



Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych



Individual differences in online spoken word recognition: Implications for SLI

Bob McMurray^{a,*}, Vicki M. Samelson^b, Sung Hee Lee^c, J. Bruce Tomblin^d

^a Dept. of Psychology and the Delta Center, University of Iowa, United States

^b Dept. of Communication Sciences and Disorders, University of Wisconsin, Eau Claire, United States

^c Dept. of Special Education, University of Washington, United States

^d Dept. of Communication Sciences and Disorders and the Delta Center, University of Iowa, United States

ARTICLE INFO

Article history:

Accepted 16 June 2009

Keywords:

Spoken word recognition
Individual differences
Connectionist models
Visual world paradigm
Specific language impairment
Lexical decay

ABSTRACT

Thirty years of research has uncovered the broad principles that characterize spoken word processing across listeners. However, there have been few systematic investigations of individual differences. Such an investigation could help refine models of word recognition by indicating which processing parameters are likely to vary, and could also have important implications for work on language impairment. The present study begins to fill this gap by relating individual differences in overall language ability to variation in online word recognition processes. Using the visual world paradigm, we evaluated online spoken word recognition in adolescents who varied in both basic language abilities and non-verbal cognitive abilities. Eye movements to target, cohort and rhyme objects were monitored during spoken word recognition, as an index of lexical activation. Adolescents with poor language skills showed fewer looks to the target and more fixations to the cohort and rhyme competitors. These results were compared to a number of variants of the **TRACE model** (McClelland & Elman, 1986) that were constructed to test a range of theoretical approaches to language impairment: impairments at sensory and phonological levels; vocabulary size, and generalized slowing. None of the existing approaches were strongly supported, and variation in lexical decay offered the best fit. Thus, basic word recognition processes like lexical decay may offer a new way to characterize processing differences in language impairment.

© 2009 Elsevier Inc. All rights reserved.

* Corresponding author. Address: E11 SSH, Dept. of Psychology, University of Iowa, Iowa City, IA 52242, United States. Fax: +1 319 335 0191.

E-mail address: bob-mcmurray@uiowa.edu (B. McMurray).

1. Introduction

The incoming speech signal is variable and noisy, arrives at a high rate of input, and maps onto a potentially vast number of lexical candidates. The question of how listeners recognize words given these challenges represents an important problem in the language sciences. Research over the last 30 years has led to a remarkable consensus for several principles that broadly characterize the processing architecture underlying spoken word recognition. These include immediate incremental processing, graded activation, parallelism, and competition.

These core principles provide descriptions of average performance in a variety of word recognition tasks and represent fundamental commonalities across listeners. However, there has been little work addressing differences between listeners. Such research has the promise for refining current models of word recognition by revealing which aspects vary freely, which are more constant, and which may be important for language processes beyond word recognition. It may also help diagnose and treat listeners at the low end of the language ability scale, listeners commonly characterized as specific- or non-specific-language-impaired (SLI or NLI).

The present paper begins to address this by using fine-grained measures of the temporal dynamics of word recognition, and testing hypotheses for underlying causes of these differences using a current model of word recognition, TRACE (McClelland & Elman, 1986). We start by reviewing the research that established the current consensus on word recognition. We then motivate the individual differences that we chose to study here: gradations in overall language ability that are commonly associated with language impairment. Next, we present an experiment that uses the visual world paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to assess online word recognition. Finally, we use variants of the TRACE model (McClelland & Elman, 1986) to test hypotheses about the underlying processing dimension(s) that may account for the individual differences we observed.

1.1. Principles of spoken word recognition

There is considerable consensus for several core principles that characterize the processes of real-time spoken word recognition: (1) words are activated *immediately* upon the receipt of the smallest amount of perceptual input; (2) activation is updated *incrementally* as the input unfolds; (3) activation is *graded*; (4) multiple words are activated in *parallel*; and (5) these words actively *compete* during recognition.

Immediacy was initially revealed by gating paradigms (Grosjean, 1980; Tyler, 1984) and later by priming and eye-movement measures (Allopenna et al., 1998; Marslen-Wilson & Zwitserlood, 1989; McMurray, Clayards, Tanenhaus, & Aslin, 2008; Zwitserlood, 1989). Given a minimal amount of information at word onset, listeners activate the set of all words compatible with this partial information. These words are maintained in *parallel*, until they can be ruled out by additional acoustic material as it accumulates *incrementally* (Dahan & Gaskell, 2007; Frauenfelder, Scholten, & Content, 2001; Marslen-Wilson, 1987). Early work focused on the kinds of words that are considered during processing, showing specifically that onset-competitors (or cohorts, such as *beetle* when hearing *beaker*) receive significant activation early, but that offset-competitors (rhymes, e.g. *speaker*) are also active (Allopenna et al., 1998; Connine, Blasko, & Titone, 1993; Marslen-Wilson, Moss, & Van Halen, 1996).

Lexical activation is clearly a *graded* phenomenon: frequency affects activation (Dahan, Magnuson, & Tanenhaus, 2001; Marslen-Wilson, 1987), as does phonetic match (Marslen-Wilson et al., 1996). Words that match at onset (but mismatch later in the word) receive more activation than those that mismatch at onset but match later (e.g. rhymes) (Allopenna et al., 1998). Moreover, activation is a function of the lexicon as a whole (not just a word's match to the input). The number of similar words (to the target word) affects recognition (Luce & Pisoni, 1998) and specific competitor words can delay activation to a target (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994). Thus, there is an active process by which lexical items *compete* with other items. This is underscored by the fact that well after a word's uniqueness point (the point where there is sufficient

information to unambiguously identify the word), there are still changes in lexical activation dynamics (Dahan & Gaskell, 2007), and activation for competitors (Luce & Cluff, 1998).

Given these principles, the field has developed several computational models that characterize word recognition as arising out of the properties of simple connectionist units. Interactive activation accounts (e.g. TRACE: McClelland & Elman, 1986; PARSYN: Luce, Goldinger, Auer, & Vitevitch, 2000; see also Shortlist: Norris, 1994) posit that words are recognized via a hierarchical dynamical system. In the most detailed account, TRACE, the input is represented as a series of units corresponding to acoustic features like voicing. These units pass activation from features to phonemes to words along a set of variable-strength connections (weights). Units compete with each other within layers (e.g. active phonemes inhibit less active ones) and (in the case of TRACE) can receive activation from higher level units as well (e.g. lexical activation feeds back to phonemes). Activation cycles through the model until the model settles on an interpretation of the input, usually a single active word.

There are important differences between interactive architectures that should not be ignored. These include issues like feedback (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland, Mirman, & Holt, 2006; Norris, McQueen, & Cutler, 2000); whether competition effects derive from bottom-up or lateral inhibition (Frauenfelder et al., 2001; McClelland & Elman, 1986; Norris, 1994); and the nature of sublexical representations and processes (Gaskell, Quinlan, Tamminen, & Cleland, 2008; McMurray, Tanenhaus, & Aslin, 2009; McQueen, Cutler, & Norris, 2006). Finally, non-interactive distributed models (e.g. Gaskell & Marslen-Wilson, 1997; Magnuson, Tanenhaus, et al., 2003) and Bayesian approaches (e.g. Norris & McQueen, 2008) offer completely different formulations of the basic problem. However, despite these debates, all of the models implement the idea that candidate words are identified immediately, considered in parallel, and compete in some way.

One particularly useful property of interactive models is the fact that these models include free parameters which control the dynamics of the system (though this can create interpretation issues: see Pitt, Kim, Navarro, & Myung, 2006). For example, in TRACE one can vary the strength of lateral inhibitory connections between words (competition), the strength of excitatory connections between phonemes and words (bottom-up activation flow), or the ability of feature units to maintain activation. Thus, this framing of spoken word recognition processes provides the opportunity to model individual differences as differences in these parameters.

1.2. Individual differences

There is little evidence suggesting that individuals vary in their gross adherence to the principles of word recognition (e.g. there is no evidence that some listeners process words in serial and others in parallel). However, within this framework there is room for individual differences in the rate at which words accumulate activation, the strength of competition effects, and other factors. But why should we care about individual differences?

First, characterizing which dimensions of word recognition vary may constrain models or help in model selection. For example, in distributed models like Gaskell and Marslen-Wilson (1997) competition emerges implicitly as the overlap between representations (in contrast to TRACE which includes explicit inhibitory connections). If variation in inhibition is not attested between individuals, this might favor the distributed architecture, which can better account for such constancy. On the other hand, if inhibition does vary, this may force us to consider how this could be implemented in a distributed representation. Similarly, Bayesian models (Norris & McQueen, 2008) have fewer free parameters (since optimality constrains the computation). Robust individual differences of a particular form, then, could challenge the few degrees of freedom in such models, or support them (if they are sufficient to capture it).

Second, identifying the components of word recognition that vary between individuals may alert us to new questions. Work in word recognition has focused on theoretically important components like feedback, inhibition and activation processes. However, if variation in such processes is not related to individual differences in language ability, it may be important to examine other components of word recognition.

Finally, an understanding the full range of individual differences could have applications in the diagnosis and treatment of language disorders. The processing parameters that vary between

individuals could be similar to those that underlie disorders (though at more extreme values). Similarly, it may be useful to compare the parameters associated with language impairment to those that differ over development. By understanding language impairment at the level of online processing, we may be able to design more targeted diagnostics and interventions.

The present study approaches these questions from the context of *functional individual differences*. That is, we seek to study individual differences in word recognition that are (1) measurable using other instruments than our primary empirical measure (for example, measures of overall vocabulary learning, or productive language ability), (2) stable across individuals, and (3) potentially applicable to a wide range of theoretical and applied problems.

One approach that fits these criteria is to use general measures of non-verbal, and verbal ability to establish differences among participants. Such measures are independent of laboratory assessments of spoken word recognition, though there is some evidence that they are related to it (Dollaghan, 1998; Montgomery, 1999, 2002; Stark & Montgomery, 1995). These measures of non-verbal and verbal skill are reliable across time, both in terms of test–retest reliability and longitudinally over development (Tomblin, Zhang, Buckwalter, & O'Brien, 2003), suggesting that they offer a stable way to identify meaningful variation between participants. Finally, these measures are diagnostic of specific and non-specific language impairment (SLI and NLI), and therefore are of clinical significance. By pinpointing the relationship of word recognition processes to these measures, we may be able to relate variation in lexical processing to processing at other levels of language (e.g. syntax, speech).

General language and non-verbal abilities are clearly continuous and partially correlated. In this regard they can be treated as dimensions along which individual differences exist. Often within individual differences research, these dimensions are carved into groups and this practice is common in research on language impairment. Despite this, there is a growing consensus that SLI and NLI are arbitrary categories on underlying continuous dimensions (Leonard, 1987, 1991; Tomblin & Zhang, 1999). Fig. 1A, for example, shows the language and cognition scores from 527 children in the Iowa Longitudinal study and clearly illustrates the continuous distributions of these measures. This sample reinforces the high correlation between language and non-verbal IQ. More importantly, there do not appear to be any clusters that could be identified as qualitatively distinct groups.

Nonetheless, it is useful to discuss these dimensions in terms of four diagnostic categories (Fig. 1B). Children with SLI, who comprise 6–7% of the population (Tomblin et al., 1997), are defined as having language abilities at least one standard deviation below the mean for their age, despite non-verbal cognitive abilities that are not more than one standard deviation below the mean and no other obvious causal factors such as hearing impairment, articulatory disorders, developmental disorders, or neurological problems. Similarly, we define non-specific language impairment (NLI) as having language and non-verbal scores that are both lower than one standard deviation below the mean (and no other gross impairments). In the present study, we also define children with normal language skills (better than –1 SD for age) but impaired non-verbal intelligence as Specific Cognition Impaired or SCI. Our use of such labels is primarily for ease of discussion and does not imply any categorical nature to their deficits. Importantly, both of these continuous measures of ability and (to a lesser extent) the diagnostic category (SLI, NLI, etc.) have been found to be very stable over a long period of time (Tomblin et al., 2003), suggesting that these measures represent a robust dimension of individual differences with important practical implications. However, it is not clear how such measures are related to word recognition.

1.3. Word recognition in SLI

Children with both SLI and NLI have deficits in a range of language domains including morpho-syntactic skills (Leonard et al., 2003; Leonard, Deevy, Miller, Rauf, et al., 2003; Leonard, Eyer, Bedore, & Grela, 1997), phonological processes (Bishop & Snowling, 2004; Joanisse & Seidenberg, 2003; Sussman, 1993), perceptual ability (Bishop, Adams, Nation, & Rosen, 2005; Tallal & Piercy, 1974) and word learning (Dollaghan, 1985; McGregor, Friedman, Reilly, & Newman, 2002; McGregor, Newman, Reilly, & Capone, 2002). They also have generalized deficits in speed of processing (Kail, 1994).

SLI and NLI have been associated with deficits in spoken word recognition, although no qualitative differences have been observed. Stark and Montgomery (1995) compared SLI and normal performance in word-spotting tasks with time-compressed and low-pass filtered sentences. Participants with SLI

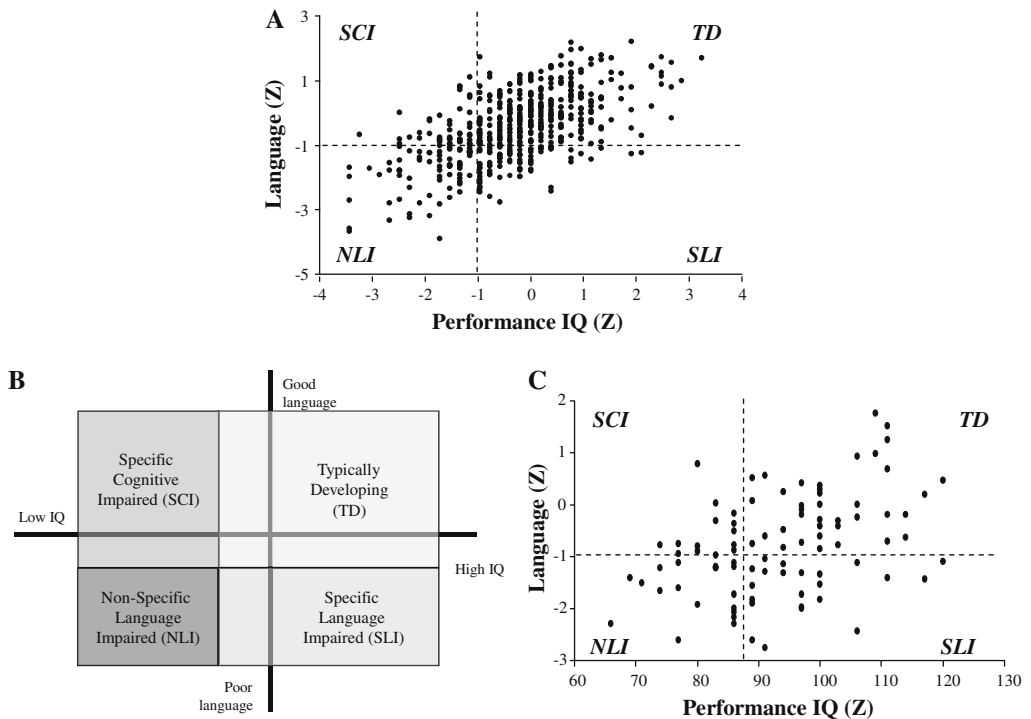


Fig. 1. Individual differences in language and cognitive ability. (A) Scatter plot showing the relationship between language and non-verbal cognitive abilities in 527 children taken from Tomblin and colleagues (1997). (B) Diagnostic criteria for specific language impairment (SLI), non-specific language impairment (NLI), specific cognitive impairment (SCI) and normal (N). (C) Scatter plot showing the relationship between language and non-verbal cognitive abilities in the present sample.

were slower overall (compared to normal controls), but were not differentially affected by either manipulation. This suggests a deficit in word recognition that was not specific to temporal perceptual processes. Similarly Montgomery (2002) examined a word-spotting task where the embedded target words contained mostly stop consonants or few stops (since stops are hypothesized to be difficult for listeners with SLI). Again, no effect of sentence type was seen, but children with SLI responded slower than controls.

Both studies were motivated by the idea that a perceptual impairment cascades to affect word recognition. This was not supported, so it is possible that such effects arise from differences in word recognition processes such as inhibition. Before reaching such a conclusion, however, two alternatives must be considered. First, Kail's (1994) generalized slowing hypothesis suggests that children with SLI are simply slower in a range of processes. Second, both studies used words embedded in sentences, raising the possibility that morpho-syntactic deficits may cause slowing (e.g. Montgomery & Leonard, 1998).

In order to reveal the unique contributions of word recognition, a number of studies have used isolated words. Edwards and Lahey (1996), for example, examined children with SLI in an auditory lexical decision task and found that language impairment was associated with a slower responses, but that the magnitude of slowing did not predict the severity of the impairment (see also Lahey, Edwards, & Munson, 2001).

While such results could be accounted for by generalized slowing, gated stimuli (e.g. Grosjean, 1980) offer a glimpse into the time-course of processing that is less sensitive to speed of processing. Dollaghan (1998) examined listeners with SLI and typically developing listeners as they heard progressively longer portions of isolated familiar words and newly learned (unfamiliar) words. There was no group difference for familiar words, but children with SLI required more acoustic material to recognize newly learned words. This suggests a deficit in the process of winnowing the competitor set, although this may only

be apparent for less active words. Additionally, in the shortest gates, children with language impairment were less likely to guess a word containing the correct onset phoneme, again raising the possibility that low-level perceptual deficits may be play a role in language impairment.

