Submitted for publication in J. Cole and J. Hualde (eds.) Laboratory Phonology 9. Mouton de Gruyter.
(Please do not cite without permission of the authors until it is in print.)

Vocabulary Growth and the Developmental Expansion of Types of Phonological Knowledge

Mary E. Beckman, Benjamin Munson, Jan Edwards

Abstract

A growing body of evidence on adult phonological processing supports the idea that phonological knowledge emerges through generalization over the experience of acquiring and using words. Some of this evidence suggests that knowledge is hierarchical, with generalization occurring at several different levels of abstraction away from the raw sensory input. Each familiar word-form has a distributed representation in the parametric phonetic space, which captures relevant generalizations over an individual's experience of hearing and saying specific tokens of the same word, but a parallel coarser-grained representation can be composed on the fly to process novel forms in terms of generalizations over the neighborhood of different word-forms in the individual's mental lexicon. Results of several studies of two clinical populations suggest that these different types of phonological knowledge can develop separately. Children with phonological disorder resemble younger children with typical phonological development in terms of measures of the robustness of parametric phonetic representations, whereas children with specific language impairment look like children with smaller vocabularies in terms of their processing of nonwords.

1. Introduction

Studies of phonological processing over the last decade strongly support models of adult mental lexicons in which the phonological form of each word is encoded in at least two ways. First, each familiar word-form is encoded in terms of episodic memory traces of very fine-grained parametric representations of the auditory and articulatory patterns that are experienced in hearing or saying specific instances of the word. Positing such a fine-grained parametric encoding can account for indexical effects, as suggested by Docherty (2004). These include the effects of seeing a simultaneously presented gender-stereotypical or atypical face in a speeded word-repetition task (Strand 2000), the effects of hearing a word repeated in a familiar voice on recall accuracy and reproduction at a delay (e.g. Goldinger 1996), and the interaction between familiarity with a speaker's voice and the parsing of contextual allophony in a word spotting task (Smith 2002). Positing an instance-based parametric encoding also can account for effects of word token frequency on phonological behavior, such as those described by Bybee (2000), Jurafsky, Bell, and Girand (2002), and Munson and Solomon (2004).

The literature on effects of phonological type frequency, on the other hand, suggests a different, second encoding of each word's form, in terms of coarser-grained generalizations about sublexical phonological patterns that recur across words. For example, a large and growing literature shows that sequences of phonemes attested in many words (high-probability diphones) are perceived and produced differently from those attested in few or no words (low-probability diphones). For example, diphone probability has been found to affect repetition latency for both real words and nonwords, but in an opposite direction (Vitevitch and Luce 1999). Real words are repeated more slowly if they contain high-probability diphones, suggesting a perceptual competition effect from other similar words that have rich instance-based encodings at the parametric phonetic level. Nonwords, by contrast, are repeated more quickly if they contain high-probability diphones. This asymmetry suggests that robust access to the fine-grained parametric phonetic representations relevant for producing a novel form is dependent on coarser-grained generalizations about the phonological structure of real words, as in Pierrehumbert's (2003) model of the "phonological grammar" as a step in the "ladder of abstractions" that leads from the immediate sensory image of the speech signal to the representation of morpho-syntactic relationships. In such a model, processing a novel form invokes an encoding in terms of categories such as "all strong syllables" or "all word-initial velar stops before a front vowel". This encoding would necessarily be coarse-grained relative to the individual word shapes over which the categories are abstracted, which are themselves each an abstraction over different tokens of the same word in the parametric phonetic space.

Positing such two different encodings is consistent with the literature on auditory form-based priming. For example, Goldinger et al. (1992) show that phonological similarity at the coarser-grained level (e.g., priming *bull* with *beer*) facilitates word identification and lexical decision even at long delays between the prime and target. By contrast, similarity only at the finer-grained parametric level (e.g., priming *bull* with *veer*) inhibits word identification and lexical decision. Moreover, the inhibition is particular to primes with low token frequencies, and it is very short-lived.

Describing the higher-order representations as abstractions over word types is consistent also with the observed patterns of variation in adult speakers' grammaticality judgments. Adult speakers are more likely to judge a nonword to be a possible word if it contains diphone sequences that occur in many words of their language, and mean ratings of nonwords on word-likeness scales are correlated with the total probability of the component diphones (e.g., Frisch, Large, and Pisoni 2000). Moreover, individual subjects differ in the threshold probability below which all forms are judged to be absolutely bad, and this threshold is correlated with the subject's estimated vocabulary size (Frisch et al. 2001). Such variation is to be expected if the phonological grammar for encoding nonword forms is an emergent level of abstraction over an individual speaker's experience in incorporating many initially novel word-forms into the lexicon.

In short, we find the literature on adult phonological processing easiest to interpret if we assume Pierrehumbert's (2003) hierarchical model of phonological encoding. If this assumption is correct, we should expect to see variation in the processing of nonwords over the course of lexical expansion in childhood. We also should not be surprised to see evidence for some dissociation between the two types of encoding in phonological development. Given the hierarchical relationship between them, we would not expect this dissociation to be completely symmetrical, but we would expect at least to see deficits in the parametric phonetic encoding which are independent of any deficit in the coarser-grained generalizations. We might also expect to see greater persistence for deficits in the coarser-grained generalizations after remediation of seemingly concomitant deficits at the lower level of abstraction. In this paper, we will review some relevant findings on phonological development in children, focusing on several studies in which we looked at the processing of high-probability diphones versus low-probability diphones in several groups of children, including children with two common types of "functional" (i.e., seemingly not organic) language or speech disorder. We will then conclude by discussing in a bit more depth the question of why there should be (at least) these two types of encoding at different levels of abstraction.

