Published in final edited form as:
J Speech Lang Hear Res. 2012 April ; 55(2): 596–608. doi:10.1044/1092-4388(2011/10-0113).

Lexical and Phonological Effects in Early Word Production

Anna V. Sosa and
Department of Speech and Hearing Sciences, University of Washington

Carol Stoel-Gammon
Department of Speech and Hearing Sciences, University of Washington

Abstract

Purpose—This study examines the influence of word frequency, phonological neighborhood density (PND), age-of-acquisition (AoA), and phonotactic probability on production variability and accuracy of known words by toddlers with no history of speech, hearing, or language disorders.

Method—Fifteen toddlers between 2;0 and 2;5 produced monosyllabic target words varying in word frequency, PND, AoA, and phonotactic probability. Phonetic transcription was used to determine (1) whole-word variability and (2) proportion of whole-word proximity (PWP) (Ingram, 2002) of each target word produced.

Results—Results showed a significant effect of PND on both proximity and variability (words from dense neighborhoods were closer to the adult targets and less variable than those from sparse neighborhoods), a significant effect of word frequency on variability (high frequency words were less variable) but not proximity, and a significant effect of AoA on proximity (earlier acquired words were farther from the adult target than later acquired words) but not variability.

Conclusions—Results provide new information regarding the role lexical and phonological factors play in the speech of young children; specifically, several factors are identified that influence variability of production. Additionally, by examining lexical and phonological factors simultaneously, the current study is able to isolate differential effects of individual factors that have often been conflated in previous work. Implications for our understanding of emerging phonological representations are discussed.

Keywords

normal phonological development; word frequency; PND; age-of-acquisition; phonotactic probability; intra-word variability

There is growing evidence to suggest that children are sensitive to many of the same lexical and phonological factors that influence speech perception and production in adults. In recent years, effects of lexical characteristics such as word frequency, phonological neighborhood density (PND), and age-of-acquisition (AoA) have been investigated in relation to children's perception of real words (Garlock, Walley, & Metsala, 2001; Metsala, 1997), production of real and nonsense words (Garlock et al., 2001; Munson, Swenson, & Manthei, 2005; Newman & German, 2002), and acquisition of real words (Storkel, 2004a; Storkel, Maekawa, & Hoover, 2010). Phonological characteristics such as phonotactic probability have been investigated in relation to children's perception, production, and acquisition of

Correspondence should be addressed to Anna V. Sosa, Department of Communication Sciences and Disorders, Northern Arizona University, 208 E. Pine Knoll Dr., Bldg. 66, Rm. 302, Flagstaff, AZ 86011. anna.sosa@nau.edu.
Anna V. Sosa is now at Department of Communication Sciences and Disorders, Northern Arizona University.

both real and nonsense words (Beckman, Edwards, & Munson, 2000; Jusczyk, Luce, & Charles-Luce, 1994; Munson, 2001; Munson, Edwards, & Beckman, 2005; Munson, Swenson, et al., 2005; Storkel, 2004b; Storkel et al., 2010; Storkel & Rogers, 2000; Zamuner, 2009; Zamuner, Gerken, & Hammond, 2004).

Variability in Early Speech Production

The present study continues this line of work by investigating whether these lexical and phonological factors may affect the production of known, real words by typically developing 2-year-old children. Specifically, we are interested in determining whether these factors may help us understand, and perhaps explain, a ubiquitous phenomenon in early phonological development, variability. The speech of young children is often quite variable; as Ferguson and Farwell (1975) observed, the acquisition of a new sound may be gradual, with correct production in one word but not in another (*inter-word variability*), or a child might produce the same word multiple ways (*intra-word variability*). Production variability has been attributed to a number of different factors, including phonological complexity of the word (Leonard, Rowan, Morris, & Fey, 1982; McLeod & Hewett, 2008), immaturity of the speech motor control system (Kent, 1992; Smith & Goffman, 1998; Walsh & Smith, 2002), instability of phonological rules (Menn, 1971; Smith, 1973), and lack of segmental detail in the child's underlying phonological representation (Sosa & Stoel-Gammon, 2006).

Variability has often been ignored in studies of language development, either intentionally because it does not "fit" the theoretical account, or by omission because the study analyzes only a single production of a word, thereby ignoring any potential intra-word differences. Two recent studies of intra-word variability (Sosa & Stoel-Gammon, 2006; Stoel-Gammon, 2004) document considerable variability in typically developing children between 18 and 30 months of age (see also Ertmer & Goffman, 2011, McLeod & Hewett, 2008, and Holm, Crosbie, & Dodd, 2007). An important finding of these studies was that variability was word specific; an individual child produced some words with 100% consistency (the same way each time) and others with 100% variability (a different way each time). While phonetic factors (i.e., syllable shape, consonant complexity) likely influence variability, it is also possible that variability is affected by a child's experience with the individual word and the sounds that make up the word. For example, word frequency, PND, age-of-acquisition (AoA), and phonotactic probability may influence variability and accuracy of production. The current study is intended to explore this possibility.

Background to Lexical and Phonological Factors

Recent models of phonological representation posit at least two levels of representation; one reflecting the word as a whole unit (e.g., /bɑl/ for 'ball') and another that represents the individual sounds and sound sequences that comprise the word (e.g., /b/, /ɑ/, /l/, /bɑ/, /ɑl/) (Luce, Goldinger, Auer, & Vitevitch, 2000; Storkel & Morrisette, 2002). Lexical factors and phonological factors are characteristics of words that correspond to these proposed levels of representation. Word frequency, PND, and AoA are factors that are thought to derive from the lexical representation of the whole word, and the influence of these factors is anticipated during tasks that involve accessing the lexical representation. Phonotactic probability is defined by properties at the level of the phonological representation and the effect is expected to be observed during tasks that involve accessing individual sounds and sound patterns.

Lexical Factors

Word frequency

Word frequency refers to the number of times a particular word occurs in a written or spoken language corpus. In general, high frequency facilitates both word recognition and word production (summary in Ellis, 2002), presumably due to strengthening of the lexical representation and access pathways as a result of repeated access in both perception and production (Bybee, 2001). Evidence for the role of word frequency in children's linguistic processing is less robust; however, a few studies have found facilitative effects of word frequency in perception (Metsala, 1997), production (Anderson, 2007; Gierut & Storkel, 2002; Leonard & Ritterman, 1971; Morrisette, 1999; Ota, 2006; Tyler & Edwards, 1993), and word learning (Storkel, 2004a; but see Storkel, 2009). There is also some evidence that word frequency may be an important variable in the treatment of children with documented speech delays. Use of high frequency words as treatment targets for children with phonological delay was found to facilitate generalization of treated sounds to untreated words (Gierut, Morrisette, & Champion, 1999). Given these findings, it is predicted that if an effect of word frequency is found, the effect will be facilitative.

