ONLINE SUPPLEMENT TO

Speech categorization develops slowly through adolescence.

Bob McMurray
Dept. of Psychological and Brain Sciences
Dept. of Communication Sciences and Disorders, Dept. of Linguistics, Dept. of Otolaryngology
University of Iowa

Ani Danelz
Dept. of Communication Sciences and Disorders
University of Iowa

Hannah Rigler
Dept. of Psychology
University of Iowa

Michael Seedorff
Dept. of Biostatistics
University of Iowa

Supplement S1: What causes a shallower identification slope?

Much of the work on speech perception in older children (or impaired populations) uses a task in which participants hear tokens from a continuum (e.g., /b/ to /p/) and make a two alternative forced choice (2AFC) decision as to the phoneme identity. These create identification functions that look like Figure S1, with typically shallower slopes for younger children. The typical interpretation of such data is that is that the shallower slopes of younger listeners likely reflects a more gradient (or less categorical) form of responding. By this view, development leads to a more categorical mode of speech perception.

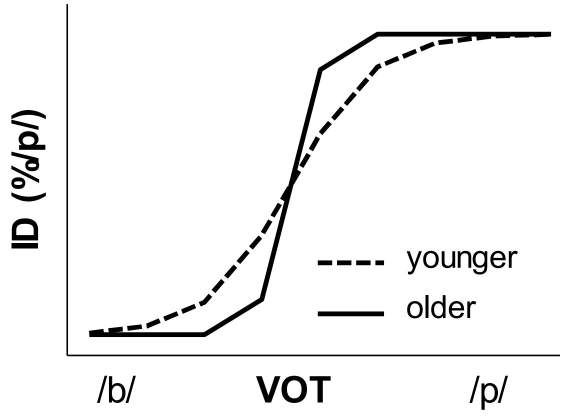

Figure S1: Schematic results from phoneme decision experiments. Here participants heard tokens from a VOT continua spanning /b/ and /p/ and decided whether each one was /b/ or /p/. Typical studies observe that the slope of the function (at the transition region) gets steeper (more step like) with age.

Even in something as simple as 2AFC categorization, there are multiple loci where a developmental change could lead to a shallower or steeper categorization slope. The goal of this study was to use a specific version of the Visual World Paradigm that can isolate changes in one factor – the underlying structure of the category – from some of those alternatives. Here, we describe a simple conceptual model to illustrate the range of possibilities in more detail than could be afforded in the main text.

The Model

Consider a simple model of speech perception (Figure S2) in which continuous acoustic cues like VOT are encoded as a location along the "maps" at the bottom of the model. That is when a VOT of 18 msec is heard it activates nodes in the center of the map, and a VOT of 42 msec activates nodes on the right side[1]. This cue-level encoding is then mapped to categories using the vertical connections. These are set up such that different cue values or locations along this map (e.g., a VOT of 5 or 10 msec) may differentially activate /b/ or /p/. This mapping implicitly encodes the boundary where cue-values stop mapping to one category and start mapping to the other.

Given this model, what are the possible loci by which identification functions are made to be steeper or shallower?

Discrete Categories: Steep Slopes

Under this model, sharp categories (as in older children) can be achieved if the mapping of cues to categories was discrete such that every cue-value on one side of the boundary is mapped to only one category. This is shown in Figure S2A where the VOTs near the boundary (18 and 24 msec) have only a strong connection to only one category – there is no ambiguity about which phoneme they map to. This can only lead to sharp categorization if cue values are encoded with

---

[1] Note that we've illustrated this in the form of a simple connectionist or associative network, but the logic of these mechanisms applies to any categorization model. We chose this form for ease of explication only.

low noise so that a given input only activates a single unit on the map (e.g., if the listener heard a VOT of 18 msec, this is the only node that was activated). Otherwise, this noise (e.g., a VOT of 18 being heard as 24) could cause listeners to activate the other category

