RUNNING HEAD: Nonword Repetition and Phonological Disorders

Relationships between Nonword Repetition Accuracy and other Measures of Linguistic Development in Children with Phonological Disorders

Benjamin Munson

Department of Speech-Language-Hearing Sciences, University of Minnesota

Jan Edwards

Department of Speech and Hearing Science, Ohio State University

Mary E. Beckman

Department of Linguistics, Ohio State University

Please send correspondence to:

Benjamin Munson
Department of Speech-Language-Hearing Sciences (Formerly Communication Disorders)
University of Minnesota
115 Shevlin Hall
164 Pillsbury Drive, SE
Minneapolis, MN 55455
Phone: (612) 624-0304
Email: Munso005@umn.edu

Abstract

A growing body of research has documented effects of phonotactic probability on young children's nonword repetition. This study extends this research in two ways. First, it compares nonword repetitions by 40 young children with phonological disorders with those by 40 age peers with typical phonological development on a nonword repetition task in which the frequency of embedded diphone sequences was varied. Second, it examines the relationship between the frequency effect in the nonword repetition task and other measures of linguistic ability in these children. Children in both groups repeated low-frequency sequences less accurately than high-frequency sequences. The children with phonological disorders were less accurate overall, but showed no larger disadvantage for the low-frequency sequences than their age peers. Across the group, the size of the frequency effect was correlated with vocabulary size, but it was independent of measures of speech perception and articulatory ability. These results support the hypothesis that the production difficulty associated with low-frequency sequences is related primarily to vocabulary growth, rather than to developments in articulatory or perceptual ability. By contrast, production problems experienced by children with PD do not appear to result from difficulties in making abstractions over known lexical items. Instead, they may be associated with difficulties in building representations in the primary sensory and motor domains.

Relationships between Nonword Repetition Accuracy and other Measures of Linguistic Development in Children with Phonological Disorders

A number of recent studies have examined the influence of the likelihood of occurrence of phoneme sequences (*phonotactic probability*) on the processing and production of nonsense words. These studies have found that nonwords containing sequences of phonemes attested in many words (*high-probability* nonwords) are processed and produced differently from those containing sequences of phonemes attested in few or no words (*low-probability* nonwords). Phonotactic probability has been found to affect repetition latency, accuracy, and duration; novel-word learning; ratings of wordlikeness, and both immediate and long-term memory (e.g., Frisch, Large, & Pisoni, 2000; Munson, 2001; Storkel, 2001; Vitevitch & Luce, 1999). The effects of phonotactic probability have been found to change throughout development. For example, the effect of phonotactic probability on segment durations and repetition accuracy declines with age (Munson, 2001). These findings prompted us to explore *how* and *why* the effects of phonotactic probability change during the course of phonological development.

Edwards, Beckman, and Munson (2004) examined the influence of phonotactic probability on nonword repetition accuracy in typically developing children aged 3 to 8, and in adults. Edwards et al. replicated Munson's finding that the difference in repetition accuracy between high- and low-frequency sequences of phonemes declines with age. Moreover, they showed that this decline was correlated more strongly with vocabulary growth than with age. That is, there was an interaction between the magnitude of the difference in repetition accuracy between high- and low-frequency sequences (which Edwards et al. termed the *frequency effect)* and vocabulary size: the frequency effect was smaller in children with larger vocabularies, even when the effects of age were controlled. These results suggest a model in which children's representation and productive control of sublexical fragments is related to their vocabulary growth. In such a model, "phonemes" — i.e., representations of each of the consonant and vowel categories of the language robustly abstracted away from the acoustic/auditory and articulatory representations of the words in which they appear—are presumed to develop gradually throughout childhood.

Nonword repetition is a task that can be used to assess the robustness of these abstractions. In nonword repetition, the speaker must assemble production plans for individual phonemes into a larger unfamiliar sequence. This may be more difficult when the nonwords contain infrequent sequences of phonemes because children cannot bootstrap from information in known lexical items. Edwards et al. (2004) hypothesized that young children with smaller vocabularies repeated nonwords less accurately when they contained sequences that were not attested in the lexicons because they did not have access to robustly abstracted phonemic representations that they could combine into novel sequences. Older children with larger vocabularies showed a smaller effect because their representational systems had grown to include robust representations of phonemes separate from the lexical items in which they occur. This interpretation is consistent with a variety of other recent investigations, which have converged on the notion that early phonological representations are relatively holistic, and that lexical growth influences the ability to segment words into phonemes, and to produce and to process these phonemes independently of the words in which they occur (Edwards et al., 2004; Metsala, 1999; Storkel, 2002). These phonemic representations are categorical units that link to the more primary acoustic and articulatory representations of whole word-shapes that children accrue by producing and perceiving words. These phonemic categories supplement children's knowledge of the complex relationship between articulatory movements and their acoustic consequences (i.e., the *forward model* described by Bailley, Laboissiere, & Schwartz, 1991; Jordan, 1990) that have already begun to develop in infancy during babbling.

The creation of autonomous phonemic categories would facilitate subsequent word learning, as suggested by Lindblom (1992), Beckman and Edwards (2000), and Fisher et al. (2001), among others.

Such categories would allow for a ‘fast mapping’ of the phonological structure of a new word. This mapping can be considered analogous to the fast mapping of novel words that has been studied previously (Dollaghan, 1987). That is, the fast mapping that these categories afford would allow the language learner to make an automatic association between an immediate acoustic representation of a newly encountered word, and the articulatory movements required to produce the same acoustic output. It could be that the increase in repetition accuracy noted by Edwards et al. (2004) reflects the influence of the development of categorical phonemic representations: individuals with larger vocabularies repeat low-frequency sequences of phonemes more accurately because they can more easily parse nonwords into their constituent phonemes, and recombine the associated articulatory representations into novel vocal motor schemes (constellations of articulatory gestures analogous to those discussed by McCune & Vihman, 1987 in younger children) for fluent speech production.

This view of normal phonological acquisition can potentially inform our understanding of the causes of phonological disorders (PD) in children. Children with PD demonstrate a variety of speech-production errors in the absence of an obvious medical etiology. Within this model of the relationship between phonological development and vocabulary growth, children with PD might have difficulty with building categorical phonemic representations that link lexical, articulatory, and acoustic representations to each other. That is, the delayed articulatory development of these children may be due to their not having access to the abstract representations needed both to parse newly encountered words into phoneme-like units during word learning, and to make correct associations to the speech movements required to accurately reproduce these newly learned words. As a consequence of this inability, children with PD may develop more poorly specified phonological representations of words. Although the specific errors that would result from these poorly specified representations are not the focus of this investigation, this account predicts that these children might default to the auditory and perceptual patterns that are well established in representations of other words; to patterns that are articulatorily easy; and/or to patterns that afford the most perceptual distance between words in the productive vocabulary. Over time, these error patterns would become gradually more entrenched, and might give the impression that a child has developed a systematic rule or constraint that operates uniformly over his or her lexicon. If so, the difficulty of children with PD might be similar to the difficulties encountered by the children with small vocabularies in Edwards et al (2004). That is, the children with PD might show greater differences in repetition accuracy between high- and low-frequency nonwords because of their poorly developed categorical phonemic representations. This view differs fundamentally from accounts of children's sound-production errors as resulting from higher-level phonological rules or constraints that are specified in a module of the grammar separate from the lexicon, and that are possibly innate and universal (e.g., Barlow & Dinnsen, 1998; Barlow & Gierut, 1999; Bernhardt, 1994; Bernhardt & Stoel-Gammon, 1996; Ingram, 1976; Shriberg & Kwiatowski, 1994). However, it is similar to these accounts in that phonological disorder is seen as a deficit in higher-level representations instead of (or in addition to) difficulty with the primary articulatory and auditory domains.

However, there is a second account of the relationship between vocabulary growth and the frequency effect that was not considered by Edwards et al. (2004). The increase in repetition accuracy with vocabulary growth observed by those investigators might simply reflect more robust generalizations about the different sources of variability in acoustic and/or articulatory representations. That is, the children with larger vocabularies might have been more accurate in repeating the nonword stimuli simply because their acoustic representations of the component sound patterns are better developed from having heard these patterns in many different real words. Alternatively, their motor representations of the articulatory gestures needed to produce the component sound patterns might be better developed, from more experience with producing them in many different real words in different prosodic and sentential contexts; under different environmental conditions; and in the presence of different degrees of afferent feedback.

This second account of the developmental change in the frequency effect also has implications for our understanding of PD. Children with PD might simply have difficulty in either or both of these primary representational domains, and these difficulties might be either a proximal or distal cause of their phonological disorders. That is, their deficits might be explained by their difficulties in building peripheral perceptual and articulatory representations. Such an account of PD would be consistent with views in which PD stems from deficits in articulatory-motor and/or acoustic-auditory representations (e.g., Edwards, 1992; Edwards, Fourakis, Beckman, & Fox, 1999; Forrest, Chin, Pisoni, & Barlow, 1995; Gibbon, 1999; Towne, 1994).