Phonological neighborhood density

The adult lexicon is thought to be organized around groups of words that share similar phonological properties, with similar words referred to as phonological neighbors. Most often, phonological neighbors are defined as words that differ from each other by one phoneme substitution, deletion, or addition in any word position (Luce & Pisoni, 1998); for example, the words *bat, mat, pit, pet, past, spat, at,* among others, would be neighbors of the word *pat*. PND is reported as the number of neighbors a particular word has; words with many neighbors will be referred to as high neighborhood density (ND) words, while words with few neighbors will be referred to as low neighborhood density (ND) words. In general, high neighborhood density is thought to inhibit word recognition in adults due to lexical competition effects among phonologically similar forms (Luce & Pisoni, 1998). However, some facilitative effects of PND have been found in production tasks (Vitevitch, 2002); adults produced fewer errors during a speech-error elicitation task for high ND words versus low ND words. Vitevitch (2002) reconciles his findings regarding a facilitative effect of PND with the usual finding that lexical neighbors compete with each other by invoking task specific differences. Lexical competition effects have been found in priming tasks, gating tasks, and tasks that measure response latencies, while facilitative effects have been found in more natural speaking contexts such as spontaneous errors (Vitevitch, 1997) and error-elicitation tasks (Vitevitch, 2002).

Studies investigating the role of PND in perception, production, and word learning in young children are limited. There is some evidence that high neighborhood density facilitates word learning in young children (Storkel, 2004a, 2009), and that children receiving treatment for phonological delay may produce treated sounds more accurately in high ND words rather than low ND words (Gierut & Storkel, 2002; Morrisette, 1999). Studies looking for the presence of lexical competition effects in children have found that while high neighborhood density may interfere with lexical access in children older than 7 (Garlock et al., 2001; Munson, Swenson, et al., 2005; Newman & German, 2002), no competition effects have been identified in younger children, suggesting that lexical competition may emerge with development (Garlock et al., 2001; Munson, Swenson, et al., 2005). Given that lexical competition due to high neighborhood density has not been found in children under 7, evidence that high neighborhood density facilitates production in tasks that are designed to elicit speech errors in adults, and limited evidence that dense neighborhood structure may facilitate the lexical diffusion of sound change in children receiving therapy for phonological delay, it is hypothesized that if an effect of PND is found in the present study, it will be a facilitative effect, with high ND words produced less variably and with higher proximity to the adult targets than low ND words.

Age-of-acquisition

An additional lexical factor that has been investigated in relation to the perception and production of words by adults and children is AoA, the age at which a word enters a child's vocabulary. AoA and word frequency together are thought to contribute to the overall familiarity of a word, which may affect an individual's ability to retrieve and produce that word (Garlock et al., 2001). In general, adult studies have shown that earlier acquired words are recognized and produced faster and more accurately than later-acquired words (Barry, Johnston, & Wood, 2006; Brown & Watson, 1987; Gerhand & Barry, 1998; Morrison & Ellis, 1995). Direct evidence for the role of AoA in the speech production of young children is limited (but see Anderson, 2008). AoA, word frequency, and PND are known to be highly correlated, with high frequency and high ND words acquired earlier than low frequency and low ND words (Storkel, 2004a). It is, therefore, difficult to tease out the independent effects of each of these factors. AoA is included as a variable of interest in the present study in order to determine whether the length of time a word has been in a child's vocabulary affects how that word is produced, independent of any potential effects of word frequency or PND.

Phonological Factors

Phonotactic probability

The relative frequency with which individual sounds and sound sequences occur in syllables and words is referred to as phonotactic probability (Jusczyk & Luce, 2002). Words or nonsense words that contain frequently occurring sounds and sound sequences are considered high probability words while words containing sounds and sound sequences that rarely occur in the lexicon are considered low probability words. A number of studies have shown that infants as young as 8 months are sensitive to the phonotactic probabilities of their native language and that they may use this information to help them segment possible words from a continuous speech stream (Aslin, Saffran, & Newport, 1998; Juscyzk, Luce, & Charles-Luce, 1994). In production, a consistent facilitative effect of phonotactic probability has been found in children as young as 1;8 (Beckman & Edwards, 2000; Edwards et al., 2004; Leonard & Ritterman, 1971; Munson, 2001; Munson, Edwards, et al., 2005; Zamuner, 2009; Zamuner et al., 2004). Most of these studies have investigated the role of phonotactic probability on children's accuracy and fluency of production of individual sounds and sound sequences during nonword repetition tasks, finding that children produce frequently occurring sounds and sound sequences more accurately and more fluently than infrequent sequences.

Nonsense words do not have a lexical representation. Thus, phonological facilitation observed during nonword repetition occurs at the level of the phonological representation. The current study explores the role of phonotactic probability during the production of real, known words, which involves accessing the lexical representation for the word as a whole. We are not aware of any studies that have looked at the influence of phonotactic probability on the production of real words by young children. Given the consistently robust findings that phonotactic probability facilitates production in children during nonword repetition tasks, however, it is hypothesized that if the present study identifies an effect of pattern frequency, the effect will be facilitative, with high probability words produced more accurately and less variably than low probability words.

The purpose of the present study is to investigate whether word frequency, PND, AoA, and phonotactic probability affect production variability and accuracy of known, real words in typically developing 2-year-olds. If effects are found, it is anticipated that high word frequency, dense neighborhoods, and high phonotactic probability will all facilitate production. This study extends previous work by looking at the role these factors play in the production of real, known words that are produced during a relatively naturalistic speaking context as opposed to during the production of nonsense words during word imitation tasks. Furthermore, with a few exceptions (Zamuner, 2009; Zamuner et al., 2004), studies investigating the role of these factors in speech production have focused on preschool age children and older; this study seeks to determine whether these same lexical and phonological characteristics influence the speech of even younger children.

Method

Participants

Participants in the study were 15 children (7 female, 8 male) between the ages of 2;0 and 2;5 (*M* = 2;3). As part of a larger investigation of typical phonological development, the children participated in six monthly data collection sessions, starting at either 1;7 or 2;0 with the final session at either 2;0 or 2;5; only data from the final data collection session for each child are included in the current analysis. All procedures were approved by the Institutional Review Board at the University of Washington. An initial telephone interview established that potential participants met the following criteria: (1) raised in a monolingual English speaking household, (2) no current or past parental concerns regarding physical, social, or cognitive development, and (3) no history of problems associated with speech, language, or hearing. Parents were then mailed the MacArthur-Bates Communicative Development Inventories: Words and Sentences form (CDI) (Fenson et al., 1993), provided instructions, and asked to send it back to the researchers. Children with CDI productive vocabulary scores at or above the 10^{th} percentile for their age and gender were invited to participate in the study (1 potential participant was excluded due to a CDI score below the 10^{th} percentile). Parents updated the CDI prior to each monthly session. Table 1 provides descriptive information for the participants, including CDI raw expressive vocabulary scores and percentile scores at the time of the data collection session; raw productive vocabulary scores ranged from 281 to 674 words (*M* = 504) and percentile scores ranged from the 30^{th} to the 99^{th} percentile (*M* = 67^{th}). Raw CDI expressive vocabulary scores did not differ between the younger children and the older children (2;0 *M* =453 , 2;5 *M* = 560), $t(13) = -1.706$, $p=.112$. The inclusionary criteria were established as an attempt to include only children who are likely to exhibit typical language development and to avoid potential effects of linguistic and dialectal differences. While it is possible that some of these children, all of whom displayed age-appropriate expressive vocabulary abilities at the time of the study, may later be identified with a speech and /or language disorder, there was no evidence or suspicion of this at the time of the study.

