers were used in the single talker condition. These talkers were balanced across listener and presentation order of the experimental blocks. The listeners in the multiple talker group heard the 132 words spoken by different talkers. All five talkers were used in the multiple talker condition. The talkers were randomly assigned to individual words so that both the talker and the experimental block in which any given word appeared were randomly selected for each listener.

The listeners were seated at personal computers equipped with matched and calibrated Beyerdynamic DT100 headphones. The listeners heard the words presented one at a time over headphones at a comfortable listening level (approximately 75 dB SPL). The volume output was set using an RMS voltmeter in conjunction with the Beyerdynamic DT100 output specifications. In the open-set block, the listeners were asked to type in the word that they thought they had heard using a standard keyboard. In the closed-set (before) block, one second prior to the onset of the auditory presentation of the words, the six response alternatives were presented in random order in a single row on the computer screen. After hearing the word, the listeners were asked to use the mouse to select which one of the six response alternatives they thought they had heard. In the closed-set (after) block, the six response alternatives were presented on the screen one second after the end of the auditory

presentation of the word. After the presentation of the response alternatives, the listeners were asked to use the mouse to select which one of the six response alternatives they thought they had heard. In all three response conditions, the experiment was self-paced and the next trial was initiated by clicking on a "Next Trial" button on the computer screen with the mouse.

Prior to data analysis, the open-set responses were corrected by hand for obvious typographical errors in cases where the response given was not a real English word and for homophones such as "pear" for "pair."

Results

Figure 1 shows the percent correct responses in each of the three experimental blocks as a function of lexical competition. Figure 2 shows the percent correct responses in each of the three experimental blocks as a function of talker variability. A repeated measures ANOVA with response format (open-set, closed-set [before], closed-set [after]) and lexical competition (easy, hard) as within-subject variables and talker variability (single, multiple) as a between-subject

variable revealed a significant main effect of response format ($F [2, 116] = 385.9, p < .001$) and a significant main effect of lexical competition ($F [1, 116] = 4.8, p = .033$). The main effect of talker variability was not significant. None of the interactions reached significance.

Planned post hoc Tukey tests on response format revealed that performance was significantly better in the closed-set (before) block than in either of the other two experimental blocks (both $p < .05$). Performance was also better in the closed-set (after) block than in the open-set block ($p < .05$).

As shown in Figure 1, while the overall effect of lexical competition was significant, planned post hoc paired sample t-tests on lexical competition within each response format condition revealed a significant effect of "easy" versus "hard" words only in the open-set block ($t[59] = 2.6, p = .011$). In the open-set condition, performance was better for the "easy" words than the "hard" words. The lexical competition effect was not significant in either of the two closed-set blocks.

The results shown in Figure 2 reveal that while the overall effect of talker

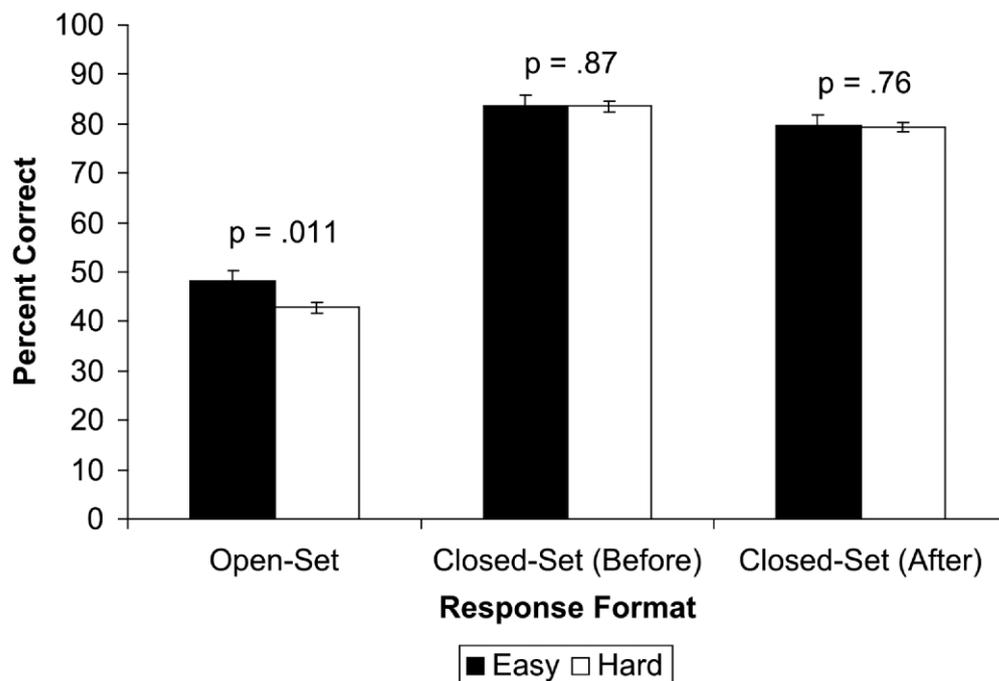


Figure 1. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for easy and hard words, collapsed across talker variability, for the listeners in the six-alternative low confusability experiment (Experiment 1).

variability was not significant, planned post hoc t-tests on the single versus multiple talker groups for each response format condition revealed a significant effect of talker variability in the open-set condition ($t[118] = 2.2, p = .035$). In this condition, performance was better for the single talker group than the multiple talker group. The effect of talker variability was not significant in either of the two closed-set blocks.

Discussion

Two primary findings emerged from Experiment 1. First, we were successful in replicating results from the earlier studies by Mullennix et al (1989) and Sommers et al (1997), which found main effects of lexical competition and talker variability in open-set word recognition tasks but not in closed-set tasks. Second, the delayed response condition did not produce either of the effects found in the open-set condition, suggesting that a one-second delay in a closed-set task is not sufficient to create lexical competition effects that are similar to those found in open-set word recognition tasks.