Later gating studies have been mixed. [Montgomery \(1999\)](#) used only highly familiar words in a similar paradigm and found no differences between language-impaired and typically developing children on either measure. Similarly, [Mainela-Arnold, Evans, and Coady \(2008\)](#) compared children with SLI and typically developing children's performance in a gating task in which both lexical frequency and neighborhood density (the number of phonologically similar words in the lexicon) were manipulated. Neither the recognition point nor the likelihood of recognizing the first sound varied as a function of language status, nor did language status interact with either frequency or density. However, at late gates, the children with language impairment vacillated between correct and incorrect responses (even though they had responded correctly at earlier gates). This would seem to implicate lexical, not perceptual factors, and suggests that lexical targets receive less activation, competitor words are more active, or some combination of both. Interestingly, neither frequency nor neighborhood density interacted with language, weakening evidence for lexical involvement.

1.4. Overview

The foregoing work suggests quantitative differences in word recognition between language-impaired and normal listeners, and the potential for broader individual differences. Conflicting results may be due to issues with the behavioral tasks. As discussed by [Montgomery \(1999\)](#) and [Dollaghan \(1998\)](#), gating tasks require metalinguistic judgments and may reflect differences in participants' conscious awareness of word recognition in addition to underlying recognition processes. It uses an untimed response, and cannot measure temporal properties unrelated to the temporal uptake of the signal (e.g. effects of late competition). Finally, the gated stimulus itself may not be treated as "part of a word", but rather, as a strange word in and of itself (see [Alloppenna et al., 1998](#); [Warren & Marslen-Wilson, 1987](#)).

The visual world paradigm or VWP ([Tanenhaus et al., 1995](#)) offers an alternative. In the VWP, participants hear spoken instructions to manipulate one or more objects in a visual environment. The set of objects represents competing interpretations of the auditory stimulus. As the subject completes the task, eye movements to each object are monitored to provide a real-time estimate of how strongly each interpretation is considered at any given time.

As applied to spoken word recognition (e.g. [Alloppenna et al., 1998](#); [Dahan, Magnuson, Tanenhaus, 2001](#); [Dahan, Magnuson, Tanenhaus, et al., 2001](#)), objects in the display typically include phonological competitors. [Alloppenna et al. \(1998\)](#), for example, presented participants with screens containing a target (e.g. *sandal*), a cohort (onset) competitor (*sandwich*), a rhyme competitor (*candle*), and an unrelated item (*parrot*). At around 200 ms after the onset of the word, participants began making more eye movements to the target and cohort than to the other two items. Because it takes approximately 200 ms to plan and launch an eye-movement, these early eye movements are indicative of immediate activation. About 200 ms after the point of disambiguation, looks to the cohort decreased, and some looks to the rhyme competitor were seen. By about 1000 ms after the onset of the word virtually all fixations were directed to the target.

This graded pattern of fixations could be fit to activation from the TRACE model ([McClelland & Elman, 1986](#)) by transforming the activations across the lexicon into fixation probabilities for each of the four objects with a simple linking function. This accounted for over 95% of the variance, suggesting that the millisecond-by-millisecond pattern of eye movements can be used to understand the online dynamics of lexical activations.

Subsequent research has demonstrated that this paradigm is sensitive to many of the factors that affect word recognition. These include frequency ([Dahan, Magnuson, Tanenhaus, 2001](#)), subphonemic mismatch ([Dahan, Magnuson, Tanenhaus, et al., 2001](#)), neighborhood density ([Magnuson, Dixon, Tanenhaus, & Aslin, 2007](#); [Magnuson, McMurray, et al., 2003](#); [Magnuson, Tanenhaus, et al., 2003](#)), sensitivity to continuous detail ([McMurray, Tanenhaus, & Aslin, 2002](#)), and semantic priming ([Huettig & Altmann, 2005](#); [Yee & Sedivy, 2006](#)).

Thus, the VWP is sensitive to real-time activation dynamics and can be readily mapped to available computational models. Additionally, it is a useful tool for impaired populations as it offers a natural

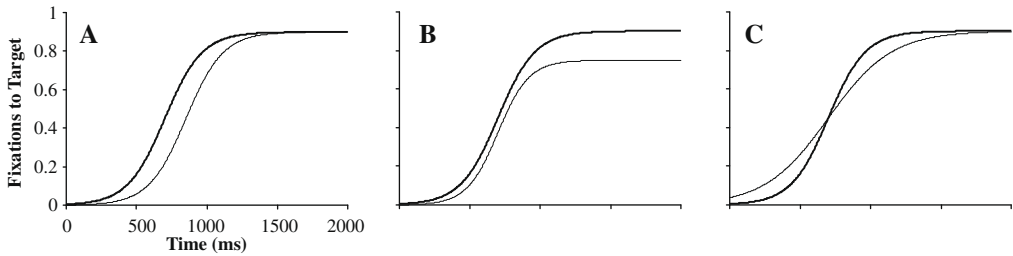


Fig. 2. Some possible patterns in the pattern of fixations to the target over time. (A) Variation in the delay before the function begins climbing; (B) variation in the peak amount of fixations; (C) variation in slope.

task that does not require metalinguistic judgments and has a low memory load. It has been used with a variety of populations including children (Arnold, Novick, Brown-Schmidt, & Trueswell, 2001), dyslexics (Desroches, Joanisse, & Robertson, 2006), autistic children (Brock, Norbury, Einav, & Nation, 2008; Campana, Silverman, Tanenhaus, Bennetto, & Packard, 2005), and aphasics (Yee, Blumstein, & Sedivy, 2004, 2008). It has also been applied to specific language impairment for sentence processing (Nation, Marshall, & Altmann, 2003).

In the present study we used the VWP in a task similar to that of Allopenna and colleagues (1998) to measure activation for the target word, a cohort competitor, and a rhyme as a function of time. These temporal dynamics were examined as a function of general language and cognitive abilities to determine how they differ between individuals. Finally, we used variants of the TRACE model to test hypotheses about the underlying processing dimensions (e.g. strength of lexical inhibition, rate of phoneme activation) that give rise to these differences.

1.4.1. Hypotheses

Work using the VWP to study lexical activation has shown that the proportion of fixations to the target word as a function of time generally follows a sigmoidal curve. Initially few eye movements are made to the target (or any other object), and this gradually ramps up to asymptote. Similarly, looks to phonological competitors (e.g. cohorts and rhymes) rise to a peak, and then fall to a baseline level. Both curves offer multiple dimensions of potential variability. For example, the function representing looks to the target can vary in at least three different ways (Fig. 2). It could be delayed but proceed at the same rate (Panel A), the upper asymptote could be reduced (Panel B) or its rise time could be slowed (Panel C). Similarly the cohort and rhyme functions (Fig. 3) could vary in the location of peak activation (Panel A), the downward slope (Panel B), the asymptote (Panel C), or the peak height (Panel D).

Thus, across both targets and competitors there at least eight degrees of freedom that can vary and may map onto specific hypotheses of language impairment. For example, both the generalized slowing approach (Kail, 1994), and the speeded word-spotting tasks (e.g. Montgomery, 2002), predict that listeners on the low end of the language scale (SLI and NLI) will be slower to recognize the target words, resulting in a change in the slope or the cross-over points. Moreover, work by Fernald and colleagues using a similar (but simpler) task (e.g. Fernald, Perfors, & Marchman, 2006) has demonstrated that one of the primary developmental changes is the speed at which participants fixate (or activate) the target word. If children with SLI or NLI are simply lagging their peers developmentally we may observe this pattern of delayed initial looks to the target. Alternatively, the gating results (e.g. Dollaghan, 1998) suggest that fixations to cohort competitors will be either higher overall (late), or will decay more slowly.

Computational modeling can test hypotheses about the nature of the underlying deficits in a different way. Some TRACE parameters map simply to known theoretical constructs. Phoneme inhibition, for example, gives rise to categorical perception or the sharpness of the phonological representation (McClelland & Elman, 1986; McMurray, Tanenhaus, et al., 2009). Similarly, parameters at the feature level such as decay and input noise can test hypotheses that low-level perceptual deficits or auditory memory processes are responsible for language impairment (Tallal & Piercy, 1974). While no single task or model can completely rule out any of these hypotheses, the use of a multi-levelled process

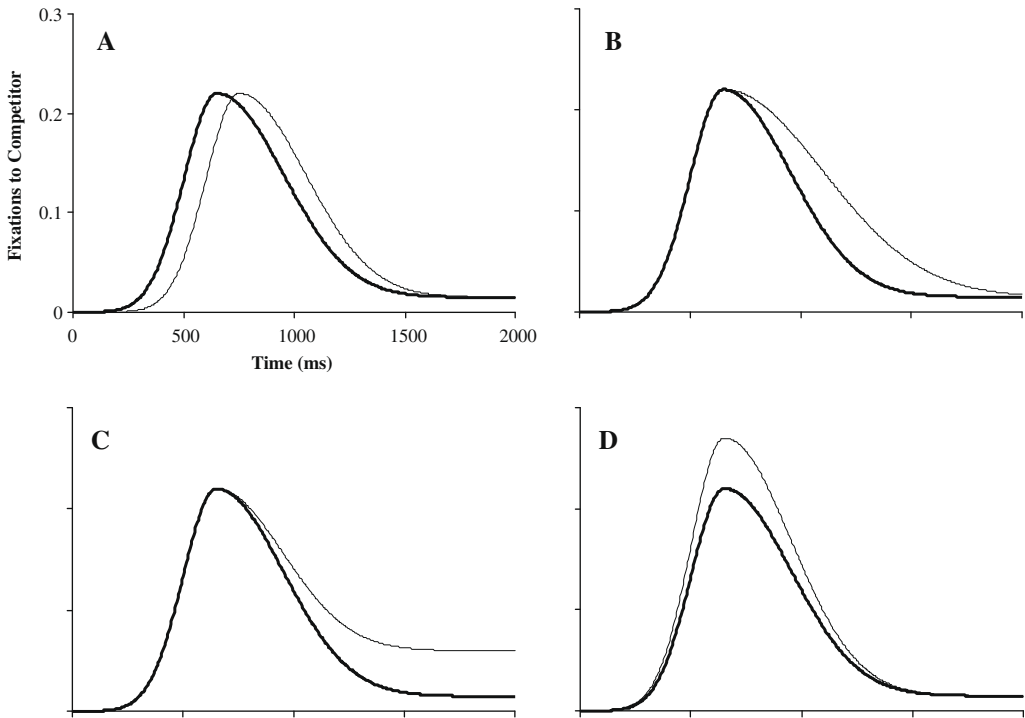


Fig. 3. Some possible patterns in the pattern of fixations to the cohort and rhyme competitors over time. (A) Variation in the delay before the function begins climbing; (B) variation in the slope as the function falls; (C) variation in the baseline it falls to; (D) variation in the peak fixations.

model like TRACE coupled with a behavioral task with this degree of specificity may offer some insight into competing theoretical accounts of the deficits faced by children with SLI, or push the research in a new direction. It can also test the validity and flexibility of TRACE to account for the kinds of differences we observe empirically.

2. Experiment

Participants were presented with screens containing a target, a cohort competitor, a rhyme competitor and an unrelated object and were asked to click on the target with a computer mouse. Eye movements were monitored throughout the process to examine activation dynamics.

We recruited a subject population with a range of variation in language and cognitive abilities, by targeting groups of participants from each of the four diagnostic groups identified earlier (SLI, NLI, SCI and TD) (Fig. 1B). These groups allowed us to determine the effect of general impairment (either IQ or language) and to isolate language- or IQ-specific deficits.

2.1. Methods

2.1.1. Participants

Ninety-three adolescents participated in this study (a subset of the longitudinal study conducted by the University of Iowa Child Language Research Center; Tomblin et al., 1997, 2003; Tomblin, Zhang, Weiss, Catts, & Ellis-Weismer, 2004). Participants were excluded if they did not have English as their primary language or if they had a history of sensory impairment, mental retardation, autism or neurological disorder. Diagnostic classification was based on eighth grade scores, since the longitudinal

Table 1

Age, language scores and performance IQ of the participants.

Group	N	Age (SD)	Performance IQ (SD)	Lang comp z-score (SD)
N	40	17; 1.3 (6.9 mo)	101.95 (8.20)	–.029 (.69)
SCI	16	17; 1.5 (8.2 mo)	82.25 (3.87)	–.607 (.505)
SLI	20	17; 2.9 (7.4 mo)	97.85 (9.26)	–1.686 (.481)
NLI	17	17; 1.7 (7.9 mo)	79.24 (6.71)	–1.722 (.468)

study clearly established the stability of these language scores with respect to age (Tomblin, 2005; Tomblin et al., 2003). Participants averaged 17 years, 1.75 months of age at the time of testing.

To achieve a roughly balanced design (while maintaining substantial variability along both dimensions) subject recruitment was targeted to the specific groups we have discussed. We collected data from 20 SLI (Specific Language-Impaired), 17 NLI (Non-specific-Language-Impaired), 16 SCI (Specific Cognition Impaired), and 40 typically developing teenagers (TD) (see Table 1). There was significant variability within groups but no clear clusters (Fig. 1C).

These groups (though arbitrary) are diagnostic in clinical speech-language pathology. The diagnostic criteria we employed are as follows. Participants in the SLI group passed the performance IQ criterion and failed the language criterion. SCI was defined as passing the language criterion but failing the performance IQ criterion. The NLI group failed both language and performance IQ criteria, and the typically developing (TD) group passed both criteria.

2.1.1.1. Language measures. Our language assessment was based on the epiSLI criteria for language impairment (Tomblin, Records, & Zhang, 1996). Listening and speaking ability in lexical, sentence and discourse contexts was measured using the language diagnostic battery described in Table 2. Raw scores were converted to standard scores and combined into a single composite score for each child by summing all subtest z-scores and then dividing this by the square root of the sum of the variances and the covariances across the subtests (Crocker & Algina, 1986). Language impairment was assigned when a child's score was less than –1.14 on this composite z-score. This cut-off approximates the probability of failing two of five subtests where failing is set at the 10th percentile.

2.1.1.2. Performance IQ measures. Participants were classified as cognitively impaired (SCI and NLI) if their performance IQ fell below one standard deviation below the mean. This was assessed with the Block Design and Picture Completion subtests of the Wechsler Preschool and Primary Scale of Intelligence-Revised (Wechsler, 1989). If the sum of these scaled scores was greater than or equal to 16, participants were classified as having normal non-verbal IQ, reflecting a performance intelligence score of 87.5 or more.

Table 2

Language measures used for determination of language status.

Test	Language domain
PPVT-R	Receptive vocabulary
CREVT – expressive	Expressive vocabulary
CELF-III: concepts and directions	Sentence comprehension
CELF-III: recalling sentences	Sentence production: imitation
Q R inventory – II/3	Listening: discourse comprehension
Qualitative reading inventory – 3	Discourse recall: total words

Peabody Picture Vocabulary Test – R (PPVT-R) (Dunn, 1981).

Comprehensive receptive and expressive vocabulary test (CREVT) (Wallace & Hammill, 1994).

Clinical evaluation of language fundamentals – III (CELF-III) (Semel, Wiig, & Secord, 1995).

Qualitative reading inventory – II and 3 (QRI-II/3) (Leslie & Caldwell, 1995, 2001).

2.1.2. Materials

2.1.2.1. Item selection. To ensure that the words used in this study would be familiar to the participants, a self-rated word familiarity survey was mailed to 249 10th-graders who were part of the original Tomblin and colleagues (1997) study. This survey was comprised of 175 items drawn from the Peabody Picture Vocabulary Test (PPVT-R, Form L) (Dunn, 1981), as well as a number of items that were potentially appropriate for the present study. Two sublists, Form A with 85 items and Form B with 90 items, were constructed. The participants received one of the two sublists and were asked to rate their familiarity with each of the words on a five-point scale, ranging from 1 (“I do not know what this word means”) to 5 (“I know for sure what this word means”). A total of 135 adolescents (28 SLI, NLI and SCI, and 107 ND) completed and returned the survey, for a response rate of 54.21%. Significant positive correlations between scores on the familiarity survey and PPVT raw scores for both language-impaired ($r(25) = .668, p < .01$) and typically developing ($r(105) = .473, p < .01$) respondents indicated that our survey could be used to assess vocabulary knowledge in adolescents (Lee, Samelson, & Tomblin, 2006).

Any words for which 70% or less of the respondents gave a rating of 4 or 5 were excluded from the possible stimulus set used in the eye-tracking study. From this final set, 41 groups of four nouns were selected for use in the experiment (Appendix A). These 41 sets included the original eight “target-cohort-rhyme-unrelated” two-syllable sets used by Allopenna and colleagues (1998),¹ as well as 33 new sets created specifically for this study: 12 sets of two-syllable words and 21 sets of one syllable words.

2.1.2.2. Stimulus pictures. For each noun, 10 pictures were downloaded from a commercial clip art database. A team of experimenters including graduate students and undergraduates examined each set of pictures to select the most representative exemplar. These selected pictures were then edited if necessary to yield a consistent level of coloring and brightness within each set, eliminate distracting elements, and make minor modifications to the picture. The final pictures were then approved by the first author, who has extensive experience with VWP tasks.

2.1.2.3. Auditory stimuli. The 163 auditory stimuli² were recorded by a male speaker in an anechoic chamber, using a Marantz PMD-670 solid-state digital recorder. The speaker produced three exemplars of each word, and the clearest was selected. One hundred milliseconds of silence was added to the onset of each word, and the entire set was amplitude normalized.

2.1.3. Procedure

Participants were seated at a computer and the eye-tracker was calibrated. Instructions were presented on the screen, followed by a verbal summary of the instructions and an opportunity to ask questions. The adolescents then performed four practice trials while the experimenter watched and corrected them if necessary. Participants were given a second opportunity to ask questions. At that point testing began.

On each trial, the four pictures corresponding to the target, cohort, rhyme, and unrelated item appeared on the screen, along with a blue circle at the center. After 750 ms, the circle turned red, and participants were instructed to click on it. This served to force the mouse and the eyes to the center of the screen before hearing the auditory stimulus. After clicking on the red circle, participants heard the target word and clicked on the appropriate picture.