Target Words

Possible target words were selected on an individual basis for each child based on parental report on the CDI that the child produced the word; whenever possible, the same target words were selected for all children. The goal was to have thirty target words per child. All words were monosyllabic words; some were CVC words (e.g., *boat*), some contained either an initial or final consonant cluster (e.g., *clock, milk)*, and some words contained post-vocalic /r/'s that could either be considered as coda consonants or as rhoticized vowels in the nucleus (e.g., *chair, fork, car*). Since target words were real words selected on the basis of whether they were present in the child's productive vocabulary, it was not possible to carefully control for the phonetic characteristics of the target words, as is often done in this type of study. It is hoped, therefore, that the results will have greater external validity as they reflect processes of phonological acquisition that occur in a more natural context rather than what occurs in a carefully controlled experimental task. Nonetheless, it is very likely that overall variability and accuracy of consonant production is influenced by the consonants in the words; for example, words containing later acquired consonants may be produced more variably and less accurately than words with early acquired consonants. In order to account for the varying levels of phonetic complexity, an average consonant age-of-acquisition score was calculated for each word using data from Smit, Hand, Freilinger, Bernthal, & Bird (1990). Consonant age-of-acquisition was defined as the age at which 75% mastery was attained for each target consonant or cluster in the target position. For example, the average consonant age-of-acquisition score for the word *spoon* is 3.75 (age-of-acquisition for initial /sp/ is 4.5, age-of-acquisition for final /n/ is 3, which averages to 3.75). Post-vocalic /r/ was considered part of a rhotic diphthong (i.e., *car* [kɑɚ]) for purposes of determining age-of-acquisition. Consonant age-of-acquisition for each target word is provided in the on-line supplemental materials.

Procedures

Hour-long data collection sessions were conducted in a sound-treated therapy room at the University of Washington Speech and Hearing Clinic. Sessions were video and audio recorded using a high quality, wireless lapel microphone hidden in a vest worn by the child. During the first 10 minutes, the child and caregiver were left alone in the room, which was equipped with age-appropriate toys and books, to engage in free play. The investigator then joined the pair and attempted to elicit five productions of each target word during play, book reading, object naming, and picture naming. Emphasis was on spontaneous production of words and children were never asked to just repeat words produced by the experimenter (e.g., “Say ball”). In some cases, a child may have shown no interest in an object that was intended to elicit a specific target word, but spontaneously produced a possible target word that had not originally been selected for that child. When this happened, the investigator encouraged multiple productions of the new target word. In this manner, the selection of target words for each child was somewhat opportunistic, in that words that the child actually produced were used in place of the originally selected words. All target words used in the analysis are provided in the supplemental materials.

Predictor Variables

The goal of the present study was to investigate the role of lexical and phonological factors in young children’s productions of known words. These factors are most often determined using written corpora or corpora based on adult speech, although some researchers have begun incorporating measures of word frequency and PND that are derived from child speech corpora (Zamuner et al., 2004; Storkel, 2004a). As this study represents a first look at the role these factors play in young children’s production of real words, the traditional adult-based measures were used in order to make the current results more comparable to those of existing research. Future research should determine whether different results are obtained when child-based measures are used for studies involving very young children.

Word frequency was determined for each target word using Kucera and Francis (1967), and values reflect the number of times the word occurs in a one-million word corpus of written American English. Due to skewed distribution, log frequency values were used in the analysis and were obtained from a website maintained by Mitch Sommers at Washington University in St. Louis (http://128.252.27.56/Neighborhood/Home.asp). PND was determined using the 20,000 word *Webster’s Pocket Dictionary*, which is searchable through the same on-line website and is based on earlier work conducted by Pisoni and colleagues in the creation of the Hoosier Mental Lexicon (Pisoni, Nusbaum, Luce & Slowiacek, 1985).

AoA for each target word is reported in months based on normative data from the CDI (Dale & Fenson, 1996). Acquisition is defined as the age in months at which 50% of the approximately 1,800 children in the normative sample were reported to produce the word. Average positional segment frequency and average biphone frequency was calculated for each word using the online database created by Michael Vitevitch at the University of Kansas (http://www.people.ku.edu/~mvitevit/PhonoProbHome.html) (Vitevitch & Luce, 1994).

Possible inter-correlations among predictor variables—In the adult literature it is often noted that word frequency and PND are correlated; high frequency words tend to reside in dense neighborhoods while low frequency words tend to have fewer neighbors (Garlock et al., 2001). Similarly, early acquired words may be high frequency and reside in dense neighborhoods, while later acquired words may be low frequency and reside in low density neighborhoods (Storkel, 2004a). Phonotactic probability is also correlated with neighborhood density, with high probability words having more neighbors than low probability words (Munson, Swenson, et al., 2005; Storkel, 2004b). While this is generally the case based on analysis of a large, adult lexicon, it is not necessarily true of the target words selected for this study. Therefore, the relationships among the lexical and phonological predictor variables for the set of target words included in this study were investigated. The correlation matrix for all predictor variables, including consonant age-of-acquisition, is presented in Table 2. The two most highly correlated variables are the two measures of phonotactic probability ($r = .584$). The only other significant correlation is between consonant-age-of-acquisition and PND ($r = -.489$). This is likely due to the fact that consonant clusters are acquired later than singletons, and longer words (i.e., target words with consonant clusters) have fewer neighbors than shorter words. Thus, the consonant age-of-acquisition metric is capturing both age-of-acquisition of the individual consonants as well as overall length of the target word. Multicollinearity is considered a risk when correlations among independent variables exceed .80 (Mertler & Vannatta, 2005), thus the moderate correlations that are present are not considered problematic for the planned analysis.

Glossing and transcription—All tokens of possible target words were identified and coded according to three parameters: 1) sound quality, 2) position in utterance, and 3) imitation or spontaneous production. Sound quality was rated as good, fair, or poor based on the listener's subjective judgment of how easy it was to hear the child's production. Factors that contributed to fair or poor sound quality included quiet or whispered speech, overlapping speech of an adult in the room, background noise from toys, and rustling noise from contact with the microphone. Next, each token was coded as occurring in isolation, in one of a small set of two-word phrases, or in word initial, medial, or final position of a multi-word utterance. The two-word phrases that were coded as "fixed phrases" included determiner followed by the target word (e.g., *the ball*) and *bye/bye bye* or *hi* followed by the target word (e.g., *hi ball*). These were coded separately as it was felt that the presence of the preceding word added little linguistic or motoric complexity to the utterance and that they fell somewhere in between an isolated production of a word and the production of a word in a true, combinatorial utterance. Finally, each token was coded as a spontaneous production or an imitated production. Imitations were defined as productions that occurred within two seconds of an immediately preceding adult utterance that contained the same target word.

All tokens of target words were transcribed by the first author using broad phonetic transcription. In addition to standard American English phonemes, a number of other symbols that represent sounds that often occur in the speech of young children were used. These sounds include bilabial fricatives, labio-dental glides, alveolar affricates, palatal fricatives, and velar fricatives. Whispered productions and productions with very poor sound quality were not transcribed. A total of 1,822 word tokens were transcribed. Approximately 20% of the words (393 out of 1,822 words) were transcribed by a second transcriber, a doctoral student experienced in both clinical and research transcription of the speech of young children. Approximately 20 words from each child, representing all target words produced by that child, were selected for reliability transcription. Point-to-point transcription reliability for consonants was calculated. Mean point-to-point inter-transcriber reliability was 82%.

