Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye-movements

Bob McMurray,
Dept. of Psychology, Dept. of Communication Sciences and Disorders and Delta Center, University of Iowa

Cheyenne Munson, and
Dept. of Psychology, University of Illinois

J. Bruce Tomblin
Dept. of Communication Sciences and Disorders, and Delta Center, University of Iowa

Abstract

Purpose—This study examined speech perception deficits associated with individual differences in language ability contrasting auditory, phonological or lexical accounts by asking if lexical competition is differentially sensitive to fine-grained acoustic variation.

Methods—74 adolescents with a range of language abilities (including 35 impaired) participated in an experiment based on McMurray, Tanenhaus and Aslin (2002). Participants heard tokens from six 9-step Voice Onset Time (VOT) continua spanning two words (*beach/peach, beak/peak*, etc), while viewing a screen containing pictures of those words and two unrelated objects. Participants selected the referent while eye-movements to each picture were monitored as a measure of lexical activation. Fixations were examined as a function of both VOT and language ability.

Results—Eye-movements were sensitive to within-category VOT differences: as VOT approached the boundary, listeners made more fixations to the competing word. This did not interact with language ability, suggesting that language impairment is not associated with differential auditory sensitivity or phonetic categorization. Listeners with poorer language skills showed heightened competitors fixations overall, suggesting a deficit in lexical processes.

Conclusions—Language impairment may be better characterized by a deficit in lexical competition (inability to suppress competing words), rather than differences phonological categorization or auditory abilities.

Recognizing spoken words is challenging. Listeners must cope with variability due to talker, dialect, speaking rate and coarticulation, in order to match a temporally unfolding acoustic signal to a lexicon containing tens of thousands of words. And they must do so in a few hundred milliseconds. It is commonly assumed that this happens effortlessly for most

Corresponding Author Bob McMurray, E11 SSH, Dept. of Psychology, University of Iowa, Iowa City, IA 52242, bob-mcmurray@uiowa.edu, 319-335-2408 (office), 319-335-0191 (fax).

people. But while this may characterize modal performance, subtle deficits in speech perception and spoken word recognition are common in developmental language impairment (LI).

Diagnoses of LI are typically based on global measures of receptive and expressive language across vocabulary, sentence use and discourse (Tomblin, Records, & Zhang, 1996) rather than deficits in speech perception or word recognition. However, impairments in these processes, particularly speech perception, have played prominent roles in accounts of the cause of the higher level language deficits that drive LI (Dollaghan, 1998; Joanisse & Seidenberg, 2003; McMurray, Samelson, Lee, & Tomblin, 2010; Montgomery, 2002; Stark & Montgomery, 1995; Sussman, 1993; Tallal & Piercy, 1974; Ziegler, Pech-Georgel, George, Alario, & Lorenzi, 2005). Such deficits could arise at several levels of processing: auditory encoding of phonetic cues, phonological categorization, or lexical access. It is currently unclear which level best describes the deficits shown by children with LI.

There has been considerable work examining this topic dating back at least 50 years (Eisenson, 1968; Lowe & Campbell, 1965; Tallal & Piercy, 1974). However, despite this extensive research, as we describe below, there has been little movement toward a uniform perspective. This lack of clarity is partially due to the inconsistent evidence for such deficits (Coady, Evans, Mainela-Arnold, & Kluender, 2007; Coady, Kluender, & Evans, 2005; McArthur & Bishop, 2004; Rosen, 2003) which may derive from methodological or interpretive issues. At a more fundamental level, however, we argue that research on this problem may have been impeded by the theoretical framing speech perception used to address the problem.

Much of the research on speech perception problems in LI has implicitly assumed a traditional theoretical framing of speech perception. In this model, a small number of critical acoustic cues (e.g., voice onset time, formant transitions, etc) are used to access discrete phonemes. These cues are continuous, but they contain considerable variation due to factors like talker and context, which is viewed as noise. Consequently, the speech perception system must strip away this noise to generate a categorical perceptual product that serves as a precursor to word recognition. These categories have distinct boundaries over which acoustic cue variation yields different phoneme percepts, and regions within categories where variation is treated as noise. As a result the phonemic representation is viewed as a discrete representation that no longer contains the acoustic fine structure of speech. After these categories are identified, lexical access proceeds as a subsequent stage based on this phonemic representation.

In this model, lexical activation and listening comprehension can be perturbed if low level processing of the speech signal incorrectly encodes cue values for these highly critical cues. Alternatively deficits could emerge if cue values are encoded correctly but mapped onto phonological categories poorly, as it is this representation that drives lexical selection. This type of model has driven much of the research concerned with low-level perceptual deficits as contributors to poor language learning. In particular, it has focused research on low-level processes such as frequency and/or temporal discrimination, temporal order etc. as these were viewed as necessary processes to building the phonological representation.

In contrast to this traditional perspective, a new view suggests that speech is based on a combination of many phonetic cues which are encoded in fine, continuous detail, and integrated directly with temporally unfolding lexical representations (Goldinger, 1998; McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray & Jongman, 2011; Nygaard, Sommers, & Pisoni, 1994; Pisoni, 1997; Port, 2007; Salverda, Dahan, & McQueen, 2003). In this model, it is not necessary to derive an abstract phonological representation before lexical access. Indeed lexical access and phonological categorization may occur simultaneously: continuous acoustic detail survives to affect lexical activation (Andruski, Blumstein, & Burton, 1994; McMurray, et al., 2002) and lexical dynamics feedback to affect perception (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland & Elman, 1986; McClelland, Mirman, & Holt, 2006). This changes what we think of as the functional goal of speech perception. Within this framework, speech perception directly serves and is embedded in complex lexical activation dynamics that are a key part of language comprehension; and low-level graded auditory information is used during lexical activation rather than being stripped away (Pisoni, 1997). However at the same time, this complex process enables category-like function to emerge from the graded information in the signal (Goldinger, 1998; Pierrehumbert, 2003) with functional properties like prototypes and fuzzy boundaries (J. L. Miller, 1997; Oden & Massaro, 1978).

In this account, impairments in speech perception could still arise from deficits in low level speech processing and if so, should be seen in the details of lexical activation since this information is driving this process. Alternatively, speech perception could be impaired if the mapping between these cues and words or categories is poorly defined such that functional phonemic categories are impaired. However, there may also be different framings of deficits at these levels: listeners could pay too much or too little attention to fine-grained detail. Moreover, as speech perception is embedded within lexical activation, the problem could lie in processes that drive and resolve lexical activation. Critically, the research supporting this model suggests that it is difficult to measure auditory and speech skills independently of their consequences for lexical processes (Gerrits & Schouten, 2004; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; Schouten, Gerrits, & Van Hessen, 2003), and that we must reevaluate the role of the perception of fine-grained cues, not in terms of discarding variance, but in terms of how listeners are sensitive to this information and use it.

The purpose of this study is to examine speech perception in children with and without LI through the lens of real-time spoken word recognition to identify whether speech perception deficits in LI arise from low-level auditory processing, deficits in the way these cues are mapped to categories or words, or the processes involved in the activating items in the mental lexicon. To do this, we build on eye-tracking work that has in part motivated this new paradigm (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; McMurray, Aslin, et al., 2008; McMurray, et al., 2002). In the rest of the introduction, we review the literature on speech perception in LI to frame it in terms of this new model of speech perception.

Auditory and Speech Perception Impairments in Language Impairment

Auditory Accounts

Auditory accounts suggest that LI derives from impairments in the auditory encoding of important speech cues or in maintaining their temporal order. Evidence for these deficits is mixed. Some studies report that listeners with LI struggle to discriminate non-speech sounds that are short, close together, or presented quickly (Neville, Coffey, Holcomb, & Tallal, 1993; Tallal, 1976; Tallal & Piercy, 1973, 1975; Wright et al., 1997). Researchers have also proposed deficits in specific types of auditory inputs (Corriveau, Pasquini, & Goswami, 2007; McArthur & Bishop, 2004; McArthur & Bishop, 2005; Tallal & Piercy, 1974). However, other studies have not supported an auditory deficit (Bishop, Adams, Nation, & Rosen, 2005; Bishop, Carlyon, Deeks, & Bishop, 1999; Corriveau, et al., 2007; Heltzer, Champlin, & Gillam, 1996; Norrelgen, Lacerda, & Forssberg, 2002; Stollman, van Velzen, Simkens, Snik, & Van den Broek, 2003), and attempts to study such deficits in the context of word recognition have not supported these hypotheses: artificially manipulating the speech input to instantiate these hypotheses does not differentially impair recognition (Montgomery, 2002; Stark & Montgomery, 1995), challenging the importance of such deficits to downstream language processes.

Some auditory accounts assume a disruption in some critical class of information (e.g., formant transitions for place of articulation). However speech perception is extremely robust: listeners can understand speech composed entirely of sine-waves (Remez, Rubin, Pisoni, & Carrell, 1981), constructed from broad-band noise, (Shannon, Zeng, Kamath, & Wygonski, 1995), or with only envelope cues (Smith, Delgutte, & Oxenham, 2002). This is likely because for every phonetic distinction, there are dozens of cues (e.g., Lisker, 1986; McMurray & Jongman, 2011). Given the robustness and redundancy of speech perception, then higher-level language may be relatively insensitive to differences in the perception of any specific cues.

Phonological Deficits and Categorical Perception

Phonological accounts of LI emphasize differences in the process of mapping acoustic cues to categories. Most studies examine this through variants of the categorical perception (CP) paradigm (Liberman, Harris, Hoffman, & Griffith, 1957; Repp, 1984). CP is established when participants exhibit a steep transition between phoneme categories, and their discrimination of sounds that cross this boundary is better than sounds within the category. By relating phoneme categories to discrimination, CP investigates the contribution of phonology to auditory processing. It was originally motivated by the idea that speech perception is a process of eliminating unnecessary [within-category] variability in the signal to support phoneme identification.

The many studies examining CP in LI report inconsistent results. Some studies report differences in identification (Shafer, Morr, Datta, Kurtzberg, & Schwartz, 2005; Sussman, 1993; Thibodeau & Sussman, 1979; Ziegler, et al., 2005) that others have not found (Coady, et al., 2005). Others find differences in only a subset of continua or conditions (Burlingame, Sussman, Gillam, & Hay, 2005; Coady, et al., 2007; Robertson, Joanisse, Desroches, & Ng,

2009). With respect to discrimination, some studies report heightened within-category discrimination (Coady, et al., 2007), others find only poor between-category discrimination (Coady, et al., 2005; Robertson, et al., 2009), and yet others find no differences (Robertson, et al., 2009; Shafer, et al., 2005; Sussman, 1993; Thibodeau & Sussman, 1979). Thus, there is no single profile of impaired CP, leading Coady et al. (2007) to suggest that differences across studies may derive from task and stimulus differences rather than differences in CP.

However, substantial work with typical listeners suggests that CP is not a valid measure of phonological processing: it is an artifact of task and does not describe typical speech perception. Discrimination tasks are memory intensive as listeners must remember multiple tokens to make a comparison. Listeners may minimize memory demands by discarding an auditory encoding in favor of something more categorical, and when less memory intensive discrimination tasks are used, there is little evidence for enhanced discrimination near the boundary (Carney, Widin, & Viemeister, 1977; Massaro & Cohen, 1983; Pisoni & Lazarus, 1974). . Consequently, group differences in CP could thus derive from differences in memory not perception. Similarly, discrimination tasks can be biased to emphasize categorical or continuous differences, and eliminating such biases can cause CP effects to disappear (Gerrits & Schouten, 2004; Schouten, et al., 2003). In this vein, the same/different, AX and AAXX tasks used in the studies of LI that measured discrimination are perhaps the most problematic as listeners can adopt any criterion they want for what counts as "different". As a result differences between groups may not derive from differences in perception, but rather differences in how listeners with LI approach the task. Underscoring this is a recent study of reading disabled children which found little concordance among a variety of discrimination tasks (Messaoud-Galusi, Hazan, & Rosen, 2011)

More broadly, as Pisoni and Tash (1974) point out, discrimination is not a pure measure of auditory encoding. If listeners use both auditory and category-levels as the basis of discrimination, behavioral profiles that look like CP will emerge even without any perceptual warping. Within-category tokens will differ in one respect (at the auditory but not the category level), while between-category tokens differ at both levels. So even if auditory encoding is unaffected by categories, between-category discrimination should always be easier. This severely challenges the case for CP by predicting the same behavioral profile in the absence of auditory warping (categorical perception).