The purpose of this paper, then, is twofold. First, it explores the alternative explanation for the frequency effect that was not considered by Edwards et al. (2004). This is done by examining interactions between performance on the nonword repetition task and measures of speech perception and speech articulation, as well as measures of vocabulary size. This comparison allows us to determine the relative extent to which development of more primary sensory representations, as distinct from growth in vocabulary size, contribute to the frequency effect. Second, the paper compares the performance of typically developing children to that of children with phonological disorders. This comparison allows us to examine whether the deficits of children with PD can be related to difficulties in building categorical phonemic representations that link among acoustic, articulatory, and lexical representations, or whether they are due to deficits in building lower-level perceptual and motor representations.

Our specific experimental questions are as follows. First, what independent measures predict the magnitude of the frequency effect in typically developing children, and in children with PD? To address this question we examined regressions predicting the frequency effect from measures of speech perception, phoneme-production accuracy in real words, and vocabulary size. A finding that the frequency effect is predicted equally well by all three measures would argue against Edwards et al.'s conclusion that its decrease during development is related to the influence of vocabulary growth on the development of categorical phonemic units. Instead, such a result would suggest that developmental changes in the frequency effect are related more broadly to the development of knowledge about the articulatory and acoustic structure of words. A finding that only vocabulary-size measures predict this would strengthen Edwards et al.'s claim.

Our second question was whether children with PD demonstrate a larger effect of phoneme-sequence frequency than typically developing children matched for age. A finding that children with PD demonstrate a larger frequency effect would suggest that they have difficulty building an intermediate layer of phonemic representations that mediate among acoustic, articulatory, and lexical representations, and that this difficulty might be causally related to their disorder. Conversely a finding that children with PD differ only on measures of perceptual ability and overall production accuracy in words and nonwords would suggest that their difficulties in production are not related to any problems with this intermediate layer of phonemic representations.

Method

Participants

The participants were 40 children with phonological disorders (PD), ranging in age from three to six years, and 40 age controls with typical phonological development (TD). All children were participating in a larger study on the etiology of phonological disorder. The 40 children with TD were a subset of the children who participated in the study described in Edwards et al. (2004). All of the participants, both the children with PD and the children with TD, met the following criteria: (1) normal hearing, as evidenced by passing a hearing screening at 20 dB at 500, 1000, 2000, and 4000 Hz (ANSI, 1989); (2) normal non-verbal intelligence, as evidenced by a standard score no more than one standard deviation below the mean on the *Columbia Mental Maturity Scale* (CMMS, Burgemeister, Blum, & Lorge, 1972); and (3) normal structure and function of the peripheral speech mechanism, as evidenced by a standard score no more than one standard deviation below the mean on the oral movement subtest of the *Kaufman Speech Praxis Test for Children* (KSPT, Kaufman, 1995). Two children with phonological disorders had standard scores below 85 on the KSPT but were still included in the study because their low scores were not related to problems with neuromotor control. One child received a 78, with errors related to a short lingual frenum, and the other child received an 81, with errors related to his refusal to touch his tongue to his lips because of severely chapped lips. The two groups differed in KSPT scores. This was true when the two outliers were included (PD mean = 98, TD mean = 106, $t[73] = 3.2$, $p < 0.01$) and when they were omitted (PD mean = 101, TD mean = 106, $t[71] = 2.9$, $p < 0.01$). We are confident that this difference is not indicative of group differences in overall speech praxis, given that all but two of the children scored within normal limits, and that the scores for the two outliers were due to exceptional circumstances. All of the children with PD scored at or below the 10th percentile on the *Goldman-Fristoe Test of Articulation* (GFTA, Goldman & Fristoe, 1986), while all of the children with TD scored at or above the 40th percentile on the GFTA and were individually matched to the children with PD on the basis of age (within 6 months), gender, and non-verbal IQ score on the CMMS (within 10 points – i.e., twice the standard error of measurement). In this study, as in Edwards et al. (2004), we examined the relationship between the effect of sublexical sequence frequency on accuracy and vocabulary size. We used the *Peabody Picture Vocabulary Test-III* (PPVT-III, Dunn & Dunn, 1997) to estimate receptive vocabulary size, and we used the *Expressive Vocabulary Test* (EVT, Williams, 1997) to estimate expressive vocabulary size. These two tests are co-normed for participants aged 2 through 90.

Table 1 gives descriptive information for the 40 children in each of the two groups. One-way ANOVAs revealed significant group differences on GFTA percentile rankings ($F[1, 78] = 441.35$, $p <$

0.001, $\eta^2 = 0.85$). While all children with PD scored within normal limits (no more than one standard deviation below the mean) on both measures of vocabulary size, there was also a small but significant group difference on PPVT-III standard scores ($F[1, 78] = 7.54$, $p < 0.01$, $\eta^2 = 0.08$). Consequently, PPVT-III scores were used as a covariate in all parametric analyses of group differences. There were no significant differences between the two groups on CMMS or EVT standard scores.

Insert Table 1 about here

Nonword repetition

A detailed description of the stimuli for the nonword repetition task, the procedure for data collection, and the methods used for transcribing the data and for scoring segmental accuracy is provided in Edwards et al. (2004). Briefly, the stimuli consisted of 22 two-syllable nonwords and 22 three-syllable nonwords, presented in Table 2. One-half of the nonwords contained a high-probability diphone sequence, and one-half contained a low-probability diphone sequence. Each of the high-frequency diphone sequences was matched to a phonetically similar low-frequency sequence. For example, the high-frequency sequence /ju/ was matched to the low-frequency sequence /jaʊ/. The target sequences were created by combining earlier-acquired sounds (Smit, Freilinger, Bernthal, Hand, & Bird, 1990) to make up seven pairs of CV sequences, seven pairs of VC sequences, and eight pairs of CC sequences. The sequences were embedded in larger wordshapes, with the same stress and syllable structure framing both target diphone sequences in each pair. These frames were not identical in phonetic content, as pilot work had shown that the use of identical frames induced a practice effect. The transitional probability of the frames was matched as closely as possible across the two word lists. Previous research has shown that wordlikeness is positively related to repetition accuracy (Gathercole, Willis, Emslie, & Baddeley, 1991). As discussed in Edwards et al. (2004), the wordlikeness of the high- and low-probability stimuli used in this study was not equivalent. However, the differences in wordlikeness between the high- and low-probability stimuli were small. Moreover, regression analyses presented in Edwards et al. showed that the differences in wordlikeness between the two lists of stimuli were due to small differences in the frequency of the phoneme combinations making up the non-target portions of the nonwords, rather than to the large differences in frequency between the target diphone sequences.

Insert Table 2 about here

Pre-recorded production prompts were used in the nonword repetition task. Three pseudo-randomized lists of the nonword stimuli were created. In each list, (a) the two members of each nonword pair were separated by at least two other stimuli; (b) all of the two-syllable nonwords were presented prior to the three-syllable nonwords; and (c) equal numbers of nonwords containing high-frequency sequences were presented prior to the paired word containing a low-frequency sequence, and vice versa. These were played from the hard-drive of a laptop computer through external speakers. Participants were instructed to repeat the nonwords as accurately as possible. Children wore a head-mounted microphone attached to a DAT recorder, and their repetitions were recorded for later analysis.

Accuracy of nonword repetition was scored as in Edwards et al. (2004). Each target sequence could receive a maximum of six points for segmental accuracy, or three points per component phoneme. Consonants received one point each for correct place, correct manner, and correct voicing, while vowels received one point each for correct front-back position, correct vowel height, and correct "length" (i.e., tense or lax for a monophthong target and monophthong or diphthong for a diphthong target). For example, a child's production of [sg] in [busgit] for the target sequence [fk] in the nonword [bufkit] would have been scored as 4 (2 points for correct manner and voicing in the s-for-f substitution, and 2 points for the correct place and manner for the g-for-k substitution).

A single researcher transcribed all of the responses of the children with PD and age controls. The recordings were transferred from the DAT to a digital file on a PC and the productions were transcribed in the International Phonetic Alphabet at the level of a careful, broad phonemic transcription. The transcriber used a waveform editor, so that she could easily listen to each word as many times as necessary. A second transcriber independently transcribed 15 percent of the data from the children with PD and the children with TD (two three-year-olds, two four-year-olds, and two five-year-olds from each group). Phoneme-by-phoneme inter-rater reliability ranged from 87 to 92 percent for data from individual participants with PD, with a mean of 89 percent. Inter-rater reliability for the children with TD was 92 percent, with a range of 86 to 99 percent for individual participants.