Selection of tokens for analysis and data trimming—The goal was to identify five useable productions of each target word to include in the analysis. If more than five productions were available, the following method was used to select the tokens to be included. The first five productions that were coded as spontaneous, occurring in isolation, and having good sound quality were used. If there were not enough of these tokens, then additional tokens were used based on the following criteria: 1) only tokens coded as good or fair sound quality were used, 2) phrase position (isolation>fixed-phrase>phrase final>phrase initial>phrase medial) was the highest ranked parameter, 3) spontaneity of utterance (spontaneous>imitation) was the next most important parameter, and 4) sound quality (good>fair) was the lowest ranked parameter. Target words with fewer than 5 useable tokens, while initially transcribed, were not included in the final analysis. Furthermore, target words that were only produced by a single child were not included and two words that were coded and transcribed as target words were removed from the analysis because log frequency (*sun*) or AoA (*snake)* information was not available. A total of 1,615 word productions were analyzed, yielding accuracy and variability data for 323 word tokens, representing 32 word types. An average of 22 different target words for each child (range = 14–26) was included in the analysis.

Outcome Measures

Proximity—Proximity between the child's production and the target form was quantified using the proportion of whole-word proximity (PWP), a whole-word measure introduced by Ingram (2002). PWP is a global measure of proximity that incorporates both consonant accuracy and word-shape approximation; thus, it captures both segmental and prosodic accuracy. PWP is reported as a ratio, with 1 representing an exact match to the adult target. Proximity scores for the five productions of each target word were averaged to determine the overall proximity of the word. The following formula is used to determine the PWP for each word:

$$\frac{\text{number of segments in child production + number of correct consonants}}{\text{number of segments in target word + number of consonants in target word}}$$

For example, if a child produces the word *cat* as [tæt], the PWP would be calculated in the following manner:

$$\text{PWP} = \frac{3\ (\text{\# of segments produced}) + 1\ (\text{\# of correct consonants})}{3\ (\text{number of segments in target}) + 2\ (\text{\# of consonants in target})} = \frac{4}{5} = 0.8$$

If a child produces extra segments (more than are in the adult target word), they are not counted in the first term in the numerator. Importantly, only consonantal accuracy is included in this proximity measure. The child receives a point for producing a syllable nucleus where one exists in the target word, but the quality of the vowel produced does not contribute to the proximity score.

Variability—Numerous methods for quantifying production variability have been proposed. Some measures look at variability of production of individual segments in a specific word position (Ertmer & Goffman, 2011), while others, such as the Inconsistency Assessment (Dodd, 1995) and the Proportion of Whole-word Variation (PWV; Ingram, 2002) quantify whole-word variability. Ingram's PWV measure can be applied to any language sample and does not require the elicitation of a predetermined set of words and was therefore considered appropriate for use with very young children producing language spontaneously. The variability measure employed in the current study is based on the PWV; instead of being reported as a ratio of the number of different productions of a word to the total number of productions of that word, however, production variability was quantified as the number of articulatory forms used for the five productions of each target word and could range from 1 to 5 for each word. For example, one participant used the following forms for the word *cheese*: [sis], [tʃis], [tis], [tʃis], and [tis]. This word was assigned a variability score of 3, since 3 different articulatory forms (i.e., [sis], [tʃis], [tis]) were used. The broad phonetic transcriptions were used to determine variability. If two productions varied in terms of the consonants used or the overall word shape, they were considered different. Consonant differences reflecting acceptable allophonic variation (e.g., aspiration or release/unrelease of final consonants) were not considered variable. Glottal stops alternating with nothing in initial or final position (e.g. [hæʔ] vs. [hæ] for *hat*) were also not considered different. And finally, variation in vowel quality did not contribute to the variability scores. Thus, as with the measure of proximity, this variability measure incorporates both segmental and prosodic variability, but variation in vowel quality does not factor into the variability score. Variability scores were calculated individually for each target word produced by each child. Transcribed examples from each child are provided in the supplemental materials.

Results

Variability and proximity scores for 323 word tokens (32 different target word types) were analyzed. Descriptive statistics providing mean proximity and variability, as well as values for lexical and phonological characteristics for all words analyzed are provided in the on-line supplemental materials.

Mean variability and proximity for each child is shown in Figure 1. Mean variability by child was 2.43 (range = 1.68 to 3.43), indicating that children used an average of 2.43 different articulatory forms in the 5 productions of each word; mean PWP was .76 (range = .65 to .88). Correlations among mean proximity, mean variability, and vocabulary scores for all children were calculated. There was a strong, significant, negative correlation between CDI raw score and variability ($r = -.81$), indicating that children with larger productive vocabularies exhibited less intra-word production variability than children with smaller vocabularies. There was no significant correlation between expressive vocabulary size and production proximity ($r = .22$). Production variability and proximity were moderately negatively correlated ($r = -.47$), but this was not a statistically significant relationship. Chronological age could not be included in the correlational analysis because there was not sufficient heterogeneity of values (children were either 2;0 or 2;5). In order to determine whether chronological age was related to either outcome measure, variability and proximity for the two age groups were compared; there were no significant differences in production variability (2;0 $M = 2.60$, 2;5 $M = 2.23$), t(13) = 1.53, $p = .60$, or proximity (2;0 $M = .74$, 2;5 $M = .78$), t(13) = −1.14, $p = .17$, by age. When boys and girls were examined separately, there were no significant differences in production variability (Boys $M = 2.51$, Girls $M = 2.34$), t(13) = −.657, $p = .523$, or proximity (Boys $M = .75$, Girls $M = .77$), t(13) = .636, $p = .536$, by gender.

Effect of Lexical and Phonological Factors on Production Proximity and Variability

In order to answer the main question addressed in this study, three lexical factors (word frequency, PND, and AoA), two measures of phonotactic probability (positional segment average and biphone average), and a phonetic complexity measure (average consonant age-of-acquisition) were submitted to a multivariate linear regression analysis to determine which variables, if any, were significant predictors of variability and proximity in young children's production of monosyllabic, known words. Since each child was measured on multiple words (therefore data points cannot be considered independent) and there is likely a strong "child-to-child" effect, the child was entered into the model as a random factor. The model assumes that the regression coefficients can vary from person to person, thereby accounting for inter-child variability. Separate analyses were conducted for each outcome variable: 1. proportion of whole-word proximity *(PWP)* and 2. variability. All predictor variables were entered simultaneously in each regression analysis, allowing the independent effect of each variable to be determined. Significant results indicate that the effect of the particular variable is present when all other predictor variables are controlled. For example, if neighborhood density emerges as a significant predictor of variability, this can be interpreted to mean that when all other variables are held constant, density has an observable effect on production variability.

Variability—The regression model for variability was statistically significant, t(6,322) = 3.18, $p < .01$, $r^2 = .31$. Results of the regression analysis are presented in Table 3. Consonant age-of-acquisition, word frequency, and PND all emerged as significant predictors of production variability. There was a positive relationship between consonant age-of-acquisition and variability; words with later-acquired consonants and consonant clusters were produced with more variability than words with earlier-acquired consonants. Word frequency and PND were both inversely related to variability; as word frequency and PND increased, production variability went down. The relationship between variability and neighborhood density and word frequency for each target word type is plotted in Figure 2. Neither phonotactic probability nor AoA of the word emerged as a significant predictor of variability.

Proximity—The regression model for production proximity (PWP) was statistically significant, t(6, 322) = 10.20, $p < .0001$, $r^2 = .49$. Results of the regression analysis are presented in Table 3. Consonant age-of-acquisition, PND, and AoA all emerged as significant predictors of production proximity. There was an inverse relationship between consonant age-of-acquisition and proximity; words with earlier-acquired consonants were produced more accurately than words with later-acquired consonants and consonant clusters. PND and AoA were both positively related to proximity; as either neighborhood density or AoA increased, proximity also increased. Neither phonotactic probability nor word frequency emerged as a significant predictor of proximity.