Periodically during the experiment a computer-generated drift correction was completed, at which point the participant could take a break if desired. At each drift correction, participants were also re-

¹ Note that several of the original Allopenna et al. (1998) unrelated items were replaced with new words. This was done to eliminate the use of words in more than one set and to allow us to use some of these as targets or cohorts in new sets.

² While our design calls for 164 words (41 sets \times 4 words/set), there were only 163 target words. The Allopenna et al. (1998) study, on which this one is based, used several items as both unrelated and competitor objects (dolphin, dollar, speaker, beaker and parrot). We eliminated this in our present design. However, they also used candle was used as both a target (in the candle/candy/handle set) and a cohort competitor (sandal/sandwich/candle). Since this was an experimental set we opted to maintain this for the sake of continuity.

minded to listen to each word and click on the matching picture. Experimental sessions were monitored by the examiner via closed circuit video camera, to ensure that the headset was not bumped and that participants remained focused on the task.

For each item-set, each word appeared as the auditory stimulus an equal number of times. This led to four trial-types. Consider the item-set consisting of *candle*, *candy*, *handle*, and *button*. A full competitor set (a Target–Cohort–Rhyme, or TCR, trial) occurred when the auditory stimulus was *candle*: *candy* served as cohort, *handle* was a rhyme, and *button* served as a phonologically unrelated picture. However, when *candy* was the auditory stimulus, this trial had a cohort competitor (*candle*) and two unrelated objects (*handle*, and *button*), but no rhymes (a Target–Cohort, or TC trial). Likewise, if *handle* was the auditory stimulus, there would be a rhyme competitor (*candle*) but no cohort (a Target–Rhyme, or TR trial). Finally, if *button* was the stimulus, all of the items were unrelated (TU).

2.1.3.1. Eye-movement recording. Eye movements were recorded with an Eyelink II head-mounted eye tracking system (SR Research, Ltd.) at 250 Hz. The standard 9 point calibration procedure was used and participants had no trouble completing it. Calibration quality was estimated for each subject as the error (in degrees of visual angle) after the final validation. In order to verify that there was no systematic difference in the quality of the eye-track between our subject groups, these error measurements were subjected to a 2 (language status) \times 2 (non-verbal cognitive status) ANOVA. We found no effect of language status ($F(1,89) = 2.67, p > .1$), no effect of non-verbal cognitive ability ($F < 1$), and these factors did not interact ($F < 1$). Thus, basic eye-movement control and measurement was not a specific issue for any of our subject groups.

Fixations were recorded from onset of the central fixation point. Recording stopped when the subject clicked on the selected picture. The eye-track was automatically parsed into events (saccades, fixations, blinks), using the eye-tracker's default "psychophysical" parameter-set. Similar to prior work (McMurray et al., 2002), adjacent saccades and fixations were combined into "looks" on which subsequent analyses were performed.

2.1.3.2. Design. There were 41 item-sets, and each of the four members of the set appeared as the auditory stimulus once (four trial-types). This led to 164 trials. When repeating fixed sets of items a small number of times, however, there is a danger that participants may employ a process of elimination strategy. For example, if participants have previously clicked the *candy* or the *candle*, on a subsequent trial when they see this item-set again, they may assume that *button* or *handle* is the correct target before hearing it. In order to minimize this, an additional 3 blocks of 41 trials were included in which the target was selected completely randomly. Thus, some words could be used up to four times in the course of an experiment and others only once. This led to 287 trials, of which approximately one fourth fell into each trial-type.

2.2. Results

Our analysis began with an analysis of the mouse-click (identification) and reaction-time data to demonstrate that our participants were able to complete the task successfully. We then analyzed the eye-tracking data. Consistent with our emphasis on individual differences, we used linear regression to relate continuous language and non-verbal IQ scores to these data. However, the performance of the diagnostic groups may be of interest to researchers in communication disorders. ANOVAs using clinical groups can be found in [Supplementary material](#) (note 1).

2.2.1. Mouse click (Identification) results

Our first analyses examined the mouse click responses (1) to verify that this was a feasible task for our impaired groups of participants and (2) to verify that the words and pictures were familiar. Participants performed quite well in this experiment, averaging 98.8% correct ($SD = 1.15\%$), an average of 3.4 missed trials out of 287 (see [Table 3](#)). Thus, this appears to be a reasonable task for use with this population.

In order to determine if language or non-verbal IQ was related to performance, hierarchical regression was used. The first regression examined percentage correct. On the first step, language and IQ

Table 3

Mean performance and RT.

Subject group	N	% Correct (SD)	RT in ms (SD)
Normal	40	99.2 (.83)	1429 (148)
SLI	20	98.3 (.87)	1450 (163)
SCI	16	99.0 (1.10)	1493 (167)
NLI	17	98.2 (1.64)	1635 (168)

accounted for 18% of the variance ($F_{\text{change}}(2,90) = 10.14, p = .0001$). However, only language was significant individually ($T(1,90) = 3.9, p = .00018$)—participants with lower language scores performed slightly worse than those with high scores (the diagnostically language-impaired listeners scored 98.3% correct compared to 99.1% for unimpaired). Performance IQ was not significant ($T < 1$). On the second step of the model, the interaction of language and non-verbal IQ was marginally significant, accounting for an additional 3% of the variance ($F_{\text{change}}(1,89) = 3.13, p = .08$). Listeners who were poor in both language and cognitive abilities performed slightly worse than those who were only poor in language.

The next regression examined reaction time. On the first step, language and IQ were added to the model and accounted for 19% of the variance ($F_{\text{change}}(2,90) = 10.41, p = .0001$). However, only language was significant individually ($T(1,90) = 2.9, p = .005$)—participants with lower language scores were slower than those with high scores (the diagnostically language-impaired listeners averaged 1542 ms compared to 1462 ms for unimpaired). This was not unexpected and extends the word spotting results described earlier (Montgomery, 2002; Stark & Montgomery, 1995) to single word recognition. Performance IQ was marginally significant ($T(1,90) = 1.98, p = .051$), also due to the fact that adolescents with cognitive impairment ($M = 1564$ ms) performed worse than those without cognitive impairment ($M = 1439$ ms). On the second step, the interaction of language and non-verbal IQ was added and did not account for significant additional variance ($R^2_{\text{change}} = .01, F_{\text{change}}(1,89) = 1.4, p > .1$).

The significant group differences in both analyses do not undermine our use of the VWP with this population. The actual size of the effect on overall performance (in terms of percentage correct) was small (less than 1% between language-impaired and non-language-impaired groups), and each individual subject performed better than 95% correct. Thus, even adolescents with language and cognitive impairments are fully capable of recognizing these words. The RT difference is larger, though not unexpected. It is potentially more problematic, since averaging eye movements across trials of different lengths can create apparent group differences even if there are none. Thus the analyses conducted on the fixations were constructed with this in mind.

2.2.2. Fixation proportions

Fig. 4 shows the probability that the participants fixated each object every 4 ms. Trials in which the adolescents failed to select the correct target were excluded from this analysis, since the participants were not expected to make eye movements to the target if they failed to recognize it. Additionally, a number of participants initiated saccades during the 100 ms period between the disappearance of the red circle and the onset of the auditory stimulus. Such eye movements were not the result of any linguistic processing, and were removed from the data set and replaced with a “no data” code.

Panel A shows the time course of fixations for TD participants. Shortly after word onset (the first response to which occurs at 300 ms in these curves), the curves representing the target and cohort deviate from the unrelated and rhyme, as at this point in time the auditory stimulus has unfolded to the point where some disambiguating information is available. Three-hundred ms reflects the 200 ms it takes to plan and launch an eye-movement and the 100 ms of silence between trial onset and the auditory stimulus. Somewhat later, fixations to the rhyme deviate from the phonologically unrelated object, and by the end of the trial, participants are only fixating the target.

This overall pattern was not qualitatively disrupted in any of the three impaired groups (Fig. 4B–D). However, there are quantitative differences. Fig. 5A shows the proportion of looks to the target for each of the four groups. An effect of language impairment can be seen, with SLI and NLI listeners showing a lower likelihood of fixating the target than TD and SCI children at the end of the time-

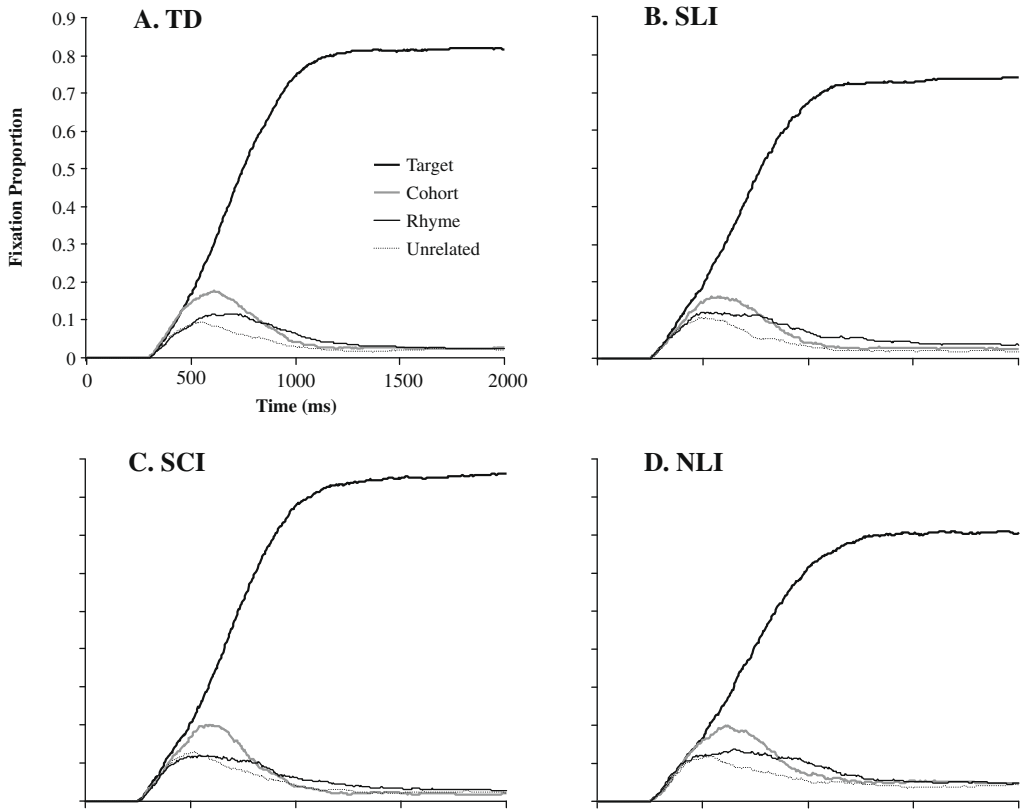


Fig. 4. Looks to the target, cohort, rhyme and unrelated objects as a function of time for each of the four groups.

course. Likewise Panels B (cohort fixations) and C (rhyme fixations) suggest that language-impaired adolescents may have difficulty suppressing looks (activation) to competitor objects, though this effect seems greater for children with NLI than SLI. All three effects are primarily late in the time-course, when fixations have stabilized. Given that these figures are based only on trials in which the subject selected the correct word, they lead to the unexpected conclusion that on at least a small proportion of trials, language-impaired adolescents are actually fixating on the cohort or rhyme *while they are clicking on the target*.

This pattern of results does not support the idea of a deficit that involves early processes in word recognition. Variation in language ability also does not slow the time-course of processing. Rather, at first glance, adolescents with lower than average language skills show decreased target activation and increased competitor activation. However, this must be verified statistically, and we must rule out other potential causes.

Our statistical analyses were designed to address three issues. First, we needed to validate this description of our data, demonstrating that this late pattern of fixations is related to language ability (and potentially to cognition). Second, it is possible that averaging across participants masks (or creates) effects. For example, variability between participants in the timing of the function could average to yield what looks like a difference in slope. Thus, in order to uncover any other differences that may relate to language impairment, we needed to characterize each subject's time-course of processing parametrically. Finally, these late asymptotic effects were somewhat unexpected and have not been attested in other VWP studies. Thus, the third goal of our analyses was to verify that this effect arises from language status, and not something less interesting (e.g. differences in how eye movements are deployed).

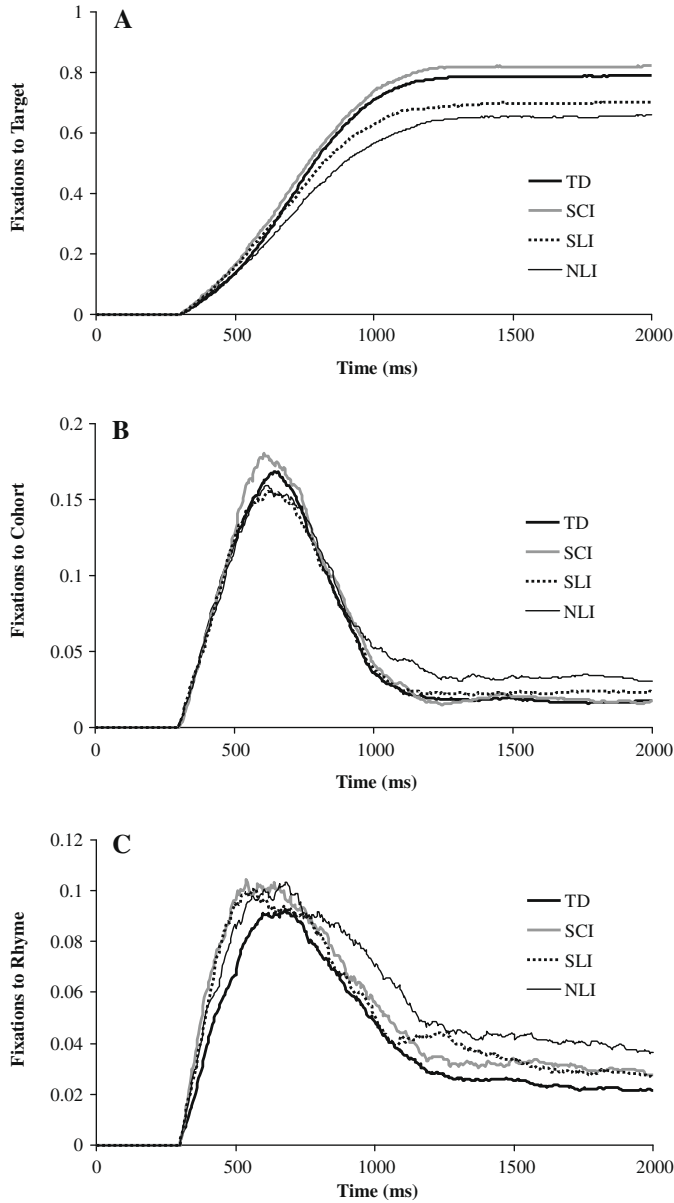


Fig. 5. Looks to the target (panel A) as a function of time and group membership. (B) Looks to the cohort as a function of time and group membership. (C) Looks to the rhyme as a function of time and group membership.

Typical analyses used with the VWP simply take the area-under-the-curve over specific time windows (and online note 1.2 reports our preliminary analyses with this approach). This is insufficient for our purposes since it does not characterize the time course of processing and would violate the independence assumptions of ANOVA (see McMurray, Tanenhaus, Aslin, & Spivey, 2003; Mirman, Dixon, & Magnuson, 2008). This makes it difficult to detect individual differences in functions that may be masked by averaging across time and participants. Moreover, this analysis requires that all partici-

pants have data in the time-window of interest. This creates problems when the time-window is late, as not all participants will have data at late points in time and some may not have completed some of the trials yet. Thus, we needed an analysis which implicitly takes time into account.

Mirman and colleagues (2008) demonstrated the ability of mixed effects models to accomplish this. Their approach fits each participant's time-course of fixations to a polynomial function of time. These are then combined in a mixture model to evaluate differences across participants. This solves a number of the problems outlined above. Time is represented explicitly in the polynomial function, and the parameters describe the change over time, not the quantity of fixations. Each participant has his or her own parameters for this function, allowing differences to be characterized precisely. Finally, because time is represented as part of the function, curves can be fit to each subject over only the time window in which they have data, eliminating the problems created by variable reaction times.

This approach can solve most of the problems raised by the area-under-the-curve method. However, polynomial functions may not be ideal for fitting this sort of data. To fit the sigmoidal target functions, for example, a 5th or 6th order polynomial is needed (without arbitrarily cutting off the asymptotic portion that is of interest). The bell-shaped cohort function can be fit by a lower-order polynomial, but only in a restricted range. In particular, we would have to exclude the asymptotic portion of the curve, as it is not possible for a reasonably small polynomial function to approximate asymptotic behavior. Finally, while the coefficients of polynomial functions mathematically describe the function, they do not offer any simple interpretation because they do not describe immediately observable properties of the function.

Thus, we adopted the spirit of Mirman and colleagues' approach, but used non-linear functions to fit each participant's data. Starting with the time course data (e.g. Figs. 4 and 5), non-linear functions were fit to each participant's results (separately for target, cohort and rhyme fixations). The parameters describing these functions were then examined with linear regression to determine precisely how participants' language and cognitive abilities influenced the specific aspects of the time course that were described by each parameter (see online note 1.3 for analogous analyses in the ANOVA framework).

We used a logistic function (Fig. 6A) to model fixations to the target, and a function based on the Gaussian to model competitor fixations (Fig. 6B). Both functions offer parameterizations in which individual parameters describe readily observable aspects of the function. This transparency allows a straightforward description of our data.³ However, this approach has a few limitations. First, the parameters of these functions cannot be independently estimated (like the coefficients of a polynomial). For example, the slope of the logistic function cannot be estimated until the cross-over is identified. However, that does not mean that any given function can be multiply determined – there is only one set of parameters that gives rise to any specific function. Thus, the lack of independence does not create ambiguity in interpreting parameter values (though it may create difficulties in estimating them), as long as the estimates are globally optimal. To ensure this was the case, we used a constrained gradient descent method to estimate the parameters of each function, and subsequently compared these fits to the empirical data visually for each subject, to verify that they did not arise from a local minimum.