In Experiment 2, we manipulated the degree of lexical competition between response alternatives in the closed-set task by increasing the number of alternatives from six to 12. By increasing the number of response alternatives within the response set, we predicted that we would also increase the effects of lexical competition and talker variability in lexical access.

EXPERIMENT 2

Methods

Listeners

Fifty-three participants (12 male, 41 female) from the Indiana University community were recruited to serve as listeners in Experiment 2. Prior to the data analysis, data from five participants were removed for the following reasons: three reported a history of a hearing or speech disorder at the time of testing, one was bilingual, and one did not complete the entire

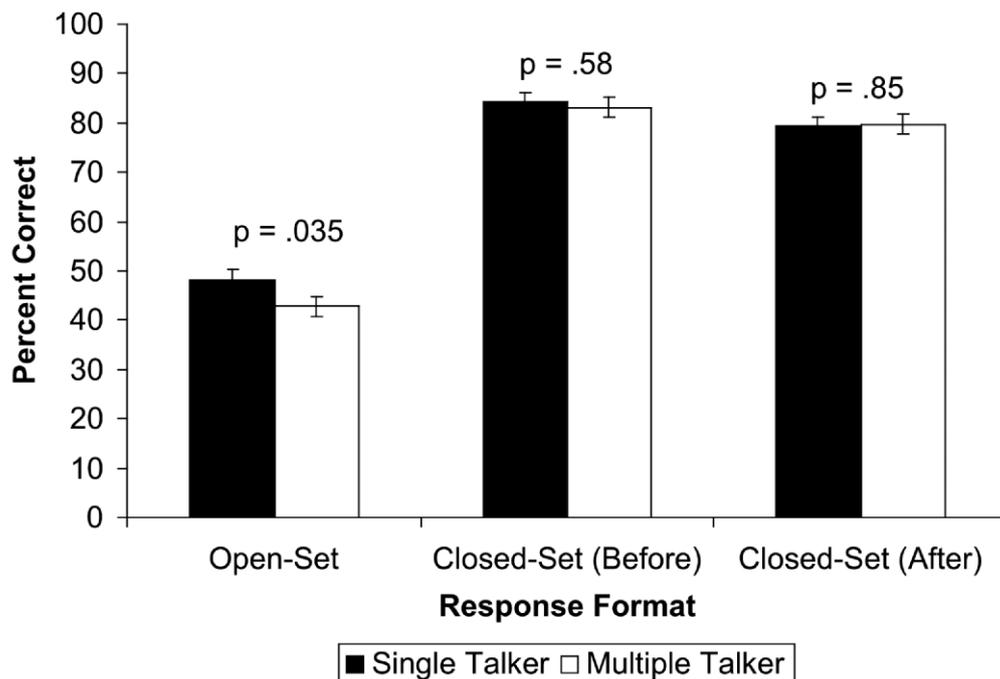


Figure 2. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for the single and multiple talker listener groups, collapsed across lexical competition, for the listeners in the six-alternative low confusability experiment (Experiment 1).

experiment. The remaining 48 listeners were all 18- to 25-year-old monolingual native speakers of English with no hearing or speech disorders reported at the time of testing. The listeners were randomly assigned to one of two listener groups: the single talker group ($N = 24$) or the multiple talker group ($N = 24$). The listeners received partial course credit in an introductory psychology course for their participation in this experiment.

Stimulus Materials

The same talkers and stimulus materials were used in this experiment as in Experiment 1. In the closed-set conditions, however, 12 response alternatives were presented to the listeners. Five of the foils were identical to those used in Experiment 1. The remaining six foils were selected from the listeners' confusion errors in the open-set condition in Experiment 1 and in several pilot studies (see Clopper and Pisoni, 2001).

Procedure

The procedure in this experiment was identical to the procedure used in Experiment 1, except that 12 response alternatives were presented to the listeners in both the closed-set (before) and the closed-set (after) experimental blocks. The response alternatives were presented on the computer screen in random order in three rows of four words each.

Results

Figure 3 shows the percent correct responses in each experimental block as a function of lexical competition. Figure 4 shows the percent correct responses for each of the three experimental blocks as a function of talker variability. A repeated measures ANOVA with response format (open-set, closed-set [before], closed-set [after]) and lexical competition (easy, hard) as within-subject variables and talker variability (single, multiple) as a between-subject variable revealed a significant main effect of response format ($F[2, 92] = 42.2, p < .001$), a

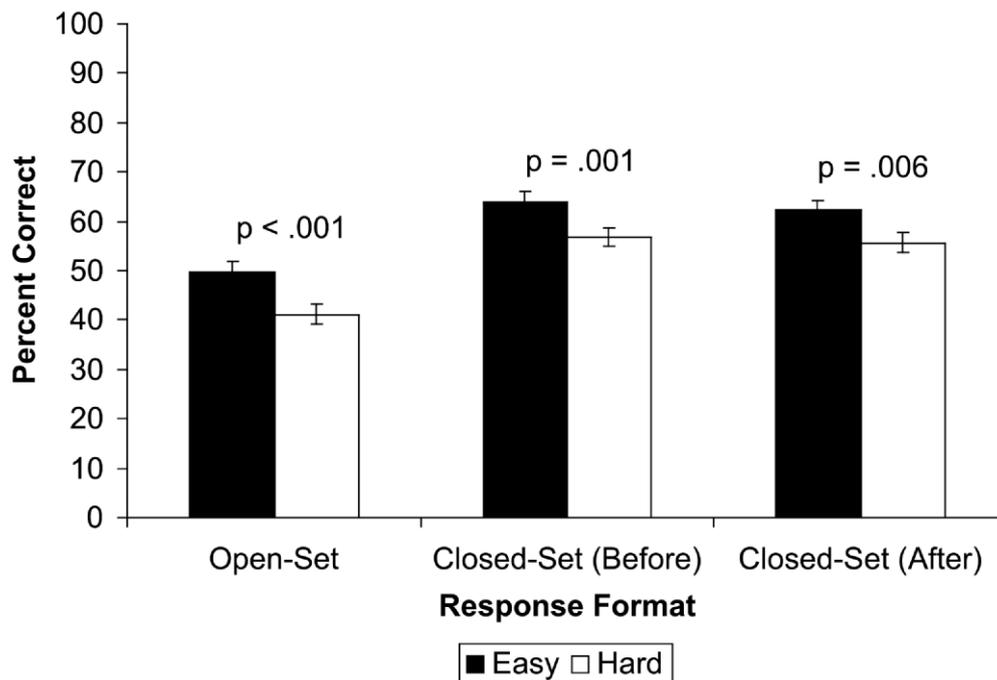


Figure 3. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for easy and hard words, collapsed across talker variability, for the listeners in the 12-alternative high confusability experiment (Experiment 2).