Discussion

This study investigated the role of word frequency, PND, AoA, and phonotactic probability on variability and proximity of 2-year-olds' productions of known words. The lexical factors of word frequency, PND, and AoA were all found to influence production variability or proximity in typically developing children; high ND words were both less variable and had higher proximity than low ND words, high frequency words were less variable (but did not have higher proximity) than low frequency words, and early AoA words had higher proximity (but not less variability) than late AoA words. Neither measure of phonotactic probability (positional average or biphone average) emerged as a significant predictor of variability or proximity.

Results presented here expand on our existing knowledge regarding the role these factors play in the speech production of children by extending the investigation to younger children (2-year-olds) and by looking at the effect of these lexical and phonological factors on the production of real, known words drawn from a relatively naturalistic speaking context, as opposed to nonword repetition tasks. Furthermore, by including all lexical and phonological factors in a single analysis, the current study was able to disentangle the relative contributions of individual variables (e.g., PND and phonotactic probability). The implications of these findings for our understanding of intra-word variability, the roles of PND and phonotactic probability in the speech production of young children, and the important relationship between lexical and phonological development will be discussed below.

Explaining Intra-Word Variability

The results of this study confirm that intra-word variability is indeed a common feature of typical phonological development; the children produced an average of 2.43 forms in 5 productions of each word. Furthermore, the present findings are consistent with previous work showing that this variability is often word specific (Sosa & Stoel-Gammon, 2006); that is, a child may produce one word consistently and another word variably. The results of this study suggest that this variability is not random; a number of factors were identified that influence whether a word is likely to be produced with variability. Not surprisingly, the strongest predictor of variability was a measure of phonetic complexity; young children exhibit more intra-word variability when attempting to produce words that contain later developing consonants and syllable structures (i.e., consonant clusters).

Word frequency and PND were also significant predictors of production variability. The finding that word frequency facilitates production stability in these 2-year-olds suggests that the ubiquitous frequency effect that strengthens lexical representations and access pathways to high frequency words emerges quite early in development. Whether the frequency effect is a result of perceptual learning, motor practice, or a combination of the two (Richtsmeier, Gerken, Goffman, & Hogan, 2009) is not definitively answered by the current study. For example, it is possible that motor practice is responsible for the observed frequency effect; that is, repeated production of a word may lead to the constituent articulatory gestures becoming part of a well-practiced motor routine that is less susceptible to production variability. It is also possible that frequent exposure to multiple exemplars of a specific word in the input contributes to lexical strengthening and enhances abstraction of phonological knowledge, which may also result in more stable productions. The fact that word frequency affected variability but did not systematically influence the proximity with which the children produced words argues in favor of a motor practice explanation; while high frequency words may be more stable due to a well-practiced motor routine, they would not necessarily be more accurate if the motor routine was learned incorrectly relative to the adult target. On the other hand, a perceptual learning explanation that implicates enhanced abstraction of phonological information would predict that high frequency words would be both more stable and more accurate. The relative contribution of perceptual learning versus motor practice may be investigated in future studies by including multiple measures of word frequency. It may be, for example, that the effect of word frequency would be even stronger if frequency is derived from a corpus that is more representative of the output frequency of an individual child, which would suggest that motor practice is the primary mechanism contributing to the frequency effect.

Phonological Neighborhood Density and Phonotactic Probability

To our knowledge, this is the first study that directly investigates the independent effects of PND and phonotactic probability on the speech production of young children. Consistent with previous work, no evidence was found for the existence of lexical competition in production due to PND in these very young children; in fact, PND facilitated both proximity and stability of production. These results are consistent with findings of studies of speech production in adults (Vitevitch, 2002) and support a model of speech production in which phonologically similar words are activated simultaneously during word production, resulting in a greater amount of activation for the target word, thereby facilitating production. An alternative, but not necessarily contradictory, explanation for the observed facilitative effect of PND is that words with many phonological neighbors may have more detailed, segmental phonological representations than words with fewer neighbors. It has often been hypothesized that young children rely more on whole-word representations during speech perception and production (Ferguson & Farwell, 1975; Lindblom, 1992; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Vihman & Croft, 2007), and that detailed knowledge of the individual segments that make up a word develops as the child adds words to his vocabulary. Specifically, it has been suggested that more detailed phonological representation may develop first in high ND words due to potential confusion in both perception and production with other similar sounding words (Charles-Luce & Luce, 1990; Garlock et al., 2001; Storkel, 2002). Furthermore, the presence of intra-word variability in young children may reflect a lack of segmental detail in underlying representations (Sosa & Stoel-Gammon, 2006). The finding that high ND words are both less variable and exhibit higher proximity to the adult target than low ND words supports a view of phonological development in which representations become more segmental as children add words to their lexicon and that this reorganization takes place on a word-by-word basis that is dependent, in part, on lexical factors such as phonological neighborhood density.

In contrast to previous work, which has found a consistent facilitative effect of phonotactic probability in the speech of young children (Zamuner et al., 2004; Zamuner, 2009), phonotactic probability did not influence either production variability or proximity in the current study. This work differs from earlier studies in that the task involved production of real, known words as opposed to repetition of nonwords. Furthermore, possible effects of both phonotactic probability and PND were analyzed simultaneously. The results suggest that for children at this early stage of language development, lexical facilitation due to PND supersedes phonological facilitation due to phonotactic probability when producing known words. Again, this is consistent with a model of phonological development in which phonological representations become more segmentally specified as the lexicon grows and that phonological facilitation due to phonotactic probability would emerge with development.

Relationship between Lexical and Phonological Development

Recently, researchers have become more and more interested in the interaction between different areas of language development. Specifically, the relationship between phonological and lexical development has been addressed extensively (see summary in Stoel-Gammon, 2011). The current findings confirm a strong association between vocabulary size and phonological knowledge. Intra-word variability was strongly correlated with the size of the child's expressive vocabulary (children with larger vocabularies exhibited less intra-word variability), but did not differ according to chronological age, suggesting that vocabulary size is a better predictor of phonological knowledge at multiple levels of representation than is chronological age. This is consistent with findings from studies of nonword repetition that have found that vocabulary size was a better predictor of nonword production accuracy than either chronological age or general articulation ability (Edwards et al., 2004; Munson, Edwards, et al., 2005; Zamuner, 2009). Currently, there are no studies that have investigated production variability and nonword repetition accuracy in the same group of children. It may be that in young children, both metrics assess the same construct, the degree of abstract phonemic knowledge.

While not the primary focus of the current study, one additional finding is worth noting. AoA influenced the proximity with which the children produced the words; later acquired words were produced with greater proximity to the adult target than earlier acquired words. Given that later acquired words tend to be composed of consonants and word shapes that are mastered later in the developmental process (Fletcher et al., 2004; Stoel-Gammon, 1998), one would predict that later acquired words would be less accurate due to the effect of phonetic complexity, which was not the case. A possible explanation for the observed effect of AoA on proximity has to do with the interaction between the state of an individual child's phonological system and the addition of new words. If a word enters the lexicon at a more advanced stage of phonological development, then it may be more likely to conform to any new developments in the child's phonology, rendering it more accurate than previously existing words that may not yet have been affected by recent changes in the phonological system. Detailed, longitudinal data from individual children would need to be analyzed to confirm this.

