

Relationships Between Vocabulary Size and Spoken Word Recognition in Children Aged 3 to 7

Benjamin Munson

The Ohio State University, Columbus¹

S

Spoken word recognition is a multifaceted process whereby a listener associates an acoustic speech signal with a mental represen-

ABSTRACT: Spoken word recognition was investigated in a group of children aged 3:0 to 7:11 (years:months) to assess the relationship between five measures of language development and spoken word recognition accuracy. Two spoken word recognition tasks, gated words and noise-center words, were used. In the gating task, participants were asked to identify words whose final consonants had been removed. In the noise-center task, participants were asked to identify words whose medial vowel had been replaced by broadband noise. In both tasks, participants provided a nonverbal picture-pointing response. Five measures of language development were examined as possible predictors of spoken word recognition accuracy: expressive vocabulary, receptive vocabulary, pre-literacy skills, phonological awareness, and articulation accuracy, each of which was measured using a standardized, norm-referenced test. In the gating task, children required acoustic evidence of the final stop burst in consonant vowel consonant (CVC) words for accurate recognition. In the noise-center task, children required at least 40% of the vowel to be present in CVC words for accurate recognition. Of the five language measures, expressive vocabulary and receptive vocabulary were found to predict a significant proportion of variance in spoken word recognition scores. Results are discussed in terms of factors that influence spoken word recognition accuracy and the relationship between vocabulary size and linguistic development.

KEY WORDS: word recognition, vocabulary size, children

tation of a lexical item stored in long-term memory (Lively, Pisoni, & Goldinger, 1994). Some mechanisms involved in identifying a word include peripheral sensory mechanisms, which are used to build the immediate transient sensory representations of spectral and other acoustic information in the speech signal; working memory, which is used to hold information while the association is being made to the evoked representation of the lexical item in long-term memory; long-term phonological representations, against which immediate sensory representations are compared; and motor skills, which are used to establish aspects of the longer term representation relevant for reproducing the word as a response.

The development of spoken word recognition underlies the development of spoken language. Before acquiring a complex grammatical system, children must accurately perceive lexical items in the ambient language. Adult-like spoken word recognition emerges gradually in development, beginning in the pre-linguistic stage and continuing through adolescence. A large body of research suggests that infants and young children are able to perceive words and other linguistic structures well in advance of being able to produce them. Studies have shown that infants begin life with prodigious speech perception skills and are able to perceive many phonetic contrasts that do not appear in the ambient language (see Jusczyk, 1992 for a review). The attenuation of this ability appears to coincide with the onset of receptive vocabulary development and the beginning of variegated babbling at the end of the first year of life (Werker & Stager, 2000). Later refinement of spoken word recognition involves the improvement of accuracy

¹ Currently affiliated with the University of Minnesota, MN.

under difficult conditions, such as word recognition in noise or in multiple-talker settings (Ryalls & Pisoni, 1997).

The current study uses two stimulus types, noise-center words and gated words, to measure spoken word recognition in children. The *noise-center* technique (more often called the *silent-center* technique because the medial vowel is typically replaced with silence; Jenkins, Strange & Edelman, 1983) involves the identification of words whose medial vowel has been removed and replaced with broad-band noise. The *gating* technique (Grosjean, 1980) involves recognition of words whose final consonant has been removed in successive steps. The results of a number of previous studies (Elliot, Hammer, & Evan, 1987; Munson, Edwards, & Fox, 1996; Murphy, Shea, & Aslin, 1989; Walley, 1988) suggest that children require more acoustic information than adults in order to recognize gated and noise-center words accurately.

Walley (1988) used the gating paradigm to examine whether young children (mean age = 5:3 [years:months]) required more acoustic information to recognize a spoken word successfully than adults do. This was motivated by research suggesting that early vocabulary development might prompt a change from a holistic phonological representation to one in which sublexical units, such as phonemes, are represented explicitly. This more detailed, segmental phonological representation would support fast and accurate recognition on the basis of partial input (Walley, 1988, p. 138). That is, a child with a holistic phonological representation for the word *cat* (perhaps in the form of an acoustic exemplar of this word) would not be able to make a transient parse of the word *ca* (i.e., *cat* with a portion of the final consonant deleted or masked) in such a way that it matched the phonological representation of the word *cat* in long-term memory. Walley found that children required more acoustic information than adults to recognize words accurately. Similar results were found by Elliot, Hammer, and Evan (1987), who used only familiar monosyllabic words, and Munson, Edwards, and Fox (1996), who used a pointing, rather than a speaking, response.

Murphy, Shea, and Aslin (1989) examined the ability of 3-year-olds and adults to categorize auditorily presented words when the medial vowel had been either removed or replaced with silence. The words *bad* and *bud* were synthesized, and glottal pulses were removed to create stimuli with 0%, 10%, 35%, 65%, and 90% of the medial vowel replaced with silence. The adult participants all performed close to ceiling on each condition. Children's performance fell into one of two groups: those who performed like adults and those performing at chance, suggesting that discrimination accuracy was not linearly related to the amount of acoustic information present. In a second experiment presenting the 10% condition with white noise instead of silence, all participants performed close to ceiling. A third experiment comparing performance on the 10% silence versus the 10% noise condition elicited the dichotomous performance for both tasks. Age differences in the perception of noise-center and silent-center words were also found by Munson, Edwards, and Fox (1996), who used an identification task rather than a discrimination task.