Second, the relationship between the parameters of a non-linear function and the function's behavior can be non-linear. This can cause problems when trying to relate the parameters to factors like language ability using linear regression. However, it is important to note that for these particular functions, the parameters describe readily observable properties (e.g. the peak fixations), so in effect we are simply relating measurable components of the data to language status – curve-fitting simply gives us a way to extract such values objectively.

Since curves were fit to each participant's data, it was important maximize the amount of data used in each fit. Thus, we averaged across trial-types. Looks to the target consisted of looks to whichever

³ The use of these non-linear functions (particular the asymmetric Gaussian) makes it difficult to employ them in a complete mixed model (as Mirman et al., 2008, did). Thus our approach does not take into account within-subject error in the between subject effects. This is only a factor if the quality of the fits differed systematically as a function language or cognitive ability. In all of the analyses presented here, this was not the case – the least-squared error of the fit was not significantly correlated to language, intelligence or their interaction (Target: $R^2_{\text{model}} = .03$; Cohort: $R^2_{\text{model}} = .02$; Rhyme: $R^2_{\text{model}} = .04$).

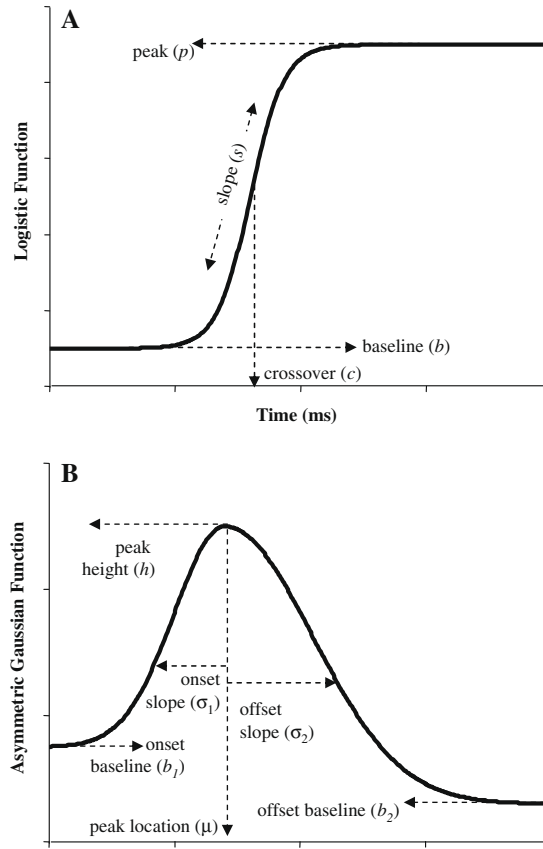


Fig. 6. The functions and parameters used for curve-fitting. (A) To fit looks to the target, a 4-parameter logistic function was used. This function is defined by upper and lower asymptotes, the cross-over point, and the slope of the transition. (B) To fit looks to the cohort and rhymes, an asymmetric Gaussian was used. This function is defined by the location and height of the peak, and the slope and asymptotes at the onset and offset components.

object was targeted across all four trial-types. Looks to the cohort collapsed across TCR and TC trials; and looks to the rhyme collapsed across TCR and TR trials. (Trial-type was assessed in separate analyses using area-under-the-curve, and reported in online note 1.2).

A final concern was how to deal with variability in the amount of time it took participants to complete the trial (both within- and across-participants). In most prior work (e.g. Allopenna et al., 1998; Dahan, Magnuson, Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, et al., 2001; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, et al., 2009), this is dealt with by choosing an arbitrary endpoint (e.g. 2000 ms) and either truncating trials that are longer, or extending the final fixation for trials that are shorter. However, in this study the effects were largely late in the time course. Adopting prior techniques could thus allow a few participants that responded relatively quickly to dominate the dataset and create the appearance of a relatively flat function. This could also mean that differences in the number of early and late responders in each group would give rise to seeming differences in the asymptotic performance.

Thus, for each subject we computed that subject's mean reaction time across trials. We then truncated (or extended) each trial relative to that subject's reaction time, prior to curve-fitting. In effect, we used the usual procedure, only with a stop-time that was unique to each subject. Since this was in fact their mean, for each subject approximately 50% of their trials would be truncated and 50% would be

extended trials to the dataset. Thus, the differing response latencies across participants could not give rise to between subject differences.

2.2.2.1. Analysis of fixations to target. To analyze the target fixations we used the **four-parameter logistic function** (Eq. (1); Fig. 6A) as the basis of our fits.

$$P(\text{target}) = \frac{p - b}{1 + \exp\left(4 \cdot \frac{s}{p-b} \cdot (c - t)\right)} + b \quad (1)$$

Time in milliseconds is represented by t . The upper asymptote is represented by p , and the lower by b . The cross-over point (in ms) of the function (where the function's rate of change is at maximum) is denoted by c . The derivative of the function at that point is s , the slope.

A priori, we would expect to see little difference in b (onset asymptote), as this reflects the number of eye movements to the target prior to the onset of the auditory stimulus. Changes in s (slope) would reflect the growth of activation and changes in c reflect when activation starts growing. Finally, changes in p would indicate a change in the peak amount of activation. Given Fig. 5A, we expect effect of language ability primarily on p , and no effect of cognitive ability.

We computed, for each subject, the proportion of trials at which the subject was fixating the target at each point in time (collapsing across all four trial-types) from 0 ms to the mean reaction time for that subject. Trials in which the subject clicked on the incorrect object were excluded. Eq. (1) was then fit to each subject's data using a constrained gradient descent technique that minimizes the RMS (root mean squared) difference between the function and the raw data (but constrains the possible fits to be within reasonable bounds). Fits were good, with an average R^2 of .998 (SD = .0027, Max = .999, Min = .973) and an average RMS of .016 (SD = .005). Thus, the logistic function was a good choice for this dataset.

We next used linear regression analyses to examine the effect of language and non-verbal cognitive status on each parameter of this function. All regressions were conducted hierarchically, with both main effects entered on the first step and the interaction on the second. The first analysis examined the baseline (b). As expected, we found no effect of language or cognitive abilities ($R^2_{\text{change}} = .02$, $F_{\text{change}} < 1$), and no interaction ($R^2_{\text{change}} < .001$, $F_{\text{change}} < 1$). Likewise, there was no effect on the cross-over point (c) of either language, cognitive ability ($R^2_{\text{change}} = .001$, $F_{\text{change}} < 1$), or the interaction ($R^2_{\text{change}} = .01$, $F_{\text{change}} < 1$). Thus, language and non-verbal cognitive status did not affect baseline looking to the target or the midpoint of the function describing the increase in looks to the target.

The third regression examined the upper asymptote (p). Here, the main effects entered on the first step significantly accounted for 9% of the variance ($F_{\text{change}}(2,90) = 4.38$, $p = .02$). This was driven by an effect of language status ($T(90) = 3.0$, $p = 0.004$) with language-impaired participants exhibiting lower peaks ($M = .737$, $SD = .22$) than unimpaired participants ($M = .848$, $SD = .135$). Non-verbal cognitive status was not significant ($T < 1$). Similarly, the interaction of language and cognition did not significantly account for any variance ($R^2_{\text{change}} = .02$, $F_{\text{change}}(1,89) = 1.5$, $p > .2$).

The final regression examined slope (s). Language and cognitive abilities together accounted for 15% of the variance ($F_{\text{change}}(2,90) = 8.0$, $p = .0006$). This was driven by a significant effect of language status ($T(90) = 3.8$, $p = .002$) and no effect of non-verbal IQ ($T < 1$). For language-impaired adolescents, the curve grew at a slower rate ($M = .0011$, $SD = .00023$) than for unimpaired participants ($M = .0014$, $SD = .00034$). On the second step the interaction was added and was not significant ($R^2_{\text{change}} = .02$, $F_{\text{change}}(1,89) = 1.7$, $p > .1$).

Thus, language ability specifically affected both the peak fixations and the slope of the function, while cognitive ability had no effect on target fixations.

2.2.2.2. Analysis of fixations to cohorts and rhymes. The fixations to the cohort and rhymes (Fig. 5B and C) require a more complex function. Rather than one transition, competitor fixations briefly rise and then fall. The rise and fall could each have different slopes, and did not always baseline at 0. We were not aware of any standard function with the necessary degrees of freedom, so we adopted a hybrid function based on two partially independent Gaussians (we call it the *Asymmetric Gaussian*, see Fig. 6B). These two Gaussian functions were not constrained in terms of either baseline or area (as

is the classic Gaussian), but were constrained to “line-up”—they had to have the same peak location and height.

$$P(\text{competitor}) = \begin{cases} \exp\left(\frac{(t-\mu)^2}{-2\sigma^2}\right)(p-b_1) + b_1 & \text{if } t \leq \mu \\ \exp\left(\frac{(t-\mu)^2}{-2\sigma^2}\right)(p-b_2) + b_2 & \text{if } t > \mu \end{cases} \quad (2)$$

The upper function describes the first half of the time course, and the lower one describes the second. μ and p (the location of the peak and its height, respectively) are the same across both Gaussians, while σ and b are specified independently for each. Thus, there are six parameters—the onset baseline (b_1), the onset slope (σ_1), the location of the peak in milliseconds (μ), the height of the peak (p), the offset slope (σ_2) and the offset baseline (b_2). This function can model most types of variation in the shape of cohort and rhyme functions, using meaningful parameters. It is also well-behaved—despite the hybrid formalization, the derivative is continuous. As before, we used a similar constrained gradient descent method to achieve our fits.

Our first fits with the asymmetric Gaussian examined fixations to the cohort. Overall fits were good, with a mean R^2 of .985 (SD = .011, Min = .913, Max = .996) and a mean RMS error of .0086 (SD = .0022). Thus, this function described these data well. As before, each of the six parameters was entered into hierarchical regressions to relate them to language and cognition.

Four parameters showed no effect of language, non-verbal cognitive status or an interaction: the onset baseline, b_1 (Language and Cognition: $R_{\text{change}}^2 = .011$, $F_{\text{change}} < 1$; Interaction: $R_{\text{change}}^2 < .001$, $F_{\text{change}} < 1$); the peak location, μ (Language and Cognition: $R_{\text{change}}^2 = .038$, $F_{\text{change}}(2,90) = 1.8$, $p > .1$; Interaction: $R_{\text{change}} = .003$, $F_{\text{change}} < 1$); the peak height, h (Language and Cognition: $R_{\text{change}} = .027$, $F_{\text{change}}(2,90) = 1.3$, $p > .2$; Interaction: $R_{\text{change}} < .001$, $F_{\text{change}} < 1$); and the onset slope, σ_1 (Language and Cognition: $R_{\text{change}} = .044$, $F_{\text{change}}(2,90) = 2.1$, $p = .13$; Interaction: $R_{\text{change}} = .018$, $F_{\text{change}}(1,89) = 1.7$, $p = .19$).

Offset slope (σ_2) did show significant effects. On the first step, language and cognition accounted for 8.3% of the variance ($F_{\text{change}}(2,90) = 4.1$, $p = .02$). This was driven by the fact that language was significant ($T(90) = 2.86$, $p = .005$), but not non-verbal IQ ($T(90) = 1.1$, $p > .2$). As language scores decreased, σ_2 increased suggesting that poor language users showed more late fixations to the cohort. On the second step of the analysis, the interaction was added and did not significantly account for any variance ($R_{\text{change}} \leq .001$, $F_{\text{change}} < 1$).

Offset baseline (b_2) showed a similar pattern. Language and cognitive abilities accounted for 7.4% of the variance on the first step ($F_{\text{change}}(2,90) = 3.6$, $p = .031$). Language was marginally significant ($T(90) = 1.88$, $p = .064$) and intelligence was not ($T < 1$). As with offset slope, participants with poor language abilities showed increased baseline values, suggesting that they were making more late eye movements to the cohort. The interaction was marginally significant ($R_{\text{change}} = .036$, $F_{\text{change}}(1,89) = 3.6$, $p = .06$): listeners with both language and cognitive deficits had marginally higher values for b_2 than those without cognitive deficits.

The analysis of rhyme fixations revealed a similar pattern. Overall fits were good, with a mean R^2 of .965 (SD = .023, Min = .88, Max = .99) and a mean RMS error of .008 (SD = .002). As before, no effects were seen on the onset baseline, b_1 (Language and Cognition: $R_{\text{change}} = .05$, $F_{\text{change}}(2,90) = 2.4$, $p > .1$; Interaction: $R_{\text{change}} = .01$, $F(1,89) = 1.3$, $p > .2$); peak location, μ (Language and Cognition: $R_{\text{change}} = .001$, $F < 1$; Interaction: $R_{\text{change}} = .029$, $F(1,89) = 2.7$, $p > .1$); peak height, h (Language and Cognition: $R_{\text{change}} = .006$, $F < 1$; Interaction: $R_{\text{change}} = .005$, $F < 1$); or onset slope, σ_1 (Language and Cognition: $R_{\text{change}} = .002$, $F < 1$; Interaction: $R_{\text{change}} = .032$, $F(1,89) = 2.96$, $p = .088$).

As with the cohort, the two late parameters described most of the individual differences. Offset slope, σ_2 was not significant on the first step of the analysis (Language and Cognition: $R_{\text{change}} = .01$, $F < 1$) but the interaction was marginally significant on the second ($R_{\text{change}} = .035$, $F(1,89) = 3.3$, $p = .07$). Children with both language and cognitive deficits showed a higher offset slope than any of the other three groups. The offset baseline (b_2) was more sensitive to language alone. In this analysis, language and cognition accounted for 8.5% of the variance ($F_{\text{change}}(2,90) = 4.2$, $p = .017$). Individually, non-verbal cognitive ability was not significant ($T(90) = 1.1$, $p > .2$), but language was ($T(90) = 2.0$, $p = .045$), with higher baselines for language-impaired ($M = .031$, SD = .026) than unim-

paired participants ($M = .022$, $SD = .017$). The interaction did not significantly account for any new variance ($R_{\text{change}} = .007$, $F_{\text{change}} < 1$).

Thus, the analysis of the rhyme fixations showed a similar pattern to the cohorts. Significant effects emerged in the late components of the curve: the offset baseline and the offset slope. These were primarily effects of language, suggesting that poorer language abilities increased activation for competitors like cohorts and rhymes toward the end of processing.

In both analyses, language did not show strong interactions with non-verbal IQ. This was surprising given the apparent difference between SLI and NLI adolescents observed in Fig. 5B and C. However, this grouping is relatively arbitrary and it is possible that this apparent interaction was driven by the slightly more severe language deficits faced by the children with NLI than those with SLI in this sample ($M_{\text{SLI}} = -1.69$; $M_{\text{NLI}} = -1.72$).

2.2.2.3. Asymptotic effects? The foregoing analyses suggest that the primary effects of language ability were on the asymptotes of the curves. This was surprising since this component of the curve has not been observed to vary systematically in other experiments, and has not been used as a primary dependent measure in any study that we are aware of. More importantly, toward the end of the trial a number of factors (beyond language) could affect the fixation record. These factors range from averaging artifacts introduced by variability in response latency, to the possibility of strategic looking after the response has been planned. Thus, it is important to determine if this effect arose from sources unrelated to language ability.

The first factor we considered is whether the language-impaired listeners knew the words. If on some proportion of the trials, the LI listeners did not know the correct target, their eye movements would be expected to be scattered randomly across the four objects. Thus, differential rates of correct responding could give rise to the differences we observed. However, the foregoing analyses excluded any such trial, so this is not the likely source.

An additional factor that has not been addressed is the possibility that participants on the lower end of the language scale may not have known the names of the *competitor* objects. To address this we discarded any item-set for which the subject was unable to identify all four objects correctly 100% of the time. Results were unchanged.

Having ruled out accuracy as the source of the difference, we looked next at the possibility that differences in response latency could give rise to the effects. We have discussed how averaging eye movements across participants with different latencies could allow late effects to be driven by a few participants. Our curve-fitting analysis deals with this potential confound by using each subject's mean reaction time as the endpoint of the trial. However, this averaging across trials may still mask important effects. For example, participants may engage a strategy by which they fixate the target prior to initiating the mouse movement, and then, while moving the mouse, return to the center to prepare for the next trial. On any given trial, this would look like a “dip” in the time-course, but averaged across trials, it would appear as lower fixations at the end of the trial. If language-impaired participants were more likely to use this strategy, this could give rise to apparent asymptotic effects.

To cope with this, we normalized time for each subject by recoding the onset of each individual eye-movement in terms of its percentage of the response latency. If the above strategy was present, but masked by averaging, it should appear in this analysis. As Fig. 7A shows, this was not observed in any of the four groups. In addition, inspection of individual participant's data did not reveal any instances of this (Fig. 7B shows the first 10 participants).

The final, and perhaps most important, factor to consider is whether the language-impaired listeners are also impaired in their deployment of eye movements (though its not clear why this would be related to language and not non-verbal IQ). We have already ruled out basic measurement issues: language-impaired participants do not have poorer calibrations than typically developing participants. However, they may still deploy eye movements differently—perhaps participants with lower language skills were simply “looking around” more, distributing more eye movements to every object (and hence fewer to the targets).