Conclusion

In summary, the results presented here shed new light on an often observed, but rarely investigated phenomenon in early phonological development, production variability. In addition to the important role of phonetic complexity (words with later acquired consonants and syllable structures are produced more variably), both word frequency and PND were found to influence the amount of production variability observed in the speech of 2- year-old children. Observed results also have implications for our understanding of the nature of children's underlying phonological representations and their development of abstract phonemic knowledge; when producing real, known words, young children appear to rely more on the whole word lexical representation, as evidenced by the observed facilitative effect of PND, and less on the categorical phonological representations that may be emerging, suggested by the lack of a phonotactic probability effect. Further research should continue to examine these effects in children of varying ages and stages of lexical and phonological development, as well as with children diagnosed with or at risk for speech sound disorders. Results will contribute greatly to our understanding of how these lexical and phonological factors influence speech production and how this information might be used in the diagnosis and treatment of children with atypical phonological development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is based on the unpublished doctoral dissertation of the first author.

This research was supported by National Institute on Deafness and Other Communication Disorders Grant 1 F31 DC008016-01 and by the American Speech-Language-Hearing Foundation Student Research Grant in Early Childhood Language. The following individuals contributed to data coding and transcription: Karen Alexander, Amber Franklin, Carlos Nye, Vanessa Wilken, and Ru-Shin Shieh.

References

Anderson J. Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. Journal of Speech, Language, and Hearing Research. 2007; 50:229–247.

Anderson J. Age of acquisition and repetition priming effects on picture naming of children who do and do not stutter. Journal of Fluency Disorders. 2008; 33:135–155. [PubMed: 18617053]

Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. Psychological Science. 1998; 9:321–324.

Barry C, Johnston RA, Wood RF. Effects of age of acquisition, age, and repetition priming on object naming. Visual Cognition. 2006; 13:911–927.

Beckman, ME.; Edwards, J. Lexical frequency effects on young children's imitative productions. In: Broe, MB.; Pierrehumbert, JB., editors. Papers in Laboratory Phonology: Vol. V. Cambridge: Cambridge University Press; 2000. p. 208-218.

Brown GDA, Watson FL. First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. Memory and Cognition. 1987; 15:208–216.

Bybee, J. Phonology and Language Use. Cambridge: Cambridge University Press; 2001.

Dale PS, Fenson L. Lexical development norms for young children. Behavioral Research Methods, Instruments, and Computers. 1996; 28:125–127.

Dodd, B. Differential Diagnosis and Treatment of Children with Speech Disorders. London: Whurr; 1995.

Edwards J, Beckman ME, Munson B. 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. 2004; 47:421–436.

Ellis N. Frequency effects in language processing: A review with implications of implicit and explicit theories of language acquisition. Studies in Second Language Acquisition. 2002; 24:143–188.

Ertmer DJ, Goffman L. Speech production accuracy and variability in young cochlear implant recipients: comparisons with typically developing age-peers. Journal of Speech, Language, and Hearing Research. 2011; 54:177–189.

Fenson, L.; Dale, P.; Reznick, S.; Thal, D.; Bates, E.; Hartung, J., et al. MacArthur Communicative Development Inventories. San Diego, CA: San Diego State University; 1993.

Ferguson CA, Farwell CB. Words and sounds in early language acquisition. Language. 1975; 51:419–439.

Fletcher P, Chan CW, Wong PT, Stokes S, Tardif T, Leung SC. The interface between phonetic and lexical abilities in early Cantonese language development. Clinical Linguistics and Phonetics. 2004; 18:535–545. [PubMed: 15573489]

Garlock VM, Walley AC, Metsala JL. Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. Journal of Memory and Language. 2001; 45:468–492.

Gerhand S, Barry C. Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1998; 24:267–283.

Gierut J, Morrisette M, Champion A. Lexical constraints in phonological acquisition. Journal of Child Language. 1999; 26:261–294. [PubMed: 11706466]

Gierut J, Storkel H. Markedness and the grammar in lexical diffusion of fricatives. Clinical Linguistics and Phonetics. 2002; 16:115–134. [PubMed: 11987493]

Holm A, Crosbie S, Dodd B. Differentiating normal variability from inconsistency in children's speech: normative data. International Journal of Language and Communication Disorders. 2007; 42:467–486. [PubMed: 17613100]

Ingram D. The measurement of whole-word productions. Journal of Child Language. 2002; 29:713–733. [PubMed: 12471970]

Jusczyk PW, Luce PA. Speech perception and spoken word recognition: Past and present. Ear and Hearing. 2002; 23:2–40. [PubMed: 11881915]

Jusczyk PW, Luce PA, Charles-Luce J. Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language. 1994; 33:630–645.

Kent, R. The biology of phonological development. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological Development: Models, Research, Implications. Timonium, MD: York Press; 1992. p. 65-90.

Kucera, F.; Francis, W. Computational Analysis of Present Day American English. Providence, RI: Brown University Press; 1967.

Leonard L, Ritterman S. Articulation of /s/ as a function of cluster and word frequency of occurrence. Journal of Speech and Hearing Research. 1971; 14:476–485. [PubMed: 5163881]

Lindblom, B. Phonological units as adaptive emergents of lexical development. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological Development: Models, Research, Implications. Timonium, MD: York Press; 1992.

Luce PA, Goldinger SD, Auer ET, Vitevitch MS. Phonetic priming, neighborhood activation, and PARSYN. Perception and Psychophysics. 2000; 62:615–625. [PubMed: 10909252]

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

McLeod S, Hewett SR. Variability in the production of words containing consonant clusters by typical 2- and 3-year-old children. Folia Phoniatrica et Logopaedica. 2008; 60:163–172. [PubMed: 18434739]

Menn L. Phonotactic rules in beginning speech. Lingua. 1971; 26:225–241.

Mertler, CA.; Vannatta, RA. Advanced and Multivariate Statistical Methods. 3rd ed. Glendale, CA: Pyrczak Publishing; 2005.

Metsala JL. An examination of word frequency and neighborhood density in the development of spoken-word recognition. Memory and Cognition. 1997; 25:47–56.

Morrisette M. Lexical characteristics of sound change. Clinical Linguistics and Phonetics. 1999; 13:219–238. [PubMed: 21275569]

Morrison CM, Ellis AW. Roles of word frequency and age of acquisition in word naming and lexical decision. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1995; 21:116–153.

Munson B. Phonological pattern frequency and speech production in adults and children. Journal of Speech, Language, and Hearing Research. 2001; 44:778–792.

Munson B, Edwards J, Beckman ME. Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. Journal of Speech, Language, and Hearing Research. 2005; 48:61–78.

Munson B, Swenson CL, Manthei SC. Lexical and phonological organization in children: Evidence from repetition tasks. Journal of Speech, Language, and Hearing Research. 2005; 48:108–124.

Newman R, German D. Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. Language and Speech. 2002; 45:285–317. [PubMed: 12693688]

Nittrouer S, Studdert-Kennedy M, McGowan RS. The emergence of phonetic segment: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. Journal of Speech and Hearing Research. 1989; 32:120–132. [PubMed: 2704187]

Ota M. Input frequency and word truncation in child Japanese: Structural and lexical effects. Language and Speech. 2006; 49:261–295. [PubMed: 17037124]

Pisoni D, Nusbaum H, Luce P, Slowiacek L. Speech perception, word recognition, and the structure of the lexicon. Speech Communication. 1985; 4:75–95.