2. Diphone probability effects in children with SLI

In one recently completed experiment (Munson, Kurtz, and Windsor 2004), we used a non-word repetition task to compare the effects of diphone probability on production accuracy across three groups of school-aged children. The primary target group was 16 children with specific language impairment (SLI), aged 8-13 years. SLI is a syndrome that involves a delay or deficit in the development of a large range of grammatical skills. It is assessed using omnibus tests such as the *Clinical Evaluation of Language Fundamentals-3* (CELF-3; Shames, Wiig, and Secord 1997), which measure a variety of expressive and receptive skills in different structural domains, including morphology, syntax, and semantics. While children with SLI have quite variable deficits in different language skills, they also generally have smaller-sized vocabularies than their typically achieving peers. This is in keeping with the considerable body of evidence that children with SLI have difficulty repeating nonsense words (e.g., Gathercole and Baddeley 1990; Dollaghan and Campbell 1998; Edwards and Lahey 1998), and that they have difficulty acquiring new words during both structured and implicit word-learning tasks (e.g., Dollaghan, 1987). That is, given a model of the phonological grammar as structural knowledge that emerges in the course of parsing novel word-forms, many children with SLI could have problems with the coarser-grained representations that (we suggested) are necessary both for processing nonwords and for robust acquisition of new words.[1] Such children should have even more difficulty repeating nonwords that are less grammatical – i.e., less like real words – nonwords such as the stimuli composed entirely of low-probability diphones in Frisch, Large, and Pisoni (2000). We might also expect to see a correlation between severity of SLI and the size of the diphone probability effect.

The stimuli we used were a subset of the 3- and 4-syllable nonwords used in Frisch, Large, and Pisoni (2000), chosen to match word-likeness judgments between the overall high- and overall low-probability words as closely as possible. For example, /mesɚʃem/ and /gufegɛd/ were judged to be fairly similar in wordlikeness by the adult subjects in that earlier study, but /me/ and /sɚ/ occur in 134 and 194 words in the HML,[2] whereas /gu/ and /fe/ occur in only 14 and 46 words. The children were asked to repeat these "funny made-up words" in response to a digitized adult production. One of the experimenters transcribed each production phonetically from the recording, and another did a second independent transcription of 15% of the productions, yielding a 92% agreement. In addition to the 16 children with SLI, we also tested two control groups of children with typical language skills: 18 children who were matched by chronological age (CA group) and 16 younger children (aged 6-10 years) who were matched by their scores on a test of expressive vocabulary size (VA group).

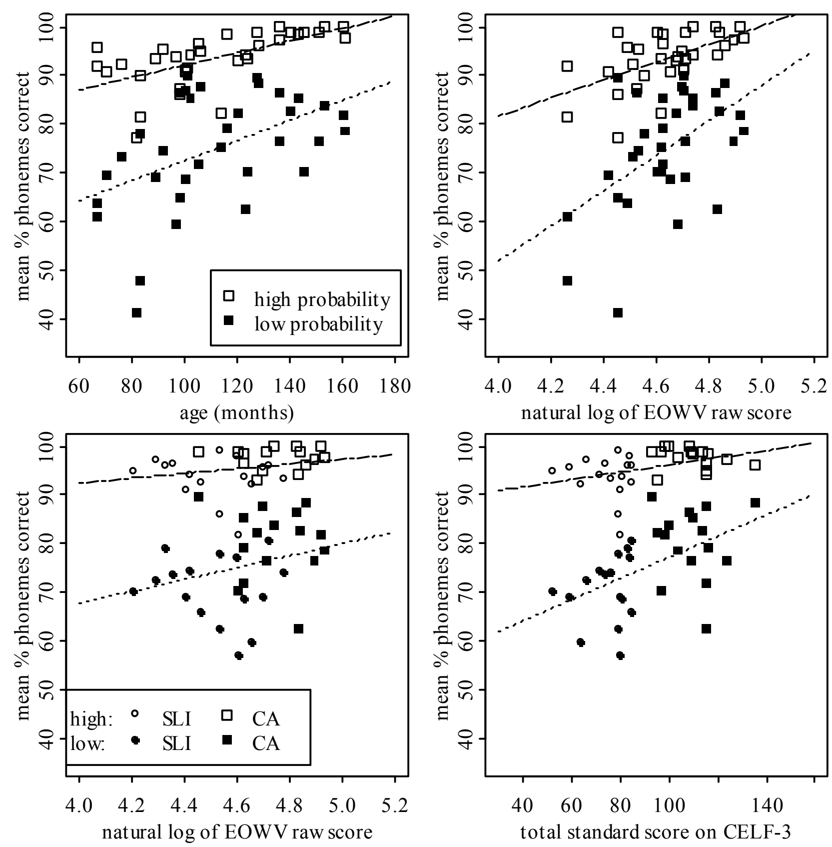

Figure 1. Percent of phones produced correctly in stimuli composed only of high-probability diphones (open symbols) or only of low-probability diphones (closed symbols) as a function of the child's age (upper left), expressive vocabulary size (upper right and low left), and severity of language impairment (lower right) in Munson, Kurtz, and Windsor (2004).

Each data point plots the depicted measures for one of the 34 children with typical development in the upper two panels, or for one of the 16 children with SLI (circles) or an age peer (squares) in the lower two panels.

Fig. 1 shows the results. Each of the panels in the figure plots mean repetition accuracy for a child's productions against some other predictive measure, such as the child's age. Means are calculated separately for non-words containing high-probability and those containing low-probability sequences, so that each child is represented by two datapoints. Data from the 34 typically developing children are shown in the upper two panels. The upper-left panel plots repetition accuracy against age and the upper-right panel plots it against vocabulary size. Age in months predicts a significant proportion of variance in both high- and low-probability nonwords. However, the slope of the regression line for the low-probability nonwords is steeper, so that the two lines converge on the right. Thus, age predicts not only overall accuracy, but also the difference in repetition accuracy between high- and low-probability forms; older children show a smaller effect of phonotactic probability than younger children.

The measure of vocabulary size in the upper-right panel is the log-transformed raw score on the *Expressive One-Word Picture Vocabulary Test* (EOWPVT; Gardner, 2000). This measure also predicts a significant proportion of variance in the repetition of the high- and low-probability items, and it predicts the difference between them. This result is not surprising, since vocabulary size is also correlated with age. To tease apart the contributions of these two measures of development, therefore, we did two step-wise multiple regression analyses that entered one or the other measure as the first independent variable. These analyses showed the EOWPVT raw score to be the stronger predictor of the diphone probability effect. This is essentially the same result as in other studies of nonword repetition which include younger children, such as Munson (2001) and Edwards, Beckman, and Munson (2004).