significant main effect of lexical competition ($F[1, 92] = 42.8, p < .001$), and a significant main effect of talker variability ($F[1, 92] = 6.4, p = .015$). None of the interactions were significant.

Planned post hoc Tukey tests on response format revealed that performance in the closed-set (before) and closed-set (after) blocks was significantly better than performance in the open-set block (both $p < .001$). Performance did not differ significantly between the two closed-set blocks.

As shown in Figure 3, planned post hoc paired sample t-tests on “easy” versus “hard” words within each of the response format conditions revealed a significant effect of lexical competition for all three experimental blocks ($t[47] = 4.7, p < .001$ for open-set, $t[47] = 3.4, p = .001$ for closed-set [before], and $t[47] = 2.9, p = .006$ for closed-set [after]). In all three cases, performance was better on the “easy” words than the “hard” words.

As shown in Figure 4, planned post hoc t-tests on the single versus multiple talker conditions for each of the three response format conditions revealed significant effects of talker variability for the open-set ($t[94] = 2.04, p = .044$) and closed-set (after) ($t[94] =$

$2.7, p = .009$) blocks. The talker variability effect was also marginally significant for the closed-set (before) block ($t[94] = 1.8, p = .077$). In all three blocks, performance by the single-talker group was better than performance by the multiple-talker group.

Discussion

The results of Experiment 2 replicated the lexical competition and talker variability effects observed in the open-set condition found in Experiment 1. In addition, these two effects were also significant in both the closed-set (before) and closed-set (after) conditions, although only marginally so for the talker variability effect in the closed-set (before) condition. These results are consistent with previous research on the effects of set size on recognition performance (e.g., Sumbly and Pollack, 1954; Verbrugge et al, 1976) and confirm that competition at the level of the response alternatives is an important factor distinguishing between open-set and closed-set word recognition tasks. In particular, 12 highly confusable alternatives can induce lexical competition and talker variability effects, whereas six less confusable

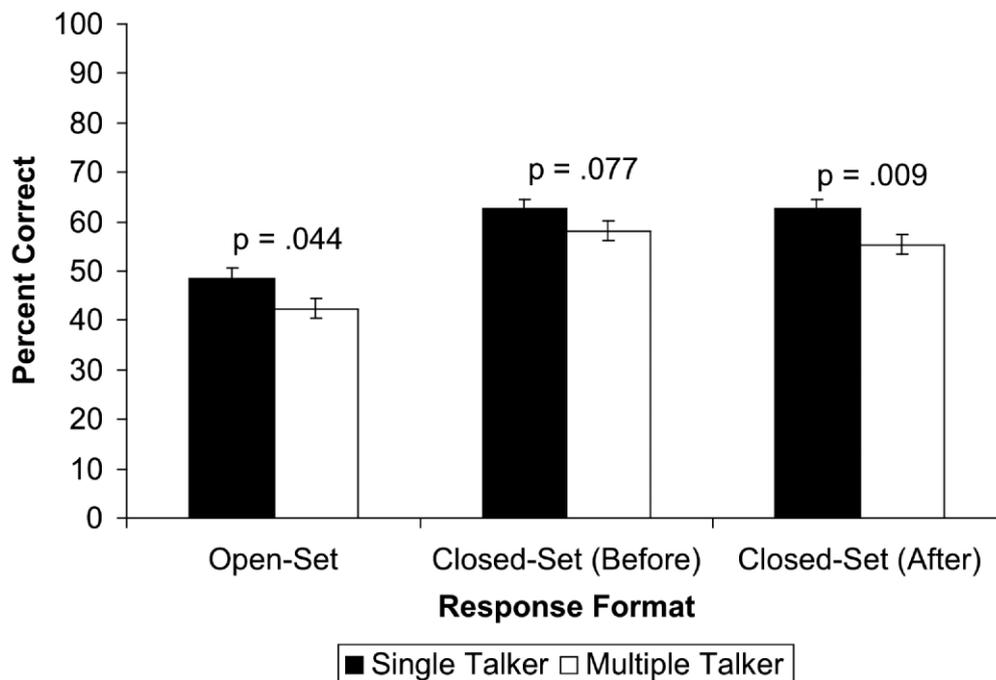


Figure 4. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for the single and multiple talker listener groups, collapsed across lexical competition, for the listeners in the 12-alternative high confusability experiment (Experiment 2).

alternatives, such as those used in Experiment 1, did not.

The findings from this experiment suggest that increased task difficulty due to a greater number of response alternatives led to greater competition effects during lexical access and spoken word recognition. Recall, however, that the original five foils were selected using an objective phonetic similarity metric and that the six additional foils in the current experiment were selected based on open-set confusions. The results of Experiment 2 clearly demonstrate that lexical competition and talker variability effects can be obtained under certain closed-set conditions, but it is not clear whether these effects are due to the total number of response alternatives (i.e., set size), to the phonetic confusability of the response alternatives (i.e., competition in lexical access), or both. Experiments 3 and 4 were designed to tease apart the relative contribution of these two factors. In Experiment 3, we used 12 randomly selected response alternatives to examine the effect of set size, independent of phonetic confusability, on closed-set word-recognition performance. In Experiment 4, we used six highly confusable response alternatives to examine the effect of phonetic confusability, independent of set-size, on performance in the closed-set word recognition tasks.