Speech perception ability

One of the goals of this project was to examine whether accuracy of repetition of high- and low-frequency sequences was affected by speech perception ability. The measure of speech perception ability in this article comes from a speech perception task which all of the children had completed as part of their participation in the larger study (Edwards, Fox, & Rogers, 2002). In this task, they were asked to choose between two CVC words that differed only in the identity of the final consonant, in both a gated condition (i.e., a condition in which the burst or the burst and most of the VC formant transition into the closure of the final consonant had been removed), and a condition in which the entire word was played. Due to technical problems, data were missing for two children with PD. In this experiment, children listened to audio files of pre-recorded, digitized stimuli played from a computer. On each trial, children were asked to identify which of two pictures the stimulus named. Two different pairs of words were used, *cap*/*cat* and *tap*/*tack*. The experiment was blocked by word pair; the participants always made a two-alternative forced-choice identification. For example, on a particular trial, a child might hear a full or gated version of the word *cat*, and point to a picture of either a cat or a cap. Further details about the administration of this task can be found in Edwards et al. (2002).

The measure we chose from this experiment was the d-prime values from the ungated condition for the two stimulus sets, averaged together. This served as a gross indicator of the participants' ability to attend to fine phonetic detail in speech perception, since successful performance on this task requires children to attend to subtle acoustic cues for final consonant place of articulation. Edwards et al. (2002) selected this measure for use in their examination of the relationships among speech perception, age, vocabulary size, and articulatory ability, because the TD participants in every age group and the children with PD all performed above chance only in the ungated condition. They had found that this measure was correlated with both receptive vocabulary size, as measured by PPVT-III raw score, and with articulatory ability, as measured by GFTA raw score.

Assessing accuracy of individual phoneme segments

One additional goal of this project was to examine whether accuracy for high- and low-frequency sequences was related to children's phoneme-production accuracy in real words. To examine this, we measured the number of times each sound from the target diphone sequences was produced correctly on a non-norm-referenced Phonetic Inventory test (PI). The PI is a picture-naming task that we developed to provide a more detailed probe of articulation skills than the GFTA (Isermann, 2001). Like the GFTA, the PI used colored line drawings of familiar words to elicit single word productions of each consonant of English in initial, medial, and final word positions. The PI provided a more in-depth assessment of a child's articulation skills than the GFTA because it elicits each target consonant in three different words for each target word position, and because it also assesses vowel production. Due to recording problems, data were lost for two children with PD from the PI.

Each production of a target sound was scored as correct or incorrect, and the number of correct productions was summed to get an accuracy rate for that segment. Vowels were elicited in the stressed syllables of three words and the consonants /h/, /j/, and /w/ were elicited in word-initial position in three words. Thus, the maximum accuracy rate for each of these sounds was 3, indicating that the child produced the sound correctly in all three target words. Nearly all of the other consonants were elicited in three different positions (initial, medial, and final) yielding a maximum accuracy rate of 9.

The only exceptions were /**θ**/, which was not elicited in word-medial position, and /**ð**/, which was not elicited in word-final position. These sounds are very infrequent in these positions, and it was difficult to find pictureable words to elicit them.

The researcher who had transcribed a subset of the nonwords for the reliability analysis transcribed the words elicited with the PI for all of the children, using the same methodology as was used for the nonwords. Another researcher (who was the primary transcriber for the nonwords) then independently transcribed the words elicited with the PI for a subset of the children comprising two three-year olds, two four-year olds, and two five-year olds, with one child with PD and one child with TD for each of these three age ranges. Phoneme-by-phoneme inter-rater reliability ranged from 89 to 95 percent, with a mean of 92 percent.

Summary Measures of Real-Word and Nonword Accuracy

In the current study, we used the productions elicited with the PI to assess whether differences in the accuracy scores for phonemes on the nonword repetition task were due to the phonotactic probabilities of the high- and low-probability sequences, or to more general difficulty producing the sounds that made up the high- and low-frequency sequences. To make this assessment, we calculated total segmental accuracy scores for the segments in the high- and low-frequency target sequences on the nonword task, and we converted the accuracy rates for the same phonemes elicited as targets on the PI into weighted total scores. That is, we summed (a) the segmental accuracy scores for the phonemes in the two sequence types on the nonword task, and (b) the accuracy scores for the *same* phonemes in real words, as elicited by the PI. Since we expected vowels to be generally more accurate than consonants, we calculated separate totals for the consonants and for the vowels, to make two sums for the segments that were in the high-frequency target sequences and two sums for the segments that were in the low-frequency target sequences.

For the nonword task, these four total segmental accuracy scores (consonant and vowel scores in high- and low-frequency nonwords) were simply the sums of the accuracy scores for individual phonemes in each category. That is, the scores for the high-frequency nonwords was simply the sum of the 3-point segmental accuracy scores for all of the consonants in the high-frequency diphone sequences.

As described above, three points was the highest segmental accuracy score that a target phoneme could receive on the nonword task. There were 14 vowels in the high- and low-frequency sequences, so the maximum possible score for the vowels was 42. There were 30 consonants in the high- and low-frequency sequences, so the maximum possible score for the consonants was 90. In the results section, we refer to these as the *nonword accuracy sums*.

To make comparable totals for the same phonemes on the PI, we needed to weight the accuracy rates for the target phonemes elicited in the real words by the number of times each sound occurred in the target sequences in the nonword stimuli. To do this, we treated the number of correct productions of the target phoneme on the PI as an analog to the three-point segmental accuracy score on the nonword repetition task, and multiplied this "accuracy score" by the number of times the segment occurred in the relevant target sequence category in the nonword stimuli. For example, the consonant /f/ appears in three low-frequency target sequences on the nonword repetition task in the stimuli /**auftəgɑ/**, **/bufkit/**, and **/næfkətu/**. Therefore, we multiplied the PI accuracy score for /f/ by 3 in calculating a child's total PI accuracy for consonants in low-frequency sequences. Similarly, the vowel /æ/ occurs in two high-frequency target sequences in the stimuli /mæbɛp/ and /bedæ**g**/. Therefore, we multiplied the PI accuracy score for /æ/ by 2 in calculating a child's total PI accuracy for vowels in high-frequency sequences. The consonants and vowels making up the high- and low-frequency sequences are listed in Table 2 in the bottom two rows. For example, the total segmental accuracy score for consonants making up the high-frequency diphone sequences would be calculated as the segmental accuracy sum of /**f**/ multiplied by two; /**g**/ multiplied by 4; /**m**/ multiplied by 3; /**v**/ multiplied by 2; /**j**/; /**n**/ multiplied by 2; /**k**/ multiplied by 4; /**p**/ multiplied by 3; /**t**/ multiplied by 6; /**w**/ multiplied by 2; and /**d**/.

As described above, the PI accuracy scores for vowels ranged from 0 to 3 because each vowel was elicited three times. The total possible PI accuracy score for the vowels in each of the two sequence types therefore is 42, the same as for the total accuracy score for the vowels in the nonword task. The consonant types that were used in the 8 CC and 14 CV and VC sequence pairs

were **/p, t, k, b, d, g, m, n, f, v, j, w**/. For most of these consonants, the maximum number of correct productions was 9, for the three elicitations of each in word-initial, medial, and final position. The two exceptions were /w/ and /j/, which occurred in several CV sequences in the nonword stimuli. These were elicited only in word-initial position on the PI. We multiplied the accuracy scores for these two consonant types by 3 before multiplying each consonant type by the number of times it occurred in the target sequences so that the accuracy scores for /w/ and /j/ would be weighted similarly to the accuracy scores for all of the other consonant types. The highest possible PI accuracy score for consonants was thus 270 – i.e., 9 (for the maximum adjusted PI score) * 30 (for the 30 consonants in the high- and low-frequency sequences). In the results section, we refer to these as the *real-word accuracy sums*. These data were used for two purposes. First, they were used in regression analyses in which we predicted children's accuracy in phoneme production in nonwords from their accuracy in producing the same phonemes in real words. Second, we used in them in analyses of covariance, to determine whether there were group differences in participants' accuracy in producing the sounds that made up the high- and low-frequency diphone sequences. Table 3 summarizes the measures that we used in our analyses, including the tasks from which they were computed.

Insert Table 3 about here

Results

This section is divided into two parts. The first part uses analyses of covariance to examine group differences in the mean accuracy of high- and low-frequency sequences, and differences in the types of errors that the children produced. The second part uses a variety of regression analyses to examine predictors of each child's overall accuracy, as well as predictors of the difference in accuracy between high- and low-frequency sequences for each child's productions.

Analyses of Covariance

Mean Nonword Repetition Accuracy

The first analysis examined group differences in mean accuracy scores for the high- and low-frequency sequences in the nonword repetition task, averaged across the 22 types. Figure 1 shows mean accuracy scores for high- and low-frequency sequences for the children with PD and their TD peers. Since the two groups differed somewhat in their PPVT-III standard scores, these were used as covariates in this analysis. An α-level of 0.05 was used to determine significance for all analyses. A two-way (frequency by group) mixed-model ANCOVA revealed significant effects of frequency ($F[1,77] = 10.2, p < 0.01, \eta^2 = .12$) and group ($F[1,77] = 47.8, p < 0.01, \eta^2 = .37$). The PPVT-III was a significant covariate ($F[1,77] = 9.5$ $p < 0.01$, $\eta^2 = .11$). As expected, the PPVT-III interacted significantly with frequency ($F[1,77] = 5.0$, $p < 0.05$, $\eta^2 = .06$). This essentially replicates Edwards et al.'s finding that vocabulary size differentially predicted production accuracy of high- and low-frequency sequences. This interaction is explored further with multiple regression analyses below. Finally, group and frequency interacted significantly ($F[1,77] = 4.8$, $p < 0.05$, $\eta^2 = .06$). This interaction can be seen by comparing the difference in bar heights in Figure 1.