The predictive power of vocabulary size is illustrated also in the bottom-left panel of Fig. 1. This graph plots repetition accuracy for high- and low-probability sequences against expressive vocabulary size for the 16 children with SLI and the 18 older typically developing children who were their chronological age peers. Here, age variation is controlled so that the plot emphasizes the relationship between accuracy and vocabulary size. The circles lie generally to the left of the squares, showing the typically smaller vocabularies of children with SLI relative to their age peers, and the regression curves for the high- and low-probability stimuli have positive slopes, showing that the children with SLI have more difficulty overall with the task, as in the studies by Gathercole and Baddeley (1990) and others. Moreover, the two curves converge on the right, showing that the children with SLI are disproportinately more affected by the difficulty of producing the low-probability sequences. Although neither curve is as steep as in the graph that includes the younger children, there is clearly a relationship between the difficulties that children with SLI have with nonwords and their generally smaller vocabularies.

On the other hand, the data in the bottom-right panel suggest that this is not a direct causal relationship. While EOWPVT raw score is a better predictor than age of the effect of phonotactic probability, it is not as good a predictor as the severity of the language disorder, as measured by the children's standard scores on the CELF-3. That is, the best predictor of the size of the phonotactic probability effect in the Munson, Kurtz, and Windsor (2004) study of children with SLI is this global measure of grammatical knowledge, including procedural knowledge of inflectional morphology, syntax, and word meanings, as well as the phonological knowledge that supports lexical expansion. In other words, even at this age, after the children with SLI are no longer making the kinds of errors on real words that Macken (1995) characterizes in terms of "acquisitional rules" based on perception and articulation, their productions of nonwords are still different from their age peers in being more drastically affected by the demands of parsing a novel input form and composing the corresponding novel output form.

Together, these results suggest that SLI is associated with difficulties in making coarse-grained phonological generalizations over the store of episodic representations of lexical items (akin to Macken's "acquisitional rules" based on "generalization" rather than constraints on perception and articulation). Consequently, children with SLI are poorer than their chronological age peers at generalizing correct phoneme production to unfamiliar or unattested sequences in a nonword repetition task. This same deficit must be partially responsible for their deficits in acquiring new words: children with SLI do not have the robustly abstracted coarse-grained phonological representations needed to efficiently parse unfamiliar patterns such as novel phone sequences. Consequently, they have difficulty learning novel word-forms and not just difficulty associating new forms with poorly generalized morpho-syntactic and semantic categories. Thus, we might characterize their lower accuracy on the nonword repetition task as a problem with the phonological grammar per se – i.e., the higher-order coarse-grained generalizations and not the lower-level encoding of the word in terms of finely-detailed parametric phonetic representations.

3. Parametric representations in children with phonological disorder

Another population that we can compare to children with SLI is children with phonological disorder (PD). PD is a syndrome of idiopathic habitual age-inappropriate misarticulation of consonant sounds as assessed by standardized tests such as the *Goldman Fristoe Test of Articulation* (GFTA; Goldman and Fristoe 1986). We have looked at diphone probability effects in children with PD as part of a larger ongoing investigation of the etiology of the syndrome. The general motivation and design of the first stage of the investigation are described in Edwards et al. (1999). We had three sets of tasks, which were intended to provide derived measures to evaluate potential differences across groups of children in their ability to form and use phonological representations at three different levels, as summarized in Table 1.[3] We compared younger children with typical phonological development (TD) to older children with TD. We also compared children with PD to their age peers with TD.

Table 1. The three levels of representation described in Edwards et al. (1999), and the tasks and evaluative measures used to assess them.

fine-grained representations in the parametric articulatory space
1. picture naming; consistency of correct or incorrect production of target consonants in the elicited real words (Isermann 2001)
2. multiple repetitions of nonwords containing lingual stops; spectral analyses of the stop bursts (Edwards, Gibbon, and Fourakis 1997; White 2001)
fine-grained representations in the parametric auditory space
3. gated word identification; accuracy of final consonant identification at different gate conditions (Edwards, Fox, and Rogers 2002)
categories for mapping between the two parametric domains
4. repetition of nonwords containing high- versus. low-probability diphones; size of frequency effect on accuracy (Edwards, Beckman, and Munson 2004; Munson, Edwards, and Beckman 2005)

Our motivation for including the comparison across age groups is that one account of PD describes it as delay in phonological development. This description is plausible because the misarticulations typically made by children with PD resemble errors frequently observed in younger children with TD. That is, the error patterns look very similar when we use coarse-grained observational tools such as the GFTA. This standardized test provides colored line drawings and prompts that let the speech language pathologist sample one target production of each English consonant in word-initial, medial, and final position, as well as in most initial clusters. The speech language pathologist uses the pictures to elicit spontaneous productions of the words, re-prompting with a "That's ____; can you say ____?" if the child does not seem to know the word, and then transcribes the child's productions of the target consonants, counts the number of errors and compares this error rate to a table of age-normed scores. The errors transcribed for children with PD on such tests typically involve substitutions and deletions that are virtually identical to the types of errors transcribed for younger children with TD. Table 2 lists some examples.

Table 2. Examples of some typical misarticulation patterns in words produced by children with PD, from Isermann (2001)

error type	target	adult form	child form	ID sex age (yr; mo)
"stopping"	*socks*	/sɑks/	[datʰ]	p137 M 4;4
	sheep	/ʃip/	[ti]	p112 F 5;4
	cheeze	/tʃiz/	[ki]	p103 F 5;9
"fronting"	*cake*	/keɪ̯k/	[teɪ̯k]	p106 F 5;7
	brush	/bɹʌʃ/	[bwʌs]	p106
	shoe	/ʃu/	[su]	p124 M 4;11

The substitution errors are similar not only at this coarse-grained level of observation, but also in the patterns of "covert contrast" that are revealed in finer-grained instrumental analyses. For example, a cross-sectional study by Kewley-Port and Preston (1974) and a longitudinal study by Macken and Barton (1980) showed that when English-acquiring children with TD are first mastering the contrast between word-initial /p, t, k/ and /b, d, g/, they typically produce only voiceless unaspirated sounds, which will be transcribed as [b], [d], or [g] by the English-speaking speech language pathologist. Many children will then begin to differentiate the two sets by producing somewhat longer VOT values in their target /p, t, k/, although it may take a month or more of further experience before their aspiration intervals are long enough for the stops to be reliably transcribed as [p], [t], or [k]. This phenomenon of covert contrast has been observed for English-acquiring children with PD as well, not only for "voicing" of voiceless stops (e.g., Gierut and Dinnsen 1986; Scobbie et al. 2000), but also for "fronting" of /s/ to [θ] (Baum and McNutt 1990), /ʃ/ to [s] (Hardcastle, Gibbon, and Scobbie 1995), and /k/ to [t] (e.g., Forrest et al. 1990; Edwards, Gibbon, and Fourakis 1997; White 2001). We interpret such results as indicating that a good number of children with PD have subtle motor and/or perceptual deficits that affect the children's ability to form a robust encoding of the more successful articulatory patterns that they have experienced in earlier babbling and word production.[4]