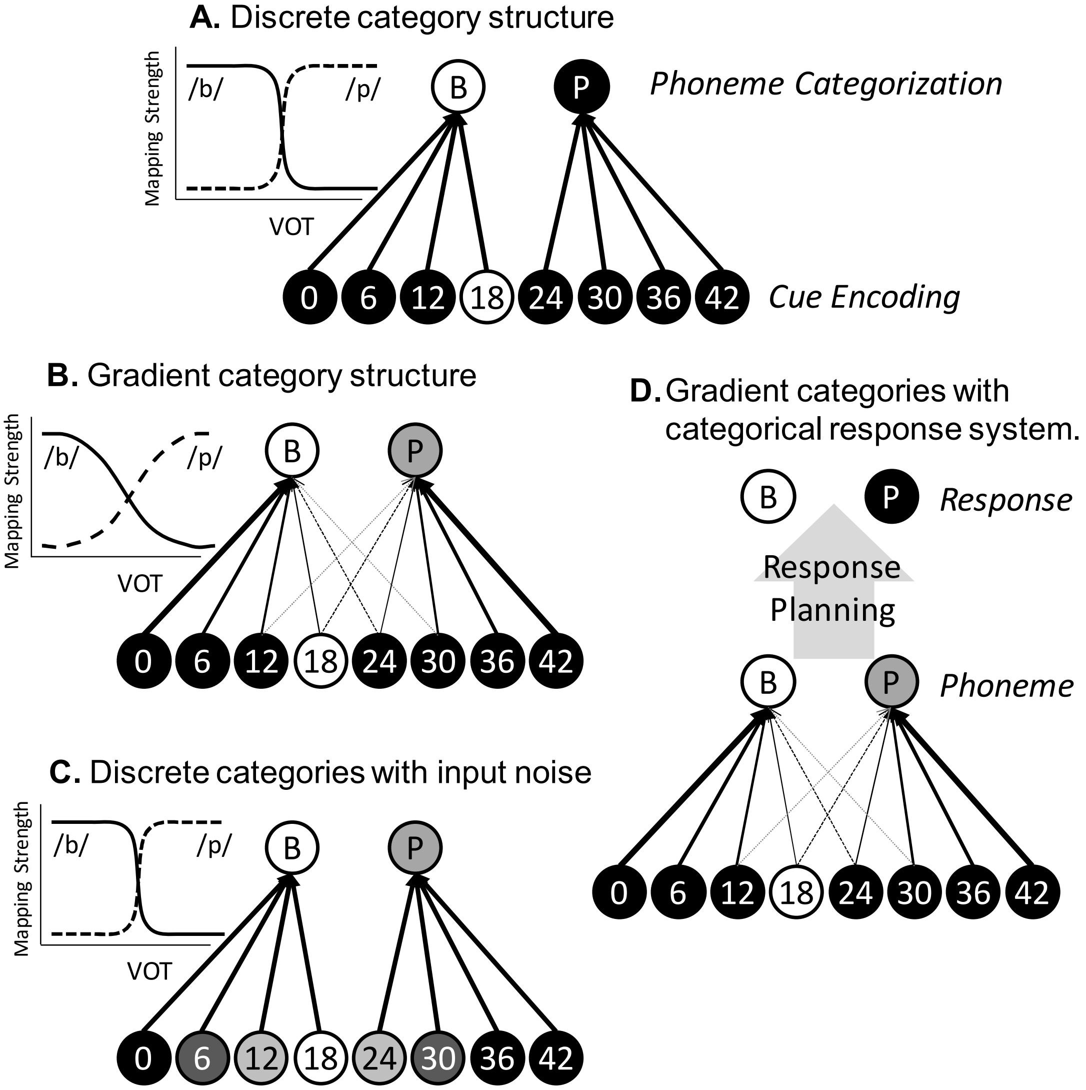

Figure S2: Schematic model illustrating potential loci for change in 2AFC categorization decisions. Cues like VOT are encoded an activated region of a "map". These are then mapped to phoneme categories (indicated by the associative links, and by the inset figures which show the strength of mapping across the input map to each category). This mapping is the learned structure of the category. A) Sharp categorization is achieved if each location on the map links discretely to one phoneme. B) One way to achieve shallower slopes is with a more gradient category structure in which regions of the input space (map) are partially associated with both categories. C) Even if the mapping is discrete, if multiple locations of the map are activated by a given VOT (noisy cue encoding), both categories are partially activated, particularly near the boundary. This makes the slope shallower. D) Here we add a response system which maps activation for phonemes to responses in the experimental task. If internal mappings are gradient, this response system could create a steeper slope in the identification task by responding perfectly consistently (any time /b/ is more active than /p/, respond /b/), rather than matching the probability of a response to the level of activation.

Category Structure: Shallower Slopes
One typical interpretation of a shallower slope in a 2AFC task is that the representation of the categories is more gradient and more overlapping. In this case, the mappings could be somewhat gradient (as a function of VOT), with tokens near the boundary being partially mapped to both categories (Figure S2B). Consequently, when a VOT of 18 (near the boundary) is heard, it will partially activate both /b/ and /p/; however, when a VOT of 0 msec (a good /b/) is heard, it will only activate /b/. With development, then it is assumed that the structure of the category moves from a gradient/overlapping one to the more discrete mapping in Figure S2A.