EXPERIMENT 3

Methods

Listeners

Fifty participants (18 male, 32 female) from the Indiana University community were recruited to serve as listeners in Experiment 3. Prior to the analysis, data from two participants who reported a history of a hearing or speech disorder were discarded. The remaining 48 listeners were all 18- to 25-year-old monolingual native speakers of English with no history of a hearing or speech disorder reported at the time of testing. The listeners were randomly assigned to one of two listener groups: the single talker group ($N = 24$) or the multiple talker group ($N = 24$). The listeners received \$6 for their participation in this experiment.

Stimulus Materials

The same stimulus materials were used in this experiment as in Experiment 2. The set of response alternatives from Experiment 2 were

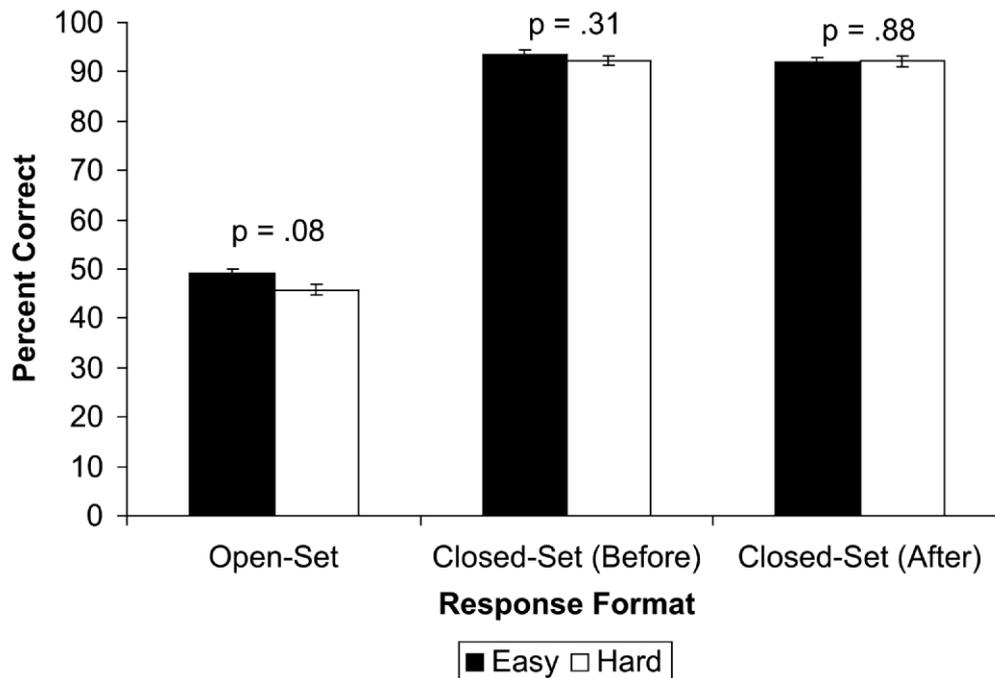


Figure 5. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for easy and hard words, collapsed across talker variability, for the listeners in the 12-alternative low confusability experiment (Experiment 3).

also used in this experiment, but the 11 foils on each trial were randomly selected from the total set of possible foils. Foils that were identical to targets were excluded. The foils were assigned randomly without replacement to the closed-set trials for each individual participant.

Procedure

The procedure was identical to the procedure used in Experiment 2.

Results

Figure 5 shows the percent correct responses in each of the three experimental blocks as a function of lexical competition. Figure 6 shows the percent correct responses in each of the three experimental blocks as a function of talker variability. A repeated measures ANOVA with response format (open-set, closed-set [before], closed-set [after]) and lexical competition (easy, hard) as within-subject variables and talker variability (single, multiple) as a between-subject variable revealed a significant main effect of

response format ($F[2, 92] = 633.9, p < .001$), a marginally significant main effect of lexical competition ($F[1, 92] = 3.7, p = .06$), and a marginally significant main effect of talker variability ($F[1, 92] = 3.5, p = .07$). The response format \times talker variability interaction was also marginally significant ($F[2, 92] = 2.9, p = .06$). None of the other interactions were significant.

As expected, planned post hoc Tukey tests on response format revealed that performance was better in the closed-set (before) and closed-set (after) experimental blocks than in the open-set block (both $p < .001$). Performance did not differ between the closed-set (before) and closed-set (after) blocks.

As shown in Figure 5, planned post hoc paired sample t-tests on easy versus hard words within each response format condition revealed a marginally significant main effect of lexical competition for the open-set block ($t[47] = 1.8, p = .08$). Performance on the “easy” words was better than performance on the “hard” words. The lexical competition effect was not significant for either the closed-set (before) or the closed-set (after) block.

Figure 6 shows that planned post hoc t-tests revealed a significant effect of talker variability