At a theoretical level, modern theories of speech perception also eschew the notion that within-category variation is noise that listeners must ignore (Fowler & Smith, 1986; Goldinger, 1998; McMurray & Jongman, 2011; Oden & Massaro, 1978; Port, 2007), instead arguing that it is a crucial source of variation for coping with coarticulatory and talker variation. There is now substantial evidence that listeners are highly sensitive to within-category detail (Carney, et al., 1977; Massaro & Cohen, 1983) and that early perceptual representations encode speech cues veridically, not in terms of categories (Frye et al., 2007; Toscano, McMurray, Dennhardt, & Luck, 2010). Thus, phonological representations may function quite differently from the sort posited by CP, and may have to be measured in substantively different ways.

Word Recognition in LI

Auditory and phonological processes must ultimately support language-level processes like word recognition. Given the uncertain evidence for lower level deficits, it may be more fruitful to investigate word recognition, a clearer link between perception and language. It is well known that from earliest moments of a word, typical listeners activate multiple words in parallel that match the portion of the input has been heard at that point (Marslen-Wilson & Zwitserlood, 1989). As the input unfolds, these words compete (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998), and the set of active words is reduced as more information arrives until only a single candidate word is active.

Given this theoretical framing a number of recent studies show evidence for impairments in these processes. Dollaghan (1998) used a gating task to show that children with LI recognized familiar words as well as TD children, but needed longer portions of input to do so. In a similar task, Mainela-Arnold, Evans, and Coady (2008) found that LI children vacillated more than TD children, even after there was only one word consistent with the input. Thus, the process by which words compete for recognition may be impaired in LI.

More recently, McMurray, Samelson, Lee and Tomblin (2010) examined LI listeners using the visual world paradigm or VWP (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In the VWP, listeners hear a spoken word (e.g., *wizard*) and select its referent from a display containing pictures of the target word, and phonologically related words (*whistle* and *lizard*) while fixations to each object are monitored. Eye-movements can be generated within a few hundred milliseconds of the beginning of the word, and are made continuously while listeners select the referent. As a result, how much listeners fixate each object can reveal how strongly they are considering (activating) each word over the timecourse of processing.

McMurray et al., (2010) used the VWP to compare LI and Typically Development (TD) listeners in the degree of activation for onset competitors (e.g., *whistle* when they heard *wizard*) and rhymes (*lizard*). LI listeners selected the correct word on 98% of trials. However, adolescents with LI made fewer looks to targets and more looks to rhymes and onset-competitors, particularly late in processing. This suggests LI participants may have difficulty fully activating the target and suppressing competitors. TRACE (McClelland & Elman, 1986) simulations of the data showed that these results could not be modeled by manipulating feature or phoneme level parameters like input noise, memory for acoustic features, or phoneme inhibition (from which CP derives). Rather, variation in lexical decay —the ability of words to maintain activation over time—was the single best way to model variation in language ability. This suggests an LI deficit that may be purely lexical. However, this modeling depends on the specific predictions of TRACE, and this claim needs empirical verification.

Perceptual and phonological processing through the lens of word recognition

Disentangling auditory, phonological and lexical accounts of LI is difficult for several reasons. First, if lower-levels deficits are meaningful, they must cascade to affect higher level processing. An auditory deficit that does not affect word recognition may not be

crucial for explaining LI (e.g., Montgomery, 2002; Stark & Montgomery, 1995). Second and relatedly, a word recognition deficit can come from lower level deficits making it difficult to assign cause. Third, discrimination and phoneme identification tasks used to examine phonological deficits are memory demanding and meta-linguistic and may pose unique problems for LI listeners.

Thus, an ideal assessment should measure word recognition, but do so in a way that can disentangle lower-level causes of a deficit. Recent work on typical listeners points the way by asking how lower-level perceptual changes (e.g., changes in acoustic cues within a category) cascade to affect lexical processes (Andruski, et al., 1994; Utman, Blumstein, & Burton, 2000). McMurray et al. (2008; 2002) presented participants with tokens from a continua of Voice Onset Times (VOT, a continuous cue to voicing: Lisker & Abramson, 1964) spanning two words (e.g., *beach*/*peach*) while they viewed screens containing pictures of both referents and unrelated words. Even restricting analysis to trials in which the participant overtly chose the /b/ items, and controlling for variation in participants' category boundaries, participants made increasing fixations to the /p/ word as VOT neared the boundary (and vice versa for the other category; Figure 1A). Thus, even when tokens were ultimately categorized identically, lexical activation was sensitive to continuous detail. Crucially by measuring how lexical processes differ with respect to sublexical differences in the stimulus (while simultaneously accounting for differences in categorization across VOT steps), this paradigm offers a way to independently measure lexical processes, and listeners' responses to fine grained acoustic variation.

Clayards et al. (2008) offers an example of how this can be used to study between-group changes in phonetic categorization. They employed a similar VWP design to McMurray et al. (2002) but one group of listeners was exposed to narrow distributions (low-variance) such that VOTs were consistently clustered around their prototypes while a second group was exposed to wider distributions (high-variance) that effectively destabilized listeners' categories by causing them to overlap substantially. The high-variance group showed more sensitivity to fine-grained detail (they were more gradient / less categorical), while the low-variance group was almost categorical. Thus, the nature of the categories listeners have formed can have clear effects on the sensitivity to detail seen in patterns of lexical activation.

In the Clayards et al (2008) study, changes in the degree of gradiency can be tied to the structure of phonetic categories. However, their paradigm (based on McMurray, et al., 2002) measures lexical level processes (since participants must access the meaning of a word to click on or fixate it). This suggests that the a version of the McMurray et al. (2002) paradigm may help pinpoint the locus of a perception deficit in LI. Specifically, if an ***auditory deficit*** prevents LI listeners from accurately encoding acoustic cues, we may observe a less gradient response to changes in VOT along with higher competitor activation (uncertainty) overall (Figure 1B). Listeners will be less sensitive to these small acoustic changes, even while this auditory problem makes them much more uncertain as to whether the word started with a /b/ or /p/. Here, VOT is an ideal cue in this case as it is typically short and fundamentally temporal. Alternatively, if ***phonological categories are poorly defined*** we should see a more gradient response in LI (Figure 1C), much like Clayards et

al.'s condition in which categories overlapped more. This would suggest that listeners may be overly sensitive to fine-grained detail. For either hypothesis, we must be able to disentangle differences in the response to VOT from differences (or variability) in the accuracy with which VOT is mapped to categories. This necessitates tracking the relationship between fixations and VOT for tokens that are all within the same category. This can rule out the possibility that increased (or decreased) fixations near the boundary derive from a mixture of trials in which VOT was encoded discretely, but was sometimes mapped to the voiceless category and sometimes to the voiced category. Finally, a *lexical deficit* would yield differences in competitor fixations that do not interact with VOT (Figure 1D).

To test these hypotheses, participants heard a spoken word and selected the referent from a set of four pictures. Stimuli were VOT continua. We monitored fixations to the referents as a measure of lexical activation and expected to see a gradient increase in fixations to the competitor as VOT approached the category boundary. More importantly, we asked how the nature of this effect was related to language ability and whether it indicated impairment at the auditory, phonological or lexical level.

Method

Design

The design followed McMurray et al., (2002). Nine-step VOT continua were constructed from six minimal-pairs: *bale/pail, beach/peach, beak/peak, bear/pear, bees/peas* and *bin/pin*. Participants also heard six l-initial (*lime, lake, leg, lips, leash, lace*) and six ∫-initial filler items (*shoe, shark, sheep, shack, shop, sheet)*. On each trial, participants saw pictures corresponding to each end of one continuum along with two filler items. For each subject, one /l/- and one /∫/-initial item were randomly paired with each continua (semantically related items could not be paired) and these four objects always appeared together. This maintained an equal co-occurrence frequency of the b- and p-items with the fillers. Each continuum step was repeated 10 times (6 word-pairs × 9 steps × 10 repetitions) for 540 experimental trials with an equal number of filler trials for a total of 1080 trials run across two one-hour sessions. Prior to the first session subjects underwent a short training to ensure that they knew the names of the pictures.

Participants

Participants were 74 adolescents drawn from several Iowa communities (the greater Cedar Rapids, Des Moines, Waterloo and Davenport areas) who were participating in an ongoing longitudinal study of language development (Tomblin et al., 1997; Tomblin, Zhang, Buckwalter, & O'Brien, 2003). Informed consent was obtained according to university protocols and participants received monetary compensation. All participants had normal hearing and no history of intellectual disability or autism spectrum disorder. As part of this project they were all evaluated on language skills and nonverbal IQ periodically, and we used 10th grade measures to establish language status using the epiSLI criteria (Tomblin, et al., 1996). Language measures consisted of the *Peabody Picture Vocabulary Test-R* (Dunn, 1981) and the Recalling Sentences, Concepts and Directions, and Listening to Paragraphs

subtests of the *Clinical Evaluation of Language Fundamentals-3* (Semel, Wiig, & Secord, 1995). These scores were combined to form a composite standard score for general spoken language ability. While these tests appear weighted toward receptive (rather than expressive) language, these measures do not show strong dimensionality related to modality, suggesting that this is a measure of generalized language ability (Tomblin & Zhang, 2006).

The distribution of this sample with regard to composite language is shown in Figure 2 with a mean of 90.55 (Z=−.63, SD=1.1). Our analyses treated language ability as a continuous individual difference. However, for ease of description we also divided our participants into LI and typically developing (TD) groups with a cutoff of −1 SD. Nearly half of these individuals (35 / 74) had language composite scores that were classified as LI (Table 1 for group contrasts).

Nonverbal IQ was measured with the Block Design and Picture Completion subtests of the *Wechsler Intelligence Scale for Children-III* (Wechsler, 1989). The distribution of these scores is also shown in Figure 2 (M= 95.63; SD=16.18). Twenty-one of our LI subjects could be diagnosed as SLI (a language deficit with normal non-verbal IQ); and 14 were NLI (both language and cognitive deficits). We did not attempt to construct a sample that was matched on non-verbal IQ. Our aim was to understand the processes that underlie individual differences in language. In this, we cannot assume that language is isolated from general-purpose cognitive systems. Language and non-verbal ability are strongly correlated in the population at large, in part due to common developmental origins (e.g., children from more enriched environments develop better language and non-verbal abilities). Under these conditions, a sample matched on IQ would not represent the variation in the population at large and may constrain the insight into cognitive systems important to language (Dennis et al., 2009). Thus, rather than constraining the covariance between language and non-verbal IQ, we maintained a more representation sample in which these are correlated, but we explored statistically how these factors relate to our measures.

Stimuli

Auditory Stimuli—B/p continua consisted of 9 steps of VOT from 0 to 40 ms. Stimuli were constructed from recordings of natural speech using progressive cross-splicing (Coady, et al., 2007; McMurray, Aslin, et al., 2008). In typical adult listeners, the gradient response to within-category detail has been observed for both synthetic and natural continua (McMurray, Aslin, et al., 2008). However, research on LI has suggested minimal or no deficits in categorical perception with continua constructed from natural speech (Coady, et al., 2007). Thus, our use of natural speech makes for a stronger (and substantially more ecological) test.