To assess this we examined eye movements before any language input was heard (otherwise, these fixations could be motivated by language processes), during the time from the onset of the trial until participants clicked the red circle to initiate the auditory stimulus. In this window we counted the

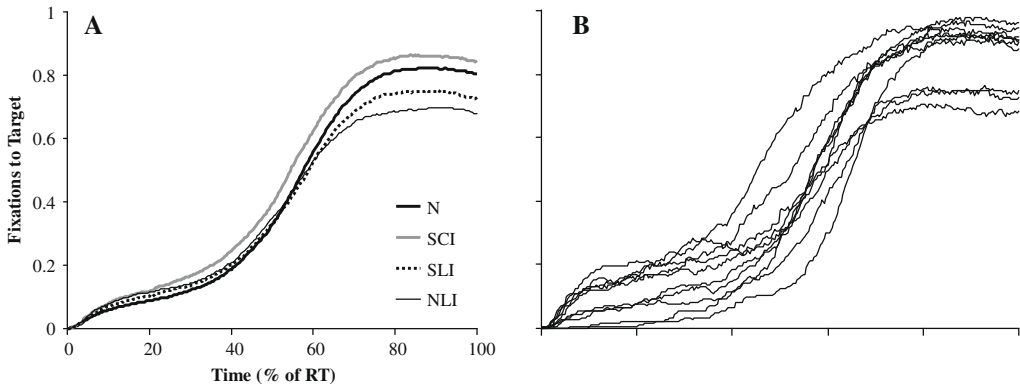


Fig. 7. Looks to the target as a function of time when time is normalized to the reaction time on *each trial*. (A) Fixations broken down by each of the four diagnostic groups. (B) The first 10 participants' individual functions.

number of fixations to the objects, to determine if language status was related to basic scanning processes. Since this time was variable (it was dependent on the participants' responses, though it did not differ systematically as a function of language or IQ, $F(2,90) = 2.1, p > .1$), this factor was also included in the regression analysis.

In the first step, the length of this scanning period accounted for 19% of the variance ($F(1,91) = 22.0, p < .0001$) in number of fixations. In the second step, language and non-verbal IQ were added and did not significantly account for any new variance ($R_{\text{change}} = .018, F_{\text{change}}(2,89) = 1.1, p > .1$). On the third step, the interaction of language and IQ accounted for no additional variance ($R_{\text{change}} = .003, F_{\text{change}} < 1$).

While this rules out scene scanning differences, fixations to each object reflect both activation and the processes that link activation to eye movements. Thus, it is possible that language-impaired listeners have similar lexical activation processes to typically developing listeners, but do not link them to the visual scene in the same way. To assess this, we conducted the same analysis, with the number of fixations to the lexical candidates *after* the auditory stimulus as a DV. In this analysis, the duration of the trial (reaction time) was entered on the first step and accounted for 23% of the variance ($F_{\text{change}}(1,91) = 27.8, p < .0001$). Language and non-verbal IQ were added on the second step and did not account for new variance ($R_{\text{change}} = .008, F_{\text{change}} < 1$). Finally, the interaction was not significant ($R_{\text{change}} = .003, F_{\text{change}} < 1$).

Thus, we did not see any effects of language on the number of eye movements participants make in general (e.g. before the trial) or during language processing. However, such patterns could be driven by other components of the eye-movement record (e.g. the duration of eye movements, the probability of staying on an object or scanning within it), all of which are averaged to create curves like Fig. 4. Thus, the strongest test would look at the fixations to the unrelated object, which should not be affected by language ability.

We used the same curve-fitting techniques as were used for the cohort and rhyme to examine looks to the unrelated object. Fits were good, with an average R^2 of .968 (SD = .024, Min = .884, Max = .992). As before, regression analyses examined the effects of language, cognition, and their interaction on each of the six components. There were no effects on any of the six parameters: peak location (μ) (Language and Cognition: $R_{\text{change}} = .038, F_{\text{change}}(2,90) = 1.8, p > .1$; Interaction: $R_{\text{change}} < .001, F < 1$), peak height, h (Language and Cognition: $R_{\text{change}} = .016, F_{\text{change}} < 1$; Interaction: $R_{\text{change}} < .001, F_{\text{change}} < 1$), onset slope, σ_1 (Language and Cognition: $R_{\text{change}} = .037, F_{\text{change}}(2,90) = 1.7, p > .1$; Interaction: $R_{\text{change}} = .002, F_{\text{change}} < 1$), offset slope, σ_2 (Language and Cognition: $R_{\text{change}} = .019, F_{\text{change}} < 1$; Interaction: $R_{\text{change}} = .002, F_{\text{change}} < 1$), or onset baseline, b_1 (Language and Cognition: $R_{\text{change}} = .005, F_{\text{change}} < 1$; Interaction: $R_{\text{change}} = .003, F_{\text{change}} < 1$). Offset baseline, b_2 , approached significance on the first step ($R_{\text{change}} = .052, F_{\text{change}}(2,90) = 2.5, p = .088$), but neither language ($T(90) = 1.66, p = .1$) nor cognitive ability ($T(90) = 1.17, p = .24$) were significant. In addition, the interaction of language and cognitive abilities was not significant ($R_{\text{change}} = .001, F_{\text{change}} < 1$).

Thus, across these analyses there is little evidence to support the hypothesis that the late effects on targets and competitors can be attributed to eye-movement processes that are unrelated to the lexical properties of the stimuli. It is still possible that language-impaired listeners differ in the way the distribution of activations across the lexicon maps onto the distribution of fixations (the “linking function” we describe later). There has been little work examining these linkages directly. While we cannot completely rule out such a model, it seems unlikely that language-impaired listeners have equivalent lexical activation processes to typically developing listeners, and differ only in the linking process, given that basic oculomotor processes appear similar, and given the previously discussed evidence for some sort of lexical deficit from studies using both open and closed-set responding (Dollaghan, 1998; Mainela-Arnold et al., 2008; Montgomery, 1999, 2002; Stark & Montgomery, 1995).

These analyses rule out variation in accuracy, response latency, and eye-movement ability (as well as associated averaging artifacts) as the cause of the effects we have observed. Thus this particular effect on fixations is most likely due to differences in language ability between participants. This is supported by modeling results we will present shortly which suggest that such differences can arise from several plausible instantiations of TRACE.

2.2.2.4. Fixation patterns: summary. Language status affected the asymptotic fixations to the target, cohort and rhyme. There was a small difference in slope, but Fig. 5A suggests that this difference was by no means the dominant component of the effect of language status. Adolescents with poor language skills activated the target less and took slightly longer to get to peak activation than TD adolescents. Similarly, effects on the cohort and rhyme competitors were also late—participants with poor language skills could not fully suppress the competitor and had higher offset baseline activation than normal adolescents.

3. Simulations

We next set out to model these results using variants of the TRACE model (McClelland & Elman, 1986). Our goals were twofold. The first was to use TRACE to validate the pattern of data we observed. Since the asymptotic differences we observed were unexpected, modeling could offer further support that this pattern derives from underlying language differences.

Our second goal was more far-reaching. Activation flow in TRACE is regulated in part by a number of free parameters which control the rate at which representations (words, phoneme and feature units) decay, the strength of lateral inhibition in each layer, the amount of noise in the input, and the strength of bottom-up activation flow. This provides an avenue for describing individual differences in terms of variants in processing dynamics.

Ideally, we could fit individual TRACE models to each subject's data and examine the relationship between specific parameter values and language and non-verbal IQ. However, TRACE has over 19 free parameters, and there is no efficient way to search the space in a reasonable amount of time. Moreover, this would be of limited theoretical value—the empirical data do not offer sufficient constraint; there may be multiple solutions for any subject; and we may end up fitting noise. It is also not clear if theory would keep pace with such a model. The specific collection of parameters values that we find may not be theoretically interpretable, so the modeling may not lead to any theoretical advance.

Nonetheless, the empirical domains we are relating (general language ability and word recognition) offer an alternative approach. General language ability is used to diagnose SLI, and there are a number of theoretical accounts of this class of individual differences. By instantiating these accounts in the TRACE model, we can at least rule out specific accounts of individual differences in language use. Thus, we instantiated twelve different versions of the model corresponding to six theoretical accounts of SLI. Our goal was to examine the variation in performance in each class of models to narrow the space of possible hypotheses.

3.1. Methods

We used the jTRACE implementation of TRACE (Strauss, Harris, & Magnuson, 2007) for all of the simulations. TRACE does not include all of the phonemes necessary for testing the entire set of items

used here. Thus, 14 analogous sets were selected that had similar degrees of overlap and deviations from rhymes and target (see [Appendix B](#)). A few liberties were taken in translating the limited set of TRACE phonemes to real words.

On each simulation, the model was exposed to a target word, and activation for the target, cohort competitor, rhyme, and an unrelated word was saved at each processing cycle. These activations are not directly reflected in the fixation proportions to each object. Rather, fixation patterns reflect activation only after it has been filtered through a linking function which converts activation across the lexicon to response probabilities. [Allopenna and colleagues \(1998\)](#) present one way of doing this by using the Luce-choice rule to convert activation to probability, a technique that was adopted here and in a number of prior studies (e.g. [McMurray, Tanenhaus, et al., 2009](#); [Dahan, Magnuson, Tanenhaus, 2001](#)). This normalizes the activation for each item by the sum of the activation for the visible items (after transforming them using the exponential function) to yield the relative probability of fixating each object at that time.

Allopenna and colleagues further scale these probabilities as a function of overall activation to allow for the fact that participants initially make few fixations to any object. We discovered irregularities in their scaling function that could give rise to apparent individual differences (or mask them) in TRACE. Thus, we modified this component of the function to use a more straightforward scaling factor, and tested several other ways of computing the scaling factors (see [Appendix C](#), and online note 3 for details).

As a whole, the linking function we adopted has 9 parameters that control factors such as the mapping of TRACE frames onto time in milliseconds (a linear mapping), the scaling of probabilities over time, and the scaling of the temperature of the Luce-choice function over time (temperature controls how peaky the resulting distribution of probabilities is). We wanted to explore the role of differences in activation processes in predicting fixations in this task, and our analysis of basic oculomotor properties did not suggest any obvious differences that co-varied with language abilities. While it is possible that participants differ in the way they link lexical activation to eye-movements, the evidence for a lexical deficit in other tasks suggests there is something to explain at this level. More importantly a deficit in this linking function has not been proposed in the literature, and our modeling attempted to reduce the degrees of freedom by restricting our analysis intended to address existing hypotheses for SLI. Thus, we held this function and its parameters constant across all participants. These parameters were found by minimizing the difference between the default activations from TRACE and the average of the non-language-impaired participants using a brute-force search of the parameter space.

This led to remarkably good fits for the average performance of the non-language-impaired participants ([Fig. 8](#)). When the predicted fixations (to the target, cohort, rhyme and unrelated objects) were compared to the empirical data we observed an $R^2 = .998$, and an RMS error of .0123 (by comparison, [Allopenna and colleagues](#) report an R^2 of .95 and an RMS of .03).

To examine which variants of TRACE could best account for the variation we observed empirically, we used the following procedure. We first identified a hypothetical cause of the variation and its implementation in TRACE. Most hypotheses could be implemented by manipulating a single parameter (e.g. phoneme inhibition to manipulate categorical perception), but for others multiple were manipulated simultaneously. We then selected a number of evenly spaced steps, both above and below the default value.

The number of steps and the size were different for different parameters, reflecting their natural scaling (see [Table 4](#)). For example, after lexical inhibition reached .05, further changes in this value resulted in small changes in activation, so we stopped at .07. Input noise could scale safely up to .9, but after that the model failed to settle on the correct word (and since all of the participants were well above 95% correct, such values were not relevant). Thus, after examining the resulting pattern of activation for a model variant, more steps were added to either increase the range (if that was possible) or to make finer-grained distinctions within the range.

After generating a reasonable range of variation in the parameter values, the output from each variant of the model was transformed into fixation proportions for comparison to the empirical data. Finally, across the range of model outputs for a given dimension we identified the single parameter value that best fit *each subject*, and computed the RMS difference between that model and the subject's data. The sum of these RMS errors, then, is the potential fit of the TRACE model if that parameter is the underlying source of the variance.

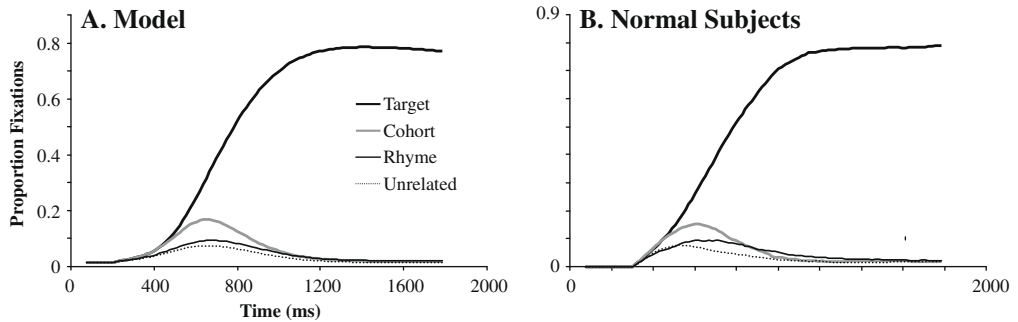


Fig. 8. Model fit. (A) Predicted fixations to the target, cohort, rhyme and unrelated objects as a function of time after the linking hypothesis had been fit to the non-language-impaired participants. (B) The corresponding empirical data.

After using this procedure to narrow down the space of possible models, we further considered two factors. First, we examined the qualitative pattern of the activation dynamics to verify that they showed the right type of effects. Second, we examined the fit of the model to the average fixations

Table 4

Parameters manipulated in TRACE.

Hypothesis	Parameter	Default	Range	Steps	Description	Theoretical construct
Lexical	Lexical-activation rate	.05	.01–.07	13	Rate that activation in lexicon builds on basis of phonemic input	Ability to activate spoken words from phonemic input; lexical access
	Lexical inhibition	.03	.005–.055	11	Strength of inhibition between words	Competition between words; neighborhood dynamics
	Lexical decay	.05	.01–.16	16	Decay of lexical activation over time	Ability to maintain words in memory
Perceptual	Input noise	0	0–.9	10	Noise in input	Noisy or inaccurate auditory processing
	Feature spread	6	2–10	9	Spread of feature representation over time	Sparseness or overlap of auditory information in sensory memory
	Feature decay	.01	0–.055	12	Decay of feature unit activation	Ability to retain representation of acoustic input after it has been heard
Phoneme	Phoneme-activation rate	.02	.0025–.035	12	Rate that phonemic activation builds in response to feature input	Ability to activate phonological representations from input
	Phoneme inhibition	.04	.005–.1	12	Strength of inhibition between features	The sharpness or gradiency of phonetic categories, categorical perception
	Phoneme decay	.03	.02–.1	9	Decay of phoneme unit activation	Ability to maintain phonological representation in memory
Vocabulary	Lexical size	100%	25–100%	4	Number of words known	Size of lexicon
Generalized slowing	Phoneme- and lexical-activation rate	.02 .05	.005–.035 .0125–.0875	10	Rate of activation accumulation at all levels	Slowing
Generalized inhibition	Feature phoneme and lexical	.04 .04	.01–.07 .01–.07	12	Inhibition between representations at all levels	A general deficit in inhibitory processes
	Inhibition	.03	.0075–.0525			

pattern of the language-impaired participants. While this necessarily collapses across a lot of interesting variation, it can help isolate only those differences that are related to functional language ability, and since TRACE was originally intended to model group data, may be a better application. We largely focus here on the individual results (but see online note 2.1 for a complete description of the qualitative and group fits).

3.1.1. Hypotheses tested

Given the vast parameter space of TRACE, our goal was to determine limited number of parameters to consider based on the existing literature. We thus examined 12 different variants of TRACE, corresponding to six theoretical accounts of language impairment. They will each be described here before we describe the results.

3.1.1.1. Lexical processes. Our first fits examined processes that control lexical activation. These were intended to ask if TRACE could show asymptotic behavior at all, since lexical processes seemed most proximal to the behavior in question. Three models were considered. The first modeled variation in language ability as variation in the rate that words accumulate activation, under the assumption that if words accumulated activation more slowly this could account for the slope effects on the target as well as the reduction in activation overall. This was partially successful and motivated us to examine two other lexical processes: the degree to which words inhibit each other; and the rate that lexical activation decays.

3.1.1.2. Perceptual processes. Our second set of hypotheses examined perceptual processes. Work on SLI suggests that low-level auditory deficits may be a source of the impairment (Tallal & Piercy, 1974) and numerous studies have assessed this (Bishop et al., 2005; Rosen, 2003; Stark & Heinz, 1996; Tallal & Piercy, 1974; Tallal et al., 1996). Sensory encoding in TRACE can be affected by a number of parameters that control the behavior of the feature units.

The simplest manipulation is to add noise to the input, something which was implemented in a later version of TRACE (McClelland, 1991). A second manipulation is the temporal spread of the input. Normally in TRACE, the feature inputs gradually ramp up to their maximum value over some number of cycles and then ramp down to zero. The feature-spread parameter controls the timing of this. If it is shorter than the default, the model is forced to rely on its maintenance of activation at feature and phoneme levels, since there will be many frames for which there is no input. If it is higher than the default, features from different points in time will overlap, creating problems in maintaining the sequential nature of the inputs. The final perceptual parameter we manipulated was feature-decay. This controls the rate at which activation values for features decay and is analogous to sensory memory of some kind.

3.1.1.3. Phonological processes. A related hypothesis is that language impairment arises from a deficit in perceiving and/or retaining phonological input. First, deficits in phonological perception have been suggested by studies showing abnormal categorical perception in children with SLI (Joanisse & Seidenberg, 1998; Sussman, 1993; Thibodeau & Sussman, 1979), although later work has failed to find these differences (Coady, Evans, Mainela-Arnold, & Kluender, 2007; Coady, Kluender, & Evans, 2005). Categorical perception in TRACE is produced by inhibition between phonemic representations (McClelland & Elman, 1986; McMurray, Aslin, & Toscano, 2009). Thus, reducing the inhibition can directly test of this hypothesis.