The current study examines whether children's spoken word recognition accuracy can be predicted by language measures that are plausibly related to the amount of segmental detail in long-term phonological representations. Accurate recognition of a noise-center or gated word depends on the ability to access a lexical item given only partial information. Walley (1988) claimed that segmental phonological representations would promote fast and accurate word recognition, given that the long-term phonological representation acts as a mediator between the transient sensory representation and the long-term knowledge of the word's meaning. The search for skills predictive of spoken word recognition accuracy in the current study is motivated by looking at language skills that would influence the degree to which phonological representations are segmental.

Five measures are presented as plausibly being related to segmental detail in phonological representations. The first two of these are *expressive and receptive vocabulary size*. Many studies suggest that lexical development is central to many disparate aspects of language development (e.g., Bates & Goodman 1999). For example, researchers have suggested that children require a "critical mass" of lexical items to discover the grammatical structure of the adult language. Whereas most of these studies have focused on morphosyntactic development (i.e., Marchmann & Bates, 1994), Lindblom (1992) discussed the relationship between lexical development and the amount of segmental detail in phonological representations. As the size of a child's lexicon increases during development, more finely grained units must be used to represent lexical contrasts. In Lindblom's theory, the phonological units with which words are represented are a function of the number and shape of phonetically similar words in the child's mental lexicon. For example, a child with a small vocabulary might possess phonological representations organized by higher level units like syllables or whole words, whereas a child with a larger vocabulary might have phonological representations organized by lower level units like phonemes or individual articulatory gestures. This theory claims that increases in the size of the lexicon will change the segmental detail in phonological representations. Lindblom presented evidence from a computational simulation suggesting that a substantive relationship exists between vocabulary growth and detail in phonological representations.

A third measure is *phonemic awareness* (also called phonological awareness). This is the ability to make metalinguistic judgments about the internal phonological structure of spoken words (McBride-Chang, 1995). In English, this has been assessed with a variety of tasks, such as rhyming, position analysis (judging the position of a phoneme relative to the position of another phoneme), or phoneme deletion (uttering a nonsense word with a phoneme missing). McBride-Chang (1995) discussed four component skills of phonological awareness: perceiving a speech segment, holding a speech segment in working memory, operating on (manipulating, deleting, identifying, etc.) a speech segment, and communicating the results to an experimenter. Intelligence, verbal memory, and speech perception all contributed roughly equally to variance

among scores on three phonological awareness tasks: phoneme synthesis, whereby listeners concatenate strings of phonemes into words; phoneme deletion, whereby listeners are presented with nonsense words and told to utter them with a phoneme deleted; and position analysis, whereby listeners make judgments about the position of phonemes in nonsense words. Because phonological awareness deals with metalinguistic knowledge of sublexical detail in phonological representations, it is necessarily related to the amount of segmental detail in these representations.

A fourth potential predictive measure is *knowledge of orthography*, a skill associated with learning to read. English uses an orthographic alphabet, in which sounds and letters have a semi-regular correspondence. Once an orthographic representation for a word has been learned, it exists in parallel with the auditory-acoustic representation. English orthography is inherently segmental (as opposed to a logographic writing system, such as Japanese kanji). The written word *cat* clearly comprises *c*, *a*, and *t*. Early teaching and learning about print typically follows a “bottom-up” pattern, whereby children first learn the alphabet, then learn to combine letters into written words. Sound-letter correspondences are learned, and good novel-word reading is typically not achieved until this skill is made automatic. A child with parallel orthographic and acoustic representations, who also has good sound-letter correspondence skills, should have more segmental phonological representations. Again, this more segmental representation should lead to better performance on the word recognition task.

McBride-Chang (1999) made a distinction between the acquisition of letter-name and letter-sound knowledge: The former includes knowing that *s* is the letter [ɛs]; the latter involves knowing that *s* represents the [s] sound. Only letter-sound knowledge was found to predict later phonemic awareness. Although no study has found a correlation between reading scores and speech perception, one study (Manis et al., 1997) found that children with developmental dyslexia were less accurate than age peers in labeling synthetic and natural speech continua, supporting the possibility that orthographic awareness is linked with speech perception ability.

A final control parameter is *articulation skill*. Given Lindblom's hypotheses (1992) that the amount of detail in phonological representations is based on contrasts present in speech production, we would expect that poor articulation skill would be associated with poor auditory word recognition.

The purpose of the current study is to examine relationships between clinical measures of these five skills—expressive vocabulary, receptive vocabulary, phonemic awareness, knowledge of orthography, and articulation skill—and spoken word recognition accuracy. This question is examined in children aged 3–7, as this is an age range during which dramatic growth occurs in all of these language areas. A finding that all of these skills are related to spoken word recognition accuracy would support the claim that spoken word recognition accuracy is supported by segmental detail in phonological representations.

METHOD

Participants

Sixty-one children served as study participants. Children were recruited from local preschools and elementary schools. All participants passed a 25 dB pure-tone hearing screening at 500, 1000, 2000, and 4000 Hz. The mean age of children was 67.7 months (5:7.7), with a standard deviation of 15.5 months (1:3.5). The range was 3:0 to 7:11. Girls comprised 47.5% of the sample. Sixty of the children were European American, and one was biracial European American/African American. All children were native, monolingual speakers of American English. Informed consent was secured from the children's parents prior to testing.