Insert Figure 1 about here

Figure 1 shows that the typically developing children produced high-frequency sequences on average .49 points higher than they produced low-frequency sequences. In contrast, the difference in accuracy for the children with PD was a somewhat smaller .35.

Mean Accuracy in Real Words

The second set of ANCOVAs examined whether there were group differences in accuracy of production of the consonants and vowels in the high- and low-frequency sequences when they occurred in real words on the PI. For this analysis, the real-word accuracy sums were transformed to percentage-correct scores. As is conventional with percentage scores, these were arcsine transformed prior to being used in statistical analyses, to minimize error variance. The arcsine-transformed scores were submitted to a two-factor mixed-model ANCOVA, with group (PD v TD) as the between-subjects factor, frequency (phonemes making up the high- vs. low-frequency sequences) as the within-subjects factor, and PPVT-III as the covariate. Vowels and consonants were examined separately.

When consonants were examined, there was no significant main effect of frequency, $F[1,73] = 2.7$, $p > 0.05$. Across the two groups, the consonants that were used to create the high-frequency and low-frequency sequences were produced with approximately equal accuracy when occurring in real words (HF mean = 85.2%, LF mean = 83.7%). Unremarkably, the main effect of group was significant, $F[1,73] = 30.4$, $p < 0.01$, partial $\eta^2 = 0.29$. Across the two sequence types, children with PD produced these consonant sounds with 75.5% accuracy; children with TD produced them with 94.4% accuracy. The PPVT-III was not a significant covariate, $F[1,73] = 2.9$, $p > 0.05$. Although the PD children showed a slightly larger difference than the TD children in production accuracy for the sounds used in creating the high- and low-frequency sequences, this interaction did not achieve statistical significance (HF mean = 76.2%, LF mean = 74.8% for children with PD; HF mean = 94.3%, LF mean = 94.6% for children with TD).

When vowel production accuracy was examined, there was no significant main effect of frequency, $F[1,73] = 1.2$, $p > 0.05$. The vowels used in creating the high-frequency sequences were produced in the real words with 93.0% accuracy; those used in created the low-frequency sequences were produced with 94.3% accuracy. There was a significant main effect of group, $F[1,73] = 14.5$, $p < 0.05$, partial $\eta^2 = 0.17$. Children with TD produced vowels more accurately than children with PD (TD mean = 97.8%, PD mean = 89.6%). Again, the PPVT-III was not a significant covariate. None of the interactions was significant, although slightly different patterns were noted for the two groups: children with TD produced vowels associated with the high-frequency sequences with 98.2% accuracy when these occurred in real words and vowels associated with the low-frequency sequences with 97.3% accuracy. In contrast, children with PD produced vowels for high-frequency sequences with 87.3% accuracy, and vowels for low-frequency sequences with 91.3% accuracy.

The results from these ANOCVAs on accuracy of phoneme production in real words suggests that the observed differences between repetition accuracy for high-probability sequences and low-probability sequences in nonwords is not due to their segmental composition: when the same sounds are assessed in real-word productions, accuracy did not differ between the two sets. Relationships between real-word accuracy and accuracy in nonwords is explored further with regression analyses below.

Low-frequency versus Zero-frequency Sequences

One factor complicating the interpretation of the frequency effect in nonword repetition is that the low-frequency sequences in this study included some non-occurring sequences (such as /**aʊk**/, which does not occur in any words in the databases used to create the stimuli) as wells as some sequences that occur in a small number of real words (such as /**gaʊ**/, which occurs in the word *nightgown*). It is possible that the children with TD repeated low-frequency sequences more accurately than the children with PD because they knew more words and therefore might have had more prior experience with saying the attested low-frequency sequences. If this were the case, we would predict that there would be larger group differences on the low-frequency sequences than on the zero-frequency sequences. Alternatively, it is possible that the children with TD were more accurate because they had more robustly abstracted representations of individual phonemes to use in creating completely novel production routines. If that were the case, then we might expect there to be larger differences on the zero-frequency sequences. To examine whether either of these two possibilities was supported by this data set, we conducted two additional ANCOVAs examining the differences between these two sequence types. The first ANCOVA examined the difference in accuracy between the low-frequency but attested sequences and their matched high-frequency sequences; the second examined accuracy differences between zero-frequency sequences and their matched high-frequency counterparts. It should be noted that the high-frequency sequences in these two analyses were different, as these ANCOVAs examined matched pairs of high- and low- or zero-frequency sequences. This was a post-hoc analysis; the original stimulus set was not designed to examine these differences. Thus, the stimuli were not counter-balanced for whether they were attested in a few words or completely unattested. Nine of the 22 low-frequency sequences were attested in between one and six words in the HML. Four of these nine sequences were used in both two- and three-syllable nonwords. This resulted in 13 nonwords containing low-frequency sequences, and nine nonwords containing zero-frequency sequences.

The first ANCOVA examined the difference in accuracy between the attested low-frequency sequences and their matched high-frequency counterparts. The within-subjects factor was frequency, and the between-subjects factor was group. The covariate was PPVT-III standard score. Significant main effects were found for group ($F[1,77] = 30.6$, $p < 0.01$, partial $\eta^2 = 0.28$) and frequency ($F[1,77] = 10.7$, $p < 0.01$, partial $\eta^2 = 0.12$). Moreover, the PPVT-III was a significant covariate ($F[1,77] =$

12.8 $p < 0.01$, partial $\eta^2 = 0.14$). As in the earlier analysis, the PPVT-III interacted with frequency ($F[1,77] = 7.5$, $p < 0.01$, partial $\eta^2 = 0.09$). However, a significant interaction between group and frequency was not found. The performance of the two groups on the low-frequency sequences is shown in Figure 2a.

The second ANCOVA examined differences in accuracy between the zero-frequency sequences and their matched high-frequency counterparts. Again, the within-subjects factor was frequency, and the between-subjects factor was group. The covariate was PPVT-III standard score. Significant main effects were found for group ($F[1,77] = 42.1$, $p < 0.01$, partial $\eta^2 = 0.35$) and frequency ($F[1,77] = 3.8$, $p < 0.05$, partial $\eta^2 = 0.05$). Moreover, the PPVT-III was a significant covariate ($F[1,77] = 5.7$ $p <$ 0.01, partial $\eta^2 = 0.07$). Unlike in the earlier analyses, the PPVT-III did not interact with frequency. A significant interaction between group and frequency was not found. The performance of the two groups on the zero-frequency sequences is shown in Figure 2b.

Insert Figures 2a and 2b about here

Group differences in accuracy were found for both sequence types. Thus, it seems unlikely that they were due simply to group differences in familiarity with the attested low frequency sequences. In neither case was there a group by frequency interaction. However, inspection of Figures 2a and 2b suggests that there was a larger difference in accuracy between the zero-frequency sequences and their matched high-frequency counterparts than between the attested low-frequency sequences and their matched high-frequency counterparts. This is not surprising, as it indicates that participants in both groups experienced greater difficulty generalizing correct production of phonemes to completely novel sequences than to sequences that were represented in the lexicon, albeit with few examples.

Regression Analyses

The second set of analyses used a variety of regression analyses to examine predictors of overall accuracy of nonword repetition, and of the frequency effect. Tables 4, 5, and 6 show simple correlations (Pearson's r) between various predictor measures and various measures of each subject's performance on the nonword repetition task, first for the entire group of subjects, then for the TD children only, and then for the PD children only. Inspection of these tables suggests that each participant's mean repetition accuracy is well predicted by a variety of variables, for both the low-frequency and the high-frequency sequences. However, the only consistent predictors of the frequency effect were measures of vocabulary size.

Insert Tables 4, 5, and 6 about here

Sequence Frequency and Accuracy in Real Words

The first regression analysis was intended to assess articulatory difficulty or general articulatory ability as a potential factor mediating the effect of sequence frequency on nonword repetition accuracy. As noted earlier, low-frequency target sequences often contain one segment that is itself lower in frequency than the analogous segment in the paired high-frequency sequence. For example, the low-frequency sequence /moɪ/ contains the vowel /oɪ/, which is very infrequent in English. The matched high-frequency sequence /mæ/ contains the vowel /æ/, which is much more frequent than /oɪ/ in English words. If we assume that low-frequency phonemes are inherently more difficult to produce than high-frequency phonemes, then we might hypothesize that the effect of phonotactic probability on repetition accuracy is due simply to the phonetic composition of the two sequence types. If this is the case, then we should find (a) that accuracy of production of that sound in a familiar word predicts accuracy of production of that sound in a nonword; and (b) that frequency of the sequence in which that sound is embedded should not predict performance beyond what is predicted by accuracy in real words. The number of participants analyzed in these regressions was smaller than in the analysis of nonword repetition accuracy, as data from the PI for two of the children with PD were lost due to recording problems.