Second, a number of studies have shown that children with SLI have poor phonological working memory (see for instance, Bishop, 2006; Conti-Ramsden & Hesketh, 2003; Gathercole & Baddeley, 1990). While TRACE has no explicit phonological working memory, we hypothesized that any such store would be based on the activation at the phoneme layer. Thus, we can simulate a deficit in phonological working memory by varying the decay parameter at the phoneme level, thus varying the model's ability to retain a phonemic code of the input.

Across this work, there has been discussion that phonological representations are simply "less robust" or more fragile in language-impaired individuals. While it is unclear exactly what this means, a simple way to model it would be to reduce the rate that activation at the phoneme level builds, by

reducing the FP-ACT parameter. This would result in slower activation growth and less stable representations (e.g. more susceptible to noise, feedback effects, or inhibition).

Decreasing this parameter also tests a second hypothesis. A number of studies model phonological categories with a Gaussian distribution of cue-values (e.g. McMurray, Aslin, et al., 2009; Vallabha, McClelland, Pons, Werker, & Amano, 2007). If SLI listeners had noisier representations, this would be modeled as the phonemes having wider distributions (e.g. the variance). At the prototypical value of the cue (the only relevant value, since the stimuli here are unambiguous) this would reduce the likelihood of the category. Thus, decreasing the rate of phoneme activation (which decreases the functional connection strength between the prototype feature value and the phoneme) allows a rough approximation of the effects of noisy phonological categories (at least at prototype values).

3.1.1.4. Vocabulary size. Children with language impairment have difficulty learning words and will thus have poorer vocabularies (McGregor, Friedman, et al., 2002; McGregor, Newman, et al., 2002; Tomblin & Zhang, 1999). They may simply know fewer words than normal children, and this in turn may account for differential phonological or lexical processing abilities (e.g. Beckman, Munson, & Edwards, 2007). We asked if this could give rise to our empirical effects by creating new lexica for TRACE. The default lexicon included 264 words (including the 56 words needed for testing). We constructed new lexica by randomly cutting 25%, 50% or 75% of these words prior to testing. Each run of the model had its own unique lexicon, and none of the test words were eliminated.

3.1.1.5. Generalized slowing. Kail and colleagues (Kail, 1994; Miller, Kail, Leonard, & Tomblin, 2001) have proposed that listeners with SLI are slower processors in a range of linguistic and non-linguistic domains. While the generalized slowing hypothesis does not offer a particular mechanism, it can be implemented in TRACE via activation rate. At all levels of TRACE, activation accumulates at a rate determined by the phoneme- and lexical-activation rate parameters. While generalized slowing would clearly have effects outside of word recognition, by simultaneously reducing these parameters we can slow the growth of activation in the model. Since each has different default values, this was implemented by setting each to a percentage of their defaults, ranging from 25% to 175% of their default values.

3.1.1.6. Generalized inhibition. Finally, there are suggestions in the literature that a deficit in inhibitory processes may be relevant to individual differences in language use, though generally at higher levels of processing (Gernsbacher & Faust, 1991; Mainela-Arnold et al., 2008; Norbury, 2005). Thus, we also examined the possibility that broad-based deficits in inhibition of some kind were responsible. To do this we started at the default values for feature, phoneme and lexical inhibition, and varied each as a percentage of the default, ranging from 25% to 175% to simulate a model in which inhibition as a whole was reduced or increased.

3.2. Results

Fig. 9A shows the average RMS for each of the models fit to individual participants. Table 5 shows the same data numerically along with the group fits to the language impaired data.

Given the individual data, three models emerged as candidates for describing the individual differences observed here. Lexical decay was the best fit, with an average RMS of .0405. The actual decay value that was optimal for each subject was significantly correlated with the language scores of these children ($R = -.273$, $p = .01$), but not with non-verbal intelligence ($R = -.193$, $p = .072$). Participants at the lower end of the language scale showed faster decay of lexical activation ($M = .071$) than typically developing participants ($M = .047$; default = .05).

The rate that phonemes accumulated activation was a close second with an RMS of .0406. Participants with lower language scores were best fit by this lower values of this parameter ($R = .209$, $p = .0504$), though this was a significantly smaller effect than what was observed for lexical decay ($T(90) = 2.0$, $p = .048$, by Cohen & Cohen's (1983) method).

Generalized slowing was also a good fit, with an RMS of .0412. Participants with poor language were best fit by models with slower rates of activation ($R = .227$, $p = .033$). This was smaller, but not significantly different than the correlation with decay ($T(90) = 1.35$, $p = .18$). It was not surprising

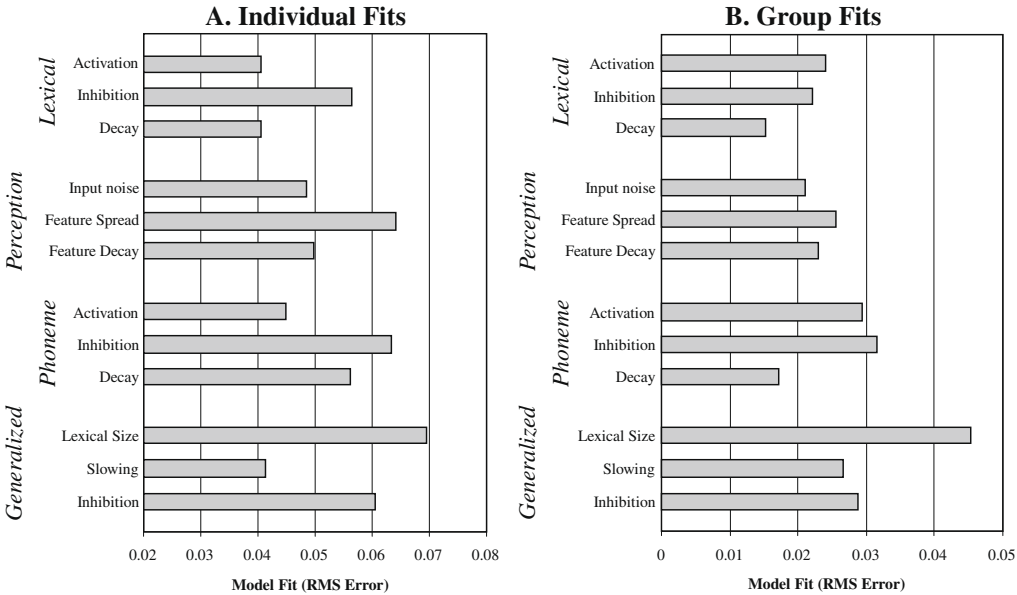


Fig. 9. A summary of the results of the simulations. Shown is RMS error (smaller error means better fits) for each simulation when fit to individual participants (A) or the mean of the language-impaired group (B). See Table 5 for numerical values.

the generalized slowing performed similarly to lexical-activation rate, since this parameter was included in the slowing manipulation.

The next best fit was phoneme-activation rate, FP-ACT (RMS = .0448, the other half of generalized slowing), followed by input noise (RMS = .0485), and feature decay (RMS = .0497). Beyond that, the remaining parameters did not offer good fits.

Table 5
Summary of model fits.

Hypothesis	RMS (individual)	RMS (group)	Correlation w/ language	Li models were
Default (N)	–	.012	–	
Default (Li)	–	.045	–	
<i>Lexical</i>				
Lexical activation	.0405	.030	.21 ⁺	Decreased
Lexical inhibition	.0563	.022	.24 ⁺	Decreased
Lexical decay	.0406	.015	–.27 ^{**}	Increased
<i>Perceptual</i>				
Input noise	.0485	.021	–.22 ⁺	Increased
Feature spread	.0640	.025	.28 ^{**}	Extremes
Feature decay	.0497	.023	–.25 ⁺	Increased
<i>Phonological</i>				
Phoneme inhibition	.0633	.032	–.13	No change
Phoneme decay	.0561	.017	–.18 ⁺	Increased
Phoneme activation	.0448	.030	.19 ⁺	Decreased
<i>Global</i>				
Lexicon size	.0695	.045	.01	No change
Generalized slowing	.0412	.027	.23 ⁺	Slower
Generalized inhibition	.0605	.029	.29 ⁺	Increased

⁺ $p < .1$.
^{*} $p < .05$.
^{**} $p < .01$.

To distinguish these best fitting parameters, we started by examining each qualitatively. Fig. 10 shows results from the top three model variants: lexical decay, lexical-activation rate and generalized slowing (for raw activations, see online note 2.2). Panels A through C show the effects of lexical decay on the predicted fixations to the target, cohort, and rhyme. All three show a close match to the group data depicted in Fig. 5. At higher levels of decay, the target shows a decreased asymptote, and the cohort and rhyme show increased fixations. Panels D–G show the target and cohort for the lexical activation rate and generalized slowing models (the results for the rhyme are similar to the cohort – see online note 2.1). Both show decreased asymptotic fixations to the target, and an increased asymptote for the cohort. Thus, they also capture our dominant effects. However, they also show a delay in the target fixations and a reduction in peak fixations to the competitors, neither of which was observed in the group data.

In fact, if we compare the fits of each of these three hypothesis to the group data, the lexical decay performs extremely well with an RMS of .0153, while lexical-activation rate and slowing perform much poorer (Lexical Activation: RMS = .024; Slowing: RMS = .026). In fact, in the group fits (Fig. 9B), decay was still the top ranked parameter, while lexical-activation rate moved down to 6th, and generalized slowing to 8th.

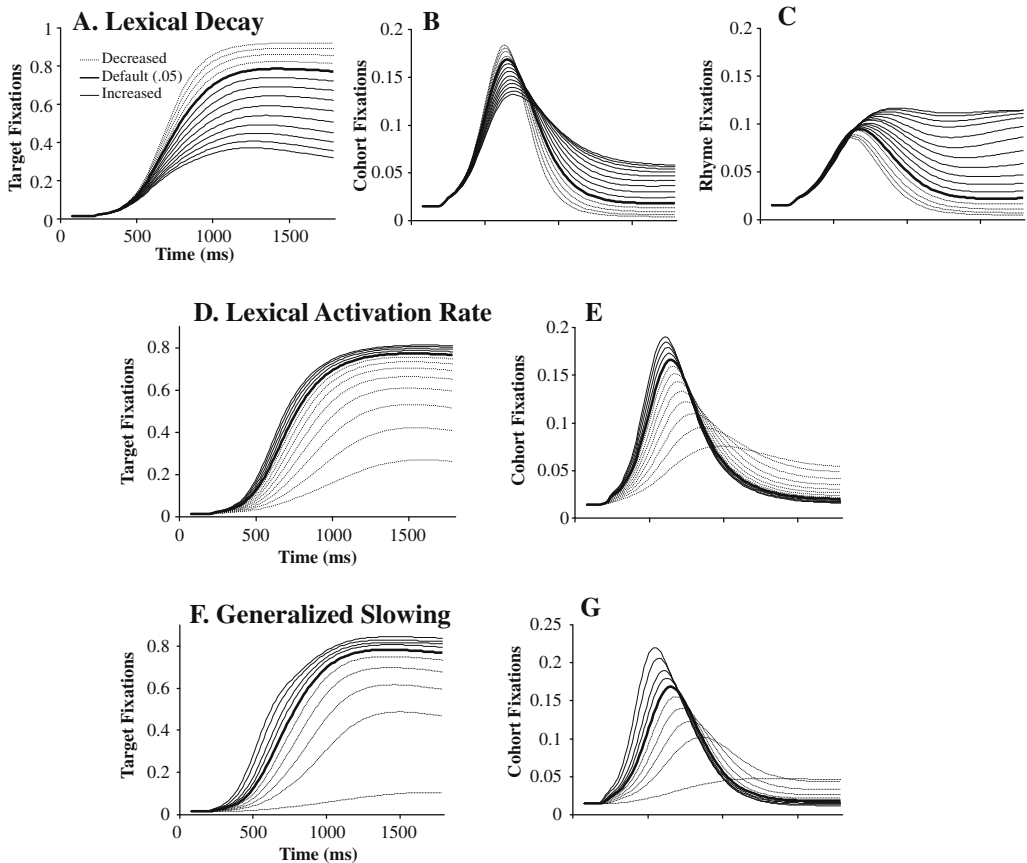


Fig. 10. Results from the top three fitting models. Note the default value of each parameter is shown in the heavy lines; dashed lines show lower values; and thin solid lines show values higher than the default. The top row shows results from variation in lexical decay. (A) Predicted looks to the target. (B) Cohort. (C) Rhymes. The second row shows results from varying the rate at which phonemes acquire activation. Note that rhymes were similar and are not displayed. (D) Predicted fixations to the target. (E) Cohort. The third row shows results from manipulations of lexical inhibition. (F) Predicted looks to the target. (G) Cohort.

So what accounts for this disparity? The reason for the good fits individually was simply that most of the parameters tested did not affect asymptotic behavior at all (see online note 2.1). Of those that did, some could only model a restricted range: lexical inhibition, for example, could get the target looks down to 0.5, but increasing it further had little effect. Thus, there was substantial range in the asymptotes in the empirical data, which lexical-activation rate and generalized slowing could capture (though at the cost of missing the timing).

Moreover, while for most participants, low rates of fixation to the target meant more looks to the cohort and rhyme, a handful of participants showed low rates of target looking with fewer looks to the cohort and rhyme. Lexical activation and generalized slowing both exhibited this effect at the extremes (decay did not), allowing them to model these participants well. In fact, when these six participants were excluded, the RMS of lexical decay (for individuals) dropped substantially (.0386) while the other two were unchanged (Lexical-Activation Rate: .0404; Slowing: .0406), suggesting that it was the failure of decay to account for these participants that resulted in nearly equal performance in the individual fits. This may have been due to strategic looking (these participants did not make many eye movements).

Finally, some participants varied in the cross-over points of the target curves ($SD = 69$ ms; Range = 477–910 ms), though this was not correlated with language status. General slowing and phoneme activation were picking up on this and modeling it successfully (which explains the weaker correlations with language ability), suggesting that these may be important differences in this aspect of word recognition (though differences that may not bear directly on functional language abilities).

Thus, there may be meaningful variation in activation rate, but variation in lexical decay is a much better descriptor of the individual differences in word recognition that are related to functional language ability as a whole. As Table 5 shows, no parameter other than decay fit the data well at both the level of individuals and groups.

This makes a strong case for a deficit in lexical decay, but in the spirit of hypothesis testing, a handful of failures were notable. Lexical size was the worst fit in both individual ($RMS = .0695$) and group ($RMS = .0453$) analyses. This was because variation in lexical size did not affect the time-course of lexical activation (Fig. 11A). Similarly, phoneme inhibition performed poorly in both individual ($RMS = .0633$, Rank = 10) and group ($RMS = .0316$; Rank = 11) analyses. While it did show asymptotic differences (Fig. 11B), this was only at very extreme values, and the range was not big enough to account for either group or individual differences.

Lexical inhibition was promising in qualitative analyses as it showed the observed effects (Fig. 11C and D), and good group level fits: $RMS = .0221$, Rank = 4). However, it reached ceiling near the default value (and could not show higher than average fixations). Even when inhibition was almost eliminated, target fixations could not get low enough to model all the subjects. As a result, it did not cover much of the range of the empirical data and showed poor fits individually.

Finally, as a whole, perceptual level parameters did not do a good job. Feature spread had little effect on activation, and mostly delayed it (Fig. 11E), and feature decay (Fig. 11F) caused a decrease in the target fixations at the end. Thus, parameters relating to the temporal organization of the input did not capture the empirical pattern of variance. Input noise, was perhaps the best of the perceptual parameters. It did affect asymptotic looks to the target (Fig. 11G). However, similar to lexical inhibition, it could not model individuals with greater than average target fixations – perhaps a default of 0 is inappropriate for this parameter. More importantly, increasing noise reduced the early activation for the cohort and made the peak later (Fig. 11F)—in some ways, the cohort was behaving like a rhyme.

Thus, it appears that lexical parameters, in particular lexical decay, offer the best fit to the pattern of individual differences observed here. However, there are two remaining issues that must be clarified related to our linking function.

First, it is unclear if the asymptotic performance of the lexical decay model could arise from the linking function, rather than the underlying activation. While our modifications of the linking hypothesis (Appendix C; online note 3) were intended to preserve transparency between activation and fixations, it is nonetheless possible that the linking function imposes this pattern of results (though as discussed in online note 3.3, all variants of the linking function show the same support for lexical

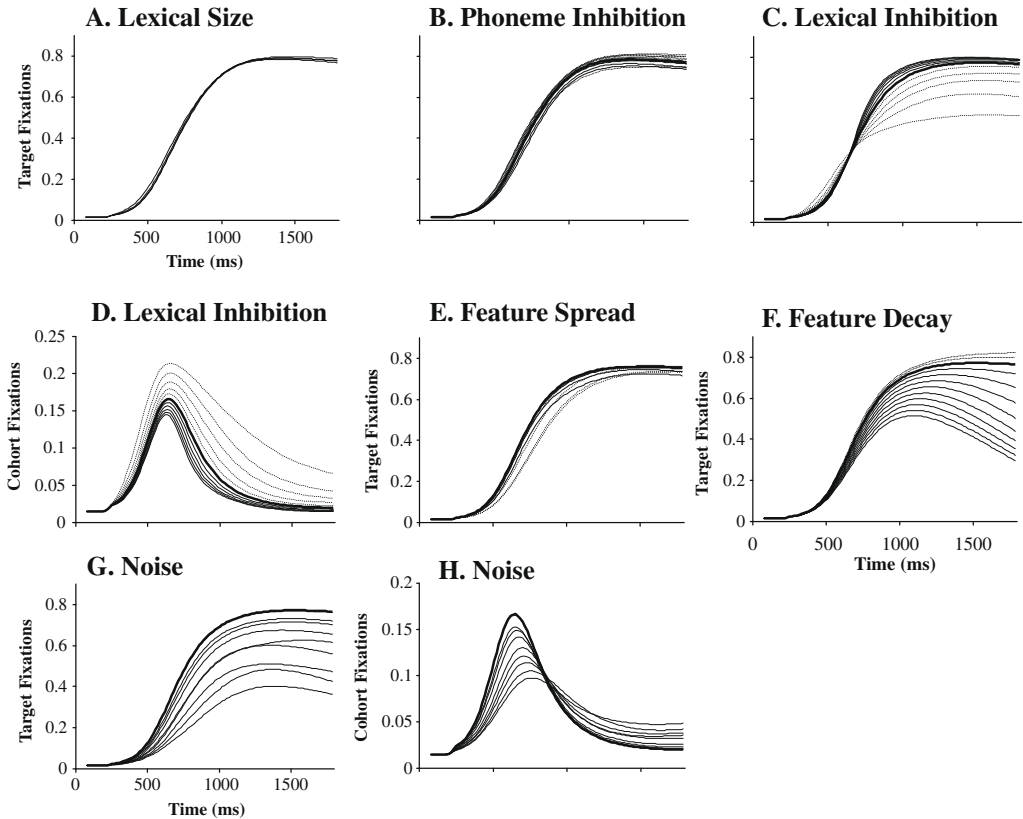


Fig. 11. Representative results from simulations that failed to fit the data. (A) Vocabulary size had almost no discernable effect on lexical activation. (B) Phoneme inhibition could only create a small range of values. (C and D) Lexical inhibition created the right pattern, but could not reach the full range of values necessary to account for the data. (E) Feature spread affected largely the timing of the function. (F) Feature decay caused target activation to decrease. (G and H) Input noise created asymptotic differences in the target and cohort, but also shifted the peak fixations to the cohort.

decay). If so, this would suggest that the mapping between activation and fixations, not differences in activation, are the fundamental individual difference here.