Cue-Encoding Noise: Shallower slopes
An alternative way to achieve shallower identification slopes (and consistent with much older thinking in psychophysics), is that both young and old listeners have a perfectly discrete category structure. However, younger children could encode cues like VOT more noisily (Figure S2C). Thus, a VOT of 18 msec may be sometimes encoded as 12 or 24 msec. When an 18 msec VOT (a /b/) is misheard as 24 this crosses the boundary, resulting in a /p/ response on some trials (or partial activation for /p/ on the same trial). In contrast, when a 0 msec is heard as 6, the response is still a /b/. Thus, this noise will flatten the function near the boundary and it is the elimination of this noise with development creates a sharper function.

Disentangling these hypotheses
Both of these above models can explain how younger listeners may show shallower slopes, but they differ quite dramatically in the nature of the categories that support this behavior, and therefore what must change to account for development.

The version of the visual world paradigm (VWP) used here (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2002) was designed to disentangle these hypotheses. In this task, analysis of the fixations is conditioned on the ultimate response. Under an encoding noise account, if a VOT of 18 msec was misheard as 24 msec, this would result in the incorrect response (a /p/ instead of a /b/) and the trial would be excluded. Thus, any differences in the fixations as a function of VOT likely derives from differences in the strength of the mapping.

The Response System: Steeper Slopes
As described in the main text, the evidence with adults is that speech categories are gradient and this may be useful for perception (Andruski, Blumstein, & Burton, 1994; Clayards, Tanenhaus, Aslin, & Jacobs, 2008; McMurray, Tanenhaus, & Aslin, 2009). If we start from the assumption of more gradient category structures, is there a way to achieve steeper (more categorical) slope over development?

One possibility is to consider a response system whose goal is to map underlying activation for various phonemes to the particular responses available in the experimental task (Luce, 1959). If we assume a response system of this sort (shown in Figure S2D), it would have several options when dealing with partial activation for multiple phonemes. One possibility would be to probability match – if /b/ is slightly more active than /p/ for a given VOT, choose /b/ on some trials, or /p/ on the others. However, the response system could also impose a more discrete rule: always choose the most active phoneme, regardless of whether it is only a little more active or a lot more active than its competitors. Indeed Nearey and Hogan (1986) have argued that this may be the optimal response strategy, as listeners basically strive for consistency across trials. Here, then categories could be gradient throughout development (beginning in infancy: Galle &

McMurray, 2014; McMurray & Aslin, 2005; Miller & Eimas, 1996), and what is developing is this response system, moving from a probabilistic to consistent mode of mapping phoneme-level activation to the response. A more subtle version of this hypothesis is that downstream lexical processes like lexical inhibition (Dahan, Magnuson, Tanenhaus, & Hogan, 2001) could play this kind of role, resolving the ambiguity more completely at the level of lexical activation.

This issue as well can be partially controlled with the version of the VWP used here allowing a clearer estimate of the category structure. Eye-movements in this task are implicit and probabilistic, and are not entirely coupled to the ultimate response – listeners can fixate a competitor (at least for some portion of the trial) while ultimately clicking on something else. While they are not immune to response-level demands, they may be less susceptible to this and reveal the underlying category structure than more explicit meta-linguistic tasks like phoneme judgements.

Supplement S2: Graphical Explanation of Fricative Construction Methods

Fricative continua were constructed via a method first developed by (Galle, Klein-Packard, Schreiber, & McMurray, submitted; McMurray, Farris-Trimble, Seedorff, & Rigler, 2016), and see https://osf.io/vz6wp/ for code and examples. Fricatives were based on recordings of naturally produced /s/- and /ʃ/-initial word. First, we excised frication portions of the /s/ and /ʃ/ recordings (Figure S3A). This was done at the last zero crossing where high frequency frication was observed. Frication segments were then equated in length by removing material at zero-crossings from the middle of the longer fricative.

Next, we computed the long-term average spectrum for each fricative (Figure S3B). The /s/ and /ʃ/ spectra were then shifted in frequency space to have the same spectral mean (Figure S3C), and the power of the two mean-centered spectra was averaged to create eight steps (Figure S3D). Next, the locations of these spectra were shifted in 8 steps from the spectral mean of a /s/ to that of a /ʃ/ to create the continuum (Figure S3E).