Stimuli: Gating

For each word recognition task, two sets of four pictures were developed for the testing phase and one for the training phase (see Edwards, Fourakis, Beckman, & Fox, 1999 for further details). For the gating task, *Pete*, *peep*, *peak*, *P*, and *Kate*, *cape*, *cake*, and *K* were used for testing and *beet*, *beep*, *beak*, and *bee* were used for training. For the noise-center task, *cape*, *cap*, *keep*, and *cup*, and *bad*, *bead*, *bed*, and *bird* were used for testing and *big*, *beg*, *bag*, and *bug* were used for training. The pictures were arranged on an 8½ x 11 piece of paper divided into four equal size units (like the picture arrangement on the Peabody Picture Vocabulary Test—III [PPVT—III], Dunn & Dunn, 1997). The pictures were colored line drawings drawn by a local artist. All of the picture names should have been familiar to young children, based on Moe, Hopkins, and Rush's (1982) data on word frequency in the speech of first-grade children.

For the gating task, four gates were chosen, each of which was defined by an articulatory-acoustic landmark. The four gates chosen were (a) 40 milliseconds (ms) before the onset of final consonant closure, (b) at the onset of final consonant closure, (c) 10 ms after the release of the final consonant, and (d) the entire word. For the consonant vowel (CV) words *K*, *P*, and *bee*, the four conditions were (a) 60 ms before the offset of voicing for the vowel, (b) 40 ms before the offset of voicing for the vowel, (c) 20 ms before the offset of voicing for the vowel, and (d) the entire word.

An adult male speaker of American English was recorded saying each of the words. These were digitized at 20 kHz and low-pass filtered at 8.7 kHz. Each token was normalized for amplitude, such that the peak amplitude of each token was the same. Waveform editing software (Hyperception, Inc., 1991) was used to view the time-amplitude waveform. Two articulatory landmarks were defined for the CVC words: the offset of voicing for the medial vowel and the onset of the burst for the final consonant. The most extreme gate was created by deleting all information following 40 ms from the first landmark; the next gate was created by deleting all information following the first

landmark; the next gate was created by deleting all information 10 ms after the second landmark; and the final gate was created using the entire word. Custom software was used to ramp the last 5 ms of each stimulus item, using cosine-squared ramping. For the CV words, the offset of voicing was defined by viewing the time–amplitude waveform. Each gate was created by removing successive amounts of acoustic information from the waveform, then ramping the last 5 ms of each stimulus using custom software.

Stimuli: Noise-Center

For the noise-center task, four stimuli for each word were created, with (a) the middle 90% of the vowel replaced by noise, (b) the middle 60% of the vowel replaced by noise, (c) the middle 30% of the vowel replaced by noise, and (d) the entire word. The same speaker was used to record both the gating and the noise-center words. Each word was digitized at 10 kHz and low-pass filtered at 5 kHz. Each token was normalized for amplitude, such that the peak amplitude of each token was the same. Waveform editing software (Hyperception, Inc., 1991) was used to view the time-amplitude waveform. From the waveform, two acoustic landmarks were defined: the beginning point in the waveform that demonstrated excitation of higher harmonics characteristic of vowel production, and the offset of this acoustic pattern. The duration of the vowel was calculated. For the 90% removed token, the duration of 90% of the vowel was calculated. The waveform was manipulated such that two files were created: one with the initial consonant plus the first 5% of the vowel, and the other containing the last 5% of the vowel plus the final consonant. The last 5 ms of the former file was off-ramped using custom software and cosine-squared ramping. The first 5 ms of the latter file was on-ramped using custom software and cosine-squared ramping. A file of speech-shaped noise was created that was identical in length to the section of vowel removed. The three files were then concatenated in a new file. The amplitude of the new file was not altered, as this may have increased the intensity of the vowel transitions, affecting their salience. The tokens with 60% and 30% of the medial vowel removed were created in a similar manner.

Procedure

Children were tested in a quiet room at their school during the school day. Pilot testing revealed that young children were often unwilling to wear headphones for the period of time needed to complete the word recognition experiments. Thus, stimuli were played from a speaker attached to a 386 personal computer with an Ariel D-A/A-D sound board. Loudness of stimulus presentation was determined each morning by the experimenter, based on his or her judgment regarding the ambient noise in the room. Once this loudness level was chosen, it remained fixed for that day unless a significant change in ambient noise occurred. For both tasks, each set of four pictures was presented separately. For each set of pictures, there was a picture name instruction process, which had three components: first, each

picture was identified by name by the experimenter (e.g. “This is a bag. You carry things in a bag.”); second, the child was then asked to name all four pictures (if a child forgot a picture name, the picture name was provided and the child was asked to repeat it); and finally, the child was asked to identify all four pictures as the experimenter said the picture names. This picture name instruction process was followed for all three sets of pictures for both tasks.

Following the picture name instruction process for the first set of pictures, the child was then trained to identify stimuli in either the gating or the noise-center condition. The child first heard the picture names of all four pictures presented on the computer in the full word condition and was asked to identify the pictures. The child was then told that the computer would sometimes “talk funny” and that the child should continue to identify the pictures, but that he or she might have to guess sometimes if the computer talked “funny.” For the set of training pictures, 10 trials were presented, including four in the whole-word condition and two in each of the other gating/noise-center conditions.