Further support for this idea comes from another resemblance between children with PD and younger children with TD. Although PD is defined as age-inappropriate misarticulations in the absence of any evidence of gross auditory problems, there are experimental results suggesting that many children with PD have subtle deficits in their ability to form a robust encoding of the acoustic patterns that they have experienced (Rvachew and Jamieson 1989; Forrest et al. 1995). Since there is also a large literature showing that younger children's auditory representations are not as robust as those of adults (e.g., Graham and House 1971; Ryalls and Pisoni 1997; Walley 1988, Nittrouer and Boothroyd 1990), these results for children with PD might also be interpreted as evidence for a delay in acquiring adult-like auditory-phonetic representations. Edwards, Fox, and Rogers (2002) report the results of a study supporting this interpretation. The study used a gated word identification task. On each trial, the child was asked to point to the appropriate picture after hearing a sample production either of *cap* versus *cat* or of *tap* versus *tack* presented in one of three conditions: the digitized ungated (whole) word, a less extreme gate in which the final stop burst was removed, and a more extreme gate in which the vowel was truncated immediately after the beginning of the F2 transition into the stop. Older children with TD could identify the word with some success even in the gated conditions, whereas younger children with TD and children with PD were at chance except in the ungated condition.

Fig. 2 shows the results for this whole-word condition for the target group of children with PD (aged 3-6 years) and their TD age peers. The two panels plot the mean discrimination score (d-prime) against one other measure. The d-prime is averaged over the whole-word condition for all four words, so that there is only one datapoint for each child. It is plotted first as a function of the child's age and then as a function of the child's raw score (number of consonant errors) on the GFTA, which is used here as a measure of severity of phonological disorder in the children with PD. The regression curves for the first panel show that the children with PD are generally less accurate on this task (the dotted line is below the dashed line), and that there is a good correlation between perceptual acuity and age for children with TD. (The larger study included 7- and 8-year-old children with TD, and showed an even stronger correlation.) The regression curve for the second panel shows that for the children with PD, accuracy on this task is strongly correlated with the severity of disorder. This result suggests that many children with PD have difficulties developing robust encodings of words in the auditory dimensions of the parametric phonetic space as well as in the articulatory dimensions.

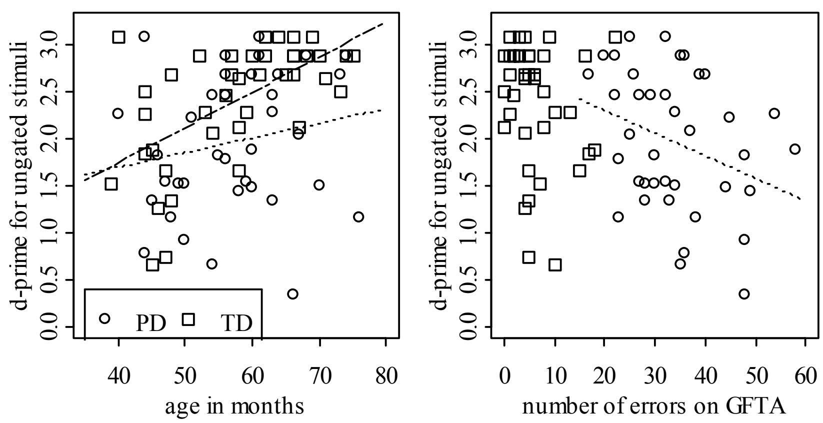

Figure 2. Auditory perceptual acuity, as gauged by the mean discrimination score for the whole word condition in Edwards, Fox, and Rogers (2002) plotted as a function of the child's age (left panel) and of the severity of phonological disorder (right panel). Each data point plots the depicted measures for one of the 40 children with phonological disorder (circles) or an age peer with typical phonological development (squares).

4. Diphone probability effects in children with phonological disorder

In another recently completed study with the same group of children (Munson, Edwards, and Beckman 2005), we looked for higher-order grammatical deficits of the sort we posit for children with SLI by assessing the size of the probability effect in a nonword repetition task. Since the children in this study are younger than those in Munson, Kurtz, and Windsor (submitted), we could not use such difficult stimuli as /gufegɛd/. Instead we devised easier nonword pairs which placed contrasting target diphones in frames that were relatively easy overall. For example, in the pair /bedæg/ versus /donug/, the high-probability target diphone /æg/ occurs as a rhyme in many words in the MHR,[5] including *bag*, *drag*, *flag*, *magnet*, and *tag*, whereas the low-probability target diphone /ug/ occurs in only one word, *cougar*, where the /g/ is arguably not part of the rhyme of the first syllable. The other CV and VC sequences in each nonword are all more probable than /ug/ and add up to a roughly equally probable preceding frame. As in our other studies using this task, the children were asked to repeat these “funny made-up words” in response to a digitized adult production. The target diphone in each production was transcribed and was scored for correct place, manner, and voicing (for consonants) or correct place, height, and length (for vowels) to make a scale from 0 to 6 points.

Fig. 3 shows the results. As in Fig. 1, the accuracy measure is averaged separately over the productions of high- versus low-probability targets, so that each child is represented in any plot by two datapoints. The upper two panels plot the mean accuracy scores against the size of each child’s expressive vocabulary, as measured by log-transformed raw score on the *Expressive Vocabulary Test* (EVT; Williams, 1997). The upper-left panel shows the children with TD from Fig. 2 and also includes a larger group of older children with TD who were tested in the Edwards, Fox, and Rogers (2002) study and in Edwards, Beckman, and Munson (2004). The correlation of nonword repetition accuracy with vocabulary size is like that in the upper-right panel of Fig. 1; multiple regression analyses show that vocabulary size rather than age is the relevant variable, since age accounts for none of the variance not accounted for by the EVT. The upper-right panel here is the analogue to the lower-left panel in Fig. 1, but the details are different. Although the children with PD are reliably less accurate in their productions (the circles lie below the squares), they do not have small vocabularies for their ages (the circles do not lie to the left of the squares). Thus, the effect of including both the children with PD and their age peers with TD in this graph is to increase the scatter above and below the regression curves, without extending the two lines leftward or increasing the distance between them in the region covered by the data.