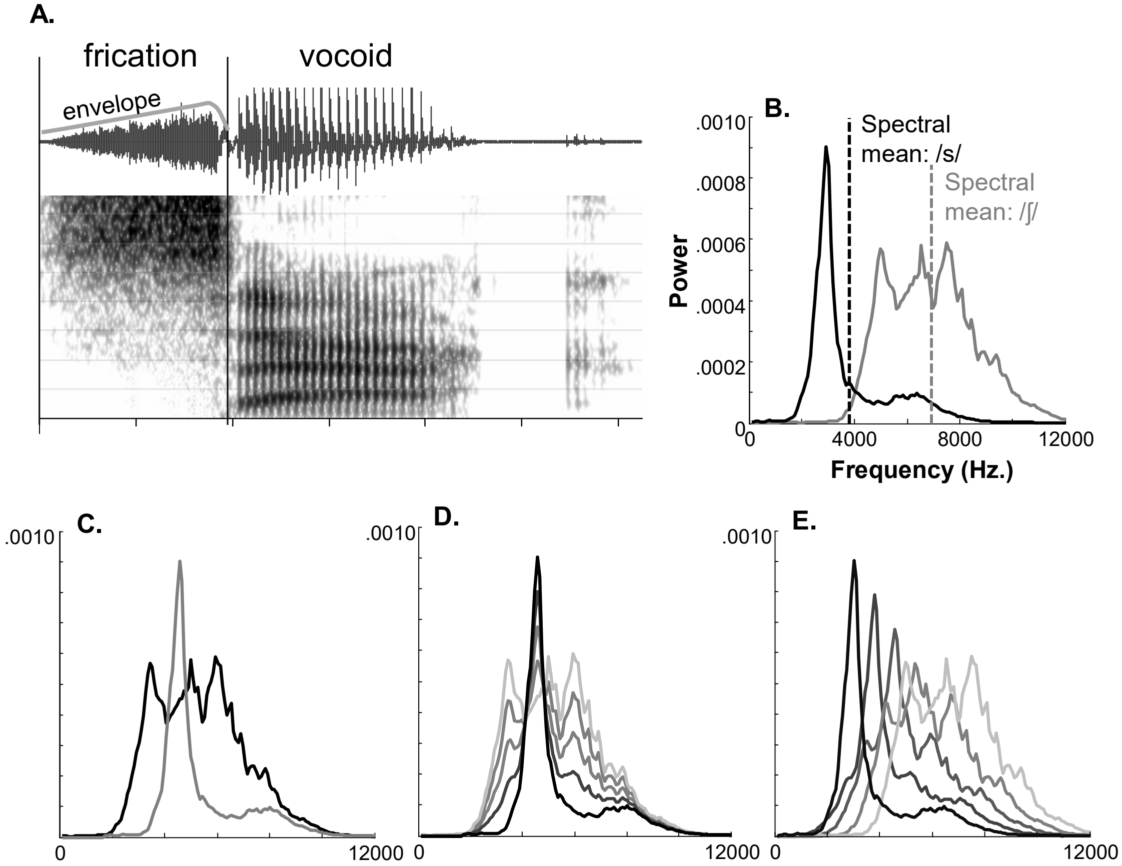

Figure S3: Fricative construction. A) Frication portions are spliced off of natural fricative recordings; B) Long term average spectra for the frication portions are extracted, and their spectral means are computed; C) Spectra are shifted to have the same spectral mean; D) Center-equated spectra are averaged in equal steps (shown are 6 steps, but the actual continua used 8 steps); E) Resulting spectra are then re-shifted in frequency space to means spaced in equal steps between that of the /s/ and /ʃ/. These spectra are then used to filter white noise, the average envelope (marked in A) is imposed on the noise, and the resulting fricative is spliced onto a neutral vocoid.

Subsequently, these spectra were used to filter white noise to create a frication noise with the appropriate spectrum. Finally, we computed the average envelope of the /s/ and /ʃ/ and imposed it on this noise to create the final fricative. These were then spliced onto a neutral vocoid (recorded in the context of an h-initial word). This process was repeated for each word pair.

Complete code for this is available at the Open Science Framework website for (Galle et al., submitted): https://osf.io/vz6wp/.

Supplement S3: Alternative Analysis of Categorization Slope

Motivation.

An alternative approach to analyzing the mouse-click responding is to fit logistic functions to each subjects' data and use the slope of the function as a dependent measure in an ANOVA or mixed model. Many prior studies using this paradigm have adopted this approach (e.g., McMurray et al., 2008; McMurray, Munson, & Tomblin, 2014).

