S1: Analysis of Unfiltered Data

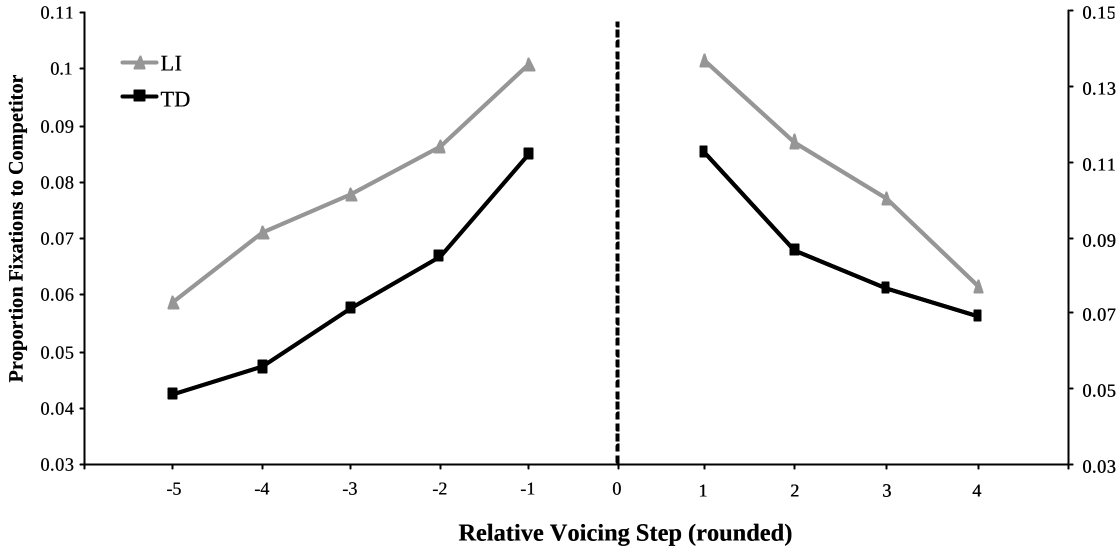
Following McMurray et al., (2008; 2002) our analysis were predicated on examining fixations for only truly within-category variation in VOT. This allows us to disentangle differences in the response to VOTs (e.g., an auditory account) from differences in how VOTs are mapped onto categories (e.g., a phonological account), since the category is held constant. The logic of focusing on within-category variation is based on several prior studies with typically developing adults (McMurray, et al., 2008; McMurray, et al., 2002) which attempted to isolate the structure of a category and the sensitivity to within-category changes in VOT from any changes in the accuracy of categorization. This is essential for observing the auditory sensitivity to differences in VOT in way that would not be affected by categories and for isolating the structure of the category (relative to VOT) in a way that is independent of the boundary between words. That is, we needed to know that the heightened fixations to the competitor for VOTs near the boundary do not result 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. By only examining fixations to trials that were all classified in the same way, we could be sure that differences in the response reflect primarily difference in the response to VOT, not differences in categorization.

To be confident in this assertion, we took the following steps. First, we computed the boundary for each participant, for each word-pair and recomputed VOT relative to that boundary (*rStep*). Second, we discarded any trials in which the VOT was on the voiced side of the boundary and the participant responded with the other category. This allowed us to quantify changes in the response to VOT even for tokens from the same category, a precise measure of the structure of the category relative to changes in VOT that is independent of changes in the category itself. Finally, to be a valid measure of lexical activation the listeners must know all of the relevant words (for a given trial), so we excluded word-pairs for which participants were not accurate at endpoint responding.

In order to achieve this, we needed to be confident in the participant/word-pair category boundary; we needed to have sufficient data on each side of the continuum to estimate the effect of VOT; and we needed to know that the participant knew both words for the measure of lexical activation to be valid. As a result, we excluded from analysis particular word-pairs for each participant based on two criteria. First, we excluded any word-pair for which we were not able to obtain a reliable fit of the logistic function from its identification data ($R^2 < .80$). This criterion ensured that we could be confident in the category boundary for each word-pair in the analysis. Second, we excluded any word-pair for which the participant could not accurately identify each end-point at greater than 80% accuracy. This ensured that participants knew both words and would have some tokens on each side of the boundary.