To evaluate this question, we ran a series of regressions predicting phoneme accuracy in nonwords from accuracy of production of the same phonemes in real words. Vowels and consonants were examined separately. The dependent measures in these regressions were based on the nonword accuracy sums described in the methods section. The dependent measure in the first regression was the average nonword accuracy sum for the sounds in the high- and low-frequency sequences in the nonword repetition task. The independent measure in this regression was the average accuracy sum for the same sounds when produced in real words in the PI . In the second set of regressions, the dependent measure was the difference in repetition accuracy between high- and low-frequency sequences, which we term the *frequency effect*. The independent measure in this set of regressions was the difference in accuracy for the phonetic inventory sums for the same sounds when produced in real words, which we call the *difference in real-word accuracy scores*. Vowels and consonants were examined separately for both sets of regressions. For both consonants and vowels, three separate regressions were run: one for the entire group of participants, one for the participants with PD, and one for the TD participants. If the effect of frequency on nonword repetition is simply due to the intrinsic difficulties of producing the sounds that make up the low-frequency sequences, then the difference in real-word accuracy sums should predict the frequency effect. If the frequency effect were related to some other aspect of development, then we would predict that there would be no relationship between those measures.

Results of these regressions are presented in Table 7. Not surprisingly, there was a high correlation between nonword accuracy sums and real-word accuracy sums, when the entire group of 80 children is examined. Children who produce sounds accurately in familiar words are also likely to produce sounds accurately in nonwords, and vice versa. However, this relationship was not present for the children with TD, either for vowels or for consonants because of the limited range of performance on both the dependent and independent measures in this population. It was robustly present for the children with PD for both phoneme types.

The correlation between the real-word accuracy sums and the nonword accuracy sums in the entire group of participants with TD and PD, and in the group with PD only contrasts strongly with the lack of a relationship between the frequency effect and the difference in real-word accuracy sums. That is, the effect of sequence frequency on repetition accuracy was not predicted by measures of the difference in accuracy for the same sounds in real words.

Insert Table 7 about here

Thus, there is an effect of sequence frequency on accuracy of consonant and vowel production, over and above the accuracy of articulation of the sounds making up the high- and low-frequency sequences in real words. This suggests that developmental decrease in the frequency effect does not index a simple developmental decrease in the difference in production accuracy for the phonemes composing the high- and low-frequency sequences, as measured by the PI.

Sequence Frequency and Measures of Speech Perception

The next set of analyses examined whether the frequency effect could be predicted by measures of the participants' skill in speech perception. In this analysis, we examined whether whole-word discrimination accuracy predicts the frequency effect. We also examined relationships between whole-word discrimination and average nonword repetition accuracy.

We used simple regression to examine these relationships. In the first set of analyses, mean accuracy (averaged across high- and low-frequency sequences) was the dependent measure. Whole-word discrimination was the independent measure. When the entire group of participants was examined, a significant relationship was found ($F[1,76] = 13.9, p < 0.01$). Whole-word discrimination accounted for 16.1% of the variance in accuracy scores. The standardized β coefficient was .401, indicating that participants with better speech-perception ability also evidenced more accurate nonword repetition. When the two groups were examined separately, a significant predictive relationship was only found for the children with PD ($R^2 = 13\%$, $F[1,76] = 5.4$, $p < 0.05$, $\beta = .36$). No relationship was found for the TD children. However, once again the children with TD in this group showed a much smaller range of performance on the speech-perception task, and the lack of a predictive relationship may have been due to this.

In the second set of analysis, the frequency effect was the dependent measure. Again, the frequency effect was the difference in mean accuracy between the high- and low-frequency sequences, using the six-point scale. Whole-word discrimination was the independent variable. When the entire group of 80 participants was examined, the relationship was not significant. The same was true when the 40 children with TD and the 40 children with PD were examined. That is, the magnitude of the frequency effect was not affected by speech-perception ability. While children with poorer speech-perception ability were less accurate than children with better speech-perception ability on repetition accuracy overall, they were not disproportionately poorer at repeating low-frequency sequences.

Sequence Frequency and Vocabulary Size

The regression analyses presented thus far found no relationship between perceptual or articulatory skills and the frequency effect. Edwards et al. (2004) observed an interaction between vocabulary size and sublexical sequence frequency. Smaller effects of sublexical sequence frequency were noted in children with larger-sized vocabularies. We wanted to examine this same relationship for the PD and TD participants in this study. As in the earlier study, we examined the relationship between phonotactic probability effects and vocabulary-size effects using hierarchical multiple regression. We ran two sets of regressions. In the first set of regressions, mean accuracy (averaged over both high- and low-frequency sequences) was the dependent measure. This set contained three individual regressions, one for the entire group of 80 participants, and one for each of the two groups. The regression had four independent measures: age, percentile rank on the GFTA, log-transformed raw score on the EVT and the PPVT-III. The variables were entered in a fixed order: age was forced as the first variable; GFTA was forced as the second variable; and PPVT-III and EVT raw scores were then entered stepwise if they accounted significantly for any additional variance ($\alpha < 0.05$). The second set of regressions examined predictors of the frequency effect, that is, the difference in accuracy between high- and low-frequency sequences. Again, three separate regressions were run; and the independent measures for these regressions were the same as for the first set. The results of these regressions are summarized in Table 8. Because age never accounted for a significant proportion of variance in either dependent measure, it is not included in this table.

Insert Table 8 about here

As Table 8 shows, GFTA percentile rank predicted a large proportion of variance in mean accuracy for both the entire group of 80 participants, and for the 40 participants with PD. In addition, PPVT-III raw score predicted an additional proportion of variance for the same two groups. No significant predictors of performance were found for the 40 TD matches. When the frequency effect was examined, GFTA was a significant predictor for the group of children with PD only. That is, despite the much larger range of GFTA scores when the two groups were combined, GFTA was not a significant predictor of the frequency effect. This is not surprising, since the children with PD as a whole showed a smaller frequency effect in the group analyses. In contrast, EVT standard score predicted a significant proportion of variance in all three regressions.

To illustrate the relationship between EVT score and the frequency effect, Figure 3 plots mean accuracy scores for the high- and low-frequency sequences against the log-transformed raw EVT scores for the whole group of 80 children (top), the children with TD (middle), and the children with PD (bottom). Figure 3 shows an interaction between vocabulary size and phonotactic probability. That is, the regression lines for the high-frequency and low-frequency sequences are not parallel. Instead, the distance between the two regression lines (which represents the effect of target sequence frequency) decreases as vocabulary size increases. This is true even for the group of children with TD, who evidence a more restricted range of performance on all of the variables than the entire group of 80 participants, or the 40 participants with PD.

Insert Figure 3 about here

One surprising finding of this series of regression analyses is that there was a significant predictive relationship between GFTA percentile rank and the frequency effect. The positive *β*-weight associated with this relationship shows that the children with PD who had lower percentile ranks (i.e., a more severe deficit) also had smaller frequency effects than children with milder disorders. Given our interpretation of the size of the frequency effect, this finding is the opposite of what we would expect to see if PD were associated with difficulties in building phonemic representations. This apparent anomaly suggests a possible progression in the specificity of the pattern that can be generalized. That is, at the very earliest stages of vocabulary expansion, the patterns which can be abstracted away from known words are very general schema such as 'labial-closure followed by an open vowel', so that something that sounds like [**ɡæʔ**] might be the word shape for *cat*, *carrots*, *car*, *gobble*, and any new word that even vaguely fits that template. There should be no difference in production accuracy between high-frequency /**kæ**/ and lower-frequency /**ɡɑ**/ until the child has accurate production routines for both of these sublexical patterns and can abstract a representation of the sequence from the following word context. Thus, over development, the frequency effect should have a floor as well as a ceiling. If this picture is correct, the seemingly anomalous finding for the relationship between GFTA and frequency effect may be due to the children with severe PD being so impaired in their speech-sound production that they experience difficulty generalizing production of two-phoneme sequences to any novel wordshape, regardless of the degree to which the sequence it was attested in lexical items.

Discussion

This study examined the ability of typically developing (TD) children and children with phonological disorders (PD) to repeat high- and low-frequency sequences of phonemes embedded in nonwords. This study replicated the earlier finding of Edwards et al. (2004) that sequences of phonemes that are attested in few or no real words are repeated less accurately than those that are attested in many real words. That is, children experience difficulty generalizing correct phoneme production to novel phonetic contexts. In addition, this study replicated the previous finding that measures of vocabulary size predict the raw difference in repetition accuracy between high- and low-frequency sequences (the *frequency effect*) beyond what is predicted by age.