In the present study, we switched to a binomial mixed model (which assumes a two parameter logistic, instead of four parameter logistic we can extract with nonlinear curvefitting) primarily because our dataset was sparser than prior work (we had fewer trials per subject). As a result, It was not possible to get separate fits by continuum (e.g., *beach/peach* vs. *bear/pear*) within each subject. Consequently, more traditional curvefits would have missed a lot of crucial item-level variance. More importantly, this can also create artificially shallow slopes. For example, if the youngest listeners had perfectly steep slopes, but had more variable boundaries between continua (e.g., between the *beach/peach* and *bear/pear*), the average across multiple continua would have appear to have a shallow slope, even as each individual continua would have a steep slope. The more traditional curvefitting approach would miss this (unless there was enough data to fit separate curves by subject and continuum).

In contrast, a mixed model working on trial–level data can account for this kind of item-variability more appropriately, by working across all of the data but with a random effect of item. The limit with this approach, however, is that a binomial model (by far the most robust non-linear mixed model) cannot handle differences at the asymptotes (since it uses a two parameter logistic). In this case, a binomial model was justified because the asymptotes in general were very close to 0 and 1 in each group for both the b/p (Lower: M_{young}= .014, M_{middle}=.007, M_{old}=.004; Upper: M_{young}=.972, M_{middle}=.991, M_{old}=.998) and the s/ʃ (Lower: M_{young}= .019, M_{middle}=.016, M_{old}=.010; Upper: M_{young}=.986, M_{middle}=.986, M_{old}=.987) continua.

Nonetheless, for consistency with prior work, we also had slope estimates (since we needed to conduct the logistic fits to compute relative step for the eye-tracking analyses. Thus, here we present an analysis of these slopes here for consistency.

Approach.

Four-parameter logistic functions (Equation 2, main text) were fit to the average proportion /p/ (or /s/) responding for each subject for each continuum type (stop voicing and fricative place). We averaged across the different items within a continua. These were the same fits used to compute relative step for the analysis of the VWP data. Resulting slope estimates were compared with ANOVA with follow-up t-tests between ages.

Results.

Table S1 shows a summary of the mean slope estimates. For the b/p slopes this shows an increase between each age. Concordant with this an ANOVA showed a main effect of age group (F(2,71)=11.02, p<.0001). Follow-up t-tests showed a significant difference between the 7-8 and 12-13 year olds (t(47)=2.75, p=.009), but not between the two older groups (t(47)=1.56, p=.127).

The ʃ/s slopes were quite a bit shallower than the b/p slopes, with large difference between the two youngest groups. The overall ANOVA was not significant (F(2,71)=1.69, p=.193). However, the difference between the two youngest groups was marginally significant (t(47)=1.89, p=.065), while the two older groups did not differ (t(47)=.67, p=.51).

Table S1: Mean Identification slope (Δ %p / step at the midpoint) for each age group. SD is given in parenthesis.

	b/p	ʃ/s
Young	0.739 (.186)	0.465 (.116)
Middle	0.880 (.172)	0.547 (.182)
Old	0.941 (.094)	0.514 (.168)

Summary.

Thus, these analyses largely mirror the effects shown with the binomial model in the main text. While that model did detect a significant effect for voicing in the older age groups (and this analysis did not), effects were in the predicted direction numerically. While we did not detect differences in the slope for fricatives, the 7-8 vs. 12-13 contrast was marginally significant (and the older contrast, which was not significant here, was also not significant in the binomial model). We suspect that this analysis was less statistically sensitive to some of these effects because some of the variation in slope (across children) likely derives from averaging across continua. While our binomial models could account for this, the curvefitting could not.

Supplement S4: Dynamics of Heightened Lexical Competition

Motivation.

The primary results of this study were to examine developmental changes in speech categorization across multi loci of potential change. In our analysis of within-category sensitivity to VOT and frication step, we found main effects of age on competitor activation suggest that younger children had more overall competitor fixations. This suggests that in addition to changes in the perceptual structure of the category, there may also be changes in the degree of lexical competition, seen across both unambiguous and ambiguous steps of the continua. However, this analysis did not address the timecourse of the effect. This raises a critical question: *What changes in lexical activation dynamics accompany (and perhaps interact with) these changes in speech perception?*

It is already established that there are changes in lexical activation dynamics between 9 and 16 years (Rigler et al., 2015). These take the form of increased initial competition to cohort and rhyme competitors and a slower attenuation of this competition, along with a slower time to resolve on the target. The present study extends this in three ways. First, minimal pairs represent a subset of tightly related rhyme competitors, offering an in depth and highly focused look at one specific class of competitors. Second, prior investigations have started at 9 years, whereas this study extends this to 7 years. Third, and most importantly, to the extent that increased lexical activation is part of the answer to how speech perception develops, precisely characterizing what about the dynamics is changing with development (e.g., the speed of suppressing competitors, the duration over which they are active and so forth) is crucial.