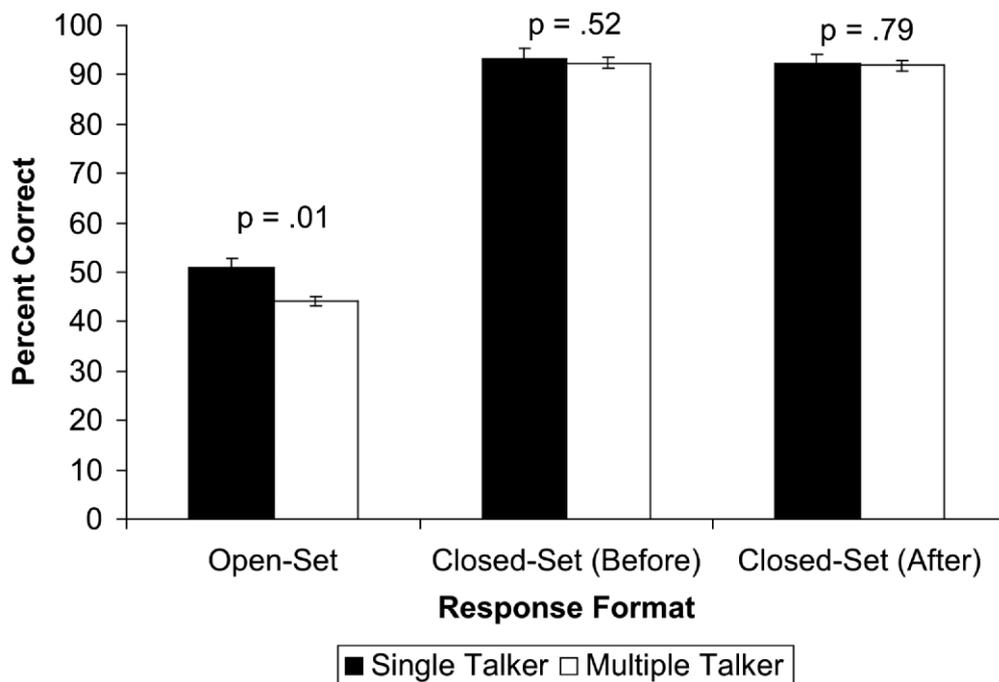


Figure 6. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for the single and multiple talker listener groups, collapsed across lexical competition, for the listeners in the 12-alternative low confusability experiment (Experiment 3).

in the open-set condition ($t(94) = 2.6, p = .011$). Performance was better for the single talker group than the multiple talker group. The effect of talker variability was not significant in either the closed-set (before) or the closed-set (after) condition. This difference in talker variability effects between the open-set block and the other two blocks is responsible for the response format \times talker variability interaction reported above.

Discussion

The results of Experiment 3 replicated the lexical competition and talker variability effects observed in the open-set condition reported in Experiments 1 and 2. As predicted, however, these two effects were not significant in either the closed-set (before) or the closed-set (after) conditions. These results are consistent with contemporary theoretical approaches to spoken word recognition that assume that lexical access is a relational process in which phonetic similarity plays a central role. However, it should be noted that performance in the closed-set (before) and closed-set (after) tasks in this experiment was near ceiling, with a mean accuracy of 92% across the two closed-set tasks. The high level of performance in these conditions suggests that the effects of lexical competition and talker variability may have been obscured by ceiling effects. Even with the degraded stimuli, the closed-set conditions were very easy in this condition due to the phonetic discriminability of the response alternatives. Thus, the lexical competition and talker variability effects might re-emerge with unrelated foils when the task is made more difficult through greater signal degradation and overall performance is lower.

The present findings suggest that the results in Experiment 2 cannot be attributed entirely to an increase in response alternative set size but must also reflect at least one other task demand that is related to the confusability of the response alternatives and/or overall task difficulty. In Experiment 4, we provided listeners with a smaller set of response alternatives that contained increased phonetic confusability to determine whether the effects of lexical competition and talker variability could be observed in a more difficult six-alternative task.

EXPERIMENT 4

Methods

Listeners

Fifty participants (16 male, 34 female) from the Indiana University community were recruited to serve as listeners in Experiment 4. They were all 18- to 25-year-old monolingual native speakers of American English with no history of hearing or speech disorders reported at the time of testing. Participants were randomly assigned to one of two listener groups: the single talker group ($N = 25$) or the multiple talker group ($N = 25$). The listeners received \$6 for their participation in this experiment.

Stimulus Materials

The same stimulus materials were used in this experiment as in Experiments 1, 2, and 3. The foils were selected from the subset of foils added in Experiment 2 that were obtained from open-set errors in previous experiments. For each trial for each listener, five of the six additional foils used in Experiment 2 were randomly selected.

Procedure

The procedure was identical to the procedure used in Experiment 1.

Results

Figure 7 shows the percent correct responses in each of the three experimental blocks as a function of lexical competition. Figure 8 shows the percent correct responses in each of the three experimental blocks as a function of talker variability. A repeated measures ANOVA with response format (open-set, closed-set [before], closed-set [after]) and lexical competition (easy, hard) as within-subject variables and talker variability (single, multiple) as a between-subject variable revealed a significant main effect of response format ($F[2, 92] = 92.7, p < .001$) and a significant main effect of