This study expanded on previous findings in a number of ways. First, regression analyses showed that there was an effect of frequency on repetition accuracy beyond what would be predicted by accuracy of production of the same sounds in real words. Second, regression analyses found that the magnitude of the frequency effect in individual participants was independent of measures of their speech perception ability. That is, typically developing children's ability to produce sounds accurately in novel contexts in nonwords does not appear to be related to their overall acumen or experience in perceiving and producing sounds in real words. Taken together, these results support a model of phonological ability in which children's flexible productive control over phonemes emerges gradually as they amass lexical items, and develop stable mappings among acoustic, articulatory, and semantic characteristics of those items. This productive control allows children to produce sounds in novel combinations in new words accurately and fluently.

It is important to note that in the regression analyses, the only significant predictors of the frequency effect were one of the measures of vocabulary size. Typically, this measure was the raw score on the *Expressive Vocabulary Test*. However, inspection of the correlation matrices shows that PPVT-III raw score was also significantly correlated with the magnitude of the frequency effect in the entire group of TD and PD participants, and the EVT standard score predicted the frequency effect for the entire group and for the group with TD only. We take this to be strong evidence that the developmental decline in the frequency effect is related to vocabulary growth. We propose that categorical phonemic representations develop that link among acoustic, perceptual, and lexical representations in response to developmental increases in the size of children's vocabularies. As these representations develop, children gain increasingly better productive control over phonemes as separate from the words in which they occur. This allows them to combine them into novel sequences in a nonword repetition task. Thus, the first hypothesis of this experiment was supported, in that we were able to support Edwards et al.'s assertion that developmental decreases in the frequency effect are due to vocabulary growth, and not to developmental changes in speech perception or production.

A second original aspect of this study is that it compared the accuracy of repetition of children with PD to that of their TD peers. Unremarkably, children with PD repeated sequences less accurately than their TD peers. However, as a group, they showed no particular disadvantage when repeating low-frequency sequences. This finding is the opposite of what we would predict if the children with PD were experiencing difficulty building categorical phonemic representations. In the current study, the children with PD were less accurate overall, as is to be expected from their GFTA scores, but they were not disproportionately less accurate on the low-frequency target sequences. Rather, they were slightly less sensitive to frequency than the children with TD. This finding contrasts with the findings of Edwards et al. (2004). In that study, the youngest children, who had the lowest repetition accuracy overall, showed the largest effect of sequence frequency; whereas the oldest children, who had highest overall accuracy, showed only a small difference in between high- and low-frequency sequences. Thus, the second hypothesis of this study was not supported. This result suggests that the problems encountered by children with phonological impairment are not related to the types of problems that are encountered by young typically developing children.

The relationship between phonological disorder and the frequency effect was explored further with regression analyses. Regression analyses found that the children with PD with severe deficits had smaller frequency effects than children with milder deficits. Taken together with Edwards et al.'s (2004) earlier findings from children with TD, this suggests a U-shaped relationship between the children's overall phonological and lexical skills and the magnitude of the frequency effect. Children with severe PD showed only a small difference in production accuracy for high- versus low-frequency sequences; both were produced much less accurately than they were by the other participants. Children with mild PD, and children with TD who had small vocabularies in Edwards et al. showed a strong, significant influence of frequency on repetition accuracy; high-frequency sequences were produced more accurately than low-frequency ones. Children with TD who had large vocabularies in the current study and in Edwards et al. showed a small difference in repetition accuracy for the two sequence types, and both were produced relatively accurately.

As in Edwards et al. (2004), we interpret the difference in the frequency effect between children with large vocabularies and children with smaller vocabularies to reflect a causal mechanism in phonological development. Specifically, we propose that developmental increases in vocabulary size prompt the language learner to create autonomous categorical phonemic representations. The fact that the frequency effect was not larger in the group of children with PD suggests that the speech production deficits in this population are not associated with deficits in the formation of these representations. Within our model, this result suggests the deficits seen in children with PD are associated with difficulties forming robust representations of the auditory/acoustic and articulatory characteristics of speech. This interpretation is consistent with the poor performance by children with PD on the PI and on the measure of whole-word discrimination. We propose that children with PD can make mappings among these poorly specified primary representations, just as children with TD are able to learn mappings among well-specified articulatory and acoustic representations. These mappings allow children the productive control needed for production of sounds in novel words. At the same time, the smaller frequency effect noted in the children with severe PD may be evidence that the articulatory and acoustic/auditory representations in these children are so poor that they do not afford the child the opportunity to make any robust generalization. The consequence of this is essentially floor performance on both sequence types.

It is important to note that the relationship between frequency effects and vocabulary size has important implications for our understanding of the nonword repetition task. Some previous research has utilized nonword repetition as a pure measure of verbal working memory (e.g., Gathercole & Baddeley, 1990). The results in this paper, along with the results of previous studies (e.g., Edwards & Lahey, 1998; Fischer et al., 2001) suggest that accurate nonword repetition depends on a variety of cognitive skills in addition to working memory, including access to representations in long-term memory. The findings presented here, as well as those in Edwards et al. (2004), show that successful repetition of a low-probability nonword relies on a rich representational system associated with having a large vocabulary. For this reason, the results of this investigation are consistent with models of cognitive processing that posit continuity between long-term knowledge and real-time processing (e.g., Fisher et al., 2001; MacDonald & Christiensen, 2002).

In summary, the results in this paper suggest that phonological acquisition involves not only the development of well practiced articulatory and acoustic/auditory representations, but also the emergence of a symbolic representation that links both of these primary phonological representations to each other as well as to representations of other types of linguistic knowledge in the lexicon. Thus, phonological development sets the stage for the even higher-order generalizations about morpheme structure and phonotactics that characterize the mature phonological system (see Pierrehumbert, 2003, for a review). These generalizations in the adult differ from those in the child’s phonology in that they do not usually result in production errors unless the system is stressed. However, for both the adult and the young child, phonological generalization cannot be divorced from the words that the speaker recognizes and produces. This result fits well with a growing body of research that sees grammar as an emergent property of the history of interactions between the language user and language events in the world. In this view, we do not need to make a sharp distinction between phonological knowledge and phonological processing because the relationship between the two is symbiotic. In these models, knowledge feeds on processing and processing feeds on knowledge. The more a child has heard and said a word, the better the child knows that word: knowledge increases because of processing. Similarly, the more words that the child has heard and said that contain a particular phonological pattern, the easier it will be for the child to abstract away that pattern to use for learning new words.

Acknowledgements

This work was supported by NIDCD grant R01 DC02932 to Jan Edwards; by NIDCD grant R03 DC005702 to Benjamin Munson; and by NIH training grant T32 DC0051 to Robert A. Fox. We thank the children who participated in the study, the parents who gave their consent, and the schools at which the data were collected. For assistance in stimuli preparation, data collection, data analysis, and manuscript preparation, we thank Molly Babel, Erin Casey, Lynn Carahaly, Lisa Draper, Melissa Epstein, Heidi Hochstetler, Maryann Holtschulte, Bridgett Isermann, Satako Katagiri, Laurie Vasicek, Amy Vitale, Pauline Welby, and S. David White.

References

ANSI (1989). *Specifications for Audiometers*. Washington, DC: American National Standards Institute.

Bailley, G., Laboissiere, R., & Schwartz, J. L. (1991). Formant trajectories as audible gestures. *Journal of Phonetics*, *19*, 9-23.

Barlow, J. A. & Dinnsen, D. A. (1998). Asymmetrical cluster development in a disordered system. *Language Acquisition: A Journal of Developmental Linguistics*, *7*, 1-49.

Barlow, J. A. & Gierut, J. A. (1999). Optimality theory in phonological acquisition. *Journal of Speech Language & Hearing Research, 42,* 1482-1498.

Beckman, M. E., & Edwards, J. (2000). Lexical frequency effects on young children’s imitative productions. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V* (pp. 207-217). Cambridge, UK: Cambridge University Press.

Bernhardt, B. (1994). The prosodic tier and phonological disorders. *First and Second Language Phonology*. 149-172.

Bernhardt, B. & Stoel-Gammon, C. (1996). Underspecification and markedness in normal and disordered phonological development. *Children's Language, 9,* 33-54.

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). *Columbia Mental Maturity Scale*. New York: Harcourt Brace Jovanovich, Inc.

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

Dunn, L. & Dunn, L. (1997). *Peabody Picture Vocabulary Test – III*. Circle Pines, MN: American Guidance Services.

Edwards, J. (1992). Compensatory speech motor abilities in normal and phonologically disordered children. *Journal of Phonetics, 20*, 189-207.

Edwards, J., Beckman, M. E., and Munson, B. (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*.

Edwards, J., Fourakis, M., Beckman, M., and Fox, R. (1999). Characterizing knowledge deficits in phonological disorders. *Journal of Speech, Language, and Hearing Research*, *42,* 169-186.

Edwards, J., Fox, R. A., & Rogers, C. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and phonetic inventory size. *Journal of Speech, Language, and Hearing Research*, *45,* 231-242.

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

Fisher, C., Hunt, C., Chambers, K., & Church, B. (2001). Abstraction and specificity in preschoolers' representations of novel spoken words. *Journal of Memory and Language, 45*, 665-687.