Following the picture name instruction process for the experimental trials, each picture name was presented four times, one in each of the gating or noise-center conditions. The words were presented in random order, with the provision that a more gated condition always occurred before a less gated condition. Children’s pointing responses were recorded on an answer sheet. The gating and noise-center tasks were run on separate days, with approximately one-half of the children receiving the gating task first and the other one-half receiving the noise-center task first.

Standardized Tests

In addition to the word recognition tasks, children were administered a battery of standardized tests, all of which provide norm-referenced scores. Additionally, the Goldman-Fristoe Test of Articulation (GFTA, Goldman & Fristoe, 1986) provides normative data separately for boys and girls. In the PPVT-III (Dunn & Dunn, 1997), children are presented with a word by the examiner and must point to the picture in an array of four that represents that word. The PPVT-III was used as a measure of receptive vocabulary size. The Expressive Vocabulary Test (EVT, Williams, 1997) involves two tasks. Younger children (3:0–4:11) are required to name familiar pictures. Older children (5:0+) and younger children who pass the naming task are required to provide synonyms for words when they are given a picture cue. A single standard score is generated for this task, regardless of which task the child completes. The EVT was used as a measure of expressive vocabulary size. In the Test of Early Reading Ability-2 (TERA-2, Reid, Hresko, & Hammill, 1991), children provide verbal or pointing responses to a variety of questions regarding symbol–meaning correspondences, knowledge of print, and knowledge of orthography, given a picture or written word stimulus. Scores on this test were used as measures of pre-literacy skill. The TERA-2 was used as a measure of orthographic knowledge. The GFTA Sounds-in-Words subtest involves children naming familiar pictures. Children’s responses are recorded phonetically using

International Phonetic Alphabet transcription conventions. Scores are based on the number of phonemes produced correctly. Scores on the GFTA were used as measures of articulation accuracy. Finally, the Test of Phonological Awareness (TOPA, Torgeson & Bryant, 1994) was administered. The TOPA has two different versions. The kindergarten version, administered to all 5-year-olds, involves two subtests. In the first, the children judge which word from a list of three contained the same initial sound as a comparison word. In the second subtest, children pick the two words from a list of four that contain the same initial sound. Both subtests provide picture cues for each word, and the children respond by pointing. The early elementary version, administered to all children 6:0 and older, involves the same tasks as the kindergarten version, except final sounds are used instead of initial ones. Scores on the TOPA were used as measures of phonological awareness. The TOPA provides age-normalized data for children 5:0 and older only. However, this test was administered to all participants, and thus children younger than 5:0 received a raw score only. Despite this limitation, the TOPA was used for this study because it is one of the few standardized tests of phonological awareness for which any age-normalized scores are available.

Children at all sites were tested in a maximum of four sessions. Each session lasted approximately 30 minutes. The number of sessions per child was not controlled, as a child's availability for testing depended on school schedules.

RESULTS

Means and standard deviations were calculated for each gate, noise-center condition, and standardized test. Not all children were available for all sessions or were able to complete all tests for various reasons. Five children withdrew from the preschool during the course of the study. Some of the younger children were unable to or refused to complete the TOPA ($n = 7$) or the TERA-2 ($n = 3$). The number for each variable is reported separately.

The CV words (K , P) were not included in the analysis for the gating task, as they had been included as foils.

Table 1 summarizes scores for standardized tests; Table 2 summarizes scores for the word recognition measures. For each of the word recognition tasks, the binomial probability formula was used to calculate the probability of obtaining chance results (Hays, 1994). In the gating task, six words were presented at each gate (omitting K and P); in the noise-center task, eight words were presented in each condition. Probability was .25 for obtaining a correct response at chance in both tasks. Based on the binomial probability formula, a participant needed to respond to greater than 3/6 (50%) of the items at each gate and to greater than 4/8 (50%) of the items at each noise-center condition in order to perform above chance at the $\alpha = .05$ confidence level. The percentage of participants performing at chance levels for each test is also reported in Table 2. Mean percentage correct at each gate and noise-center condition is plotted in Figures 1 and 2.

Analyses of Covariance

Analysis of covariance (ANCOVA) was used to assess whether each gate/noise-center condition significantly affected the mean percentage correct. Arcsine transformations were completed on all percentage correct scores prior to analysis. A one-way repeated measures ANCOVA, with either gate or noise-center condition as the repeated measure and age as the covariate, was calculated. For the gating task, a significant main effect of gate ($F [3,177] = 4.688$, $p = .004$) was found, but not of age ($F [1,59] = 3.291$, $p = .075$), nor was there a significant interaction between gate and age ($F [3,177] = 2.413$, $p = .072$). These results indicate that the amount of acoustic information has an effect on mean percentage correct, but that age does not. The lack of an interaction suggests that the effect of gate on percentage correct does not differ as a function of age. Post-hoc Tukey tests indicate that the main effect of gate is due to significant differences between the following pairs of gates (critical difference = 7.57%): 40 ms before closure

Table 1. Number of participants, type of data, mean, and standard deviation for each standardized test administered.