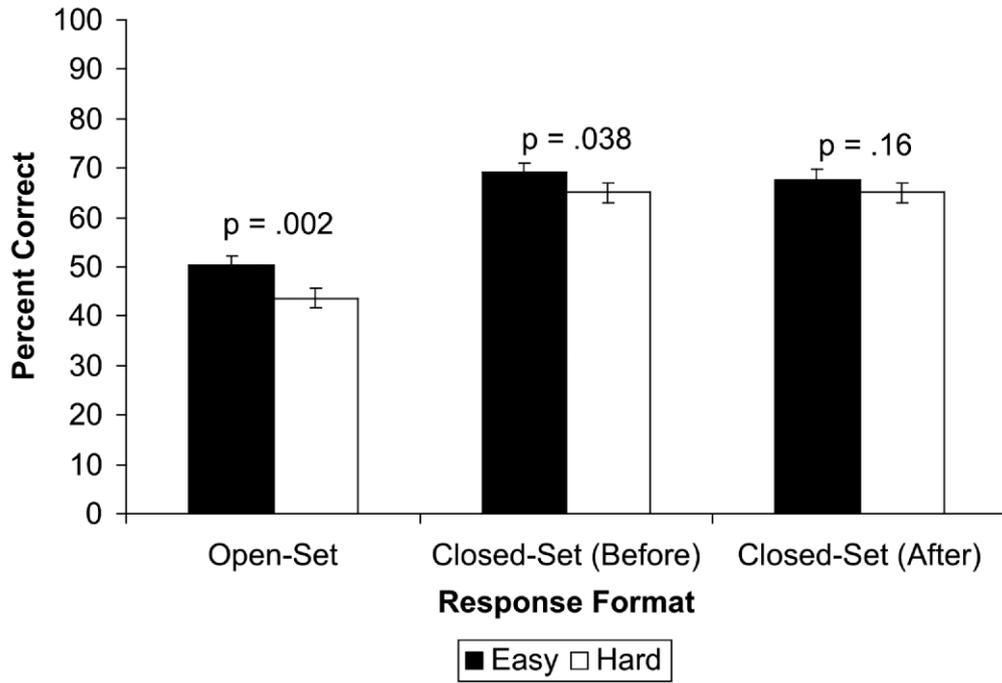


Figure 7. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for easy and hard words, collapsed across talker variability, for the listeners in the six-alternative high confusability experiment (Experiment 4).

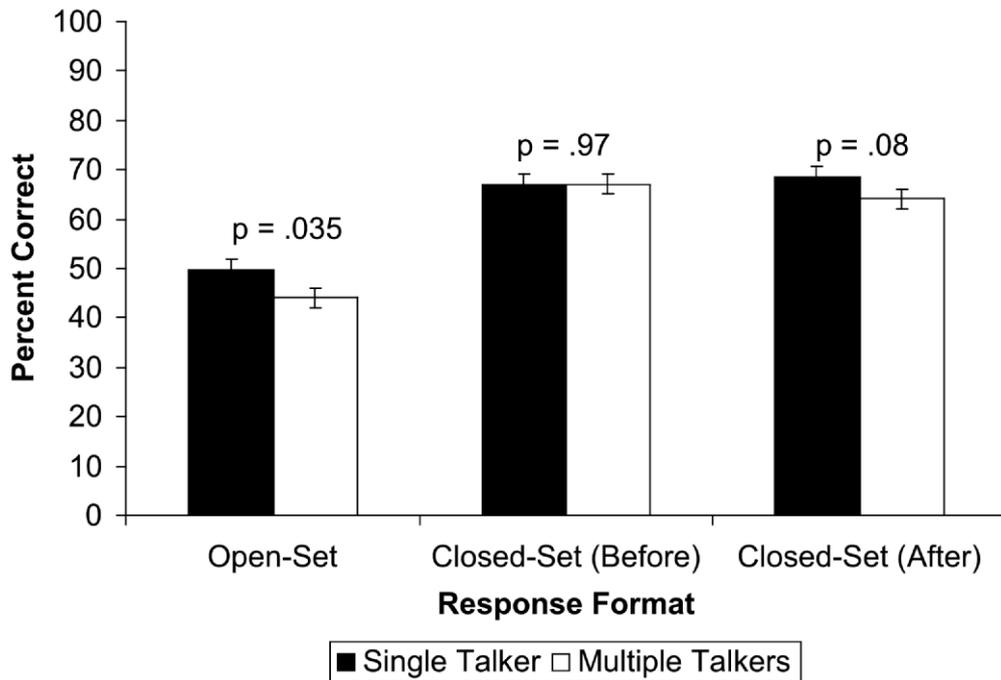


Figure 8. Proportion correct responses in the open-set, closed-set (before), and closed-set (after) response conditions for the single and multiple talker listener groups, collapsed across lexical competition, for the listeners in the six-alternative high confusability experiment (Experiment 4).

lexical competition ($F[1, 92] = 16.4, p < .001$). The main effect of talker variability was not significant. None of the interactions were significant.

Planned post hoc Tukey tests on response format revealed that performance was better in the closed-set (before) and closed-set (after) experimental blocks than in the open-set block (both $p < .001$). Performance did not differ between the closed-set (before) and closed-set (after) blocks.

As shown in Figure 7, planned post hoc paired sample t-tests on easy versus hard words within each response format condition revealed a significant main effect of lexical competition for the open-set block ($t[49] = 3.2, p = .002$) and the closed-set (before) block ($t[49] = 2.1, p = .038$). In both cases, performance on the "easy" words was better than performance on the "hard" words. The lexical competition effect was not significant for the closed-set (after) block.

Figure 8 shows that planned post hoc t-tests revealed a significant effect of talker variability in the open-set condition ($t[98] = 2.1, p = .035$) and a marginally significant effect of talker variability in the closed-set (after) condition ($t[98] = 1.8, p = .08$). In both cases, performance was better for the single talker group than the multiple talker group. The effect of talker variability was not significant in the closed-set (before) condition.

Discussion

We were once again successful in replicating the results from Experiments 1, 2, and 3 and obtained lexical competition and talker variability effects in the open-set task. We also found evidence of a lexical competition effect in the closed-set (before) task and a talker variability effect in the closed-set (after) task. It is not clear why these two results were not robustly present in both the closed-set (before) and closed-set (after) tasks. We can conclude, however, that greater confusability among the response alternatives had some effect on closed-set word-recognition strategies. Together with the results of Experiment 3, the present findings suggest that increased set size and phonetic confusability together produce a more difficult closed-set task in which lexical competition and talker variability effects are observed, although neither variable is

sufficient on its own under these conditions.

GENERAL DISCUSSION

The results of the present set of four word recognition experiments revealed robust effects of lexical competition and talker variability in open-set tests of spoken word recognition, replicating the earlier findings reported by Sommers et al (1997). Lexical competition effects were also obtained in the closed-set (before) and closed-set (after) conditions with 12 highly confusable response alternatives and in the closed-set (before) condition with six highly confusable response alternatives. Similarly, talker variability effects were obtained in both closed-set conditions in the highly confusable 12-alternative task and in the closed-set (after) condition in the highly confusable six-alternative task.