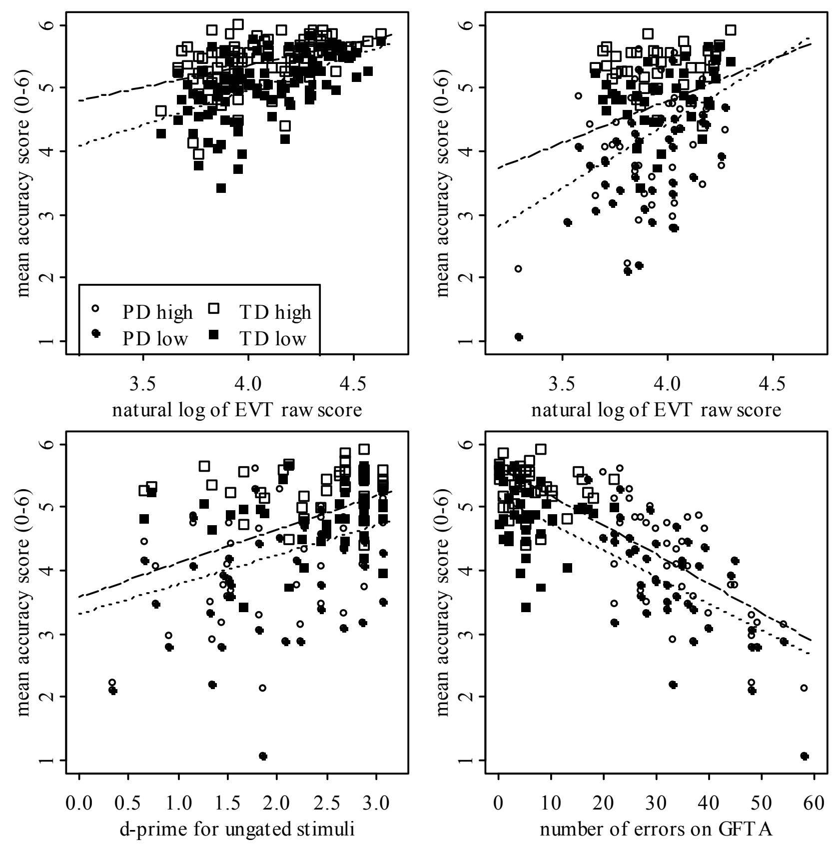

Figure 3. Mean segmental accuracy score for productions of high-probability targets (open symbols) versus low-probability targets (closed symbols) as a function of the child's expressive vocabulary size (upper two panels), perceptual acuity as measured by mean discrimination score (lower left panel), and severity of phonological disorder (lower right). Each data point in the upper left panel shows the depicted measures for one of the 104 children with typical phonological development in the study described in Edwards, Beckman, and Munson (2004). Each data point in the remaining panels shows the depicted measures for one of the 40 children with phonological disorder (circles) or an age peer (squares) in the studies described in Munson, Edwards, and Beckman (2005) and Edwards, Fox, and Rogers (2002).

The lower two panels plot mean accuracy on the nonword repetition task against two of the measures shown in Fig. 2. The left-hand panel shows the relationship to the d-prime measure that was the dependent variable in that figure, and the right-hand panel shows the relationship to the GFTA raw score that was the independent measure plotted in the lower-right panel of Fig. 2. By contrast to the pattern in the upper two panels in Fig. 3, these two lower panels differentiate the children with PD from those with TD on both dimensions, because the children with PD have generally lower perceptual acuity as measured by the mean d-prime in Edwards, Fox, and Rogers (2002) and they by definition have more errors on the GFTA. The regression curves in the graph to the left slope steeply upward, showing how overall production accuracy depends on having robust perceptual representations of the acoustic shapes of the nonword stimuli. The regression curves in the graph to the right slope steeply downward, showing that children who misarticulate consonants on the GFTA also misarticulate the target sounds on the nonword task. However, neither graph shows a convergence of the regression curve. The children with PD are no more affected by the probability of the target diphone sequence than are their age peers with TD. The result is in marked contrast to the result for the children with SLI described above.

Extrapolating from these two sets of results, then, the differences between children with PD and children with SLI support Pierrehumbert's (2003) proposal that diphone probability effects reflect an encoding at a different level of abstraction from the encoding of word-forms that gives rise to token frequency effects on production and perception. In Pierrehumbert's proposal, the token frequency effects come about because word-forms are "abstractions over the phonetic space" — i.e., over the space of positional allophones and other phonetic categories that begin to emerge already in the infancy through exposure to the ambient language as the baby looks and listens at the talking faces in the environment, and learns how to invoke more input from these faces by cooing and babbling back at them. In Pierrehumbert's model, the primary encoding of a word-form is at this level of the abstraction; a word-form is a generalization over the experienced instances of the word in the phonetic space "learned through repeated exposure to that word in speech" (p. 179). By contrast, phonotactic effects, such as the lesser accuracy with which children repeat low-probability diphones, involve a higher order abstraction. They are "generalizations over the word-forms in the lexicon, which are in turn generalizations over speech" (p. 180). In this model, then, phonotactic effects are part of a "phonological grammar" which "describes the set of possible words of a language". Abstractions at this level emerge later in development as a consequence of the expanding lexicon, because "phonology does not abstract over speech directly, but rather indirectly via the abstraction of word-forms" (pp. 179-180). Assuming this model, we can interpret our results as evidence that children with PD differ from children with TD only in terms of the robustness of the encoding of word-forms in the parametric phonetic spaces, not at the level of the phonological grammar. For these children, representations in the acoustic dimensions and/or in the articulatory dimensions of the phonetic space are still like the novice chess player's representations of the arrangement of pieces on the chessboard. They cannot as easily focus attention on the most relevant details in listening and they have more inflexible, entrenched whole-word motor patterns. The syndrome of SLI, by contrast, seems to implicate problems with the encoding of generalizations at the more abstract level of the phonological grammar. For these children, novel words pose difficulty because there is less of the elaborated abstract structure in place that allows the adult speaker to make an immediate, stereotypical parsing of the acoustic pattern and a fluent retrieval of gestures and coordination routines that can be composed into novel motor scores.