Given these findings, we thus examined the detailed timecourse of competition to ask what aspects of these dynamics change with development. This was intended to replicate our prior work but with a much more focused subset of competitors (in this case minimal pair rhymes) and to extend it as prior work using this paradigm has only examined 9-16 year olds—here, we can examine children as young as 7 and with an intermediate age group.

Approach

In order to focus on general lexical processes, rather than perceptual ones, this analysis started by considering only the unambiguous, endpoints of each continua (Steps 1,2 and 7,8). We filtered trials to include only those where the participant answered correctly. From these trials, we computed the timecourse of fixations to the target and competitors. Figure S4A and B show target fixations as a function of time and age. Most notably, the timecourse of fixating the target shows increases speed of access with age, particularly between the 7-8 and 12-13 y.o. age groups. This is consistent with prior work with adolescents (Rigler et al., 2015) and with preschoolers (Fernald, Perfors, & Marchman, 2006).

Figure S4C and D show fixations to the competitors (in this case a rhyme). Here, we've subtracted unrelated fixations from this (following Rigler et al., 2015) to isolate consideration of the competitor over and above any changes in the overall amount of fixations (directed to anything) that may differ with age. This suggests a complex pattern. For stop voicing (Figure S4C), it is important to note that very little competitor activation was observed at all. This is consistent with the difficulty in finding activation for rhyme competitors in general (Marslen-Wilson, Moss, & Van Halen, 1996), and with informal analyses from our lab showing that rhyme activation is much more detectable in multisyllabic words than in monosyllabic words (as were used here). However, within the small amount of competitor looking, competitor activation appears to build more quickly in 12-13 than 7-8 year olds. For the fricative tokens we observed