In the main paper, our analysis of the identification data suggested that 1) goodness of fit did not vary as a function of language ability; and that 2) we excluded an equal number of word-pairs for LI and TD participants. Thus, these exclusions were not biased toward (or against) individuals the low or high end of the language scale. As a result, it seems likely that these excluded word-pairs represent simply noise – words that were unfamiliar to our sample, or idiosyncratic responses to our stimuli (auditory and visual). However, at the same time, this does represent a somewhat sizeable portion of the data that was excluded. Thus, here we replicate our primary analysis using the full dataset.

The primary analysis was a linear mixed model with proportion looks to the competitor between 250 and 1750 ms as the DV. Composite language ability (continuous, between-subject)



Supplementary Figure 1: Average proportion looks to the competitor between 250 and 1750 ms as a function of rStep (rounded) and language status (binned) for all of the items/participants.

and rStep (continuous, within-subject) were the only fixed effects, and the random effects included random slopes of rStep on both subject and word-pair. Separate analyses were conducted on both the voiced and voiceless sides of the continuum. Means are displayed in Figure S1.

Results of this analysis are shown in Table S1. The maximum correlation among fixed effects was low (Voiced: R=-.006; Voiceless: R=-.009). On the voiceless side, the results corresponded quite well with the more restrictive analysis reported in the paper: a main effect of language (B=-.011, p=.0007), a main effect of rStep (B=-.014, p=.0005), and no interaction (p=.5). On the voiced side, we also found a main effect of language (B=-.01, p<.0001), and a main effect of rStep (B=.008, p=.0008). However, there was a marginal interaction (B=.001, p=.085), which took the form of a slightly reduced effect of rStep for listeners with poor language.

Table S1: Results of linear mixed effects model analysis examining looks to competitors as a function of language ability and rStep for the full dataset.

Analysis	Factor	*Estimates from full model*			*Test of model fit*	
		B	***SE***	***T***	*c^2(1)*	***p***
/p/	Language Ability	-.011	.01	3.4	11.5	.0007
	rStep	-.014	.00	5.5	12.2	.0005
	Language × rStep	.0008	.00			.5
/b/	Language Ability	-.010	.00	4.5	19.3	<.0001
	rStep	.008	.00	5.1	11.2	.0008
	Language × rStep	.001	.000	1.7	3.0	.085
/b/ *(rStep > -6)*	Language × rStep	.001	.000	1.5	2.1	.14

However, the mean values shown in Supplementary Figure 1 did not appear to support an interaction, leading us to look closer at this analysis. Here, we discovered that there were a number of participants in this analysis with rSteps as low as -8 (typically for the *bale/pail* continuum, where some participants did not recognize *pail*). Here, with rSteps this low, participants had only one token in which they heard /p/, and likely for only one word-pair, making it difficult to estimate effects at this range. Moreover, if even the boundary were valid, one would expect any gradient effects to taper off for /b/'s this unambiguous, potentially yielding a nonlinearity. Thus, we reran the model continuing to use all participants and word pairs, but this time excluding any token with an rStep less than or equal to -6 (the most extremes). In this analysis, the interaction was now not significant (p=.15).

As a whole then, this analysis confirms the story from the more restrictive analysis with a main effect of rStep and language and little evidence for an interaction.

S2: Analysis of Fixations to Unrelated Objects

Our primary finding was that children with lower language ability showed heightened fixations to lexical competitor objects, and that this did not interact with their sensitivity to within-category differences in VOT. However, it was possible that these heightened fixations may not reflect heightened activation or consideration of these items, but rather reflect a more general uncertainty or differences in visual-cognitive processes like visual search or eye-movement control. While McMurray et al., (2010) performed extensive analyses to rule this out in earlier VWP work with a similar population, it was important to address it in the context of this study.