Forrest, K., Chin, S. B., Pisoni, D. B., & Barlow, J. N. (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, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on processing of non-word sound patterns. *Journal of Memory and Language, 42*, 481-496.

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

Gathercole, S. E., Willis, C., Emslie, H, & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*, 349-367.

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, *42*, 382-397.

Goldman, R. & Fristoe, M. (1986). *The Goldman Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Services.

Ingram, D. (1976*). Phonological disability in children*. NY: Elsevier.

Isermann, B. C. (2001). *Variability and consistency of articulation in children with phonological disorders*. Unpublished master's thesis. Department of Speech and Hearing Science, Ohio State University.

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

Kaufman, N. (1995). *Kaufman Speech Praxis Test for Children*. Detroit, MI: Wayne State University Press.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, Implications* (pp.131-163).

MacDonald, M.C. & Christiansen, M.H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Science, 109*, 35-54.

McCune, L., & Vihman, M. M. (1987). Vocal motor schemes. *Papers and Reports on Child Language Development*, *26*, 72-79.

Metsala, J. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3-19.

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

Pierrehumbert, J.B. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (p. 177-228). Cambridge, MA: MIT Press.

Shriberg, L. D. & Kwiatkowski, J. (1994). Developmental phonological disorders: I. A clinical profile. *Journal of Speech and Hearing Research, 37*, 1100-1126.

Smit, A. B., Freilinger, J. J., Bernthal, J. E., Hand, L., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779-798.

Storkel, H. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321-1338.

Storkel, H. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language, 29*, 251-274.

Towne, R. (1994). The effect of mandibular stabilization on the diadochokinetic performance of children with phonological disorder. *Journal of Phonetics, 22*, 317-332.

Vitevitch, M., & Luce, P. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374-408.

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Services.

Table 1. Demographic data and standard test scores for children with phonological disorder (PD) and children with typical phonological development (TD). Standard deviations are in parentheses.

	Children with PD	Children with TD[d]
Age in months	57 (9)	58 (10)
Age range in months	40 - 76	39 - 75
Gender	13 female, 27 male	13 female, 27 male
GFTA[a] percentile ranking[b]	5 (3)	69 (19) *
CMMS[c] standard score	107 (11)	109 (11)
EVT[d] standard score	105 (13)	110 (11)
PPVT-III[e] standard score	106 (12)	113 (11)*
KSPT[f] standard score	98 (14)	106 (5)*

[a] *Goldman-Fristoe Test of Articulation* (Goldman & Fristoe, 1986).

[b] Any GFTA percentile ranking below the first percentile was entered as .5 in the computation of the mean and standard deviation.

[c] *Columbia Mental Maturity Scale* (Burgemeister et al., 1972).

[d] *Expressive Vocabulary Test* (Williams, 1997).

[e] *Peabody Picture Vocabulary Test-III* (Dunn & Dunn, 1997).

[f] *Kaufman Speech Praxis Test for Children* (Kaufman, 1995).

* indicates that scores were significantly different between the two groups using a one-tailed t-test, $p < .05$.

Table 2 . Nonword stimuli, arranged as matched pairs. Target sequences are underlined.

	Low freq.	High freq.
	/**jugoin**/	/**bogib**/
	/**moipəd**/	/**mæbɛp/**
	/**vugim**/	/**vɪdæg**/
	/**bodəjau/**	/**medəju/**
	/**vukɑtɛm/**	/**vɪtəgɑp/**
	/**gaunəpek/**	/**gitəmok/**
	/**nʊbəmən**/	/**nɪdəbɪp**/
	/**motauk/**	/**petik**/
	/**donug**/	/**bedæg**/
	/**tedaum**/	/**podaud**/
	/**auptəd**/	/**iptən**/
	/**dugnəted**/	/**tʌgnədit**/
	/**aukpəde**/	/**ikbəni**/
	/**auftəgɑ**/	/**auntəko**/
	/**nəfæmb**/	/**mɪnæmp**/
	/**pwɑgəb**/	/**twɛkɛt**/
	/**bufkit**/	/**kiften**/
	/**dogdet**/	/**tæktut**/
	/**kɛdəwəmb**/	/**fɪkətæmp**/
	/**pwɛnətɛp**/	/**twɛdəmɪn**/
	/**næfkətu**/	/**gʌftədaɪ**/
	/**dɛgdəne**/	/**tiktəpo**/
Consonants making up the target sequences	/**f**/ (3 times), /**g**/ (6 times), /**m**/ (4 times), /**v**/ (2 times), /**j**/ (1 time), /**n**/ (1 time), /**k**/ (4 times), /**p**/ (3 times), /**b**/ (2 times), /**w**/ (2 times), /**d**/ (2 times)	/**f**/ (2 times), /**g**/ (4 times), /**m**/ (3 times), /**v**/ (2 times), /**j**/ (1 time), /**n**/ (2 times), /**k**/ (4 times), /**p**/ (3 times), /**t**/ (6 times), /**w**/ (2 times), /**d**/ (1 time)
Vowels making up the target sequences	/**oɪ**/ (2 times) /**u**/ (4 times), /**aʊ**/ (7 times) /**ʊ**/ (1 time)	/**i**/ (5 times), /**æ**/ (2 times), /**ɪ**/ (3 times), /**u**/ (1 time), /ʌ/ (2 times), /**aʊ**/ (2 times)

Table 3. Summary measures on real-word and nonword repetition accuracy.

Measure Name	Description
Mean Nonword Repetition Accuracy	The average segmental accuracy scores for the high- and low-frequency sequences in the nonword repetition task. Used in ANCOVAs.
Nonword Accuracy Sum	The sum of the segmental accuracy scores on the nonword repetition task. Calculated separately for consonants and vowels in high-and low-frequency sequences. Used in regression analyses.
Real-Word Accuracy Sum	The sum of the accuracy scores on the phonetic inventory. Calculated separately for consonants and vowels making up the high- and low-frequency sequences, weighted for the frequency of occurrence of those sounds in the target sequences. Used in the regression analyses and in the ANCOVAs. When used in ANCOVAs, these values were expressed as an arcsine-transformed percentage correct.
Mean Accuracy in Real Words	The real-word accuracy sums, transformed to percentage-correct measures. Used in ANCOVAs.
Frequency Effect	The difference in average segmental accuracy scores between the high- and low-frequency sequences in the nonword repetition task. Used in the regression analyses.
Difference in Real-Word Accuracy Sums	The difference in the real-word accuracy sums between the real-word accuracy sums for sounds making up the high- and low-frequency sequences. Used in the regression analyses.

Table 4. Correlations among accuracy scores for low-frequency sequences, accuracy scores for high-frequency sequences, the frequency effect, and other measures of linguistic development, entire group

	Mean Accuracy Score, Low-frequency sequences	Mean Accuracy Score, High-frequency sequences	Frequency Effect[a]
Age	0.10	0.06	-0.10
Speech perception[b]	0.37**	0.42**	0.11
PPVT-III raw score	0.49**	0.48**	-0.21*
PPVT-III standard score	0.47**	0.38**	-0.18
EVT raw score	0.44**	0.29	-0.33**
EVT standard score	0.42**	0.28*	-0.29**
GFTA percentile rank	0.60**	0.66**	0.15
KSPT standard score	0.13	0.15	0.06

** $p < 0.01$, * $p < 0.05$

[a] Mean accuracy score for high-frequency sequences minus the mean accuracy score for low-frequency sequences

[b] d’ score on the speech perception task in Edwards et al. (2002). See text for details

Table 5. Correlations among accuracy scores for low-frequency sequences, high-frequency sequences, the frequency effect, and other measures of linguistic development, children with TD only.

	Mean Accuracy Score, Low-frequency sequences	Mean Accuracy Score, High-frequency Sequences	Frequency Effect[a]
Age	-0.03	-0.01	0.04
Speech perception[b]	0.10	0.13	-0.01
PPVT-III raw score	0.25	0.22	-0.15
PPVT-III standard score	0.37*	0.30*	-0.25
EVT raw score	0.30*	0.10	-0.36*
EVT standard score	0.39*	0.11	-0.48**
GFTA percentile rank	0.16	0.12	-0.11
KSPT standard score	0.17	0.23	-0.04

** $p < 0.01$, * $p < 0.05$

[a] Mean accuracy score for high-frequency sequences minus the mean accuracy score for low-frequency sequences

[b] d’ score on the speech perception task in Edwards et al. (2002). See text for details

Table 6. Correlations among accuracy scores for low-frequency sequences, high-frequency sequences, the frequency effect, and other measures of linguistic development, children with PD only.

	Mean Accuracy Score, Low-frequency sequences	Mean Accuracy Score, High-frequency Sequences	Frequency Effect[a]
Age	0.23	0.12	-0.22
Speech perception[b]	0.32*	0.38*	0.10
PPVT-III raw score	0.52**	0.34*	-0.35
PPVT-III standard score	0.39*	0.26	-0.25
EVT raw score	0.52**	0.32*	-0.37*
EVT standard score	0.38*	0.24	-0.26
GFTA percentile rank	0.49**	0.60**	0.20
KSPT standard score	-0.18	-0.19	-0.001

** $p < 0.01$, * $p < 0.05$

[a] Mean accuracy score for high-frequency sequences minus the mean accuracy score for low-frequency sequences

[b] d' score on the speech perception task in Edwards et al. (2002). See text for details

Table 7. Results of the regression analyses predicting repetition accuracy in nonwords from production accuracy in familiar words. See text for details.