<i>Test</i>	<i>N</i>	<i>Type of data</i>	<i>Mean</i>	<i>Standard deviation</i>
PPVT-III	56	Standard score; $M = 100$, $SD = 15$	113.8	12.5
	56	Raw score	95.2	22.9
EVT	56	Standard score; $M = 100$, $SD = 15$	108.9	11.7
	56	Raw score	66.4	14.6
TOPA	42	Standard score; $M = 100$, $SD = 15$	91.5	12.3
	54	Raw score (maximum of 20)	10.4	5.3
TERA-2	58	Standard score; $M = 100$, $SD = 15$	106.3	10.7
GFTA	61	Percentile ranking	60.9	33.3

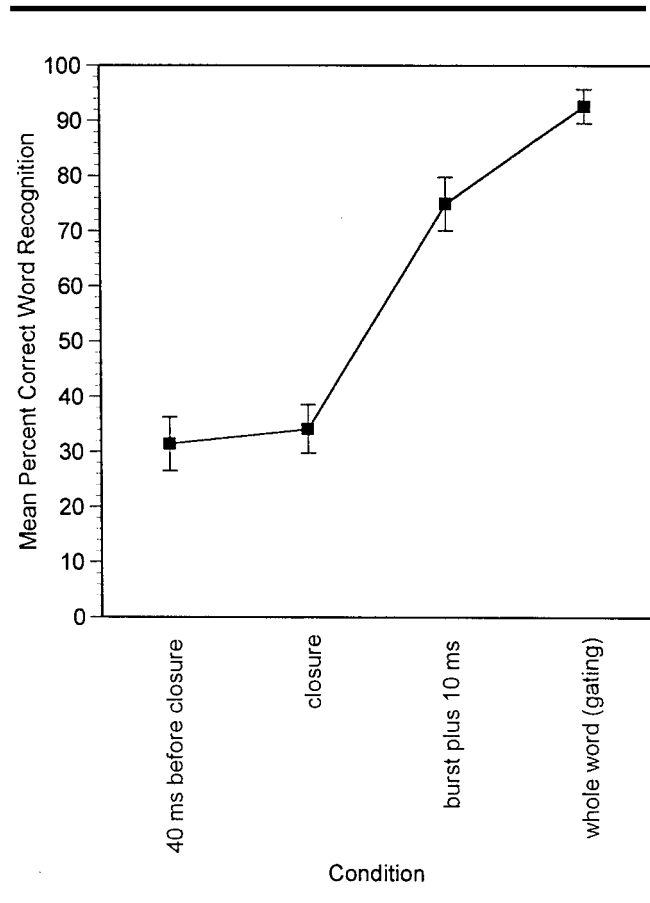
Note. PPVT-III = Peabody Picture Vocabulary Test-III, EVT = Expressive Vocabulary Test; TOPA = Test of Phonological Awareness, TERA-2 = Test of Early Reading Ability-2, and GFTA = Goldman-Fristoe Test of Articulation.

Table 2. Mean, standard deviation, and chance performance for the eight word recognition conditions ($N = 61$).

Test	Mean	Standard deviation	Chance performance ^a
40 ms before closure	31.4%	19	96.7%
Closure	34.1%	17	91.8%
Burst plus 10 ms	74.9%	19	14.8%
Whole word (gating)	92.6%	12	1.6%
90% removed	38.1%	13	93.4%
60% removed	85.0%	10	4.9%
30% removed	93.0%	10	0%
Whole word (noise center)	97.5%	6	0%

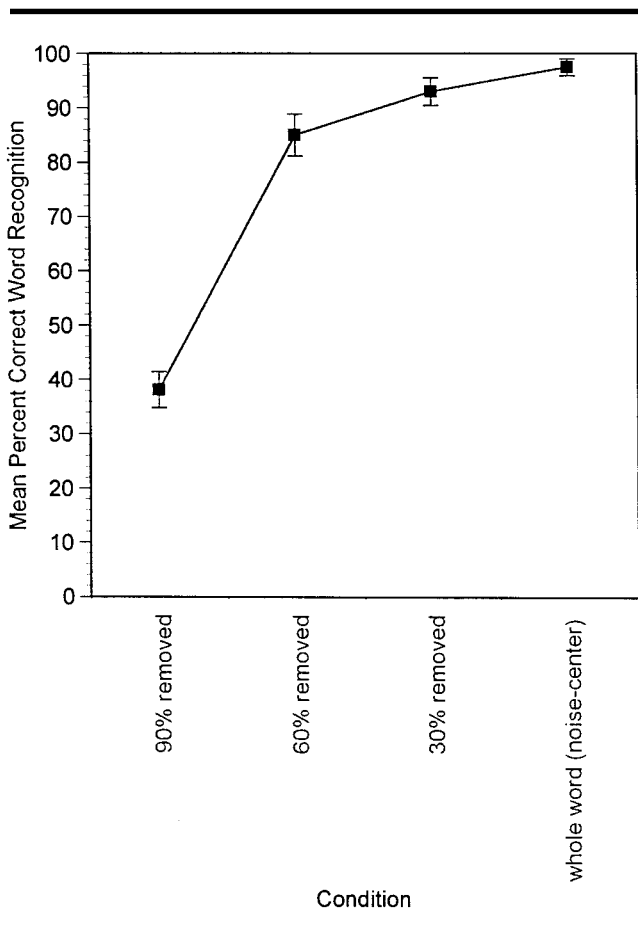
^a Percentage of participants performing at chance based on $\alpha = .05$.