Fig. 12 suggests that this is not the case. It shows the raw activations as a function of time for the lexical decay model. These preserve the qualitative pattern seen in the fixations (and the modeling)—increasing decay leads to lower asymptotes for the target, higher asymptotes for the cohort and relatively little effect early. Online note 2.2 discusses raw activations for all of the parameters examined—none show a similar pattern to decay.

Second, for the language-impaired participants there was a disparity between the excellent mouse-clicking performance, and the significantly reduced lexical activation indicated by the eye movements. This too can be handled by our modeling framework, with the simple assumption that the transformation of activation to mouse-click probabilities uses a higher temperature coefficient (e.g. how “winner-take-all” it is), than the same transformation from activation to fixation probabilities (see online note 3.2).

Thus, the modeling supports variation in lexical processes (as opposed to perceptual or phonological) as the source of the deficits observed here, and most likely, variation in lexical decay. Four points are worthy of further consideration. First, we have tested only twelve possibilities, most of them using single parameters. Multiple parameter fits may better account for this behavior. However, we have pushed the modeling as far as available theories will permit – without clearer theoretical direction, further searches would easily turn into a fishing expedition.

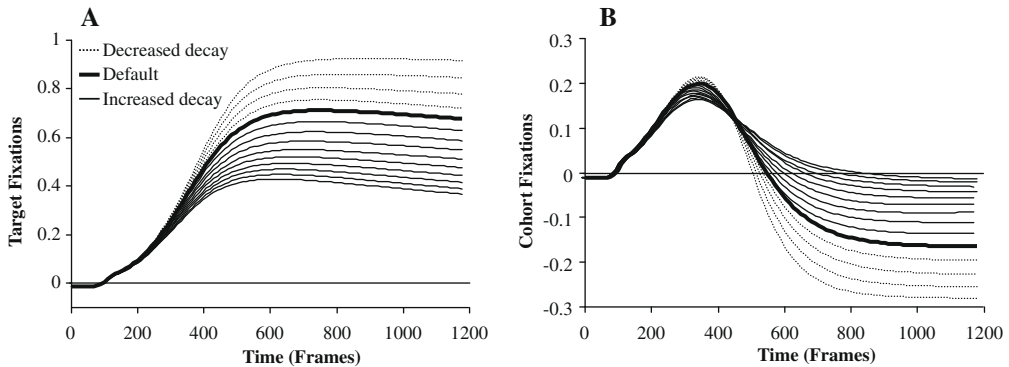


Fig. 12. Raw activations for manipulations of lexical decay. (A) Target. (B) Cohort.

Second, we assume the specific dynamics of TRACE. However, it is important to point out that most of the available models incorporating temporal dynamics at this scale rely on similar interactive activation dynamics (Merge: Norris et al., 2000; Shortlist: Norris, 1994; and Parsyn: Luce et al., 2000; though see Magnuson, McMurray, et al., 2003), and TRACE has modeled a large range of phenomena in word recognition and speech perception. It is not clear that there is an alternative more widely applicable in its scope.

Third, our results are conditioned in part on our linking hypothesis. As we discuss in Appendix C, we have made several improvements over the original Allopenna and colleagues (1998) framework, and tried several variants with similar results (online note 3.3). We have also seen similar results in raw activation. Thus, it seems unlikely that our fits are an artifact of these assumptions. Nonetheless, a better understanding of how lexical activation maps onto visual-motor behavior could improve this undertaking.

Finally, we have only used a single behavioral paradigm as the criteria for model fitting. Given the robust findings here, and the clear results from the simulations, this is not a terribly limiting factor. However, the VWP offers an array of tasks that have been mapped onto some of the same processing constructs, for example categorical perception (McMurray, Aslin, et al., 2008; McMurray, Clayards, et al., 2008; McMurray et al., 2002), frequency effects (Dahan, Magnuson, Tanenhaus, et al., 2001), competition effects (Dahan, Magnuson, Tanenhaus, 2001), cohort and neighborhood density (Magnuson, McMurray, et al., 2003; Magnuson, Tanenhaus, et al., 2003; Magnuson et al., 2007) and lexical ambiguity resolution (McMurray, Tanenhaus, et al., 2009). Future work of this kind could provide additional constraint.

4. General discussion

The behavioral data presented here suggest that variation in overall language ability can be associated with very specific changes in word recognition. These effects can be clearly dissociated from non-verbal intelligence measures. The standardized language measures we used in this study to characterize individual differences are broad and constructed from many language skills. Indeed, at the outset of this study, it was possible that variation in such a broad measure of language may have resulted in a variety of changes in the fixation patterns across participants. Among those that performed poorly in overall language, some participants could have been slower, others could have shown less peak activation, and others a delay in recognition. While there was undoubtedly variation of this type embedded in our dataset, systematicities also emerged. Listeners with poor language do not fully activate the target word and show increased activation for cohort and rhyme competitors.

The largest effects of language ability were late in the time-course. While it is possible that these effects reflect eye-movement or attention processes, this appears to not be the case for three reasons. First, it is not clear why this should be related to language and not non-verbal IQ. Second, analyses of the number of fixations and the looks to the unrelated object did not support any simple version of this hypothesis. Finally, we were able to model this behavior using variants of TRACE, in which only

the dynamics of lexical activation were manipulated (not attention or oculomotor control). Even the simplest model of individual differences (the addition of noise) can yield late effects (though it was not the best fit). Such late effects are also supported by a number of studies that converge on the idea that significant processing occurs after sufficient acoustic material has arrived to unambiguously identify a target (Dahan & Gaskell, 2007; Luce & Cluff, 1998; McLennan and Luce, 2005). Thus, it seems reasonable that individual differences might appear late in word recognition.

This also fits with the emerging picture in work on language impairment. Gating studies have shown that language-impaired children require additional gates after the uniqueness point to recognize a word (Dollaghan, 1998) or exhibit increased vacillation after the uniqueness point (Mainela-Arnold et al., 2008). Most importantly, Edwards and Lahey (1996; see also Lahey et al., 2001) found that while children with SLI were slower to respond in lexical decision tasks, their speed was not correlated with the severity of the impairment. This is easier to account for if language ability affects the overall amount of activation, not the timing.

Simulations examined the underpinning of these behavioral effects by testing 12 variants of the TRACE model corresponding to major theoretical approaches to language impairment. We found no support for different vocabulary sizes, and scattered support for a perceptual or phonological deficit. At the perceptual level, only input noise came close, showing similar qualitative changes as the behavioral data, a moderate group-wise fit, but poor individual subject fits. Adding noise to the input, however, is a far cry from the very specific auditory deficits that have been posited by the perceptual deficit hypothesis for language impairment. At the phonological level, we can safely rule out phoneme inhibition or categorical perception deficits, as well as explanations based on less robust phonological representations in general.

The fact that phoneme inhibition created very little variation in lexical processes raises an important issue. Phoneme inhibition was implemented in TRACE to force sublexical representations to arrive at a discrete representation of the input, categorical perception. However, it is now clear that categorical perception may be an artifact of the task (Carney, Widin, & Viemeister, 1977; Gerrits & Schouten, 2004; McMurray, Aslin, et al., 2008; Schouten, Gerrits, & Van Hoesen, 2003) and recent work suggests that it may prevent TRACE from modeling certain ambiguity resolution effects (McMurray, Aslin, et al., 2009). The present work suggests that eliminating or reducing inhibition does not have consequences in this simple task, and may not be necessary to account for individual differences related to overall language ability.

The rate that activation accumulated for words specifically, or in general, offered reasonable fits to individual participants. However, such variation did not capture the group effects, and may have been capitalizing on differences in word recognition (or oculomotor control) that do not ultimately affect overall language ability.

Lexical decay offered the best fit to the data. Activation for the target decayed faster in adolescents with language impairment. Decay plays an important role in the flexibility of the model (allowing it to move onto a new word when appropriate, and making it more susceptible to inhibition or new bottom-up input. By increasing the decay rate, the target becomes less stable, in effect decreasing its inhibitory power over the competitors. As a result, cohorts and rhymes have the opportunity to become more active (given bottom-up support). Thus, decay may be functionally intertwined with inhibition.

So what does this tell us about individual differences? First, the fact that lexical decay seemed to be the best fit to the data was surprising. A number of studies have examined other processes embedded in TRACE and related models; processes like inhibition (Dahan, Magnuson, Tanenhaus, et al., 2001; Frauenfelder et al., 2001), feedback (McClelland et al., 2006; Norris et al., 2000, for competing reviews), and the process of building activation on the basis of bottom-up input (Marslen-Wilson, 1987 for a review). However, few studies have examined decay as a component of word recognition. Given its potential importance in language processing as a whole, this seems fertile ground for new research.

Moreover, computational models of developmental dyslexia (e.g. Harm, McCandliss, & Seidenberg, 2003; Harm & Seidenberg, 1999) have also implicated decay as a source of individual differences. In these models decay was manipulated at the phonological (not lexical) level, a subtle difference from the present work. However, as distributed models, they did not have an explicit

lexical level, so this seems more an architectural, rather than theoretical difference. Nonetheless the parallels are striking, particularly given similarities between dyslexia and SLI (e.g. Bishop & Snowling, 2004).

Second, while this initial foray into individual differences implicates decay as an important dimension of variability, it does not rule out variation in other processes. More specific paradigms, like those mentioned above, may be needed to determine if other dimensions such as activation rate or inhibition vary as well. More importantly, however, variation in these dimensions did seem to be related to overall language ability to the same degree as decay, suggesting that higher level processes may be robust against these other forms of variation.

It was surprising that variability in overall language ability was not captured by perceptual or phonological factors. There undoubtedly is variation in these processes, but this study suggests that it is unrelated to overall language measures. This is reminiscent of the overall failure to find robust correlations between basic auditory abilities and language measures (e.g. Surprenant & Watson, 2001; Van Rooij & Plomp, 1990, 1992), and is clearly illustrated by children with Cochlear Implants, many of whom attain normal or near-normal language skills despite severely degraded input (Spencer, Barker, & Tomblin, 2003; Tomblin, Peng, Spencer, & Lu, 2008). Thus, the word recognition system must be fairly robust and capable of dealing with wide variation in perceptual ability either in real-time or developmental time.

Despite this robustness, however, there are important individual differences in lexical activation processes that are correlated with overall language. Decay seems the most plausible locus of such differences. Higher levels of lexical decay prevent the correct word from being fully active, and allow competitors to become more active than they should be. The result is that if we consider the array of activations output by word recognition as analogous to a vector of probabilities, individual differences in word recognition appear as variation in the entropy of the system. In skilled language the lexical activation vector is a low-entropy representation in which only a single word or a small number of words are highly active. Less-skilled users show greater entropy with more words active simultaneously, even at the end of processing.

In this analysis, the entropy of the lexical representation can easily be related to overall language ability. Current theories suggest that parsing and sentence comprehension are based on lexical (not syntactic) representations (Altmann & Kamide, 1999; MacDonald, Pearlmutter, & Seidenberg, 1994; MacRae, Spivey, & Tanenhaus, 1998; Tanenhaus & Trueswell, 1995; Trueswell, 1996). Moreover, syntactic and semantic processing do not wait for word recognition to finish before accessing potential interpretations (Dahan & Tanenhaus, 2004; Yee & Sedivy, 2006). As a result variation in how well the system settles on a single candidate will directly affect syntactic and semantic processes involved in sentence comprehension.

This cascade could also have consequences at lower levels of representation. There is evidence that continuous acoustic information survives perceptual processing to yield systematic effects on word recognition (Andruski, Blumstein, & Burton, 1994; McMurray, Aslin, et al., 2008b; McMurray et al., 2002); and evidence that lexical processes interact in real-time with perceptual processes (Magnuson, McMurray, et al., 2003; for a review, see McClelland et al., 2006; for an alternative, Norris et al., 2000). Lexical representations may even participate in the retention and organization of fine-grained detail (McMurray, Aslin, et al., 2008, 2009). Thus, low-level perceptual processes are not solely responsible for what we would consider speech perception, and a deficit in word recognition could easily create deficits in speech perception tasks.

5. Conclusions

Individual differences in spoken word recognition appear late in processing in the form of reduced activation for the target and increased activation for competitors. Such differences arise from differences in lexical processing and are robustly related to global language ability. These can be modeled by differences in the specific parameters that control activation flow in dynamic connectionist models like TRACE, and our work suggests lexical decay as a primary variable. Such differences seem to take the form of lower- or higher-entropy lexical representations—the degree to which word recognition can fully settle on a single interpretation.

Spoken word recognition is unparalleled in the consensus that has been built for basic processing mechanisms like immediacy, parallelism, competition and gradiency. These mechanisms only set the framework for what appears to be a complex set of individual differences that have ramifications throughout language processing.

Acknowledgments

The authors would like to thank Marcia St. Clair, Connie Ferguson, and Jaunita Limas for help with data collection; Marlea O'Brien and Dan McEchron for administrative support; Cheyenne Munson and Scott Spilger for technical assistance; and Jim Magnuson for making jTRACE available to the psycholinguistics community. We would also like to thank Joe Toscano and Keith Apfelbaum for grinding out the simulations and for discussion on the various linking hypotheses. Finally, we would like to extend our gratitude to Mike Tanenhaus and Arty Samuel for helpful comments on a previous draft and for pushing us to quantify model fit more precisely. This study was supported by NIH DC-02748 to JBT and DC-008089 to BM.

Appendix A. Items used in the experiment

Target	Cohort	Rhyme	Unrelated	Source
Dragon	Dracula	Wagon	Checkers	New
Turtle	Turkey	Hurdle	Cannon	New
Batter	Banjo	Platter	Monkey	New
Tower	Towel	Shower	Penguin	New
Funnel	Fungus	Tunnel	Window	New
Lizard	Liver	Wizard	Bottle	New
Hockey	Hotdog	Jockey	Blanket	New
Mustard	Mustache	Custard	Pencil	New
Mountain	Mousetrap	Fountain	Shutter	New
Beaver	Beehive	Cleaver	Soccer	New
Powder	Power	Chowder	Camel	New
Table	Tailor	Cable	Muffin	New
Beaker	Beetle	Speaker	Hammer	Allopenna et al.
Carrot	Carriage	Parrot	Building	Allopenna et al.
Candle	Candy	Handle	Button	Allopenna et al.
Pickle	Picture	Nickel	Robin	Allopenna et al.
Casket	Castle	Basket	Rocket	Allopenna et al.
Paddle	Padlock	Saddle	Waiter	Allopenna et al.
Dollar	Dolphin	Collar	Hamster	Allopenna et al.
Sandal	Sandwich	Candle	Necklace	Allopenna et al.
Road	Roll	Toad	Cake	New
Rose	Rope	Hose	Band	New
Coat	Comb	Goat	Badge	New
Bees	Beach	Peas	Cap	New
Rake	Rain	Lake	Toe	New
Plate	Plane	Gate	Dress	New
Lamb	Lamp	Ram	Bike	New
Snail	Snake	Tail	Web	New
Mouse	Mouth	House	Chain	New
Horn	Horse	Corn	Box	New
Ghost	Goal	Toast	Bag	New
Bell	Bed	Well	Can	New
Chips	Chin	Lips	Boat	New
Pier	Peach	Deer	Ring	New
Clown	Cloud	Gown	Pipe	New

(continued on next page)

Appendix A (continued)

Target	Cohort	Rhyme	Unrelated	Source
Bowl	Bone	Pole	Nest	New
Pen	Peg	Hen	Jam	New
Fish	Fin	Dish	Belt	New
Cat	Cab	Bat	Net	New
Bug	Bus	Rug	Cane	New
Bale	Bait	Pail	Nose	New

Appendix B. Items used in the TRACE simulations

Target	Cohort	Rhyme	Unrelated
Beaker (bik [^] r)	Beatle (bit [^] l)	Speaker (spik [^] r)	Target (targ [^] t)
Rake (raik)	race (rais)	Lake (laik)	Dust (d [^] st)
Pier (pir)	Peach (pitS)	Beer (bir)	Shock (Sak)
Bees (bis)	Beach (bitS)	Peas (pis)	Cap (kap)
Bug (b [^] g)	Bus (b [^] s)	Dug (d [^] g)	Cash (kaS)
Turkey (t [^] rki)	Turtle (t [^] rt [^] l)	Perky (p [^] rki)	Racket (rak [^] t)
Table (tab [^] l)	Tailor (tal [^] r)	Cable (kab [^] l)	Police (p [^] lis)
Pickle (pik [^] l)	Picture (piktur)	Tickle (tik [^] l)	Secret (sikr [^] t)
Chips (tSips)	Chick (tSik)	Ships (Sips)	Lock (lak)
Subtle (s [^] t [^] l)	Succeed (s [^] ksid)	Shuttle (S [^] t [^] l)	Product (prad [^] kt)
Guard (gard)	Got (gat)	Card (kard)	Leap (lip)
Dart (dart)	Dark (dark)	Tart (tart)	Greet (grit)
Legal (lig [^] l)	Least (list)	Regal (rig [^] l)	Pocket (pak [^] t)
Crash (kraS)	Creep (krip)	Trash (traS)	Plus (pl [^] s)

Appendix C. Implementation of the Luce-choice linking hypothesis

The output of TRACE is a set of lexical activations across the entire 256 word lexicon. However, our eye-tracking measure yields a probability of fixating a small set of four objects at any given time. This measure clearly taps underlying lexical activation, including the activation for non-displayed competitors (Allopenna et al., 1998; Dahan, Magnuson, Tanenhaus, et al., 2001; McMurray, Aslin, et al., 2009). However, it is not a direct measurement of activation as the pattern of fixations will be affected by oculomotor dynamics, and most importantly by the particular set of objects (and their activations).