For each continuum we recorded multiple exemplars of each endpoint from a male talker with a Midwest dialect. Tokens were recorded in a variety of carrier sentences "Move the [target word] from the … to …" (used for another experiment), with prosodic emphasis on the target word. From these, we selected the clearest token that was most similar to its cross-voicing competitor (in every other respect than VOT). Next, onset portions of the voiced (/b/) stimuli were incrementally replaced with same duration of material from the onset of the voiceless (/p/) stimuli. As a result, across the continua, vocalic portions came

from the /b/ items, and the onset contained progressively larger amounts of burst and aspiration from the /p/ items. Splicing was done in approximately 5 ms increments, but because splices were performed at zero-crossings in the waveform, continua steps were not uniformly spaced (Table 2 for actual VOTs). Filler items were recorded by the same talker in the same session. Stimuli were normalized to the same RMS amplitude and 50 ms of silence was inserted at the beginning of each file.

Visual Stimuli—Visual referents were colored line-drawings[1]. These were created in a multi-stage process to ensure each picture was a prototypical representation of the corresponding word, and to minimize differences in brightness, salience, etc., (McMurray, et al., 2010). First, we downloaded multiple candidate images from a commercial clipart database. A group of 4–8 young adults then selected the most prototypical image from each set for each word. These images were then edited based on suggestions from the group to make the pictures more prototypical, more uniform, or to remove distracting elements. The final images were approved by the first or second author or a lab manager with experience using pictures in the VWP.

Procedure

Before beginning the experiment an EyeLink II head-mounted eye-tracker was calibrated for each participant. Participants then read the instructions (Appendix A) on the computer screen and the experimenter explained them and verified that the participant understood the task. Next, participants completed two types of training before beginning the experimental trials.

Training—During the first training phase, participants were familiarized with the images of the words in the experiment. Images were shown one at a time with their names printed below. Familiarization was self-paced and participants clicked to advance to the next picture, but had to view each picture for at least 500 ms. Each of the 24 words appeared once. During the second phase, participants were familiarized with the layout of the pictures on the screen and the task. On these trials participants saw the four pictures (b-, p-, l-, and sh- initial) from a given item-set in the corners of the screen and a blue dot in the middle. After 500 ms the dot turned red. When participants clicked on the dot it was replaced with the printed word for one of the four pictures. They then clicked the matching picture to move to the next trial. There was no time limit. If an incorrect picture was selected, the trial did not advance until the correct picture was selected. Each word appeared as the target twice during this training phase for a total of 48 trials.

Testing—The experiment used a similar procedure to McMurray et al. (2002). Experimental trials were identical to the Phase 2 training trials except that instead of seeing the target word, participants heard the target word over headphones. Having participants click on the red dot to hear the auditory stimulus centered their attention and typically their gaze on the middle of the computer screen at the onset of the target word. The 500 ms delay

---

[1]Across multiple studies we have found consistently cleaner results with line-drawings over photographs as line drawings are often more prototypical, and have fewer details and distracting elements.

before the dot turned red allowed participants to see the items on the screen before hearing the target word.

Eye-Movement Recording and Analysis—Eye-movements were monitored with a head-mounted EyeLink II eye-tracker sampling gaze position every 4 msec. This system compensates for head movements so participants could move their heads normally during the experiment. When possible, both eyes were tracked (and the better used in the analysis). A drift-correct procedure was run every 30 trials to compensate for the tracker shifting on the head. Eye-movement recording began at the onset of the trial and lasted until the participant clicked on a picture. The eye-tracking data was automatically parsed into saccade, fixation, and blink events using the system's default parameters. These events were converted to "looks" to the four images on the screen. A look lasted from the onset of a saccade to a specific location to the offset of the subsequent fixation. Pictures were 250×250 pixels and these boundaries were extended by 100 pixels to accommodate noise in calibration and drift in the eye-tracker. The four regions remained distinct from one another and did not overlap. To cope with variation in the length of the trial, trials were arbitrarily ended at 2000 ms. If a trial lasted longer than that, the data were cut off; for shorter trials, the final "look" was extended to 2000 ms.

Results

Mouse-click results

Participants' mouse-click responses were first examined as a measure of overt categorization. Participants rarely clicked filler pictures (l- or sh- words) on experimental trials (0.16% of trials), and these were excluded from analysis. One TD participant showed many more such errors (6.8%) and was excluded from further analysis[2].

We first examined the mouse clicking as a function of VOT, an analogue to phoneme identification measures. This was done to determine if there are any substantive group differences (which were not predicted as the stimuli were constructed from natural speech: Coady, et al., 2007). Importantly, our analysis of the eye-movements (based on McMurray, Aslin, et al., 2008; McMurray, et al., 2002) required us to select trials that had the same overt response and were on the same side of the boundary, to ensure that we were testing within-category differences. Thus, this analysis was needed to determine if overt categorization differed.

Identification curves averaged across participants and word-pairs are shown in Figure 3A. To examine between-group and continua differences in identification we fit a four-parameter logistic function[3] (1) to the identification data for each of the six word-pairs, for each participant. This provides an estimate of the category boundary, slope, and the minimum and maximum asymptote of the curve for each participant × word-pair.

---

[2]Nine additional participants (4 LI, 5 TD) only completed one of the two 540 trial sessions. Their data were retained, though analyses of both the identification and eye-movement data without these participants yielded identical results.
[3]We used the four-parameter logistic function to capture group differences in end-point accuracy which cannot be easily modeled in a standard (two- parameter) logistic function (which assumes asymptotes at 0 and 1).

$$P(presponse)=\frac{max-min}{1+\exp\left(4\cdot\frac{slope}{max-min}\cdot(boundary-VOT)\right)}+min \quad (1)$$

Curve-fits were good overall with an average $R^2=.97$ and a RMS error of .004 (this did not differ as a function of language ability, F<1 in an ANCOVA on RMS). We examined each parameter with an ANCOVA treating language ability as a continuous covariate and word-pair as a within-subjects factor (Table 3). To summarize, there was substantial variability in *boundary* across word-pairs (range = 23.8 – 32.8 ms of VOT) and participants (range = 21.1 – 37.3 ms); there was also substantial variation in *maximum* (by word-pair: Range = .71 to .98, SD=.10; by subject: Range = .63 to .99, SD=.074) and *slope*. However, for all of these, only word-pair reached significance; these parameters were not related to language ability. In contrast, *minimum* (the lower asymptote) also showed high variability (by word-pair: Range = .012 - .060, SD=.018; by subject: Range = 0 to .15, SD=.027), but did show an effect of language and a language × word-pair interaction. These effects were driven by the *bale/pail* continuum where LI children had a hard time accurately identifying *bale* (LI: *M*=8.4% /p/; TD: *M*=4.2% /p/). When this analysis was repeated without this word-pair, there were no significant effects. Thus, the only difference in identification that can be related to language ability appears to be in recognition of the word *bale*. Nonetheless, as the logic of this experiment requires us to examine fixations relative to each participant's categories, it is crucial to account for the variation in categorization across words to be sure that we are addressing within-category effects.

Fixation Analysis

Data Preparation—A number of steps were taken to ensure that our statistical analyses closely matched our experimental hypotheses. First, for the VWP to offer a valid measure of lexical activation in this context, listeners must know both endpoints of the continuum. Thus, we excluded any word-pair if endpoint responding (*max* and *min*) was below 80% correct.

Second, to isolate categorization and auditory processes from our lexical measure, we must examine lexical activation to variation that is clearly *within a category*. For example, differences in the response to VOT could derive from variation in how the VOTs were categorized, or in how VOT was encoded. However, if we can be sure that all of the VOTs were categorized identically, and we still see variation in the listeners' response to VOT, this can be more precisely matched to pre-categorical processes. This requires us to examine VOTs for only trials which were categorized identically, and this in turn requires confidence in the category membership of each token for each participant. Thus, for each participant we did not analyze fixations for a word-pair if the R^2 of the fit to the identification data was less than 0.8. Given these two exclusions, an average of 5 of the 6 word-pairs were retained for each participant (Table 4). The number of exclusions per participant and the number of participants per word-pair was similar for LI and TD participants (TD: M=5.1, SD=.82; LI: M=5.1, SD=.83)[4]. Figure 3B shows identification curves after these exclusions.

Third, to maximize our confidence that these was a within-category response, consistent with prior studies (McMurray, Aslin, et al., 2008; McMurray, et al., 2002) for rSteps less than 0 (voiced sounds) we only analyzed trials in which the participant ultimately selected /b/, and for voiced rSteps, we only analyzed trials in which /p/ was selected.

Finally, a related concern that between-participant and between-word-pair variation in category boundaries could yield an averaging artifact that would make participants appear more gradient. Even if within a category, responding is discrete for a given participant or word-pair, if boundaries vary, the average across participants or word-pairs could look gradient. As a result, differences in gradiency as a function of language (our measure of perceptual or phonological processing) could result from differences in the variability of boundaries. Thus, we recoded VOT in terms of its distance from each participant × word-pair boundary as relative step or rStep (as in McMurray, Aslin, et al., 2008). While rStep was primarily used for analysis of the eye-movement data, to assist in visualizing this transformation, identification curves using rounded rStep are shown in Figure 3C. This panel shows a steeper slope or more categorical identification curve than the two panels since variability in category boundary between-subjects (which was not accounted for in 3A and 3B) creates what appears to be a shallower slope in these panels. Crucially, even accounting for this variability are accounted for, the two types of participants continued to perform very similarly in their overt identification.

Description—Figure 4 shows the time course of fixations for the two unambiguous VOTs. Around 250 ms, the target, competitor, and filler curves begin to diverge. For TD participants (Figure 4A) target fixations quickly surpass the other referents. As it takes 200 ms to plan and launch and eye-movement, and there was 50 ms of silence at the onset of each stimulus, this reflects the first point at which signal-driven fixations were possible. Competitor looks peaked at 500 ms, reflecting the temporary overlap between targets and competitors and stayed higher for about 1000 ms. LI participants (Figure 4B) show similar patterns of eye-movements with delayed peak target fixations, higher competitor and filler fixations, and prolonged fixations.

Looks to the competitor (e.g. looks to /b/ when the VOT indicated a /p/) was our primary measure. Figure 5 shows the timecourse of competitor looks for voiced (Figure 5A,B) and voiceless (Figure 5C,D) rSteps. For both groups, there is initially little effect of rStep, but by 600 ms there are clear gradient effects: both TD and LI participants show increasing competitor looks as rStep nears the boundary. Similar to McMurray et al. (2010), LI participants show more competitor fixations overall. Figure 6 illustrates these group differences more clearly, averaging looks to the competitor from 250 to 1750 ms. The similarity in the slope of the functions across groups suggests little difference between LI and TD listeners in the degree of gradient sensitivity to VOT. However, LI participants showed higher overall looks to the competitor, suggesting greater competitor activation. This fits with a primarily lexical locus to LI (Figure 1D).

---

[4]The *bale/pail* continuum seemed particularly bad for both groups and was excluded for more than half of the participants. Participants may have struggled with this word-pair because of the semantic ambiguity of *bale* (*bail*), or because they called the *pail* image a bucket (which starts with /b/).

Statistical Models—Our dependent variable was proportion of looks to the competitor. There were two independent variables: language ability (continuous, between-subject) and rStep (within-subject). This might warrant something like a mixed ANOVA. However, due to variation in participants' category boundaries, rSteps are not perfectly comparable across subjects: one participant might have rSteps of −.6, −1.6, and −2.6, while a participant with a different boundary might have −.3, −1.3, and −2.3. Furthermore, a participant with a boundary at step 4.5 has four steps on the voiceless side, while a participant with a boundary at 5.5 has only three, creating problems of missing data if rStep were treated as a factor (rather than a continuous covariate). The nested nature of the design and the variable nature of rStep values led us to a linear mixed model using rStep as a continuous predictor. This had the added bonus of allowing us to model the effect of rStep at the level of participant (a random slope), a more sensitive test of our effects.