Figure 1. Mean percent correct for all participants ($N = 61$) on the gating task. Error bars represent 95% confidence intervals of the mean.



and burst plus 10 ms; 40 ms before closure and whole word; closure and burst plus 10 ms; closure and whole word; and burst plus 10 ms and whole word. The majority of participants performed at chance levels on the 40 ms before closure and closure conditions, whereas the majority of participants performed above chance on the burst plus 10 ms and whole word conditions.

Figure 2. Mean percent correct for all participants ($N = 61$) on the noise-center task. Error bars represent 95% confidence intervals of the mean.



For the noise-center task, there were significant main effects for noise-center condition ($F [3,177] = 19.625, p < .001$) and age ($F [1,59] = 21.672, p < .001$), and there was a significant interaction between noise-center condition and age ($F [3,177] = 5.416, p = .004$). These results indicate that both noise-center condition and age are related to the percentage correct word recognition. In addition, the interaction indicates that the effect of noise-center condition on percentage correct differs systematically as a function of age. Post-hoc Tukey tests showed that the significant effect of noise-center is due to differences between the following pairs (critical difference = 4.54%): 90% removed and 60% removed; 90% removed and 30% removed; 90% removed and whole word; 60% removed and 30% removed; and 60% removed and whole word. The difference between the means of the 30% removed condition and the whole word condition falls just short of the minimum distance required at the .05 level. The interaction between age and noise-center condition was explored with multiple regression (see discussion below). The majority of participants performed at chance on the 90% removed condition only; most participants performed above chance on the other three conditions.

Multiple Regression

To assess the influence of each of the independent variables on word recognition, multiple regressions were calculated using arcsine transformed percentage correct scores at each of the gates and noise-center conditions as the dependent variable, and all of the standardized test scores as independent variables. For both the EVT and the PPVT-III, both raw and norm-referenced scores were entered into the equation in order to assess whether raw vocabulary size (of which the EVT and PPVT-III raw scores are an estimate), rather than age-normalized vocabulary skill, influenced spoken word recognition. All variables were entered into the regression equation stepwise. Each gate and noise-center condition was analyzed separately, rather than as a single collapsed score, for two reasons. First, because the gates and noise-center conditions did not represent equal steps, it would not be appropriate to calculate an identification slope for each participant (one frequent way of collapsing identification data). It would also be inappropriate to average scores for participants across conditions, as this would neutralize the difference between a participant with uniformly poor identification (e.g., 50% on all four conditions) and a participant with more variable identification (e.g., 0% correct at the two most extremely gated conditions and 100% at the other two). In addition, regressions were not completed for conditions in which the majority of participants were performing at chance levels (40 ms before closure and closure on the gating task, and 90% of the medial vowel removed on the noise-center task).

Results of these regressions are presented in Table 3. For three of the five conditions, raw measures of vocabulary size predicted spoken word recognition accuracy, and these were better predictors than age. For the burst plus 10 ms

condition, PPVT-III raw score accounted for 12.8% of the variance; for the whole word condition of the gating task, PPVT-III raw score accounted for 21.6% of the variance; and for the whole word condition of the noise-center task, EVT raw score accounted for 21.8% of the variance. Age accounted for a significant proportion of variance on the 60% removed and 30% removed conditions. One possible interpretation of these results is that measures of vocabulary size predict spoken word recognition indirectly, because both measures are predicted by age.

To evaluate this claim, a procedure suggested by Bates and Goodman (1999) was used, in which two sets of multiple regressions were compared, one in which age is entered into the regression equation before measures of vocabulary size, and one in which it is entered after measures of vocabulary size. Results of regression analyses forcing age as the first variable can be found in Table 4. Results show that age predicted a significant proportion of variance in word recognition scores for all five conditions in which the majority of participants performed at greater than chance levels. Variance accounted for ranged from 10.3% to 25.1%. Raw scores on the PPVT-III accounted for an additional 8.4% of the variance in the whole word condition of the gating task, and EVT raw scores accounted for an additional 8.7% of the variance in the whole word condition of the noise-center task, even after the effect of age was partialled out of the equation. Table 5 shows results for the regression analyses forcing age as the second variable. Here, regressions were only calculated for the three conditions in which vocabulary accounted for the most variance in a fully stepwise regression. In all three cases, age failed to account for a significant proportion of the variance in word recognition beyond that accounted for by measures of vocabulary size. Thus, for two spoken word

Table 3. Multiple regression analyses for variables burst plus 10 ms, whole word (gating), 60% removed, 30% removed, and whole word (noise center), with all variables entered stepwise.

<i>Dependent variable</i>	<i>Independent variable</i>	<i>Variance accounted for (R²)</i>	<i>Significance</i>
Burst plus 10 ms	PPVT-III, raw score	.128	$F[1,53] = 7.805, p = .007$
Whole word (gating)	PPVT-III, raw score	.216	$F[1,54] = 14.624, p < .001$
60% removed	Age	.171	$F[1,53] = 10.926, p = .002$
30% removed	Age	.251	$F[1,53] = 17.791, p < .001$
Whole word (noise center)	EVT, raw score	.218	$F[1,53] = 14.752, p < .001$

Table 4. Multiple regression analyses for variables burst plus 10 ms, whole word (gating), 60% removed, 30% removed, and whole word (noise center), with age forced as the first variable.