Allopenna and colleagues (1998) first described a simple linking function to map activations from models like TRACE to the probability of fixating each of the four items on the screen on any given trial (at any given time).

In this procedure, the probability of fixating any given object is given by

$$P(\text{fixating object}_i) = \frac{e^{\tau \cdot a_i}}{\sum_{j=1}^4 e^{\tau \cdot a_j}} \quad (3)$$

Here, a_i refers to the activation of $word_i$, and τ is a temperature parameter (which will be discussed shortly). This equation divides the activation (transformed through the exponential function) by the sum of the activations of the four objects on the screen to yield a probability.

The temperature parameter controls how veridical this normalization is. At high values, the word with the maximal activation tends to assume all of the resulting probability (e.g. its probability is near one and the others are very low). At low values, words are more equal. Allopenna and colleagues discuss the advantages of using a temperature that gradually increases over the course of the trial, and suggest the logistic function for this purpose. This dynamic temperature parameter simulates a system that undergoes gradual pressure to settle on a single candidate. Our fits of the linking function found optimal performance if temperature had a baseline of 2, a peak of 4.5, a cross-over of 1000 ms and a

slope of .004. However, the findings reported here are not dependent on this instantiation of temperature. These simulations were replicated using a fixed τ and resulted in the same qualitative behavior of the model.

Finally, the Luce-choice rule given in (3) is a normalized probability. As a result, the total fixations at any given time can never be 0 (which is commonly seen in the eye-movement record at the beginning of the trial when participants are not fixating any of the objects). Allopenna and colleagues (1998) handle this by scaling the probabilities as a function of the maximum activation of any word over the whole time-course. This seemed an unsatisfactory way to achieve this scaling. This procedure assumes that the oculomotor dynamics at early points in time are somehow sensitive to activation at later points in time—that is, the scaling factor at time 1 will be a function of the activation at time 1 divided by the maximum activation (that likely occurred later). This seemed implausible. Moreover, an inadvertent result of their function is that if a manipulation of the model affected the maximum activation, it could change the pattern of predicted fixations at early time points. This created challenges in interpreting the resulting functions, as a difference late in time could cancel out (or enhance) results earlier in processing.

Allopenna and colleagues, of course, were not concerned with individual differences, and so this issue never arose. However, for our purposes it was crucial to have a linking function that mapped activation onto fixation probabilities as transparently and monotonically as possible. Thus, we adopted a simpler process to achieve this goal. We computed the maximum activation (across the four visible objects) at each point in time and passed it through a simple threshold – if the activation is high, then the scaling factor was high, if it was low, the scaling factor would be low. We tried several logistic functions and ultimately settled on the generalized logistic function (or Richard's curve) which is similar to the logistic function previously defined, but adds a skew parameter so that the function can be steeper at the bottom or top.

$$\text{SCALING} = \frac{p - b}{1 + w \cdot \exp(-s \cdot (c - \text{act}_{\max}))^{\frac{1}{w}}} + b \quad (4)$$

Here, act_{\max} is the maximum activation at the current time point; p and b are the peak and baseline as before; c is the cross-over point; and w is the skew. We held the baseline and peak constant at 0 and 1 in order to minimize the degrees of freedom and to ensure that activation differences in TRACE could use the full range of possible fixation probabilities. A brute-force search found optimal fits with a slope of 4 and a skew of .0001 (steeper slopes before the cross-over than after). Importantly the cross-over was at .25, meaning that if the maximum activation at that point in time was below .25, the model generates few eye movements; once activation crosses .25 we see more. This is simple and does not rely on anything outside of the current point in time. As [Supplementary material 2.2](#) shows, it also faithfully preserves the pattern of activation in the pattern of predicted fixations.

We also ran the simulations using the traditional scaling function; two logistics without the skew parameter; and with a function derived from the overall number of fixations made by individual participants. All showed similar results. However, the generalized logistic showed better model fits, more clearly reflected underlying activation differences, and seemed more theoretically plausible than prior approaches to this. While it does add an additional three free parameters, it is important to note that these were not varied between individuals, and so could not give rise to the effects we observed in the simulations.

Finally, in order to compute the RMS error we had to map processing cycles in TRACE onto time in milliseconds. This was done with the following linear mapping:

$$\text{Time} = 16 \cdot f + 75 \quad (5)$$

In which f represents the processing cycle number.

To sum up, first activations are computed into fixation probabilities using the equation given in (3), where T is defined by a logistic function. Then these probabilities are then multiplied by the scaling factor in (4) and time is rescaled according to (5). This yields a set of predicted fixation proportions that are directly comparable to the eye-movement record.

This introduces nine parameters (the four parameters of the logistic describing T , two parameters mapping processing cycles onto time, and three parameters for the scaling factor) which were estimated from the group data and held constant across all of the simulations. Importantly, we have only

added 3 to the Allopenna and colleagues procedure, and for good reasons. Nonetheless, it is important to note that the overall patterns seen here (e.g. changes in peak activation or rise time of the target) can all be seen in the patterns of raw activation – these free parameters only affect the quality of the fit to the eye movements.

Appendix D. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.cogpsych.2009.06.003](https://doi.org/10.1016/j.cogpsych.2009.06.003).

References

- Allopenna, P., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.
- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of relative reference. *Cognition*, 73, 247–264.
- Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–187.
- Arnold, J. E., Novick, J. M., Brown-Schmidt, S., & Trueswell, J. C. (2001). Children's on-line use of gender and order-of-mention for pronoun comprehension. In *Proceedings of the 25th annual Boston University conference on language development* (pp. 59–69).
- Beckman, M., Munson, B., & Edwards, J. (2007). Vocabulary growth and the developmental expansion of types of phonological knowledge. In J. S. Cole & J. Hualdo (Eds.), *Papers in laboratory phonology* (Vol. 9, pp. 241–264). New York: Mouton de Gruyter.
- Bishop, D. V. M. (2006). Beyond words: Phonological short-term memory and syntactic impairment in specific language impairment. *Applied Psycholinguistics*, 27, 545–547.
- Bishop, D. V. M., Adams, C. V., Nation, K., & Rosen, S. (2005). Perception of transient non-speech stimuli is normal in specific language impairment: Evidence from glide discrimination. *Applied Psycholinguistics*, 26, 175–194.
- Bishop, D. V. M., & Snowling, M. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, 130, 858–886.
- Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language mediated eye-movements. *Cognition*, 108(3), 896–904.
- Campana, E., Silverman, L., Tanenhaus, M., Bennetto, L., & Packard, S. (2005). Real-time integration of gesture and speech during reference resolution. In *Proceedings of the 27th meeting of the cognitive science society*.
- Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Non categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 62, 961–970.
- Coady, J., Kluender, K., & Evans, J. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research*.
- Coady, J. A., Evans, J. L., Mainela-Arnold, E. M., & Kluender, K. R. (2007). Children with specific language impairments perceive speech most categorically when it is both natural and meaningful. *Journal of Speech, Language, and Hearing Research*, 50, 41–57.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Connine, C., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193–210.
- Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: A study of young language-learning children. *International Journal of Language & Communication Disorders*, 38, 251–263.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Chicago, IL: Holt, Rinehart and Winston.
- Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57, 483–501.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317–367.
- Dahan, D., Magnuson, J. S., Tanenhaus, M., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5/6), 507–534.
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(2), 493–513.
- Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Phonological deficits in dyslexic children revealed by eyetacking. *Cognition*, 100, B32–B42.
- Dollaghan, C. (1985). Child meets word – Fast mapping-in preschool-children. *Journal of Speech and Hearing Research*, 28, 449–454.
- Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics*, 19, 193–207.
- Dunn, L. M. (1981). *Peabody picture vocabulary test-revised*. Circle Pines, MN: AGS.
- Edwards, J., & Lahey, M. (1996). Auditory lexical decisions of children with specific language impairment. *Journal of Speech and Hearing Research*, 39, 1263–1273.
- Fernald, A., Perfors, A., & Marchman, V. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116.

- Frauenfelder, U., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, 16(5–6), 583–607.
- Gaskell, M. G., Quinlan, P., Tamminen, J., & Cleland, A. A. (2008). The nature of phoneme representation in spoken word recognition. *Journal of Experimental Psychology: General*, 137, 282–302.
- Gaskell, M. G., & Marslen-Wilson, W. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–656.
- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29, 336–360.
- Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 245–262.
- Gerrits, E., & Schouten, M. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267–283.
- Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, 7, 155–182.
- Harm, M., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 491–528.
- Huetting, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), B23–B32.
- Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, 2, 240–247.
- Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, 86, 40–56.
- Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech and Hearing Research*, 37, 418–421.
- Lahey, M., Edwards, J., & Munson, B. (2001). Is processing speed related to severity of language impairment? *Journal of Speech, Language, and Hearing Research*, 44, 1354–1361.
- Lee, S. H., Samelson, V. M., & Tomblin, J. B. (2006). The feasibility of using a self-rated word familiarity survey with adolescents. Poster session presented at the annual convention of the American Speech, Language and Hearing Association, Miami.
- Leonard, L. B. (1987). Is specific language impairment a useful construct? In S. Rosenberg (Ed.), *Advances in applied psycholinguistics. Disorders of first-language development* (Vol. 1, pp. 1–39). New York, NY, US: Cambridge University Press.
- Leonard, L. B. (1991). Specific language impairment as a clinical category. *Language, Speech, and Hearing Services in Schools*, 22(2), 66–68.
- Leonard, L. B., Deevy, P., Miller, C. A., Charest, M., Kurtz, R., & Rauf, L. (2003). The use of grammatical morphemes reflecting aspect and modality by children with specific language impairment. *Journal of Child Language*, 30(4), 769–795.
- Leonard, L. B., Deevy, P., Miller, C. A., Rauf, L., & Charest, M. (2003). Surface forms and grammatical functions: Past tense and passive participle use by children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 46, 43–45.
- Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 40, 741–753.
- Leslie, L., & Caldwell, J. (1995). *The qualitative reading inventory – II*. New York: Longman.
- Leslie, L., & Caldwell, J. (2001). *Qualitative reading inventory – 3*. New York: Longman.
- Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, 60, 484–490.
- Luce, P. A., Goldinger, S. D., Auer, E., & Vitevitch, M. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62(3), 615–625.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- MacRae, K., Spivey, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Magnuson, J., Dixon, J., Tanenhaus, M., & Aslin, R. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31, 1–24.
- Magnuson, J., McMurray, B., Tanenhaus, M., & Aslin, R. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmas past. *Cognitive Science*, 27(2), 285–298.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 133(2), 202–227.
- Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2008). Lexical representations in children with SLI: Evidence from a frequency-manipulated gating task. *Journal of Speech, Language and Hearing Research*, 51, 381–393.
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25(1–2), 71–102.
- Marslen-Wilson, W., Moss, H. E., & Van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6), 1376–1392.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101(4), 653–675.
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585.
- McClelland, J. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, 23(1), 1–44.
- McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8), 363–369.

- McGregor, K., Friedman, R., Reilly, R., & Newman, R. (2002). Semantic representation in young children. *Journal of Speech, Language and Hearing Research*, 45, 332–346.
- McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 45, 998–1015.
- McLennan, C., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31(2), 306–321.
- McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., & Subik, D. (2008). Gradient sensitivity to within-category variation in speech: Implications for categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1609–1631.
- McMurray, B., Aslin, R. N., & Toscano, J. (2009). Statistical learning of phonetic categories: Computational insights and limitations. *Developmental Science*, 12(3), 369–378.
- McMurray, B., Clayards, M., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review*, 15(6), 1064–1071.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1), 65–91.
- McMurray, B., Tanenhaus, M., Aslin, R., & Spivey, M. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface. Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, 32(1), 77–97.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113–1126.
- Miller, C., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 44, 416–433.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494.
- Montgomery, J. (1999). Recognition of gated words by children with specific language impairment: An examination of lexical mapping. *Journal of Speech, Language and Hearing Research*, 42(3), 735–743.
- Montgomery, J. (2002). Examining the nature of processing in children with specific language impairment: Temporal processing or processing capacity deficit? *Applied Psycholinguistics*, 23, 447–470.
- Montgomery, J., & Leonard, L. (1998). Real-time inflectional processing by children with specific language impairment: Effects of phonetic substance. *Journal of Speech, Language, and Hearing Research*, 41(6), 1432–1443.
- Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children’s real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, 86(4), 314–329.
- Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, 90, 142–171.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234.
- Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395.
- Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Science*, 23(3), 299–370.
- Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, 113(1), 57–83.
- Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, 31(3–4), 509–527.
- Schouten, M., Gerrits, E., & Van Hoesen, A. J. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1), 71–80.
- Semel, E., Wiig, E., & Secord, W. (1995). *Clinical evaluation of language fundamentals – 3*. San Antonio, TX: The Psychological Corporation.
- Spencer, L. J., Barker, B. A., & Tomblin, J. B. (2003). Exploring the language and literacy outcomes of pediatric cochlear implant users. *Ear and Hearing*, 24, 236–247.
- Stark, R., & Heinz, J. (1996a). Perception of stop consonants in children with expressive and receptive-expressive language impairments. *Journal of Speech and Hearing Research*, 39, 676–686.
- Stark, R. E., & Montgomery, J. (1995). Sentence processing in language-impaired children under conditions of filtering and time compression. *Applied Psycholinguistics*, 16, 137–164.
- Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). JTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, Instruments & Computers*, 39(1), 19–30.
- Surprenant, A., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *Journal of the Acoustical Society of America*, 110, 2085–2095.
- Sussman, J. E. (1993). Perception of formant transition cues to place of articulation in children with language impairments. *Journal of Speech and Hearing Research*, 36, 1286–1299.
- Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., et al (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, 271, 81–84.
- Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, 12, 83–93.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Tanenhaus, M., & Trueswell, J. C. (1995). Sentence comprehension. In P. D. Eimas & J. L. Miller (Eds.), *Handbook in perception and cognition. Speech, language and communication* (Vol. 11, pp. 217–262). Academic Press.
- Thibodeau, L., & Sussman, H. (1979). Performance on a test of categorical perception of speech in normal and communication disordered children. *Journal of Phonetics*, 7, 375–391.

- Tomblin, J. B. (2005). Adolescent outcomes of developmental language disorder in kindergarten. Paper presented at the *symposium on research in child language disorders*, Madison, WI.
- Tomblin, J. B., Peng, S. C., Spencer, L. J., & Lu, N. (2008). Long-term trajectories of the development of speech sound production in pediatric cochlear implant recipients. *Journal of Speech, Language and Hearing Research*, 51(5), 1353–1368.
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, 40, 1245–1260.
- Tomblin, J. B., Records, N., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, 39, 1284–1294.
- Tomblin, J. B., & Zhang, X. (1999). Are children with SLI a unique group of language learners? In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders: Contributions to a new framework from the cognitive neurosciences* (pp. 361–382). Cambridge, MA: MIT Press.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language impairment: Four years after diagnosis. *Journal of Speech, Language and Hearing Research*, 46, 1283–1296.
- Tomblin, J. B., Zhang, X., Weiss, A. L., Catts, H., & Ellis-Weismer, S. (2004). Dimensions of individual differences in communication skills among primary grade children. In M. Rice & S. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 5376). Mahwah, NJ: Lawrence Erlbaum Associates.
- Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566–585.
- Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, 36, 417–427.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant directed speech. *Proceedings of the National Academy of Sciences*, 104, 13273–13278.
- van Rooij, J. C. G. M., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *The Journal of the Acoustical Society of America*, 88(6), 2611–2624.
- van Rooij, J. C. G. M., & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion. *The Journal of the Acoustical Society of America*, 91(2), 1028–1033.
- Wallace, G., & Hammill, D. (1994). *Comprehensive receptive and expressive vocabulary test*. Austin, TX: Pro-Ed.
- Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41(3), 262–275.
- Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence – Revised*. New York: Psychological Corporation.
- Yee, E., Blumstein, S., & Sedivy, J. (2004). The time course of lexical activation in Broca's and Wernicke's aphasia: Evidence from eye-movements. *Brain and Language*, 91(1), 62–63.
- Yee, E., Blumstein, S., & Sedivy, J. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of Cognitive Neuroscience*, 20(4), 1–21.
- Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(1), 1–14.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.