Richtsmeier PT, Gerken L, Goffman L, Hogan T. Statistical frequency in perception affects children's lexical production. Cognition. 2009; 111:372–377. [PubMed: 19338981]

Smith, NV. The Acquisition of Phonology: A Case Study. Cambridge: University Press; 1973.

Smit AB, Hand L, Freilinger JJ, Bernthal JE, Bird A. The Iowa articulation norms project and its Nebraska replication. Journal of Speech and Hearing Disorders. 1990; 55:779–798. [PubMed: 2232757]

Smith A, Goffman L. Stability and patterning of movement sequences in children and adults. Journal of Speech, Language, and Hearing Research. 1998; 41:18–30.

Sommers, M. Washington University Speech and Hearing Lab Neighborhood Database [Searchable database]. Retrieved from http://128.252.27.56/Neighborhood/Home.asp

Sosa AV, Stoel-Gammon C. Patterns of intra-word phonological variability during the second year of life. Journal of Child Language. 2006; 33:31–50. [PubMed: 16566319]

Stoel-Gammon, C. Sounds and words in early language acquisition. In: Paul, R., editor. Exploring the Speech-Language Connection. Baltimore: Paul H. Brookes Publishing Co.; 1998.

Stoel-Gammon, C. Variability in children's speech; Child Phonology Conference 2004; May 14–15, 2004; Tempe, AZ: Arizona State University; 2004.

Stoel-Gammon C. Relationships between lexical and phonological development in young children. Journal of Child Language. 2011; 38:1–34. [PubMed: 20950495]

Storkel HL. Restructuring of similarity neighborhoods in the developing mental lexicon. Journal of child Language. 2002; 29:251–274. [PubMed: 12109371]

Storkel HL. Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. Applied Psycholinguistics. 2004a; 25:201–221.

Storkel HL. Methods for minimizing the confounding effect of word length in the analysis of phonotactic probability and neighborhood density. Journal of Speech, Language, and Hearing Research. 2004b; 47:1454–1468.

Storkel HL. Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. Journal of Child Language. 2009; 36:291–321. [PubMed: 18761757]

Storkel HL, Maekawa J, Hoover JR. Differentiating the effects of phonotactic probability and neighborhood density on vocabulary comprehension and production: a comparison of preschool children with versus without phonological delays. Journal of Speech, Language, and Hearing Research. 2010; 53:933–949.

Storkel HL, Morrisette ML. The lexicon and phonology: Interactions in language acquisition. Language, Speech, and Hearing Services in Schools. 2002; 33:24–37.

Storkel HL, Rogers MA. The effect of probabilistic phonotactics on lexical acquisition. Clinical Linguistics and Phonetics. 2000; 14:407–425.

Tyler A, Edwards J. Lexical acquisition and acquisition of initial voiceless stops. Journal of Child Language. 1993; 20:253–273. [PubMed: 8376469]

Vihman, MM. Phonological Development: The Origins of Language in the Child. Cambridge, MA: Blackwell; 1996.

Vihman M, Croft W. Phonological Development: toward a radical templatic phonology. Linguistics. 2007; 45:683–725.

Vitevitch MS. The neighborhood characteristics of malapropisms. Language and Speech. 1997; 40:211–228. [PubMed: 9509578]

Vitevitch MS. The influence of phonological similarity neighborhoods on speech production. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002; 28:735–747.

Vitevitch MS, Luce PA. A web-based interface to calculate phonotactic probability for words and nonwords in English. Behavior Research Methods, Instruments, and Computers. 2004; 36:481–487.

Walsh BM, Smith A. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. Journal of Speech, Language, and Hearing Research. 2002; 45:1119–1133.

Zamuner TS. Phonotactic probabilities at the onset of language development: Speech production and word position. Journal of Speech, Language, and Hearing Research. 2009; 52:49–60.

Zamuner TS, Gerken L, Hammond M. Phonotactic probabilities in young children's speech production. Journal of Child Language. 2004; 31:515–536. [PubMed: 15612388]

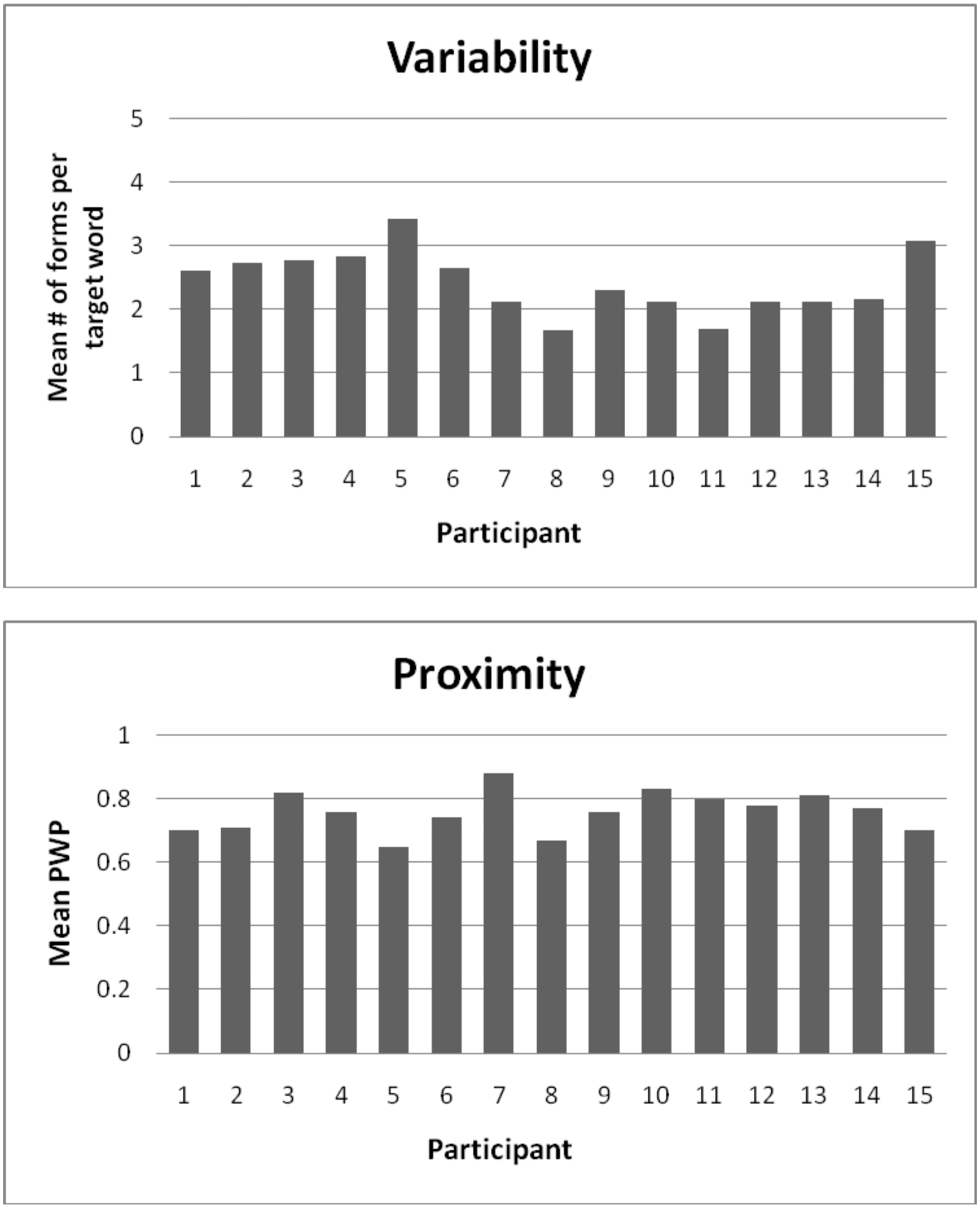

Figure 1.
Mean variability and proximity for each participant. $N = 15$.