We analyzed the data with linear mixed effects models[5] using the LME4 package of R (Bates & Sarkar, 2011). The proportion of fixations to the competitor between 250 and 1750 ms was the DV. Independent variables included two linear effects: *language ability* (between-subject, a continuous Z-score) and *rStep* (within-subject, also continuous). Both variables were centered to appropriately interpret interactions. As prior studies suggest that in some cases, the voiced and voiceless side of the continuum show different effects of VOT (McMurray, Aslin, et al., 2008), separate models were run for each. We first compared a series of models to determine the best random effects structures. We compared 1) models with only random intercepts by subjects, 2) models with subject- and word-pair-random intercepts; 3) models with random slopes of rStep on subjects (each subject has their own response to rStep), and 4) models with random slopes of rStep for both subject and word-pair. This last model offered a better fit to the data than the other models (Voiced: $\chi^2(2)=6.5$, p=.039; Voiceless: $\chi^2(2)=11.7$, p=.0029).

It is currently unclear how best to compute the significance of fixed effects in linear mixed models. There is debate about the d.f. for T-statistics, and Monte-Carlo methods have not been implemented for models with random slopes. Thus, to estimate significance, we used the χ^2 test of model comparison. For main effects, we compared models with only the effect of interest to a model with just random effects. For interactions, we compared a model with the interaction term to one with just main effects. For each fixed effect we thus report coefficients and SE from the full model with the χ^2 test of model fit from these comparisons.

In this framework, we can make clear predictions (Figure 1). All of the hypotheses predict a main effect of rStep. An auditory deficit (Figure 1B) predicts a main effect of language and a language × rStep interaction with a smaller effect of rStep for listeners with lower language ability. A phonological deficit (Figure 1C) also predicts a main effect of language and an interaction, though individuals with lower language ability should have a larger effect of rStep. Finally, a lexical deficit (Figure 1D) predicts only main effects of language and rStep.

---

[5]We've also analyzed this data with hierarchical regression and ANOVA/ANCOVA. While both of these are less suited to our data, they both yield an identical pattern of results to what we report here.

Analysis—Complete model results are in Table 5. For both voiced and voiceless models, the correlation between fixed effects was low ($R_{voiced}=-.009$; $R_{voiceless}$: $R=-.015$). Both models found highly significant main effects of rStep (Voiced: B=.0093, p=.0004; Voiceless: B=−.0158, p=.0002). Replicating McMurray et al., (2008; 2002), for voiced sounds, an increase of 1 rStep toward the boundary led to an increase of just under .01 looks to the competitor; for voiceless sounds, an increase of 1 rStep (away from the boundary) led to a decrease of just over .01 competitor looks. We also found a significant main effect of language (Voiced: B=−.0083, p=.00018; Voiceless: B=−.0110, p=.0008). As in our prior VWP on LI (McMurray, et al., 2010), people at the lower end of the language scale fixated competitors more. There was no interaction—the effect of rStep was similar across different levels of language ability. To rule out the possibility that these results derive from the choice to exclude some word-pairs, the analysis was repeated with the full dataset (Supplement S1) finding the same pattern of results.

Individual Differences vs. LI Status—We next asked whether language ability was a continuous or dichotomous predictor in this study. We replaced the continuous language score with a dichotomous diagnosis (LI vs. TD, with a −1 SD cut-off) and compared this to the previous model using Bayesian Information Criteria[6] (lower BIC indicates a preferred model). The model using a continuous language score offered a better fit (Voiced: $BIC_{continuous}=-5463$, $BIC_{diagnosis}= -5459$; Voiceless: $BIC_{continuous}=-3733$, $BIC_{diagnosis}= -3730$), suggesting that the effect of language on competitor fixations is continuous. We also asked if the relationship between language and competitor fixations was non-linear by adding additional polynomial terms to the initial model. There was no benefit for adding quadratic (Voiced: $\chi^2(2)=.6$, p=.75; Voiceless: $\chi^2(2)=.7$, p=.72), cubic (Voiced: $\chi^2(2)=.1$, p=.94; Voiceless: $\chi^2(2)=.6$, p=.73) or quartic terms (Voiced: $\chi^2(2)=3.9$, p=.14; voiceless: $\chi^2(2)=5.5$, p=.065), suggesting a strongly linear relationship. Critically, even in the polynomial analyses there was not an interaction between any of the language terms and rStep. For example, in the quartic analysis, there was no significant improvement for adding interaction terms for rStep and all four language terms (Voiced: $\chi^2(4)=2.0$, p=.72; Voiceless: $\chi^2(4)=2.5$, p=.65).

Specificity of the Effects—Our final analyses simultaneously addressed two concerns. First, we addressed the contributions of non-verbal IQ both to determine if heightened competitor fixations were uniquely due to language, and if there was any subset of children who responded differentially to changes in VOT. Second, addressed the possibility that the heightened competitor fixations derive from differences in factors like visual search, eye-movement control or general uncertainty, which are only indirectly related to language (if at all).

With respect to IQ, language and cognitive abilities do not represent orthogonal or independent measures; however, they are clearly not isomorphic making it important to attempt to account for both. In our sample, as in the real world, IQ and language ability were highly correlated (R=.7), so we could not take the simplest approach of putting both into a

---

[6]We couldn't use the χ^2 test of model fit as the models were not nested.

statistical model. Moreover, since both derive from a common developmental origins (e.g., an enriched environment), it would be inappropriate to simply partial IQ out of language—this ignores the shared variance which is a key component of how language abilities distribute in the population (Dennis, et al., 2009). Despite the collinearity, we can still ask whether listeners' sensitivity to VOT interacts with individual differences more broadly (language+IQ). Moreover, we can eliminate the collinearity between language and IQ by residualizing IQ against language to create a new variable, IQr, which describes variation in IQ over and above what would be predicted for that level of language ability. Similarly, we can conduct a second analysis using language residualized against IQ to determine if there was an effect of language over and above IQ. This can determine if there are independent effects (and if rStep interacts with either).

With respect to fixation factors, a series of analyses of the filler trials (Online Supplement S2) offered some evidence issue that differences in factors like visual search or eye-movement control concern may play a role. Participants with low language scores or non-verbal IQ were more likely to fixate the /b/- and /p/-initial items (on filler trials). However, a communality analysis found that while IQ exerted effects on unrelated fixations over and above language, language did not have an effect over and above IQ. Thus, these heightened unrelated fixations may be largely due to IQ. Nonetheless, it is important nonetheless to verify that our competitor effects survive after accounting for this, and that it is not masking a language × rStep interaction. Thus, in this final analyses we controlled for individual differences in unrelated fixations to determine if the effect of language was still be observed (and if it was independent of rStep).

To do this, we first computed for each participant the proportion of fixations to each of the /b/ and /p/ objects between 250–1750 msec. on the filler trials. This estimates how much a participant would have looked at these specific items, in this task when they were unrelated to the spoken input. Next, we conducted a regression predicting looks to those objects on experimental trials, from their looks on unrelated trials and used the residuals as a measure of how much *more* the participant looked at that object when it was relevant to the spoken input.

With this residualized looking as a DV, we conducted a mixed effects analysis with Language and rStep as fixed effects and random slopes of rStep for both subject and word-pair. The correlation among fixed effects was low (Voiced: R=−.012; Voiceless: R=−.016). This analysis confirmed the earlier results (Table 6): there were still significant main effects of rStep (Voiced: B=−.0090, p=.00043; Voiceless: B=−.0153, p=.00016), and language (Voiced: B=−.0059, p=.00012, Voiceless B=−.009, p=.0011), and no Language × rStep interaction. Thus, even controlling for differences in overall looking, LI listeners show heightened competitor fixations; listeners at both ends of the language spectrum show a gradient response to within-category VOT; and this gradiency did not differ as a function of language ability.

Using this residualized, measure of competitor fixations, we next conducted a communality analysis to determine if the primary effects of language on competitor fixations were uniquely due to language, or were due to variation shared with IQ. We first constructed a

model with only language and rStep as fixed effects and then added residualized IQ. Here, adding IQr to the model resulted in no additional benefit (Voiced: $\chi^2(1)=.12$, p=.72; Voiceless: $\chi^2(1)=.004$, p=.94). When we started with rStep and IQ, adding LanguageR resulted in a significantly better model in the voiced analysis (Voiced: $\chi^2(1)=3.9$, p=.046) and a marginal gain in the voiceless ($\chi^2(1)=2.75$, p=.097)[7]. Thus, there is some component of the increased competitor fixations that can be uniquely attributed to language, however, this increase (whether due to variation unique to language or shared with IQ) cannot be attributed to the fact that adolescents with lower language and/or IQ simply look at everything more. Critically, in both models, there was no Language × rStep, IQ × rStep or three-way interaction suggesting no subgroup (along these dimensions) was differentially sensitive to VOT.

General Discussion

We found that lexical activation is gradiently sensitive to changes in VOT within a category, replicating several studies with typical adults (Andruski, et al., 1994; McMurray, Aslin, et al., 2008; McMurray, et al., 2002; Utman, et al., 2000). This used a conservative analysis that accounted for variation in participants' boundaries and in their responding on that same trial. We also found that adolescents with poorer language fixate competitors more. This was observed despite the fact that listeners with low language abilities clicked on the correct target at identical rates to those with higher abilities, and even after accounting for differences in visual processes (that appear to derive from IQ). This effect was linearly related to language ability, not just clinical category (and see, Tomblin, 2011, for a discussion of LI as a continuous dimension). It replicates McMurray, et al. (2010) and suggests that poor language users have difficulty suppressing competing lexical candidates. Finally, these individual differences in language did not interact with sensitivity to VOT: the gradient effect of VOT was similar at both ends of the language ability/IQ scale. Listeners with both good and poor language were similarly sensitive to fine-grained variation in VOT, and their mapping of this cue to categories took the same form. Thus, listeners at both ends of the scale are equally good at encoding VOT, and they map it onto phonological categories in a similar way. This supports that the deficits in word recognition associated with LI may have a primarily lexical locus (Figure 1D).

Before discussing the theoretical implications of these findings, we address several limitations. First, while we found significant effects of language, our most important finding is the non-significant rStep × Language interaction. Of course it is possible that there is a phonological or auditory deficit and our measure was simply not sensitive enough to detect it. We cannot rule this out. However, there are some reasons to be skeptical. First, we had a large sample (N=73), and a large number of experimental trials (540) for a study of this sort. In contrast, Clayards et al., (2008) did observe an experimental-group × VOT interaction in a very similar paradigm to our own (only with a between-groups experimental manipulation of phonological category structure). However, they used only 24 participants and 228 experimental trials. Thus, our study was quite over powered compared to a similar between-groups design. Second, we used an extremely sensitive statistical technique that accounts for

---

[7]Note that this was a significant ($\chi^2(1)=4.49$, p=.034) when random effects were treated as intercepts.

multiple random effects, and models listeners' responses to rStep at individually; in contrast prior studies using this paradigm primarily use ANOVA (Clayards, et al., 2008; McMurray, et al., 2002). Third, we extended our hunt for factors that could have interacted with rStep to examine IQ, and its joint contribution with language, as well as dichotomous and non-linear effects of language ability. Across all of these analyses we still found no interactions. Fourth, we partialed out visual factors from our DV and replicated the same pattern of effects. Thus, while we cannot rule out the possibility of a lower-level effect on the basis of a null interaction, if there was an underlying interaction the conditions were favorable for observing it. Moreover, this non-significant interaction is consistent with a number of studies that also report non-significant interactions of acoustic factors and language ability (Coady, et al., 2007; Montgomery, 2002; Stark & Montgomery, 1995).

The second concern was that language and IQ were correlated in our sample (R=.7, Figure 2B). This was deliberate as we wanted to study natural variation in language ability, and non-verbal IQ is inherently a component of this. To some extent this doesn't matter. As long as our effects are not due to *uniquely* non-verbal causes, it is just as useful if they derive from shared language/IQ variation as unique language variation—under this broad brush, children who present with poor language (regardless of IQ) do activate competitors more, and are not differentially sensitive to auditory/phonological factors. However, our analyses were able to somewhat disentangle these factors: IQ played a unique role in predicting looking in general (even to unrelated items), and once we took that into account, IQ it did not uniquely predict competitor fixations and there was evidence for a unique role of language. In light of the fact that McMurray et al., (2010), showed no effect of IQ on competitor fixations with a different sample, this suggests this may be largely an effect of language on lexical processes. Perhaps more importantly, however, listeners' response to VOT did not interact with language or IQ, so our claim of a lexical-level difference is not affected by the language/IQ collinearity.