<i>Dependent variable</i>	<i>Independent variable</i>	<i>Variance accounted for (R²)</i>	<i>Significance</i>
Burst plus 10 ms	Age	.103	$F[1,53] = 6.062, p = .017$
Whole word (gating)	Age	.147	$F[1,54] = 9.287, p = .004$
	PPVT-III raw score	.085	$F[1,53] = 5.832, p = .019$
60% removed	Age	.171	$F[1,53] = 10.926, p = .002$
30% removed	Age	.251	$F[1,53] = 17.791, p < .001$
Whole word (noise center)	Age	.162	$F[1,54] = 10.406, p = .002$
	EVT, raw score	.097	$F[1,53] = 6.950, p = .011$

Table 5. Multiple regression analyses for variables burst plus 10 ms, whole word (gating), and whole word (noise center), with age forced as the second variable.

<i>Dependent variable</i>	<i>Independent variable</i>	<i>Variance accounted for (R²)</i>	<i>Significance</i>
Burst plus 10 ms	PPVT-III, raw score	.129	$F [1,54] = 7.979, p = .007$
	Age	.007	$F [1,53] = .399, p = .53$
Whole word (gating)	PPVT-III, raw score	.231	$F [1,54] = 16.200, p < .001$
	Age	.001	$F [1,53] = .038, p < .001$
Whole word (noise center)	EVT, raw score	.256	$F [1,54] = 18.629, p < .001$
	Age	.002	$F [1,53] = .162, p = .689$

recognition conditions (the whole word conditions of the gating and noise-center tasks), measures of vocabulary size were unambiguously better predictors of accuracy than age. Multiple regression analyses were used to examine the interaction between age and vocabulary size for these two conditions in order to assess whether the effect of vocabulary size on spoken word recognition was the same for the entire age range studied. Following Cohen and Cohen (1983), the interaction was studied by calculating a hierarchical multiple regression with three independent variables: age, PPVT-III/EVT raw scores, and their product. Age and PPVT-III/EVT raw score were forced into the equation before their product. A significant proportion of variance accounted for by the product would indicate a significant interaction. These results are presented in Table 6. No significant interactions were found, suggesting that the influence of vocabulary size on spoken word recognition accuracy is stable between 3 and 7 years of age.

DISCUSSION

Of the five measures hypothesized to influence spoken word recognition accuracy, only raw measures of vocabulary size were found to predict a significant proportion of variance in word recognition scores. This finding is consistent with theories in which lexical development is the primary driving force in language development (Bates & Goodman, 1999). For example, expressive vocabulary size has been shown to predict many aspects of language development, such as the onset of overregularization errors in the production of the past tense morpheme (Marchmann & Bates, 1994).

Two reasons are advanced to explain the absence of a predictive relationship between spoken word recognition measures and the other three language measures (GFTA, TOPA, and TERA-2). First, the measures of spoken word recognition accuracy were not optimal. Few words were presented in each gating/noise-center condition (6 and 8, respectively), and this prevented a normal distribution of scores. Additionally, ceiling performance was noted on many conditions, leaving only a small portion of variance to be accounted for by the independent variables. More tokens would have been needed to elicit a greater range of above-chance scores. A second possibility is that more vowel types were needed in the stimulus words, as were used in Experiment 1 in Munson, Edwards, and Fox (1996). Additionally, only one token of each word was presented at each gate, yet the analysis assumed that the stimuli were a fixed effect (Clark, 1973). Given the pervasive variability in speech production, this assumption is likely not true.

Problems may have existed in the standardized test measures as well. Children performed almost uniformly poorly on the measure of phonological awareness, the TOPA, with a mean score close to one standard deviation below the mean for the normative sample. Given that the group of children in this study scored on average at or close to one standard deviation above the mean on scores of expressive and receptive vocabulary, this result suggests that the children tested in this study differed considerably from the group used to calculate the age-normalized scores for the TOPA. The TERA-2 tested symbol-meaning relationships and knowledge of print in addition to knowledge of orthography. Only knowledge of orthography is predicted to influence the degree of segmental detail in a phonological representation. It is also not clear whether

Table 6. Hierarchical multiple regression analyses for variables whole word (gating) and whole word (noise center).

<i>Dependent variable</i>	<i>Independent variable</i>	<i>Variance accounted for (R²)</i>	<i>Significance</i>
Whole word (gating)	Age	.147	$F [1,54] = 9.287, p = .004$
	PPVT-III raw score	.085	$F [1,53] = 5.832, p = .019$
	Interaction	.001	$F [1,52] = .055, p = .816$
Whole word (noise center)	Age	.162	$F [1,54] = 10.406, p = .002$
	EVT, raw score	.097	$F [1,53] = 6.950, p = .011$
	Interaction	.047	$F [1,52] = 3.536, p = .066$

standardized tests such as the PPVT-III and the EVT are optimal measures of vocabulary size. Previous studies finding relationships between vocabulary size and grammatical development (i.e., studies reported in Bates & Goodman, 1999) used a parent report measure in which parents reported all of the words that their much younger children produced and comprehended. In these studies, moderate-to-strong relationships between vocabulary growth and grammatical development have been found. Neither of the tests used in this study were as comprehensive. However, given the obvious practical difficulty associated with measuring the size of older children's lexicons, the EVT and PPVT-III are probably the best single measures of vocabulary size currently available for this age group.