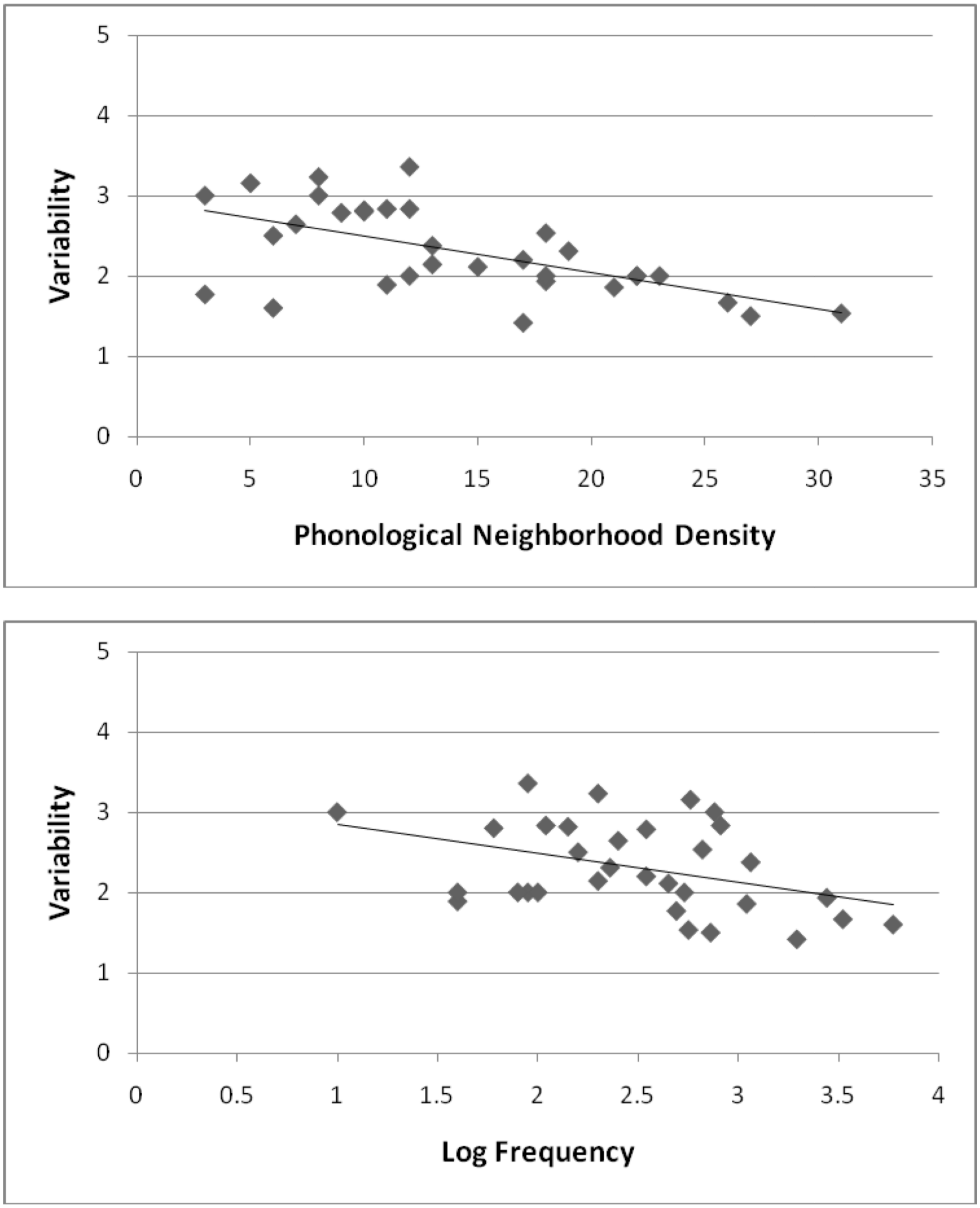

Figure 2.
Mean variability (# of forms per 5 productions of target word) for each target word type against phonological neighborhood density (top panel) and word frequency (bottom panel). $N = 32$.

Table 1

Demographic data and vocabulary scores for the 15 participants.

Participant #	Gender	Age in months	CDI Raw Score: Words Produced	CDI percentile ranking: Words Produced
1	F	24	391	55th
2	F	24	557	90th
3	F	24	281	35th
4	M	24	394	70th
5	M	24	285	55th
6	M	24	547	90th
7	M	24	518	90th
8	F	29	654	99th
9	F	29	651	90th
10	F	29	541	45th
11	F	29	674	95th
12	M	29	471	30th
13	M	29	614	80th
14	M	29	588	75th
15	M	29	387	30th

Note. CDI = MacArthur-Bates Communicative Development Inventories: Words and Sentences Form (Fenson et al., 1993)

Table 2

Pearson correlations for predictor variables of 32 test words.

Variable	Word Frequency	PND	A-o-A	Cons A-o-A	Pos Avg
PND	.24				
A-o-A	−.23	−.31			
Cons A-o-A	−.12	−.49**	.22		
Pos Avg	.12	.34	−.15	−.21	
Biph Avg	.27	.02	.01	.33	.58**

Note. PND = Phonological Neighborhood Density; A-o-A = Age-of-Acquisition of word based on CDI normative data; Cons A-o-A = average age-of-acquisition of target consonants and consonant clusters; Pos Avg = average positional segment frequency; Biph Avg = average biphone frequency.

** $p<.01$

Table 3

Pearson correlations among vocabulary scores, variability, and accuracy for 15 participants.

	CDI raw score	Variability
Variability	−.81**	
PWP	.22	−.47

Note. CDI = MacArthur Bates Communicative Development Inventories: Words and Sentences, Words produced (Fenson et al., 1996); variability = average production variability for each participant; PWP = average proportion of whole-word proximity for each participant.

** p<.01

Table 4

Results of the regression analyses predicting production variability and accuracy from lexical and phonological characteristics of words.

Variability	β Estimate	Standard Error	*t*	*p*
Word Frequency	−.245	.111	−2.20	**.029***
PND	−.024	.010	−2.39	**.018***
A-o-A	−.030	.022	−1.39	.165
Cons A-o-A	.423	.124	3.42	**<.001***
Pos Avg	−12.828	6.864	−1.87	.063
Biph Avg	10.403	28.063	.37	.711
***PWP* (accuracy)**	**β Estimate**	**Standard Error**	***t***	***p***
Word Frequency	.007	.014	.56	.577
PND	.003	.001	2.18	**.030***
A-o-A	.007	.003	2.5	**.013***
Cons A-o-A	−.149	.015	−9.68	**<.0001***
Pos Avg	−.996	.854	−1.17	.244
Biph Avg	−1.48	3.491	−.42	.673

Note. PND = phonological neighborhood density; A-o-A = age-of-acquisition of word based on CDI normative data; Cons A-o-A = average age-of-acquisition of target consonants and consonant clusters; Pos Avg = average positional segment frequency; Biph Avg = average biphone frequency.

Significant predictors are marked with * and bolded.