The third concern is that prior studies suggest that perceptual deficits may only be observed with certain tasks and synthetic speech (2007; Robertson, et al., 2009). We cannot rule out that a lower-level deficit might be observed in our paradigm with synthetic speech or in noise (or, we could also just observe a more pronounced lexical deficit). However, the VWP is more sensitive than previous CP paradigms (McMurray, Aslin, et al., 2008), and we did detect differences in lexical processes, just not the ones predicted by perceptual accounts. More importantly, a deficit that is only observable in synthetic speech may not be critical for understanding the deficits LI listeners face in everyday language use involving natural speech.

Finally, we have only studied one cue, VOT. While VOT is considered representative of phonological processes like phoneme categorization, it differs from other acoustic cues like vowel formants, fricative spectral poles, or formant transitions for place of articulation. In defense of VOT, numerous studies with both LI and RD children have shown deficits with VOT, and it is both short and temporal (important properties in auditory accounts). However, future work should take use the present paradigm to study a greater variety of acoustic cues.

Despite these concerns, our results indicate a lexical locus to the deficits faced by people with poorer language skills and against a perceptual or auditory basis. This does not rule out the possibility that LI listeners have an auditory or phonological deficit, it just does not appear to affect word recognition, and as a result is even less likely to affect downstream processes (sentence comprehension). Such a discrepancy could arise if word recognition is robust to noise at lower levels, or if there are auditory processing difficulties that do not affect language-critical processes. Our overall results confirm a pattern implied in a number of other studies of word recognition in LI. Dollaghan (1998) found no differences at short stimulus segments where there is only minimal perceptual material support lexical access (see also Mainela-Arnold, et al., 2008); neither of the Montgomery studies (Montgomery, 2002; Stark & Montgomery, 1995) found that LI listeners were differentially impaired by time-compressed or filtered speech, or by large numbers of stop consonants; and McMurray et al.'s (2010) VWP study found few differences in early in processing where phonological or perceptual impairments would play the greatest role. But all of these studies report reliable differences in word recognition between LI and TD listeners. This, converging pattern of results across several methods lends confidence to our null interaction. But more broadly this body of work as a whole suggests that if there is an auditory deficit in LI, it may not matter for word recognition.

But what is the nature of this lexical deficit? One could argue simply that poor language users are more uncertain. But that is not entirely the case. Their ultimate decision was not distinguishable from TD adolescents, even at ambiguous portions of the continuum. They clearly have access to sufficient phonetic information to recognize the correct word. So perhaps they were less confident in their decisions? If so, this lack of confidence should have been spread throughout the four objects, yet controlling for fixations to unrelated objects, there was still an effect of language on competitor fixations. Thus, their uncertainty is precise—listeners with LI have ruled out some competitors but are uncertain about others. Even under this framing, our data pinpoints the level at which poor language users are uncertain: they're certain about the acoustic cue, and how it maps to the words, but they're not certain about what the word was.

This highly specific notion of uncertainty is hard to distinguish from the concept of lexical activation, the idea that lexical candidates receive a graded degree of consideration that derives from their match to the input and the set of possible candidates; that multiple candidates are active at the same time; and that resolving this competition is how words are recognized. That is, the kind of uncertainty shown by poor language-users would be described as some words being more active than they should be (but as our data suggest, not some phonemes, or some features). Models like TRACE can go further by clarifying how deficits map onto the mechanisms by which this co-activation is resolved. McMurray et al., (2010) hypothesized on the basis of TRACE simulations of their prior VWP study that the lexical deficit associated with LI arises from the way words decay in memory. Words decay faster for listeners with LI (see Harm, McCandliss, & Seidenberg, 2003, for an analogous hypothesis in reading impairment); this slows how quickly target words build activation; and this offers less inhibition to competitors (allowing competitors to have more activation). Our results extend this account by ruling out lower level causes of this effect. However, decay may not be the only way to achieve such effects in an interactive activation framework, and

any imbalance in the competitive processes that separate targets from competitors may be relevant.

So how does a lexical account explain the idiosyncratic array of deficits (and null effects) in studies examining phonological or auditory processes? First, these studies examine a wide variation of ages and a wide variety of populations – children with SLI or NLI, and with or without concurrent reading disability (if reading was measured). Here, a focus on dimensions of individual difference, in terms of language, cognitive ability, and possibly reading, as well as their sub-components may be more helpful than focusing on a few diagnostic categories.

Second, differences in lexical processing may be the source of perceptual differences. It is fairly well accepted that lexical processes either feedback to affect perceptual ones (Ganong, 1980; Magnuson, et al., 2003; McClelland, et al., 2006; Newman, Sawusch, & Luce, 1997) or bias phoneme decision tasks (Norris, McQueen, & Cutler, 2000). Thus, phonological deficit could derive from lexical ones, and the idiosyncrasies across studies could be related to the lexical neighbors of the stimuli used in a study, and the relative robustness of the stimulus (e.g., synthetic speech may be less robustly processed, and therefore more susceptible to lexical influence). Alternatively, evidence for perceptual deficits has been largely seen in phoneme decision and discrimination task, both of which tap processes like working memory and phoneme awareness that may be difficult for children with LI. In contrast, our measure is embedded in an easy task that taps more natural forms of language processing. Our failure to find a auditory or phonological deficit suggests that some of the evidence for such differences may derive from interference from impaired lexical processes, or from factors like memory that are necessary for experimental tasks, not from perceptual differences.

Could a lexical deficit also account for higher-level deficits in things like working memory or sentence processing? While this must be speculative, there are several (non-mutually-exclusive) routes by which this could occur. First, if word recognition does not output a single candidate, this could overwhelm working memory capacity (putting more words in the same number of slots). Second, there is also evidence that partially activated lexical competitors cascade to sentence processing (Levy, Bicknell, Slattery, & Rayner, 2009). A purely lexical deficit could thus create downstream problems as syntactic/semantic processes must deal multiple candidates for each word. Finally, the observed lexical deficit may have a common source to other types of deficits. Word recognition is broadly viewed as a process of parallel competition (Marslen-Wilson, 1987; McClelland & Elman, 1986) and similar processes have also been invoked in other levels of language (Dell, 1986; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Rapp & Goldrick, 2000) as well non-verbal domains (Goldstone & Medin, 1994; Spivey & Dale, 2004). A deficit in dynamic mechanisms like decay and inhibition could thus link our lexical impairments to deficits or slowing in other linguistic and non-linguistic domains (Kail, 1994; C. A. Miller, Kail, Leonard, & Tomblin, 2001).

Beyond their relationship to higher level language, our results indicating a lexical locus to LI have important clinical implications. Interventions stressing auditory processes or phoneme

awareness may not address the fundamental problems faced by children with LI (underscored by a recent clinical trial: Gillam et al., 2008). Our data suggests that interventions and diagnostics focusing on lexical level variables may be more important. However, this may not be as simple as teaching children new words or focusing on vocabulary measures. The words in this study (and our previous study, McMurray, et al., 2010), were well known and deficits were still observed. So at the level of diagnostics tools, what may be needed are tools that measure not just how many words a child knows, but differences in the processes by which they recognize them. In this regard, Farris-Trimble and McMurray (in press) have examined the test/re-test reliability of the VWP and found fairly high reliabilities (between .5 and .8) for different components of the fixation record; however at the same time, this measure is clearly affected by other factors (e.g., visual) and only moderately correlated with language (R∼.3: McMurray, et al., 2010). Similarly at the level of treatment, we should begin investigating interventions that focus not just on teaching children new words, but on helping them resolve ongoing competition between words during comprehension.

More broadly, this work points to the fact that individual differences at the level of lexical processes may be more important for functional language ability than differences in the auditory processes that encode cues or in phonological processes that map them to categories. This highlights the robustness of word recognition against variation at lower levels and minimizes the importance of an abstract phonological representation of the signal for language processing as a whole. It suggests that the redundancy afforded by the signal and the lexicon to allows (even impaired) listeners multiple routes to achieve meaning from spoken language. Most importantly, it suggests the importance of real-time lexical processes like competition, inhibition and decay among meaningful linguistic units for functional language as a whole.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix A

The instructions for the experiment (that were displayed on the computer screen) were as follows:

> In this experiment you will match words with their pictures. On each trial you will see a blue circle with four pictures. Look at it until it turns red. At that point, click on the red circle and you will hear a word that matches one of the pictures. Your job is to click on the matching picture and go on to the next trial. This experiment has two parts: a short practice, followed by testing. You can take a short break after

> completing the first part of the experiment. Click the mouse when you are done reading this.

After participants read this, the experimenter explained this again more informally, and then asked the participant if they understood. They then monitored the first few trials (in particular) and interrupted the participant if necessary to discuss the task. For the participants with lower language abilities there was often a little conversation involved in this as they figured it out what they needed to do, but all of the participants were able to figure out the task within the first few trials, as evidenced by the excellent identification performance.

References

Andruski JE, Blumstein SE, Burton MW. The effect of subphonetic differences on lexical access. Cognition. 1994; 52:163–187. [PubMed: 7956004]

Bates D, Sarkar D. lme4: Linear mixed-effects models using S4 classes. 2011

Bishop DVM, Adams CV, Nation K, Rosen S. Perception of transient non-speech stimuli is normal in specific language impairment: evidence from glide discrimination. Applied Psycholinguistics. 2005; 26:175–194.

Bishop DVM, Carlyon R, Deeks J, Bishop S. Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. Journal of Speech Language and Hearing Research. 1999; 42(6):1295–1310.

Burlingame E, Sussman HM, Gillam RB, Hay JF. An investigation of speech perception in children with specific language impairment on a continuum of formant transition duration. Journal of Speech, Language, and Hearing Research. 2005; 48:805–816.

Carney AE, Widin GP, Viemeister NF. Non categorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America. 1977; 62:961–970. [PubMed: 908791]

Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 2008; 108(3):804–809. [PubMed: 18582855]

Coady J, Evans JL, Mainela-Arnold E, Kluender K. Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. Journal of Speech Language and Hearing Research. 2007; 50:41–57.

Coady J, Kluender K, Evans JL. Categorical perception of speech by children with specific language impairments. Journal of Speech Language and Hearing Research. 2005; 48:944–959.

Corriveau K, Pasquini E, Goswami U. Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. Journal of Speech, Language, and Hearing Research. 2007; 50:647–666.

Dahan D, Magnuson JS, Tanenhaus MK, Hogan E. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes. 2001; 16(5/6):507–534.

Dell GS. A spreading activation theory of retrieval in sentence production. Psychological Review. 1986; 93:283–321. [PubMed: 3749399]

Dennis M, Francis DJ, Cirino PT, Schachar R, Banres MA, Fletcher JM. Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. Journal of the International Neuropsychological Society. 2009; 15:331–343. [PubMed: 19402919]

Dollaghan C. Spoken word recognition in children with and without specific language impairment. Applied Psycholinguistics. 1998; 19:193–207.

Dunn, LM. Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: AGS; 1981.

Eisenson J. Developmental aphasia: A speculative view with therapeutic implications. Journal of Speech and Hearing Disorders. 1968; 33:3–13. [PubMed: 4171242]

Farris-Trimble A, McMurray B. Test/retest reliability for the visual world paradigm as a measure of real-time lexical processes. Journal of Speech Language and Hearing Research. in press.

Fowler, CA.; Smith, M. Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In: Perkell, JS.; Klatt, D., editors. Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum; 1986. p. 123-136.