Choice of independent variables might also explain the small effect sizes that were noted. In the introduction, it was stated that working memory is related to spoken word recognition, yet no measure of working memory was made. Working memory has been shown to be predictive of lexical development, in that non-word repetition at age 5 predicted vocabulary size at age 6 (Gathercole & Baddeley, 1989). Peripheral sensory mechanisms were also advanced as one component involved in spoken word recognition. Some researchers (Morongello, Kulig, & Clifton, 1984) have found developmental relationships in children's ability to sequence non-speech auditory events, with younger children requiring a greater difference between sounds than older children to sequence them in time reliably. Neither of these potential control parameters was tested, under the assumption that the control parameters influencing word recognition would be those related to the amount of segmental detail in a phonological representation.

Despite these limitations, the current experiment suggests that vocabulary size influences the development of accurate spoken word recognition. This finding contributes to the growing body of literature suggesting that vocabulary size is a driving force in language development and adult grammatical organization (Bates & Goodman, 1999). Future research should focus on the influence of all four skills hypothesized to underlie spoken word recognition: peripheral sensation, working memory, phonological representations, and motor skills, and should further examine the predictive role of vocabulary growth in the development of spoken word recognition.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) training grant 5 T32 DC0051 to Robert A. Fox and NIH grant 5 R01 DC02932 to Jan Edwards. The comments of Jan Edwards, Robert A. Fox, Jan Weisenberger, Mary Beckman, and two anonymous reviewers are gratefully acknowledged, as is the help of Lindsay Schaadt and Shurita Thomas-Tate in data collection.

REFERENCES

- Bates, E., & Goodman, J. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language* (pp. 29–79). Mahwah, NJ: Lawrence Erlbaum Associates.
- Clark, H. (1973). The language-as-fixed fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 2, 335–359.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dunn, L., & Dunn, L. (1997). *The Peabody Picture Vocabulary Test-III*. Circle Pines, MN: American Guidance Services.
- Edwards, J., Fourakis, M., Beckman, M., & Fox, R. (1999). Characterizing knowledge deficits in phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42, 169–186.
- Elliot, L., Hammer, M., & Evan, K. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception and Psychophysics*, 42, 150–157.
- Gathercole, S., & Baddeley, A. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200–213.
- Goldman, R., & Fristoe, M. (1986). *The Goldman-Fristoe Test of Articulation*. Austin, TX: PRO-ED.
- Grosjean, F. (1980). Spoken word recognition and the gating paradigm. *Perception and Psychophysics*, 28, 267–283.
- Hays, W. (1994). *Statistics*. New York: Harcourt Brace College.
- Hyperception, Inc. (1991). *Hypersignal DSP Software Users' Manual*. Dallas, TX: Author.
- Jenkins, J., Strange, W., & Edelman, T. (1983). Identification of vowels in "vowelless" syllables. *Perception and Psychophysics*, 34, 441–450.
- Jusczyk, P. (1992). Developing phonological categories from the speech signal. In L. Menn, C. Ferguson, & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 17–63). Timonium, MD: York Press.
- Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In L. Menn, C. Ferguson, & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 131–163). Timonium, MD: York Press.
- Lively, S. E., Pisoni, D. B., & Goldering, S. D. (1994). Spoken word recognition. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Peterson, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology*, 66, 211–235.
- Marchmann, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21, 339–366.
- McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, 87, 179–192.
- McBride-Chang, C. (1999). The ABCs of the ABCs: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly*, 45, 285–308.
- Moe, A., Hopkins, C., & Rush, R. (1982). *A vocabulary of first-grade children*. Springfield, IL: Thomas.
- Morongello, B. A., Kulig, J. W., & Clifton, R. K. (1984). Developmental changes in auditory temporal perception. *Child Development*, 55, 461–471.

- Munson, B., Edwards, J., & Fox, R. A.** (1996). Preschool children's use of dynamic information in word recognition [Abstract]. *Journal of the Acoustical Society of America*, *99*, 2591.
- Murphy, W., Shea, S., & Aslin, R.** (1989). Identification of vowels in "vowelless" syllables by 3-year-olds. *Perception and Psychophysics*, *46*, 375–383.
- Reid, D. K., Hresko, N. D., & Hammill, D. D.** (1991). *The Test of Early Reading Ability-2*. Austin, TX: PRO-ED.
- Ryalls, B., & Pisoni, D.** (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*, 441–452.
- Torgeson, J. K., & Bryant, B. B.** (1994). *Test of Phonological Awareness*. Austin, TX: PRO-ED.
- Walley, A. C.** (1988). Spoken word recognition by young children and adults. *Cognitive Development*, *3*, 137–165.
- Werker, J., & Stager, C.** (2000). Developmental changes in infant speech perception and early word learning: Is there a link? In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology 5* (pp. 181–193). Cambridge, MA: Cambridge University Press.
- Williams, K.** (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Services.

Contact author: Benjamin Munson, Department of Communication Disorders, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive, S.E., Minneapolis, MN 55455. Email: Munson.11@osu.edu