Any differences in purely fixation processes should appear as heightened fixations to unrelated objects, and indeed a comparison of the unrelated items in Figure 4 in the main paper (across panels A and B) suggests this may have been the case. There are two ways to think about such an effect. First, it could reflect differences in visual abilities (participants with lower language abilities are just less confident therefore look around more), and these differences (not differences in language ability) could drive the effect of language on competitor fixations. In this case, these differences must be accounted for before we can evaluate the effects of language. Alternatively, however, it is possible that enhanced unrelated fixations also derive from changes in basic word recognition processes like decay or inhibition and are another marker of our effect. Indeed, our work using the TRACE model suggests that variation in the same parameters that model LI can affect unrelated activation/fixations (McMurray, et al., 2010), so we do not want to completely dismiss such an effect.

To examine this possibility we conducted an analysis of looks to unrelated items as a function of language ability. The most intuitive approach would be to examine the /l/ and /? / initial objects on the experimental trials. However, if language-related processes are drawing fixations to the /b/ and /p/ objects, this may reduce fixations to the unrelated objects, making these trials an impure measure of basic visual processes. The filler trials may represent a better estimate of such effects as the target word (e.g., *lamp*) can be disambiguated from the competitors (*beach, peach*) from the earliest moments and there is no phonetic ambiguity to deal with. These should be less affected by language ability and looks to the unrelated objects here (in this case, the /b/ and /p/-initial objects) may offer a purer measure of visual and oculomotor processes.

We thus computed the average proportion of looks to the /b/ and /p/ items on filler trials for each participant and word-pair over the same 250-1750 ms time window and used this as the DV in a series of mixed effects models using participant and word-pair as random intercepts. We wanted to include both language and non-verbal IQ as covariates as we suspected they may play unique roles in predicting basic function. However, because these were highly collinear (R=.7) these could not be added directly. Thus, to eliminate the collinearity between language and IQ we residualized IQ against language. This yielded a new variable, IQr, which describes variation in performance IQ over and above what would be predicted for that level of language performance. We thus used both language and IQr as continuous fixed effects. These were used in a sort of communality analysis to determine if there were effects of IQ over and above language and vice versa. Since there were no within-subjects factors, random slopes were not possible and only random intercepts of subject and word-pair were included in the model. Without random slopes, we could compute p-values using the Monte Carlo Markov Chain (MCMC) procedure with 20000 iterations.

The first analysis examined two models, one with just language ability as a fixed effect, and one adding residualized IQ (IQr). We found that the model with both factors was a significantly better fit ($\chi^2(2)=12.48$, p=.0019) than the model with language alone. This suggests an effect of IQ on unrelated fixations over and above language. In this model, both language and IQr were negatively correlated with fixations (Language: B=-.0036, SE=.0015, p_{mcmc}=.0138; IQr: -.0077, SE=.-023, p_{mcmc}=.0009), and there was an interaction such that the increase in fixations was enhanced for adolescents that were low on both scales (B=.0051, SE=.0022, p_{mcmc}=.025). Thus, adolescents with lower language were more likely to fixate non-target objects in general, and over and above this, children with lower IQs showed even more fixations.

Intriguingly, when we ran the model in the other order, the picture changed dramatically. Here we used the same random effects structure with raw IQ (not residualized), and language, residualized against IQ. Here, IQ was highly significant (B=-.0058, SE=.0016, p_{mcmc}=.0003), but adding residualized language offered no additional benefit ($\chi^2(2)=.01$, p=.99). This suggests that the individual differences on fixations to *unrelated objects* may be largely due to differences in IQ.

Supplement References

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology, Human Perception and Performance*, *34*(6), 1609-1631.

McMurray, B., Samelson, V. S., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60*(1), 1-39.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33-B42.