Frye RE, McGraw Fisher J, Cody A, Zarella M, Liederman J, Halgren E. Linear coding of voice onset time. Journal of Cognitive Neuroscience. 2007; 19:1476–1487. [PubMed: 17714009]

Ganong WF. Phonetic Categorization in Auditory Word Recognition. Journal of Experimental Psychology: Human Perception and Performance. 1980; 6(1):110–125. [PubMed: 6444985]

Gerrits E, Schouten MEH. Categorical perception depends on the discrimination task. Perception & Psychophysics. 2004; 66(3):363–376. [PubMed: 15283062]

Gillam RB, Loeb DF, Hoffman LM, Bohman T, Champlin CA, Thibodeau L, Friel-Patti S. The Efficacy of Fast ForWord Language Intervention in School-Age Children With Language Impairment: A Randomized Controlled Trial. J Speech Lang Hear Res. 2008; 51(1):97–119. [PubMed: 18230858]

Goldinger SD. Echoes of Echos? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. [PubMed: 9577239]

Goldstone RL, Medin DL. The time course of comparison. Journal of Experimental Psychology: Learning Memory and Cognition. 1994; 20:29–50.

Harm MW, McCandliss BD, Seidenberg MS. Modeling the Successes and Failures of Interventions for Disabled Readers. Scientific Studies of Reading. 2003; 7(2):155–182.

Heltzer JR, Champlin CA, Gillam RB. Auditory temporal resolution in specifically language-impaired and age-matched children. Perceptual and Motor Skills. 1996; 3:1171–1181.

Joanisse MF, Seidenberg MS. Phonology and syntax in specific language impairment: Evidence from a connectionist model. Brain and Language. 2003; 86:40–56. [PubMed: 12821414]

Kail R. A method for studying the generalized slowing hypothesis in children with specific language impairment. Journal of Speech Language and Hearing Research. 1994; 37:418–421.

Levy R, Bicknell K, Slattery T, Rayner K. Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. Proceedings of the National Academy of Sciences. 2009; 106(50):21086–21090.

Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology. 1957; 54(5):358–368. [PubMed: 13481283]

Lisker L. "Voicing" in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. Language and Speech. 1986; 29(1):3–11. [PubMed: 3657346]

Lisker L, Abramson AS. A cross-language study of voicing in initial stops: acoustical measurements. Word. 1964; 20:384–422.

Lowe AD, Campbell RA. Temporal discrimination in aphasoid and normal children. Journal of Speech & Hearing Research. 1965; 8:313–314. [PubMed: 5863555]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19(1):1–36. [PubMed: 9504270]

MacDonald MC, Pearlmutter NJ, Seidenberg MS. Lexical nature of syntactic ambiguity resolution. Psychological Review. 1994; 101:676–703. [PubMed: 7984711]

Magnuson JS, McMurray B, Tanenhaus MK, Aslin RN. Lexical effects on compensation for coarticulation: The ghost of Christmash past. Cognitive Science. 2003; 27(2):285–298.

Mainela-Arnold E, Evans JL, Coady J. Lexical representations in children with SLI: Evidence from a frequency manipulated gating task. Journal of Speech Language and Hearing Research. 2008; 51:381–393.

Marslen-Wilson W. Functional parallelism in spoken word recognition. Cognition. 1987; 25(1–2):71–102. [PubMed: 3581730]

Marslen-Wilson W, Zwitserlood P. Accessing spoken words: The importance of word onsets. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15:576–585.

Massaro DW, Cohen MM. Categorical or continuous speech perception: a new test. Speech Communication. 1983; 2:15–35.

McArthur GM, Bishop DVM. Which People with Specific Language Impairment Have Auditory Processing Deficits? Cognitive Neuropsychology. 2004; 21(1):79–94. [PubMed: 21038192]

McArthur GM, Bishop DVM. Speech and non-speech processing in people with specific language impairment: A behavioural and elctrophysiological study. Brain and Language. 2005; 94:260–273. [PubMed: 16098376]

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18(1):1–86. [PubMed: 3753912]

McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? Trends in Cognitive Sciences. 2006; 10(8):363–369. [PubMed: 16843037]

McMurray B, Aslin RN, Tanenhaus MK, Spivey MJ, Subik D. Gradient sensitivity to within-category variation in words and syllables. Journal of Experimental Psychology, Human Perception and Performance. 2008; 34(6):1609–1631. [PubMed: 19045996]

McMurray B, Clayards M, Tanenhaus MK, Aslin RN. Tracking the time course of phonetic cue integration during spoken word recognition. Psychonomic Bulletin and Review. 2008; 15(6): 1064–1071. [PubMed: 19001568]

McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review. 2011; 118(2):219–246. [PubMed: 21417542]

McMurray B, Samelson VS, Lee SH, Tomblin JB. Individual differences in online spoken word recognition: Implications for SLI. Cognitive Psychology. 2010; 60(1):1–39. [PubMed: 19836014]

McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. Cognition. 2002; 86(2):B33–B42. [PubMed: 12435537]

McRae K, Spivey-Knowlton MJ, Tanenhaus MK. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language. 1998; 38:283–312.

Messaoud-Galusi S, Hazan V, Rosen S. Investigating Speech Perception in Children With Dyslexia: Is There Evidence of a Consistent Deficit in Individuals? Journal of Speech Language and Hearing Research. 2011; 54:1682–1701.

Miller CA, Kail R, Leonard LB, Tomblin JB. Speed of processing in children with specific language impairment. Journal of Speech Language and Hearing Research. 2001; 44:416–433.

Miller JL. Internal structure of phonetic categories. Language and Cognitive Processes. 1997; 12:865–869.

Montgomery J. Examining the nature of lexical processing in children with specific language impairment: a temporal processing or processing capacity deficit? Applied Psycholinguistics. 2002; 23:447–470.

Neville HJ, Coffey SA, Holcomb PJ, Tallal P. The neurobiology of sensory and language processing in language-impaired children. Journal of Cognitive Neuroscience. 1993; 5:235–253. [PubMed: 23972156]

Newman R, Sawusch JR, Luce PA. Lexical neighborhood effects in phonetic processing. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23(3):873–889. [PubMed: 9180048]

Norrelgen F, Lacerda F, Forssberg H. Temporal resolution of auditory perception and verbal working memory in 15 children with language impairment. Journal of Learning Disabilities. 2002; 35:540–546.

Norris D, McQueen J, Cutler A. Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences. 2000; 23(3):299–370. [PubMed: 11301575]

Nygaard L, Sommers M, Pisoni DB. Speech perception as a talker contingent process. Psychological Science. 1994; 5:42–46. [PubMed: 21526138]

Oden G, Massaro DW. Integration of featural information in speech perception. Psychological Review. 1978; 85(3):172–191. [PubMed: 663005]

Pierrehumbert JB. Phonetic diversity, statistical learning, and acquisition of phonology. Language and Speech. 2003; 46:115–154. [PubMed: 14748442]

Pisoni, DB. Some thoughts on "normalization" in speech perception. In: Johnson, K.; Mullennix, JW., editors. Talker Variability in Speech Processing. San Diego, CA: Academic Press; 1997. p. 9-32.

Pisoni DB, Lazarus JH. Categorical and noncategorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America. 1974; 55:328–333. [PubMed: 4821837]

Pisoni DB, Tash J. Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics. 1974; 15(2):285–290. [PubMed: 23226881]

Port R. How are words stored in memory? Beyond phones and phonemes. New Ideas in Psychology. 2007; 25:143–170.

Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. Psychological Review. 2000; 107:460–499. [PubMed: 10941277]

Remez R, Rubin P, Pisoni DB, Carrell T. Speech perception without traditional speech cues. Science. 1981; 212(22):947–950. [PubMed: 7233191]

Repp, B. Categorical perception: Issues, methods and findings. In: Lass, N., editor. Speech and Language (vol. 10): Advances in Basic Research and Practice. Orlando, FL: Academic Press; 1984. p. 244-335.

Robertson EK, Joanisse MF, Desroches AS, Ng S. Categorical speech perception deficits distinguish language and reading impairments in children. Developmental Science. 2009; 12(5):753–767. [PubMed: 19702768]

Rosen S. Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? Journal of Phonetics. 2003; 31(3–4):509–527.

Salverda AP, Dahan D, McQueen J. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition. 2003; 90(1):51–89. [PubMed: 14597270]

Schouten MEH, Gerrits E, Van Hessen A. The end of categorical perception as we know it. Speech Communication. 2003; 41:71–80.

Semel, E.; Wiig, E.; Secord, W. Clinical Evaluation of Language Fundamentals-3. San Antonio, TX: The Psychological Corp; 1995.

Shafer VL, Morr ML, Datta H, Kurtzberg D, Schwartz RG. Neurophysiological indices of speech processing deficits in children with specific language impairment. Journal of Cognitive Neuroscience. 2005; 17:1168–1180. [PubMed: 16138434]

Shannon R, Zeng F, Kamath V, Wygonski J. Speech recognition with primarily temporal cues. Science. 1995; 270(5234):303–304. [PubMed: 7569981]

Smith ZM, Delgutte B, Oxenham A. Chimaeric sounds reveal dichotomies in auditory perception. Nature. 2002; 416:87–90. [PubMed: 11882898]

Spivey, MJ.; Dale, R. On the continuity of mind: Toward a dynamical account of cognition. In: Ross, BH., editor. The psychology of learning and motivation: Advances in research and theory. Vol. Vol. 45. San Diego, CA: Elsevier; 2004. p. 87-142.

Stark RE, Montgomery J. Sentence processing in language-impaired children under conditions of filtering and time compression. Applied Psycholinguistics. 1995; 16:137–164.

Stollman M, van Velzen E, Simkens H, Snik A, Van den Broek P. Assessment of auditory processing in 6-year-old language-impaired children: Evaluacion del procesamiento auditivo en niños de 6 años con trastornos del lenguaje. International Journal of Audiology. 2003; 42(6):303–311. [PubMed: 14570237]

Sussman JE. Perception of formant transition cues to place of articulation in children with language impairments. Journal of Speech and Hearing Research. 1993; 36:1286–1299. [PubMed: 8114495]

Tallal P. An investigation of rapid auditory processing in normal and disordered language development. Journal of Speech and Hearing Research. 1976; 3:561–571. [PubMed: 979217]

Tallal P, Piercy M. Defects of non-verbal auditory perception in children with developmental aphasia. Nature. 1973; 241:468–469. [PubMed: 4705758]

Tallal P, Piercy M. Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. Neuropsychologia. 1974; 12:83–93. [PubMed: 4821193]

Tallal P, Piercy M. Developmental aphasia: The perception of brief vowels and extended stop consonants. Neuropsychologia. 1975; 13:69–74. [PubMed: 1109463]

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268:1632–1634. [PubMed: 7777863]

Thibodeau LM, Sussman H. Performance on a test of categorical perception of speech in normal and communication disordered children. Journal of Phonetics. 1979; 7:375–391.

Tomblin JB. Co-morbidity of autism and SLI: kinds, kin and complexity. International Journal of Language & Communication Disorders. 2011; 46(2):127–137. [PubMed: 21401812]

Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. Journal of Speech and Hearing Research. 1997; 40:1245–1260.

Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. Journal of Speech & Hearing Research. 1996; 39:1284–1294. [PubMed: 8959613]

Tomblin JB, Zhang X. Dimensionality of language ability in school age children. Journal of Speech Language and Hearing Research. 2006; 49:1193–1208.

Tomblin JB, Zhang X, Buckwalter P, O'Brien M. The Stability of Primary Language Disorder: Four Years After Kindergarten Diagnosis. Journal of Speech, Language, and Hearing Research. 2003; 46(6):1283–1296.

Toscano JC, McMurray B, Dennhardt J, Luck S. Continuous Perception and Graded Categorization Electrophysiological Evidence for a Linear Relationship Between the Acoustic Signal and Perceptual Encoding of Speech. Psychological Science. 2010; 21(10):1532–1540. [PubMed: 20935168]

Utman JA, Blumstein SE, Burton MW. Effects of subphonetic and syllable structure variation on word recognition. Perception & Psychophysics. 2000; 62(6):1297–1311. [PubMed: 11019625]

Wechsler, D. Wechsler Intelligence Scale for Children-Third Edition. San Antonio, TX: The Psychological Corporation; 1989.

Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. Nature. 1997; 387:176–178. [PubMed: 9144287]

Ziegler J, Pech-Georgel C, George F, Alario FX, Lorenzi C. Deficits in speech perception predict language learning impairment. Proceedings of the National Academy of Sciences. 2005; 102(39):14110–14115.

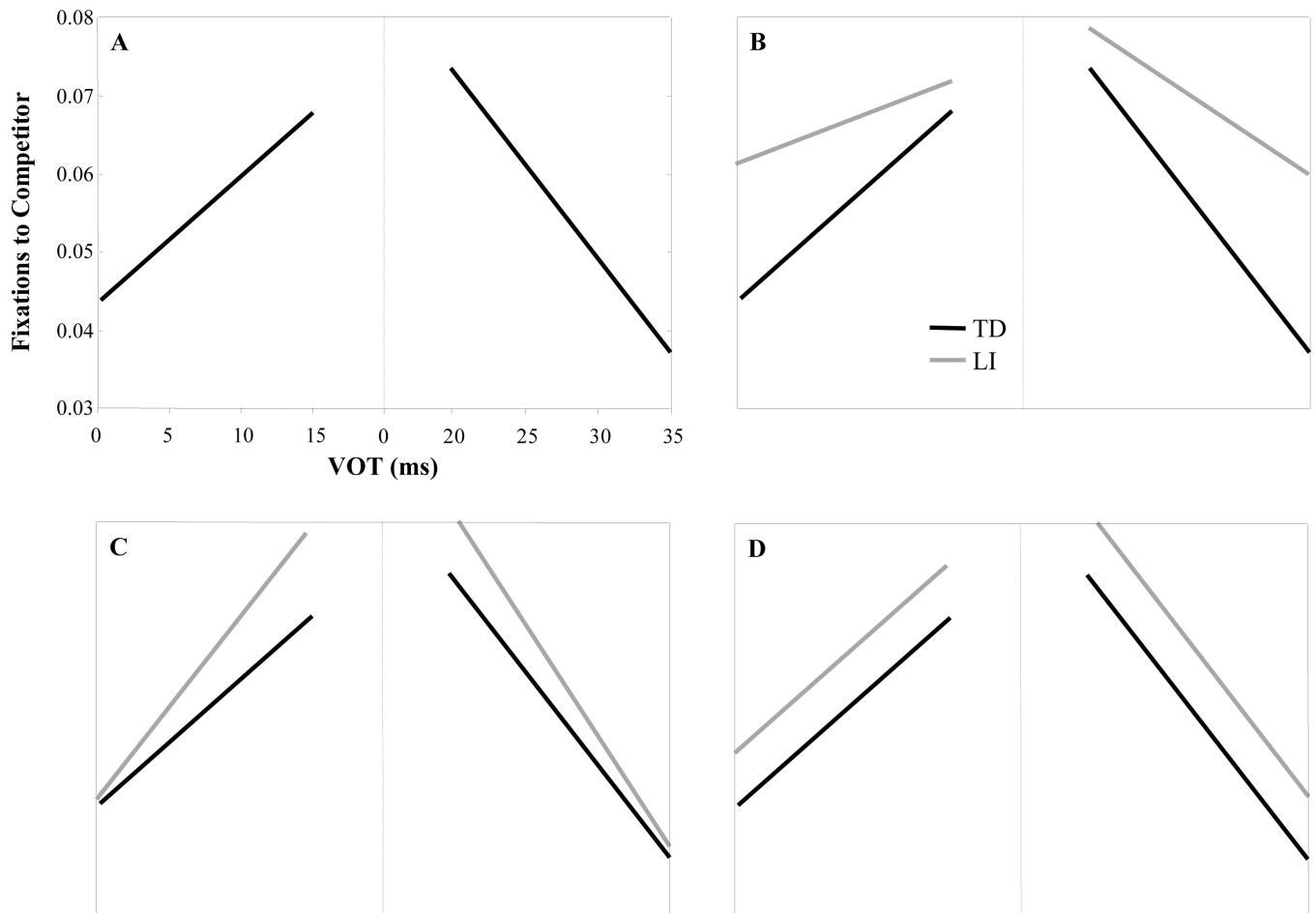**Figure 1.**
A) Schematic representation of McMurray et al. (2008a) results: looks to competitor increase as VOT approaches the category boundary. B) Results predicted by an auditory deficit that impairs encoding of VOT. C) Results predicted if LI listeners show less robust categories. D) Predicted results for a purely lexical deficit.
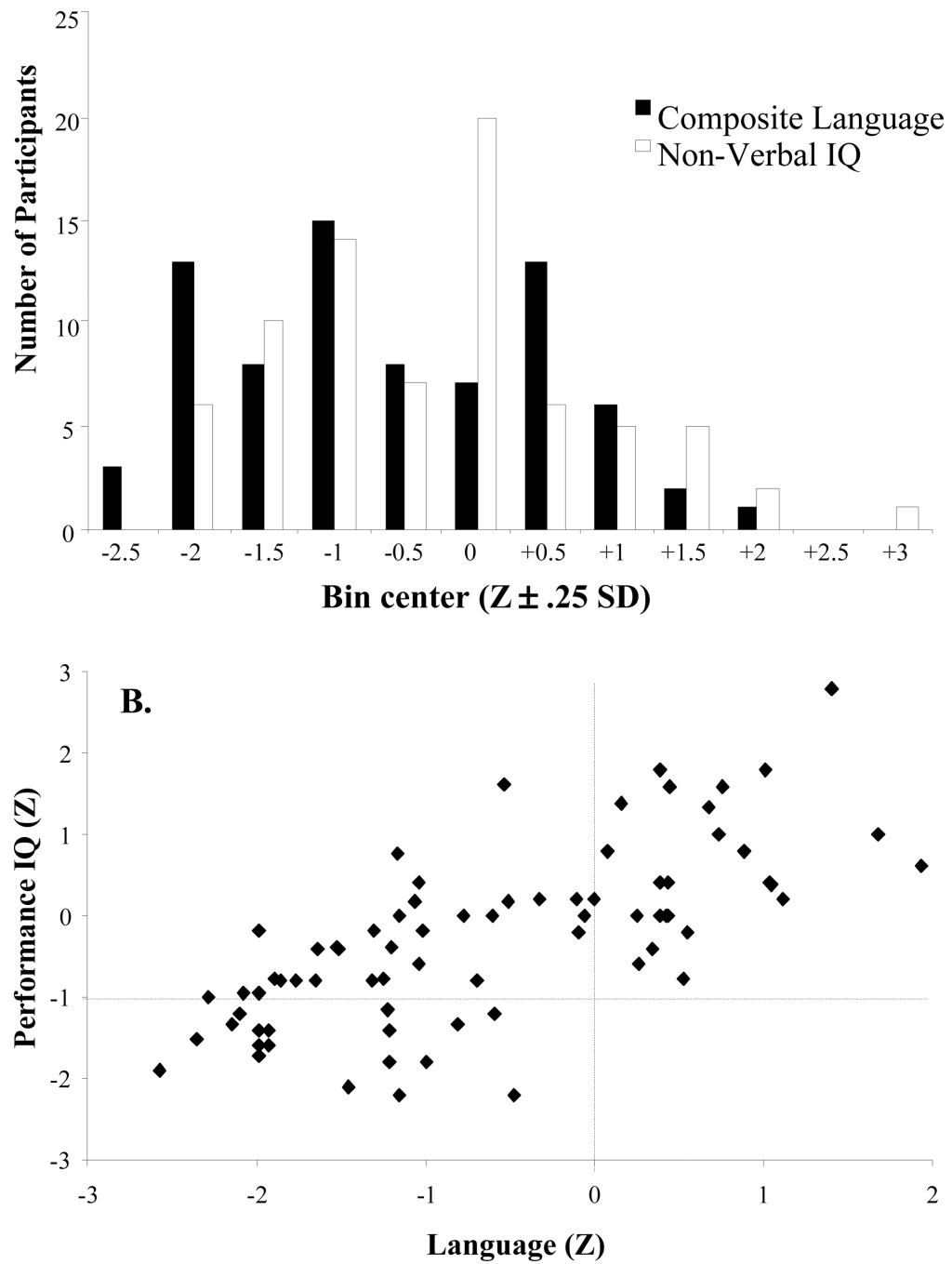
Figure 2.
A) Distribution of language and performance IQ scores. B) Scatter plot showing relationship between language and performance IQ.
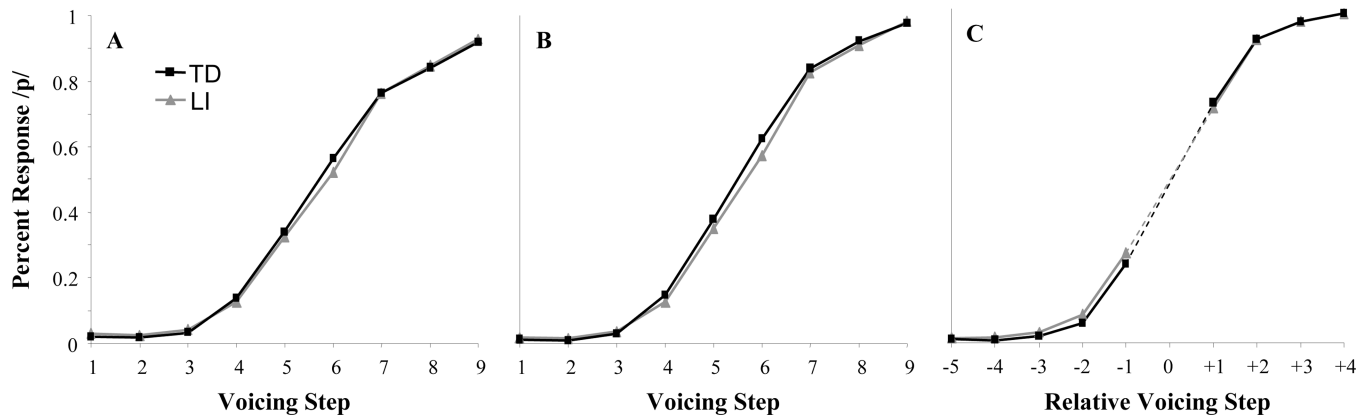
Figure 3.
Identification curves. A) Percent voiceless responding as a function of raw VOT and LI for all subjects and continua; B) Same data as A without the excluded continua. C) Percent voiceless responding as a function of rounded rStep. Note that all analyses used *language ability* as a continuous variable – data are grouped here for convenience only.
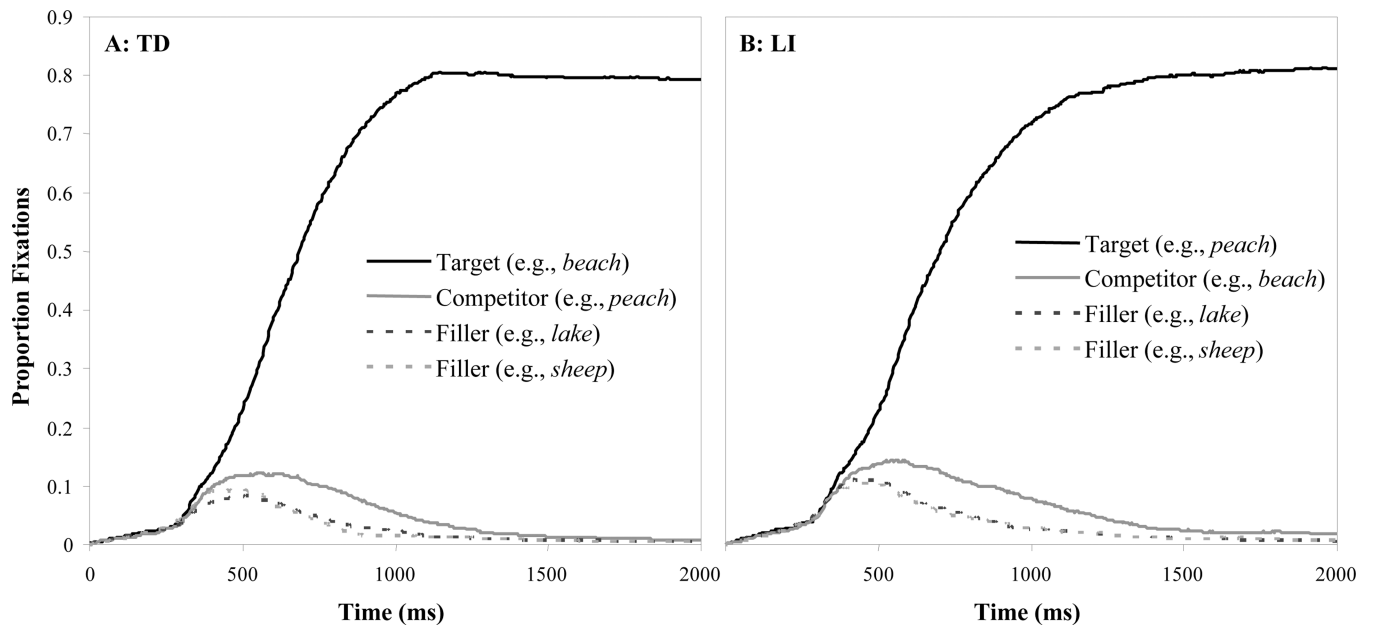
Figure 4.
Proportion fixations to the target, competitor and fillers as a function of time averaged across the two endpoints (0 and 40 ms VOTs). A) For typically developing participants. B) For participants classified as having LI.
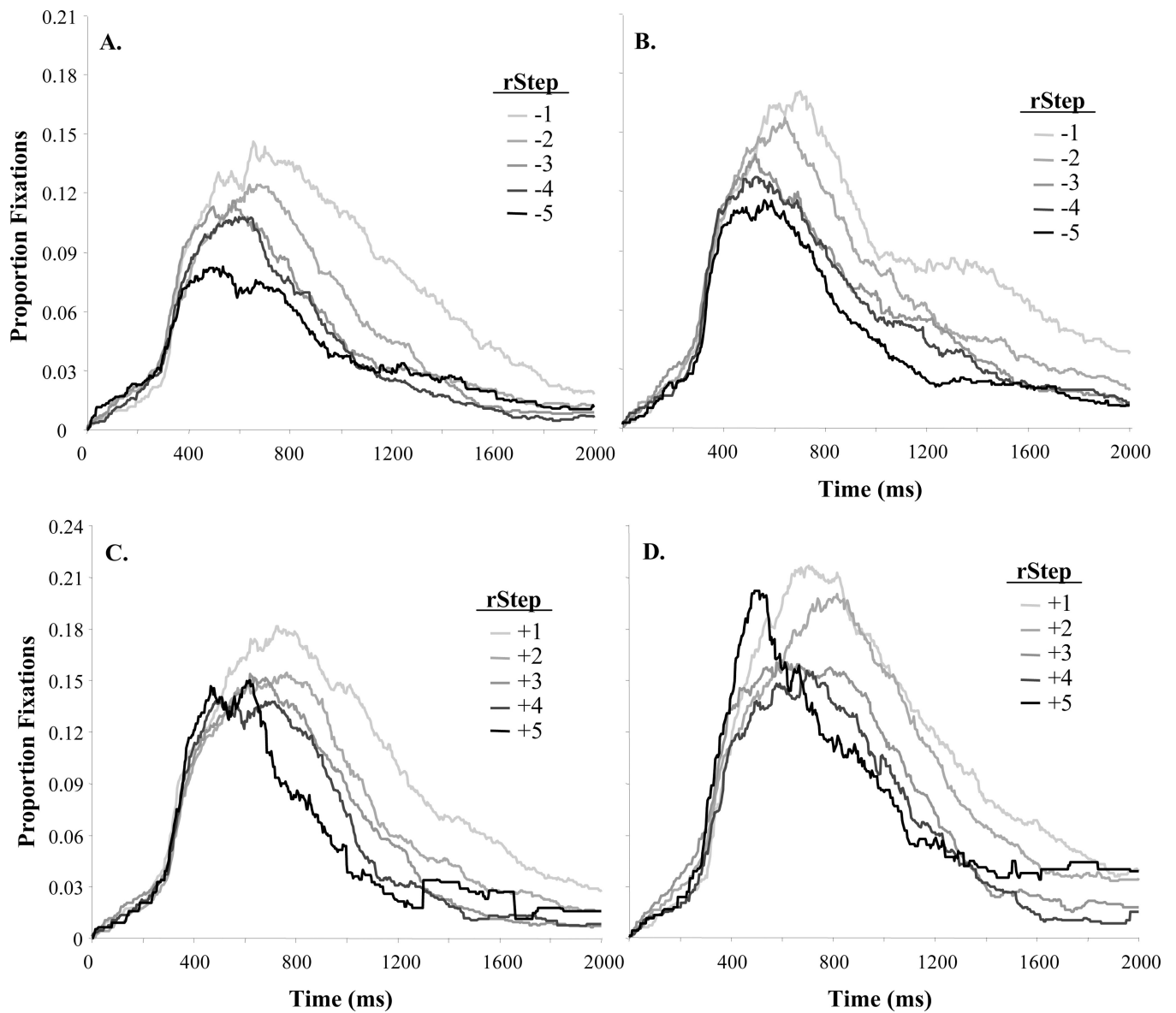
Figure 5.
Proportion competitor fixations as a function of time and rStep. A) TD participants, voiced tokens (VOTs less than the boundary; looks to /p/). B) LI participants, voiced tokens. C) TD participants, voiceless tokens. D) LI participants, voiceless tokens. Note that rStep and Language Ability were both treated as continuous variables for the analyses, even though shown as discrete here.