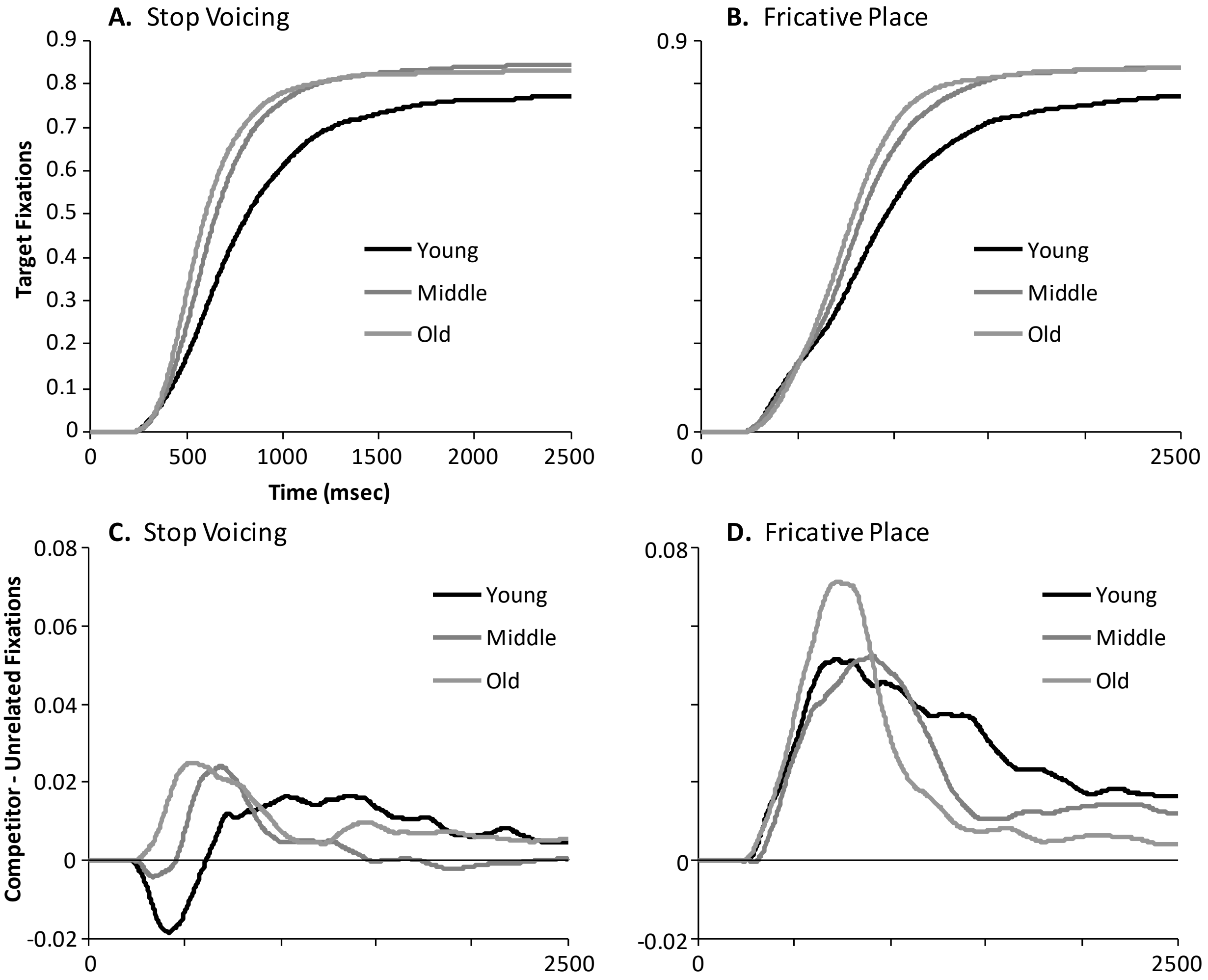

Figure S4: Lexical activation dynamics as a function of situation time and age. A, B) Proportion fixations to the target on unambiguous continuum steps (1,2, 7,8) for stop voicing trials (A) and fricative place trials (B). C, D) Proportion fixations to the competitor minus unrelated fixations for stop voicing trials (C) and fricative place trials (D).

much more competitor activation overall (than for voicing). At younger ages, peak looking was low but looks tended to persist for some time; with development there were reductions in the extent of looking, but increases in the peak.

Analysis: Targets.

To analyze target fixations, we fit the four parameter logistic function to the target curves for each subject, and compared the parameters between the young and middle groups, and the middle and old groups with t-tests (Table S2 for details). Replicating Rigler et al. (2015), we found strong effects on the timing of target fixations, but nothing on the maximum (asymptotic level). For the b/p competitors the young group had a significantly shallower slope (p<.0001) and a significantly later offset (p<.0001) than the middle group, and the middle group differed significantly from the older group on the crossover (p=.002) and had a marginally significant effect on the slope (p=.081). Similarly, for the ʃ/s competitors the young group had a significantly shallower slope (p<.0001) and a later crossover (p=.025) than the middle group; and

Table S2: Results of t-tests examining parameters of fitted curves as a function of age, stimulus-type and their interaction. P values >.2 and T<1 are not shown.

		Stop Voicing				**Fricative Place**			
		Young vs. Mid		**Mid vs. Old**		**Young vs. Mid**		**Mid vs. Old**	
		t(47)	p	t(47)	p	t(47)	p	t(47)	p
Target	Slope	6.06	<.0001	1.79	.081	4.45	<.0001	2.40	.021
	Crossover	5.74	<.0001	3.35	.002	2.32	.025	1.92	.061
	Maximum	1.60	.116	-	-	1.47	.15	-	-
Competitor	Peak	-	-	-	-	-	-	1.75	.088
	Peak Time	3.24	.002	1.47	.15	2.54	.015	2.13	.039
	Extent	2.51	.015	2.03	.048	2.10	.041	2.01	.050
	Offset	1.58	.121	-	-	1.21	-	-	-

the middle group had a significantly shallower slope (p=.021), and a marginally later crossover (p=.061) than the older group. Thus, consistent with Rigler et al. (2015) there was a clear effect of age on the speed of activating the target word that spanned all age groups.

Results: Competitors.

We next examined competitor fixations. Given the stimuli, these should be considered as rhyme competitors, though quite close ones (they are minimal pairs with the target). As unrelated fixations differed somewhat across the age groups, following Rigler et al. (2015) we subtracted the average of the two fillers from the competitor fixations, and smoothed the function with an 80 msec triangular window (Figure S4C, D). We then computed several markers for each participant. *Peak* competitor fixations was defined as the maximum value of the competitor – unrelated fixations, and scaled with the empirical logit function prior to analysis. *Peak time* was computed by taking all of the time points that were within 80% of peak, and computing the weighted average of time (weighted by the function) to obtain the approximate time at which the function reached its maximum. This was natural log scaled to achieve a more Gaussian distribution. The *extent* of competitor fixations was computed as the number of time points at which the competitor – unrelated fixations was greater than 50% of peak. Finally, the *baseline* was computed as the value of the function in the 300 msec surrounding each participant's mean reaction time. The values were then compared between adjacent age groups.