5. Why should there be two encodings?

Given this interpretation of our results, a question that comes immediately to mind is why there should be two separate levels of abstraction involved in the encoding of word-forms in the child's lexicon. We can think of at least two sets of reasons, one having to do with the auditory dimensions of the phonetic encoding and the other having to do with the articulatory dimensions.

In the first set, the most salient reason is the problems that babies must overcome in order to begin to recognize recurring patterns (i.e., word-forms) in the speech that they hear addressed to them or to other listeners in their environment. Fisher and Church (2001) group these into "the problems of word segmentation and the problems of contextual variability" and describe the constraints that these problems impose on "the development of an auditory lexicon" as follows:

> This catalog of sources of variation in natural speech has consequences for the nature of the perceptual learning mechanism we seek: Long-term representations of the sounds of words must in some sense be abstract enough to encompass variability due to voice, intonation,

> and linguistic context. Yet these representations must also include enough phonetically relevant detail to discriminate near lexical neighbors, and to permit the child to learn about the various systematic sources of variability in the sounds of words. The learning mechanisms responsible for establishing representations of the sounds of words during development must be able both to abstract over and to incorporate phonetic details and information about words' surrounding context. (Fisher and Church 2001: 53)

Rephrasing this in terms of Pierrehumbert's (2003) model, we can say that a dual encoding of each word-form promotes the development of robust word recognition because it allows the child to overcome this conflict between sufficient abstraction and sufficient detail. The detailed instance-based encoding in terms of the word's distribution of patterns in the auditory parametric space allows the child to parse the fine detail that also encodes the larger phonetic context for the word as well as the speaker's mood and identity. Experiments and modeling studies by Johnson, Strand, and their colleagues amply demonstrate that this parsing cannot be accomplished by a sequence of piecemeal "normalization" processes (see, inter alia, Johnson 1990; Strand and Johnson 1996). Rather, the auditory phonetic space simultaneously encodes every kind of speech category that the child-listener eventually learns to control, including the categories relevant to the speaker's signaling of his or her social identity and social relationship to the child as well as the positional allophones and prosodic structures that distinguish one word-form from another similar word-form. At the same time, the more abstract encoding in terms of phonological categories such as the structural properties "consonant" or "foot-initial" and the paradigmatic properties /k/ or "dorsal" function the same way that all stereotypes do: it insures efficient identification of a known word-form when the signal is noisy or incomplete (see Strand, 1999, for arguments for this application of the notion of "stereotype" to phonological categories).

This "stereotyping" function has a further advantage for speech categories because of the very large and continually expanding size of most human lexicons. That is, listeners must not only recognize known word-forms in the face of variability, they must also recognize when the signal is presenting them with a new word-form. By the time a child is 6 years of age, he or she may have learned thousands of words, and while rates of new word learning decline in maturity, it is a process that can continue throughout life.[6] At the beginning of this process, learning will be facilitated if the young child can develop a way of efficiently recognizing when a stretch of the signal does not correspond to any known word-form in the rapidly expanding lexicon. Merriman and Schuster (1991) and others have shown that when young children are presented with a novel form that is similar to a known word, they tend to assimilate it to the known word, choosing a picture of an apple over a picture of an unknown object such as a painter's palette when asked "Which one is a /dʒæpəl/?" With two-year-olds, this tendency is stronger in the small minority who say "No" when asked "Do you know what a /dʒæpəl/ is?" With four-year-olds, by contrast, the ability to acknowledge the unfamiliarity of the form is associated with the opposite tendency to point more often to the unfamiliar object when asked to choose the /dʒæpəl/. One way to interpret this difference is to posit that many four-year-olds have a more robust encoding of word-forms at the higher level of abstraction. That is, a redundant encoding both at the level of instance-based parametric representations and at the level of phonological abstractions such as the syllable and phoneme, allows the child to make an efficient comparison to phonologically similar word-forms, in order to make the "No" response in this lexical decision task. This ability to recognize when a word-form is new can only enhance the efficiency of adding the new word-form to the lexicon when it is experienced for the first time in the auditory phonetic space.

The second set of reasons that a dual encoding promotes vocabulary expansion is related to the fact most infant listeners also become speakers. This means that there must be a detailed representation of each word-form in the articulatory dimensions of the parametric phonetic space. There must also be some representation of the mapping between the encoding of a word-form in the auditory dimensions and its encoding in the articulatory dimensions. Without these two other sets of representations within the parametric phonetic level, the child will not be able to say the word-forms that he or she knows, or acquire the ability to repeat new word-forms, including the pseudo-words in the nonword repetition task. For infants with normal hearing and vision and normal propensities to human social interaction, the neural bases for these two representations are present at birth. That is, we know that neonates orient to human voices and to any image that is even vaguely like a human face, and we also know that neonates are given to imitate the orofacial gestures that they hear and see in these faces (Meltzoff and Moore 1977, 1989; Chen, Striano, and Rakoczy 2004). The vocal social interaction that these propensities support lead to a rich set of exemplars in the articulatory dimensions and in the cross-modal mapping dimensions of the parametric phonetic spaces. At least in cultures where the infant interacts first with a single caretaker, the encoding in these articulatory and cross-modal mapping dimensions of phonetic categories relevant for imitating the vowels that infants see and hear is very early, and typically precedes the encoding of social categories such as gender (Kuhl and Meltzoff 1982, 1996; Patterson and Werker 2002). This explains the effects of ambient language that can be seen already in the spectral patterns of the more sonorous portions of the vocalizations produced by 10-month-old infants (de Boysson-Bardies et al. 1989).

Despite this rich density of exemplars in the articulatory dimensions of the parametric phonetic space, however, the articulatory encoding of word-forms in the lexicons of normally hearing children can never be as richly experienced as the auditory encoding of word-forms. This disparity is probably part of the explanation for why very young children typically can hear differences among word-forms produced by adults that they cannot reliably reproduce. The child at first simply maps the adult's word-forms onto the coarse-grained "vocal motor schemes" (McCune and Vihman 1987) that the child has already established in the course of vocal exploration in the first year using the "forward plant" established in babbling (Jordan 1990, Bailley, Laboissiere, and Schwartz 1991). By the time that a child has more than fifty or so words, however, there must be some finer decomposition of the word-forms to allow the child to begin to differentiate the word /ba/ 'ball' from the word /ba/ 'block' (McCune 1992: 331). The child must find some way of interpolating among established word-forms to compose a novel articulatory representation for a new word-form. This is one of the reasons why the articulatory target that a speaker aims to produce in any given instance should be an average over nearby exemplars rather than a single specific exemplar (as proposed by Pierrehumbert, 2002). Establishing a phonological grammar of categories such as the phoneme /k/, or a more abstract front allophone of /k/ that abstracts over the vowel contexts experienced in words such as *kitty*, *cake*, and *candy*, allows for a more efficient mapping to sets of articulatory exemplars over which the child should interpolate. We invoke this view of the second-order abstraction to explain the relationship between the size of the diphone effect and the child's vocabulary size in the two studies outlined in Sections 2 and 4.