Figure 6.
Average proportion looks to the competitor between 250 and 1750 ms as a function of rStep (rounded) and language status (binned).

Table 1

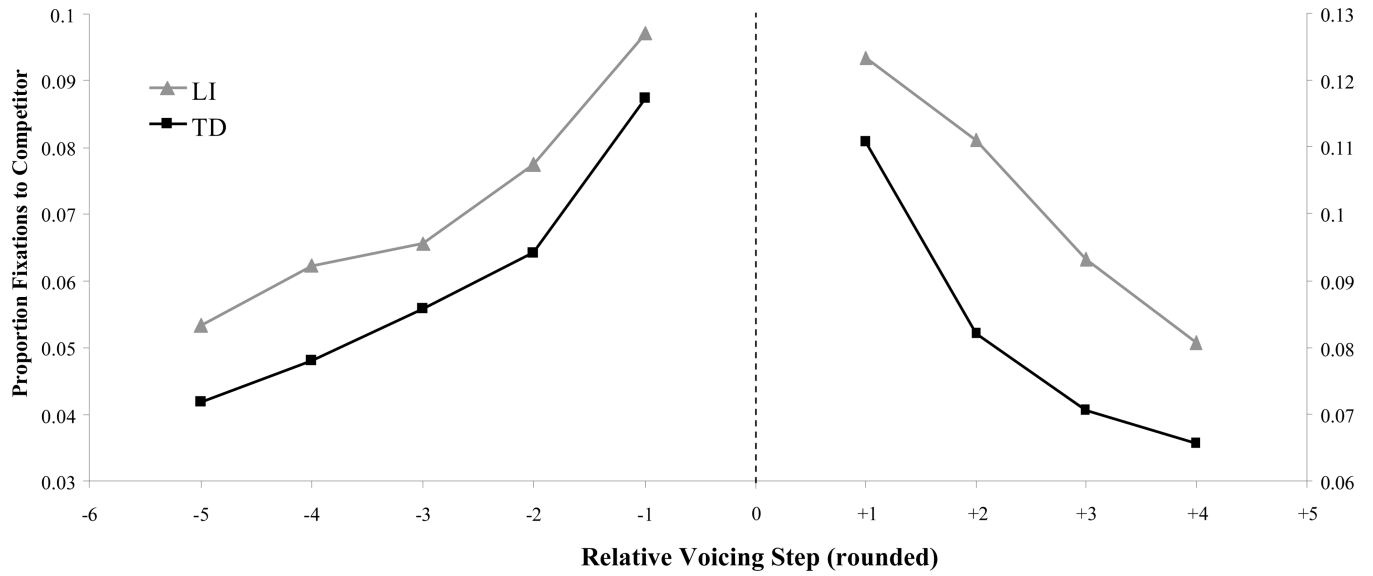Demographic and assessment data of the participants

	N	Age (Year)	Language Composite	Performance IQ	Females
TD	39	17.3 (1.44)	104.12 (10.52)	104.32 (15.38)	18
LI	35	17.99 (1.96)	75.58 (6.63)	86.59 (10.67)	15

Table 2

VOT measurements (in msec) for auditory stimuli. Continua are aligned at their boundary and each side labeled as voiced or voiceless

VOT (ms)	Bale/Pail	Beach/Peach	Beak/Peak	Bear/Pear	Bees/Peas	Bin/Pin
Voiced	0					
	5			0		
	7.9	0.7	0	4.9	0.8	3
	17.6	4.6	5.6	11.6	5.3	10.2
	21.8	11.5	12.4	18.9	13.1	15.7
	26.6	16.6	18	21.5	17.7	19.6
	32.2	20.1	20.7	27	20.9	24.4
Voiceless	34.3	24.7	24	30.7	25.6	26.3
	42.5	32.2	30.3	36.6	25.6	32.8
		33.2	35.2	41.1	30.8	36.8
		40.8	41.3		35.2	43.6
					41.3	

Table 3

Results of four ANOVAS on logistic parameters of the identification curves

Parameter	Factor	df	F	p
boundary	Continua	5, 355	36.7	<.0001
	Language	1, 71	<1	
	Language × continua	5, 355	1.0	.4
slope	Continua	5, 355	6.5	<.0001
	Language	1, 71	1.5	.2
	Language × continua	5, 355	<1	
max	Continua	5, 355	40.5	<.0001
	Language	1, 71	<1	
	Language × continua	5, 355	1.3	.2
min	continua	5, 355	4.3	.001
	Language	1, 71	4.8	.031
	Language × continua	5, 355	3.0	.012

Table 4

Percentage of participants contributing data as a function of the number of word pairs used for that participant, or for each specific word-pair

# of Word-Pairs	Percentage participants contributing data		Word-Pair	Percentage participants contributing data	
	LI	TD		LI	TD
3	9.7	0.2	bale	41.9	44.2
4	12.9	0.3	beach	100.0	100.0
5	51.6	39.5	beak	90.3	95.4
6	32.3	37.2	bear	83.9	79.1
			bees	100.0	97.7
			bin	93.6	95.4

Table 5

Results of linear mixed effects model analysis examining looks to competitors as a function of language ability and rStep

Analysis	Factor	Estimates from full model			Test of model fit	
		B	SE	T	$\chi^2(1)$	p
/b/	Language Ability	−.0083	.0023	3.57	14.0	.00018
	rStep	.0093	.0017	5.58	12.5	.0004
	Language × rStep	.0008	.0009	<1	0.8	.36
/p/	Language Ability	−.0110	.0031	3.54	11.2	.0008
	rStep	−.0158	.0024	6.32	13.9	.0002
	Language × rStep	.0014	.0013	1.07	1.2	.28

Table 6

Results of linear mixed effects model analysis examining looks to competitors after unrelated fixations had been partialed out as a function of language ability and rStep

Analysis	Factor	Estimates from full model			Test of model fit	
		B	SE	T	$\chi^2(1)$	p
/b/	Language Ability	−.0045	.0074	2.59	11.10	.00086
	rStep	.0090	.0017	5.54	12.41	.00043
	Language × rStep	.0007	.0009	.82	.65	.41
/p/	Language Ability	−.0074	.0027	2.73	6.32	.011
	rStep	−.0154	.0025	6.10	13.39	.00025
	Language × rStep	.0014	.0014	1.1	1.17	.27