The pattern of findings observed across these four experiments suggests that the crucial differences between open-set and closed-set tasks lie in the nature of the specific task demands. In open-set tests, listeners must compare the stimulus item to all possible candidate words in lexical memory, whereas in closed-set tests, the listeners need to make only a limited number of comparisons among the response alternatives provided by the experimenter. By changing the number and nature of the response alternatives in the closed-set condition, we were able to more closely approximate the task demands operating in open-set spoken word recognition tasks. This increase in lexical competition between potential candidates reflects the combined operation of both lexical activation and competition as formalized by the Neighborhood Activation Model (Luce and Pisoni, 1998) and the increased competition between talker-specific episodic representations of lexical items in memory (Goldinger, 1996).

A comparison of closed-set word recognition performance across the experiments in the current study suggests that task difficulty was an important factor in controlling lexical competition and talker-variability effects. Performance in the open-set condition was always worse than performance in the closed-set conditions, replicating the consistent finding that open-set tasks are in general more difficult than

closed-set tasks. However, imposing a memory load on the listeners in the form of a one-second retention interval between stimulus presentation and response did not affect overall task difficulty. A significant difference in performance between the closed-set (before) and the closed-set (after) conditions was found only in Experiment 1. In addition, we failed to observe any evidence of an interaction between the delay and the other two experimental manipulations. However, both the number and phonetic confusability of the response alternatives did affect task difficulty. Closed-set performance was lowest overall in Experiment 2, in which 12 highly confusable response alternatives were presented to the listeners and performance improved when the number of response alternatives and/or the similarity of the response alternatives was reduced.

Overall task difficulty also appears to be related to lexical competition and talker-variability effects. We failed to obtain either of these effects in Experiments 1 and 3 where closed-set performance was quite accurate (82% and 92% correct, respectively), but the effects returned in Experiments 2 and 4 where overall closed-set performance was poorer (60% and 67% correct, respectively). Underlying differences between the listener populations across experiments can be ruled out because performance was stable and consistent across all of the open-set conditions (45%, 45%, 47%, and 47% correct, for Experiments 1 through 4, respectively).

The results of the current set of experiments cannot be attributed entirely to differences in overall task difficulty, however. We have collected some preliminary data using the same three response conditions in a six-alternative word-recognition task using stimulus materials with greater signal degradation. While overall performance in the closed-set conditions with these materials was approximately 60% correct, we did not obtain significant effects of lexical competition or talker variability in that task. Future research should examine in more detail the relationship between competition between response alternatives and task difficulty by manipulating other aspects of the task such as the familiarity of the words, the similarity of the talkers, the intelligibility of the talkers, the degree and type of signal degradation, and different kinds of cognitive load, all of which would be expected to increase task

difficulty, independent of competition in the lexical selection process. Results of these additional studies would help further our basic understanding of the relationship between information processing task demands and lexical access in speech discrimination tasks.

The overall pattern of results obtained in the current set of experiments suggests that as the word recognition task was made more difficult due to greater competition between response alternatives, the effects of lexical competition and talker variability emerged even in closed-set tasks. In particular, we found that a simple memory load involving a one-second response delay did not affect performance but that changes in the number and nature of the response alternatives in terms of their phonetic properties combined to significantly affect the listeners' processing strategies.

The results of this study have several clinical implications. First, clinicians do not need to abandon closed-set tests of word recognition in favor of open-set tests in order to obtain accurate measures of lexical competition in speech perception tasks from children and adult patients who may struggle with open-set tests. Instead, new tests could be developed that include a larger number of more highly confusable response alternatives that more accurately simulate real-world speech intelligibility, while still providing the benefits of the closed-set tests. Such new tests might provide better and more valid assessments of the spoken word recognition skills of hearing-impaired individuals in the laboratory and the clinic. Second, the present results suggest that task demands are an important property of all spoken word recognition tests and may help explain the difference in performance between open-set and closed-set tasks. The results obtained in the current study suggest that task demands related to lexical competition between response alternatives are particularly relevant in measuring speech discrimination for young normal-hearing listeners. Further research is needed to explore the role of the number and nature of the response alternatives in affecting processing strategies in adult and pediatric hearing-impaired populations.

CONCLUSIONS

Open-set tests of word recognition consistently reveal effects on performance that reflect the contribution of several important processing variables, including lexical competition and talker variability, but traditional closed-set tests do not. In the current set of experiments, we obtained both lexical competition and talker-variability effects in closed-set tasks by increasing the number and nature of the phonetic confusability among the response alternatives. This increase in competition between potential response alternatives more closely mimics open-set word recognition performance and suggests that one of the fundamental differences between open-set and traditional closed-set tests of spoken word recognition is not simply the level of chance performance but the specific information processing demands that each task imposes on lexical access in terms of activation and competition between phonetically similar words. The present findings provide additional support for current models of spoken word recognition that assume that spoken words are recognized relationally in the context of other phonetically similar words in lexical memory by processes of bottom-up acoustic-phonetic activation and top-down phonological and lexical competition between potential responses. These findings also suggest that the basic perceptual processes used to recognize spoken words in open-set and closed-set tests are not equivalent and any comparisons of performance across tests should be interpreted with some degree of caution in both normal-hearing and hearing-impaired populations.

Acknowledgments. This research was supported by NIH NIDCD T32 Training Grant DC00012 and NIH NIDCD R01 Research Grant DC00111 to Indiana University. We thank Elena Breiter for her assistance in data collection and the audience at the 145th Meeting of the Acoustical Society of America in Nashville, Tennessee, for their comments on an earlier version of this paper.