Notes

1. That is, the word-learning deficits of children with SLI may be both a *cause* and a *consequence* of their difficulties in learning new words. They are a cause in

that the children with SLI have fewer words in memory over which to make generalizations about the categorical phonemic structure of the language. As a result, they develop more poorly specified representations of phonemes. Consequently, they have a reduced ability to make a 'fast mapping' between an unfamiliar string of phonemes and existent categories in memory during new-word learning.

2. The *Hoosier Mental Lexicon* is a list of about 19,300 lexemes with associated familiarity ratings by Indiana University undergraduates (Pisoni et al., 1985). Frisch et al. used this corpus to compute frequencies for different CV and VC sequences and then to build nonwords with a range of overall probabilities, along a continuum from very high to very low probability. They designed their materials in this way because most of their analyses were correlational. For simple two-way comparisons between high- versus low-probability forms, they divided the continuum in two at the median probability.
3. Of course, none of these tasks taps only the targeted level of generalization. Given the kind of architecture that we assume to account for the ways in which perception, production, and higher-order representations interact, we find it impossible to imagine a task that would tap only one domain or one level in any typical human older than 6 months. Rather, we chose these tasks because we think we can identify correlational patterns that pinpoint difficulties in one or another level of representation, as described in our explication of Figures 1 through 3.
4. These tasks are intended merely to differentiate deficits involving articulatory representations as a class from those involving auditory or visual representations. We do not rule out a priori an even more complex taxonomy in which some children with PD have difficulty forming robust generalizations at the level of "gestural patterning" (Saltzman and Munhall, 1989) or the like, whereas others have lower-level representational deficits implicating the kinesthetic or other somatosensory map.
5. The MHR is a list of pronunciations that we created from the 6366 most frequently occurring words in Moe, Hopkins, and Rush (1982), a corpus of transcribed spontaneous narratives elicited from first-grade children. The pronunciations are from the *Carnegie Mellon University Pronouncing Dictionary* (http://www.speech.cs.cmu.edu/cgi-bin/cmudict), which gives pronunciations from the same general dialect region as the central Ohio varieties spoken by the children in our study.
6. To illustrate, we list here the four new words that one of us (who has now spent more than fifty years expanding her first-language lexicons) acquired during the week that she wrote the first draft of this paragraph, along with the sources in which she encountered them: *cicatrise* ("Forget What Did" by Philip Larkin), *kakutoku* 'acquisition; possession' ("Acquisition of phonology and language universals" by Haruo Kubozono [*Journal of the Phonetic Society of Japan*, 7,

5-17 (2003)]), *moly* (poem of that title by Thom Gunn), *peripatetic* ("Remembering Peter Jusczyk" by Robert Remez [*Journal of Phonetics* , 31, 289-291 (2003)]).

References

Bailley, Gerard, R. Laboissiere, and Jean-Luc Schwartz
1991 Formant trajectories as audible gestures. *Journal of Phonetics* 19(1): 9-23.

Baum, Shari R., and James C. McNutt
1990 An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics* 18(1): 51-63.

de Boysson-Bardies, Bénédicte, Pierre Hallé, Laurent Sagart, and Catherine Durand
1989 A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language* 16(1): 1-17.

Bybee, Joan
2000 Lexicalization of sound change and alternating environments. *Papers in Laboratory Phonology* 5: 250-268.

Chen, Xin, Tricia Striano, and Hannes Rakoczy
2004 Auditory-oral matching behavior in newborns. *Developmental Science* 7(1): 42–47.

Docherty, G. J.
2004 Speech in its natural environment: Accounting for social factors in phonetic variability. *Paper presented at the 9th Conference on Laboratory Phonology, 24-26 June, 2004*. University of Illinois, at Urbana-Champaign.

Dollaghan**,** Christine A.
1987 Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Disorders* 52(3): 218-222.

Dollaghan, Christine, and Thomas F. Campbell
1998 Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research* 41(5): 1136-1146.

Edwards, Jan, Mary E. Beckman, and Benjamin Munson
2004 The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research* 47: 421-436

Edwards, Jan, Marios Fourakis, Mary E., Beckman, and Robert A. Fox
1999 Characterizing knowledge deficits in phonological disorders. *Journal of Speech, Language, and Hearing Research* 42(1): 169-186.

Edwards, Jan, Robert A. Fox, and Catherine L. Rogers
2002 Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and phonetic inventory size. *Journal of Speech, Language, and Hearing Research* 45(2): 231-242.

Edwards, Jan, Fiona Gibbon, and Marios Fourakis
1997 On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. *Language and Speech* 40: 203-210.

Edwards, Jan, and Margaret Lahey
1998 Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics* 19(2): 279-309

Fisher, Cynthia, and Barbara A. Church
2001 Implicit memory support for language acquisition. In *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*, Jürgen Weissenborn and Barbara Höhle (eds.) vol. 1, pp. 47-69. Amsterdam: John Benjamins.

Forrest, Karen, Gary Weismer, M. Hodge, Daniel A. Dinnsen, and Mary Elbert
1990 Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics and Phonetics* 4: 327-340.

Forrest, Karen, Steven B. Chin, David B. Pisoni, and Jessica N. Barlow
1995 *Talker normalization in normally articulating and phonologically delayed children: Methodological considerations* (Research on

Spoken Language Processing, No. 19, pp. 229-251). Bloomington: Indiana University Press, Department of Psychology.

Frisch, Stefan, Nathan R. Large, and David B. Pisoni
2000 Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42(4): 481-496.

Frisch, Stefan, Nathan R. Large, Bushra Zawaydeh, and David B. Pisoni
2001 Emergent phonotactic generalizations in English and Arabic. In *Frequency and the Emergence of Linguistic Structure*, Joan Bybee and Paul Hopper (eds), 159-179.

Gardner, Morrison F.
2000 *Expressive One-Word Picture Vocabulary Test*. San Antonio, TX: The Psychological Corporation.