		Accuracy in Real Words					
		Consonants			Vowels		
Dependent Measure	Group	β	R^2	Significance	β	R^2	Significance
Sum of accuracy in Nonwords	Both	.833	0.71	$p < 0.01$	.739	0.55	$p < 0.01$
	TD	.249	0.06	*n.s.*	.121	0.02	*n.s.*
	PD	.837	0.70	$p < 0.01$	.725	0.53	$p < 0.01$
Difference of accuracy in nonwords	Both	.098	0.01	*n.s.*	-.216	0.05	*n.s.*
	TD	-.215	0.05	*n.s.*	.047	0.00	*n.s.*
	PD	.181	0.03	*n.s.*	-.238	0.06	*n.s.*

Table 8. Results of the regression analyses predicting mean accuracy and the frequency effect from GFTA and vocabulary-size measures. The standardized β coefficients are for the regression with all variables entered. See text for details.

		GFTA			Vocabulary Size			
Dependent Measure	Group	β	R^2	Significance	Measure[a]	β	R^2	Significance
Mean Accuracy	Both	.547	0.401	$p < 0.01$	PPVT-III - RS[b]	.316	0.068	$p < 0.01$
	TD	.123	0.015	*n.s.*	none	N.A.	N.A.	N.A.
	PD	.552	0.362	$p < 0.01$	PPVT-III - RS	.348	0.094	$p < 0.05$
Frequency Effect	Both	.171	0.016	*n.s.*	EVT-RS	-.373	0.110	$p < 0.01$
	TD	-.141	0.014	*n.s.*	EVT-RS	-.563	0.221	$p < 0.01$
	PD	.342	0.096	$p < 0.05$	EVT-RS	-.355	0.098	$p < 0.05$

[a] This column shows the vocabulary-size measure (if any) that accounted for a significant proportion of the dependent measure beyond what was accounted for by GFTA.

[b] RS=raw Score

Figure 1. Mean accuracy scores for the low- and high-frequency sequences for children with PD and their TD age controls, averaged across the three sequence types. Error bars represent one standard error of measurement.

Figures 2a and 2b. Mean accuracy scores for the attested low-frequency sequences and their matched high-frequency counterparts (Figure 2b). Mean accuracy scores for the zero-frequency sequences and their matched high-frequency counterparts (Figure 2b). Error bars represent one standard error of measurement.

Figure 3. Mean accuracy scores for the high- and low-frequency sequences plotted against natural-log (LN) transformed EVT raw score for the entire group of 80 participants (top), the 40 participants with TD (middle), and the 40 participants with PD (bottom).

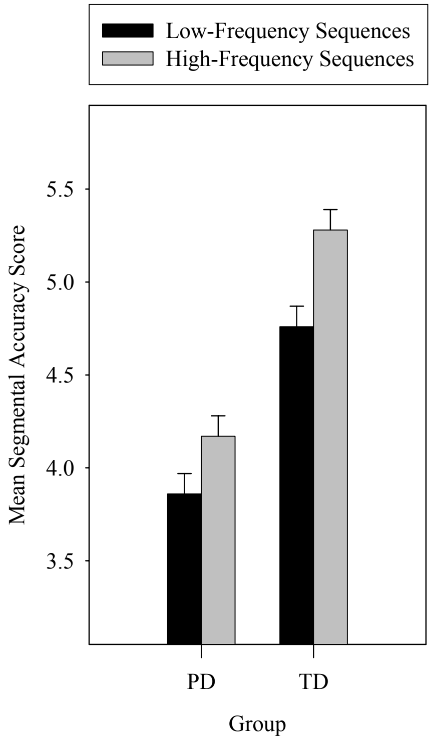
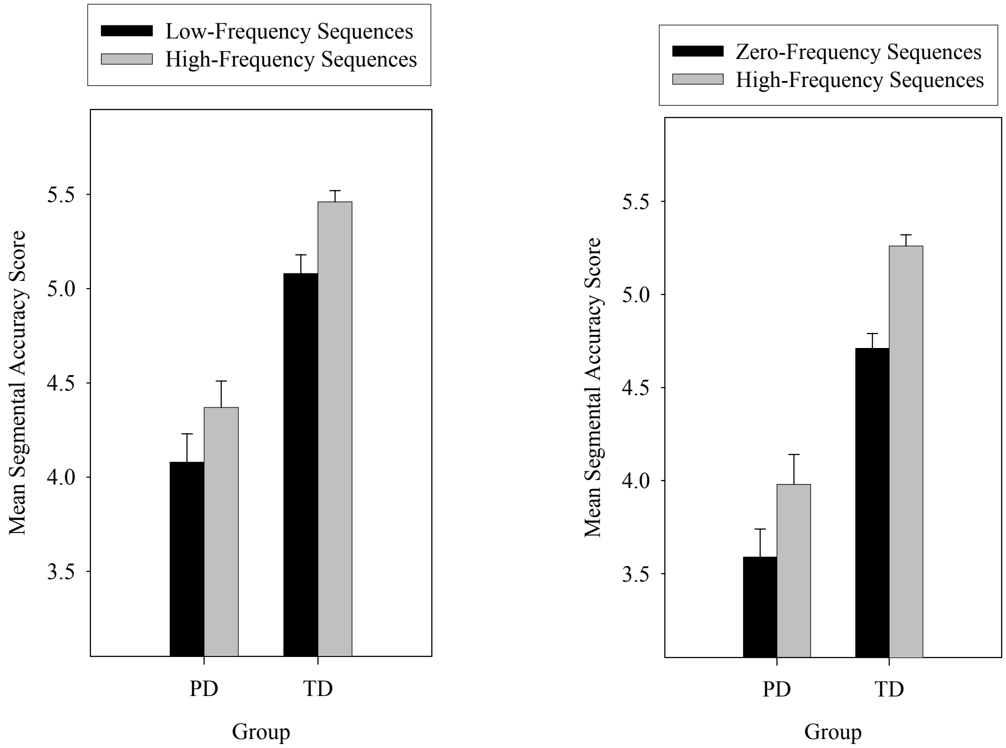

Low-Frequency Sequences
High-Frequency Sequences
Mean Segmental Accuracy Score
5.5
5.0
4.5
4.0
3.5
PD
TD
Group

Figure 1.



Figures 2a and 2b.

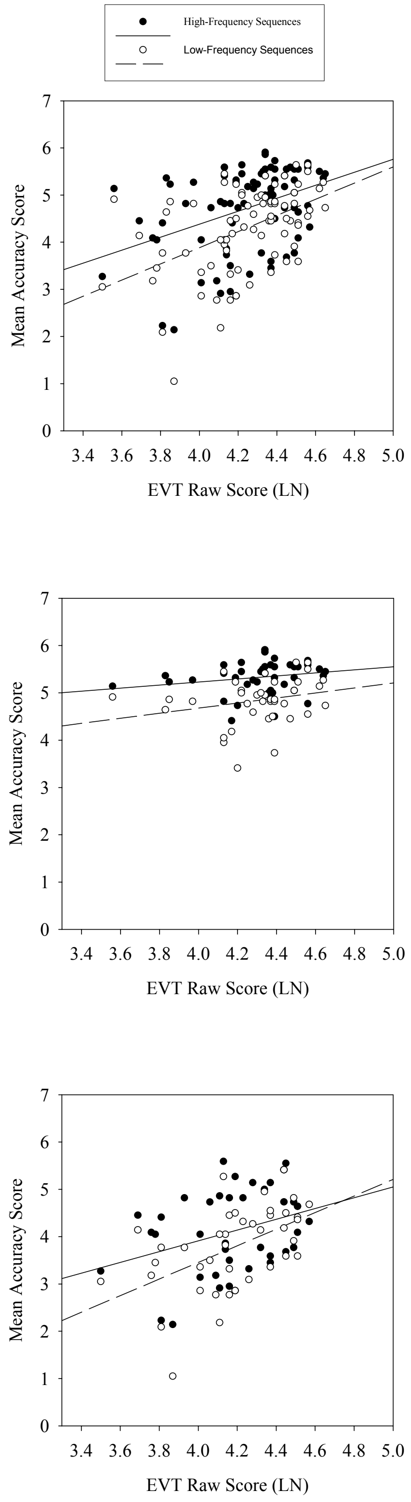

High-Frequency Sequences
Low-Frequency Sequences
Mean Accuracy Score
0
1
2
3
4
5
6
7
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.8
5.0
EVT Raw Score (LN)
Mean Accuracy Score
EVT Raw Score (LN)
Mean Accuracy Score
EVT Raw Score (LN)

Figure 3.