Results are shown in Table S2. If we start by considering the initial activation of competitors, there were several similarities between the competition dynamics for the b/p and ʃ/s competitor pair. The peak was significantly delayed in younger than middle children for both types of competitors (Stops: M_{young} = 1134 msec; M_{middle} = 839; Fricatives: M_{young} = 1101 msec; M_{middle} = 852), and for fricatives it was even earlier for the older children (Stops: M_{old} = 711; Fricatives: M_{old} = 734). This suggests that as a whole, children improve in how quickly competitors are activated (paralleling results with the targets). Unlike Rigler et al. (2015) neither class of competitors showed robust differences in the peak of activation (although there was a marginal *increase* in peak for fricatives between the middle and old groups).

With respect to the later suppression of competitors, results were more mixed. For the voicing competitors, we saw that the youngest group maintained competitor activation significantly longer than the middle group, but then the older group significantly "rebounded" to the same duration as the younger group (M_{young} = 125 msec, M_{middle} = 76, M_{old}=128). However, we are hesitant to make too much of this U shaped curve because the overall amount of rhyme activity was quite low (making it more challenging to estimate its extent). However, for fricatives (where there were substantially more competitor fixations), we saw significantly shorter extents between both age groups (M_{young} = 196 msec; M_{middle} = 151; M_{old} = 115). There were no developmental differences in the final degree of suppression (the offset).

Summary

Results suggest consistent development in basic lexical activation processes during late childhood and adolescence. In particular, there were clear effects on the speed of fixating the target, but no effects on the final degree of fixations. For competitors, peak degree of consideration was delayed in younger children, and we observed differences in the extent of competitor activation (particularly with fricatives where the amount of consideration was much greater) with older children suppressing competitors more rapidly (shorter extent) than the middle and younger age groups. All three effects are strongly consistent with our prior work (Rigler et al., 2015) on a similar age range. The one pattern of effects that was not consistent with this was the fact that we observed no differences in the peak amount of fixations while Rigler et al. (2015) report small differences in this. This may be because here we are focusing only on rhymes, and only on a very small classes of rhymes. Importantly, this suggests that the overall increase in competitor consideration observed in the analysis of within-category sensitivity (e.g., Figure 9, main text) derives primarily from the fact that younger children considered competitors for a longer duration (not a higher peak).

The lack of effects on the asymptotic levels of fixation (for either targets or competitors) is also consistent with that prior study and is consistent with an ancillary finding. Several studies using the visual world paradigm have reported that at asymptotic points in the curve (e.g., at the end of trial) listeners with language impairment show heightened overall competitor activation and reduced fixations to the target relative to typically developing peers (McMurray et al., 2014; McMurray, Samelson, Lee, & Tomblin, 2010) (and see Dollaghan, 1998; Mainela-Arnold, Evans, & Coady, 2008). The conclusion from Rigler et al. (2015) then was that this does not represent "delayed development" (since no such effects were observed in younger listeners). Our data continue to support that finding, suggesting that language impairment may exert unique effects on lexical competition dynamics from development.

References

Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*, 163-187.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804–809.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5/6), 507-534.

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics, 19*, 193-207.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98-116.

Galle, M. E., Klein-Packard, J., Schreiber, K., & McMurray, B. (submitted). What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory.

Galle, M. E., & McMurray, B. (2014). The development of voicing categories: A meta-analysis of 40 years of infant research. *Psychonomic Bulletin and Review, 21*(4), 884-906.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

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

Marslen-Wilson, W. D., Moss, H. E., & Van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1376-1392.

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95*(2), B15-B26.

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., Farris-Trimble, A., Seedorff, M., & Rigler, H. (2016). The effect of residual acoustic hearing and adaptation to uncertainty in Cochlear Implant users. *Ear and Hearing, 37*(1), 37-51.

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye-movements. *Journal of Speech Language and Hearing Research, 57*, 1344-1362.

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.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language, 60*(1), 65-91.

Miller, J. L., & Eimas, P. D. (1996). Internal structure of voicing categories in early infancy. *Perception & Psychophysics, 58*(8), 1157-1167.

Nearey, T. M., & Hogan, J. (1986). Phonological contrast in experimental phonetics: Relating distributions of measurements in production data to perceptual categorization curves. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 141–161). New York, NY: Academic Press.

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental timecourse of real-time spoken word recognition. *Developmental Psychology, 51*(12), 1690-1703.