Gathercole, Susan E., and Alan D. Baddeley
1990 Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory & Language* 29(3): 336-360.

Gierut, Judith A., and Daniel D. Dinnsen
1986 On word-initial voicing: Converging sources of evidence in phonologically disordered speech. *Language and Speech* 29: 97-114.

Goldinger, Stephen D.
1996 Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22: 1166-1183.

Goldinger, Stephen D., Paul A. Luce, David B. Pisoni, and Joanne K. Marcario
1992 Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(6): 1211-1238.

Goldman, Ronald, and Macalyne Fristoe
1986 *The Goldman Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Services.

Graham, Louella W., and Arthur S. House

1971 Phonological oppositions in children: A perceptual study. *Journal of the Acoustical Society of America* 71: 559-566.

Hardcastle, William J., Fiona Gibbon, and James M. Scobbie
1995 Phonetic and phonological aspects of English affricate production in children with speech disorders. *Phonetica* 52: 242-250.

Isermann, Bridgett C.
2001 *Variability of Consistency of Articulation in Children with Phonological Disorders*. M.A. thesis, Ohio State University.

Johnson, Keith
1990 The role of perceived speaker identity in the F0 normalization of vowels. *Journal of the Acoustical Society of America* 88: 642-654.

Jordan, Michael I.
1990 Motor learning and the degrees of freedom problem. In *Attention and Performance XIII*, M. Jeannerod (ed.), 796-836. Hillsdale, NJ: Erlbaum.

Jurafsky, Daniel, Alan Bell, and Cynthia Girand
2002 The role of the lemma in form variation. *Laboratory Phonology* 7: 3-34.

Kewley-Port, Diane, and Malcom S. Preston
1974 Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2(3): 195-210.

Kuhl, Patricia K., and Andrew N. Meltzoff
1982 The bimodal perception of speech in infancy. *Science* 218: 1138-1141.

Kuhl, Patricia K., and Andrew N. Meltzoff
1996 Infant vocalization in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America* 100: 2425-2437.

Macken, Marlys A., and David Barton
1980 The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 7: 41-74.

Macken, Marlys A.
1995 Phonological acquisition. In *The handbook of phonological theory*. G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (eds.), 671-696. Cambridge, MA: Blackwell.

McCune, Lorraine
1992 First words: a dynamic systems view. In *Phonological development: models, research, implications*, C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), 313-336. Timonium, MD: York Press.

McCune, Lorraine, and Marilyn May Vihman
1987 Vocal motor schemes. *Papers and Reports on Child Language Development* 26: 72-79.

Meltzoff, Andrew N., and M. Keith Moore
1977 Imitation of facial and manual gestures by human neonates. *Science* 198**:** 75–78.

Meltzoff, A. N., and M. Keith Moore
1989 Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology* 25: 954–962.

Merriman, William E., and J. M. Schuster
1991 Young children's disambiguation of object name reference. *Child Development* 62: 1288-1301.

Moe, Alden J., Carol J. Hopkins, and R. Timothy Rush
1982 *A vocabulary of first-grade children*. Springfield, IL: Thomas.

Munson, Benjamin
2001 Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, & Hearing Research* 44(4): 778-792.

Munson, Benjamin, Jan Edwards, and Mary E. Beckman
2005 Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research* 47: PP not yet available.

Munson, Benjamin, Beth A. Kurtz, and Jennifer Windsor

2004 The influence of phonotactic probability, vocabulary size, and wordlikeness on nonword repetition by children with and without language impairment. Ms. under review.

Munson, Benjamin, and Nancy Pearl Solomon
2004 The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47: PP not yet available.

Nittrouer, Susan, and Arthur Boothroyd
1990 Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America* 87: 2705-2715.

Patterson, Michelle L., and Janet F. Werker
2002 Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology* 81**:** 93–115

Pierrehumbert, Janet B.
2003 Probabilistic phonology: discrimination and robustness. In *Probabilistic linguistics*, Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds.), 175-228. Cambridge, MA: MIT Press.

Pisoni, David B., Howard Nusbaum, Paul A. Luce, and Louisa Slowiaczek
1985 Speech perception, word recognition, and the structure of the lexicon. *Speech Communication* 4, 75-95.

Rvachew, Susan, and Donald G. Jamieson
1989 Perception of voiceless fricatives by children with functional articulation disorder. *Journal of Speech and Hearing Research* 54(2): 193-208.

Ryalls, Brigette Oliver, and David B. Pisoni
1997 The effect of talker variability on word recognition in preschool children. *Developmental Psychology* 33(3): 441-452.

Saltzman, Elliot L., and Kevin G. Munhall
1989 A dynamic approach to gestural patterning in speech production. *Ecological Psychology* 1: 333-382.

Scobbie, James M., Fiona Gibbon, William J. Hardcastle, and Paul Fletcher

2000 Covert contrast as a stage in the acquisition of phonetics and phonology. *Papers in Laboratory Phonology* 5: 194-207.

Shames, George H., Elisabeth H. Wiig, and Wayne Secord
1997 *Clinical Evaluation of Language Fundamentals-3*. Austin, TX: Psychological Corporation.

Smith, Rachel
2002 Does memory for individual talkers help word segmentation? Paper presented at the ISCA Temporal Integration in the Perception of Speech Workshop, Aix-en-Provence, April 2002 (abstract at http://www.lpl.univaix.fr/~tips).

Strand, Elizabeth A.
1999 Uncovering the role of gender stereotypes in speech perception. Journal of *Language and Social Psychology* 18: 86-99.

Strand, Elizabeth A.
2000 Gender stereotype effects in speech processing. Ph.D. diss., Department of Linguistics, Ohio State University.

Strand, Elizabeth A., and Keith Johnson
1996 Gradient and visual speaker normalization in the perception of fricatives. In *Natural language processing and speech technology: results of the 3rd KONVENS conference in Bielefeld*, D. Gibbon (ed.), 14-26. Berlin: Mouton.

Vitevitch, Michael S., and Paul A. Luce
1999 Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40: 374-408.

Walley, Amanda C.
1988 Spoken word recognition by young children and adults. *Cognitive Development* 3: 137-165.

White, S. David
2001 Covert Contrast, Merger, and Substitution in Children's Productions of /k/ and /t/. M.A. thesis, Ohio State University.

Williams, Kathleen T.
1997 *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Services.