REFERENCES

- Bilger RC. (1984) Speech recognition test development. *Am Speech Hear Assoc Rep* 14:2–15.
- Black JW. (1957) Multiple-choice intelligibility tests. *J Speech Hear Disord* 22:213–235.
- Clopper CG, Pisoni DB. (2001) Effects of response format in spoken word recognition tests: speech intelligibility scores obtained from open-set, closed-set, and delayed response tasks. In: *Research on Spoken Language Processing Progress Report No. 25*. Bloomington, IN: Speech Research Laboratory, Indiana University, 254–265.
- Craik FIM, Kirsner K. (1974) The effect of speaker's voice on word recognition. *Q J Exp Psychol* 26:274–284.
- Creelman CD. (1957) Case of the unknown talker. *J Acoust Soc Am* 29:655.
- Dirks DD, Takayanagi S, Moshfegh A. (2001) Effects of lexical factors on word recognition among normal and hearing-impaired listeners. *J Am Acad Audiol* 12:233–244.
- Egan JP. (1948) Articulation testing methods. *Laryngoscope* 58:955–991.
- Fairbanks G. (1958) Test of phonemic differentiation: the rhyme test. *J Acoust Soc Am* 30:596–600.
- Foster JR, Haggard MP. (1987) The Four Alternative Auditory Feature test (FAAF)—linguistic and psychometric properties of the material with normative data in noise. *Br J Audiol* 21:165–174.
- Gelfand SA. (1998) Optimizing the reliability of speech recognition scores. *J Speech Lang Hear Res* 41:1088–1102.
- Gelfand SA. (2003) Tri-word presentations with phonemic scoring for practical high-reliability speech recognition assessment. *J Speech Lang Hear Res* 46:405–412.
- Goh W. (2005) Talker variability and recognition memory: instance-specific and voice-specific effects. *J Exp Psychol Learn Mem Cogn* 31:40–53.
- Goldinger SD. (1996) Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp Psychol Learn Mem Cogn* 22:1166–1183.
- Goldinger SD, Luce PA, Pisoni DB. (1989) Priming lexical neighbors of spoken words: effects of competition and inhibition. *J Mem Lang* 28:501–518.
- Grosjean F. (1980) Spoken word recognition and the gating paradigm. *Percept Psychophys* 28:267–283.
- Hirsh I, Davis H, Silverman SR, Reynolds E, Eldert E, Benson RW. (1952) Development of materials for speech audiometry. *J Speech Hear Disord* 17:321–337.
- Hood JD, Poole JP. (1980) Influence of the speaker and other factors affecting speech intelligibility. *Audiology* 19:434–455.
- Horii Y, House AS, Hughes GW. (1970) A masking noise with speech-envelope characteristics for studying intelligibility. *J Acoust Soc Am* 49:1849–1856.

- House AS, Williams CE, Hecker MHL, Kryter KD. (1965) Articulation-testing methods: consonantal differentiation with a closed-set response. *J Acoust Soc Am* 37:158–166.
- Kirk KI, Pisoni DB, Miyamoto RC. (1997) Effects of stimulus variability on speech perception in listeners with hearing impairment. *J Speech Lang Hear Res* 40:1395–1405.
- Kucera F, Francis W. (1967) *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Luce PA, Pisoni DB. (1998) Recognizing spoken words: the neighborhood activation model. *Ear Hear* 19:1–36.
- Marslen-Wilson WD. (1984) Function and process in spoken word recognition: a tutorial review. In: Bouma H, Bouwhuis DG, eds. *Attention and Performance X: Control of Language Processes*. Hillsdale, NJ: Erlbaum, 125–150.
- McLennan CT, Luce PA. (2005) Examining the time course of indexical specificity effects in spoken word recognition. *J Exp Psychol Learn Mem Cogn* 31:306–321.
- Miller GA. (1946) Articulation testing methods. In: *Transmission and Reception of Sounds under Combat Conditions*. Washington, DC: Division 7, National Defense Research Committee, 69–80.
- Mullennix JW, Pisoni DB, Martin CS. (1989) Some effects of talker variability on spoken word recognition. *J Acoust Soc Am* 85:365–378.
- Murphy GL. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
- Norris D. (1994) Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52:189–234.
- Nusbaum HC, Pisoni DB, Davis CK. (1984) Sizing up the Hoosier Mental Lexicon: measuring the familiarity of 20,000 words. In: *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Speech Research Laboratory, Indiana University, 357–376.
- Nyang EE, Rogers CL, Nishi K. (2003) Talker and accent variability effects on spoken word recognition. Poster presented at the 145th Meeting of the Acoustical Society of America, Nashville, TN.
- Nygaard LC, Sommers MS, Pisoni DB. (1994) Speech perception as a talker-contingent process. *Psychol Sci* 5:42–46.
- Owens E, Kessler DK, Telleen CC, Schubert ED. (1981) The Minimal Auditory Capabilities (MAC) battery. *Hear Aid J* (September):9.
- Palmeri TJ, Goldinger SD, Pisoni DB. (1993) Episodic encoding of voice attributes and recognition memory for spoken vowels. *J Exp Psychol Learn Mem Cogn* 19:309–328.
- Pisoni DB, Nusbaum HC, Luce PA, Slowiaczek LM. (1985) Speech perception, word recognition and the structure of the lexicon. *Speech Commun* 4:75–95.
- Pollack I, Rubenstein H, Decker L. (1959) Intelligibility of known and unknown message sets. *J Acoust Soc Am* 31:273–279.
- Schultz MC, Schubert ED. (1969) A multiple-choice discrimination test. *Laryngoscope* 79:382–399.
- Sommers MS, Kirk KI, Pisoni DB. (1997) Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners I: the effects of response format. *Ear Hear* 18:89–99.
- Sumbly WH, Pollack I. (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
- Takayanagi S, Dirks DD, Moshfegh A. (2002) Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *J Speech Lang Hear Res* 45:585–597.
- Treisman M. (1978) Space or lexicon? The word frequency effect and the error response frequency effect. *J Verbal Learn Verbal Behav* 17:37–59.
- Verbrugge RR, Strange W, Shankweiler DP, Edman TR. (1976) What information enables a listener to map a talker's vowel space? *J Acoust Soc Am* 60:198–212.
- Walden BE. (1984) Validity issues in speech recognition testing. *Am Speech Hear Assoc Rep* 14:16–18.