Published in final edited form as:

J Commun Disord. 2017 July ; 68: 70–80. doi:10.1016/j.jcomdis.2017.06.004.

Perceptual statistical learning over one week in child speech production

Peter T. Richtsmeier and

Department of Communication Sciences and Disorders, Oklahoma State University

Lisa Goffman

Callier Center for Communication Disorders, University of Texas at Dallas

Abstract

What cognitive mechanisms account for the trajectory of speech sound development, in particular, gradually increasing accuracy during childhood? An intriguing potential contributor is statistical learning, a type of learning that has been studied frequently in infant perception but less often in child speech production. To assess the relevance of statistical learning to developing speech accuracy, we carried out a statistical learning experiment with four- and five-year-olds in which statistical learning was examined over one week. Children were familiarized with and tested on word-medial consonant sequences in novel words. There was only modest evidence for statistical learning, primarily in the first few productions of the first session. This initial learning effect nevertheless aligns with previous statistical learning research. Furthermore, the overall learning effect was similar to an estimate of weekly accuracy growth based on normative studies. The results implicate other important factors in speech sound development, particularly learning via production.

Keywords

statistical learning; speech sounds; phonological development; child speech; speech production

1. Introduction

This paper addresses expressive speech sound development—or how children come to produce the sound patterns of their native language—within the framework of statistical learning. We review a large and growing literature on frequency effects in child speech. These studies, as well as closely related corpus work, suggest that statistical learning could be a prominent factor in expressive speech sound development. However, only a handful of studies have applied the statistical learning framework to child speech accuracy. Thus, there is an empirical gap separating our understanding of statistical learning on the one hand and

Correspondence should be addressed to Peter Richtsmeier, Communication Sciences and Disorders, 042 Murray Hall, Stillwater, OK 74078, prichtsmeier@yahoo.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

child speech development on the other hand. We begin with a basic and longstanding finding from the child speech literature: Children gradually become more accurate in their productions of speech sounds. The purpose of this study is to explore the relevance of statistical learning to that finding.

2. Statistical Learning in Language and Speech Development

It may be helpful to establish a working definition of statistical learning. Romberg and Saffran (2010) refer to statistical learning as the “discovery of patterns in the input” (p. 906). They apply the term widely, for example, to the operant conditioning of behaviorism and to learning algorithms implemented on a computer. This broad application of the term reflects statistical learning's apparent relevance to all linguistic patterns that are inherently statistical, or frequency-based.

Statistical learning research over the last 20 years has primarily focused on infant perceptual development. This research has adhered to a basic procedure: Infants are familiarized with an abbreviated or artificial language containing statistical regularities. Following the familiarization, infants are exposed to test items that conform to or do not conform to those regularities. Based on differences in looking times to the conforming and nonconforming items, researchers infer what infants have learned. For example, Saffran, Aslin, and Newport (1996) examined statistical learning of nonsense words like *dutaba* and *pidabu* embedded in a larger, continuous stream of syllables like *dutabatutibupidabupatubibupadababupudutaba*. Following exposure to the syllable stream, infants looked longer when hearing part-words like *tabatu* compared to words like *dutaba*, indicating that infants were sensitive to which syllables went together to form words.

Based on a large body of infant research, statistical learning appears to be relevant to many aspects of language acquisition (for in-depth reviews see Gómez & Gerken, 2000 and Romberg & Saffran, 2010). Furthermore, infant statistical learning persists over time (Gómez, 2011) and operates in cognitive domains besides language, for example, in visual pattern learning (Fiser & Aslin, 2002).

Returning to expressive speech sound development, statistical learning could be defined as the discovery of basic speech patterns, including the production of sounds and sound sequences. There is reason to believe that the type of statistical learning studied in infants is relevant to these patterns, that is, to expressive speech sound development. In fact, a considerable literature has documented how child speech accuracy is closely tied to the statistics of the child's native language. For example, Zamuner, Gerken, and Hammond (2004) reported that two-year-olds produce a coda consonant like /-n/ more accurately if that consonant is in a high probability environment, as in /nin/, compared to a relatively low probability environment, as in /von/. Munson (2001) controlled the frequency of word-medial consonant sequences in nonwords like /fæstət/ and /fæfpət/. He found that both three-year-olds and eight-year-olds were faster, more accurate, and less variable when producing high English frequency sequences like /st/ compared to low frequency sequences like /fp/. Storkel (2001) also found that children were more accurate when producing novel words composed of high probability sequences, like /hʌp/, compared to words composed of lower

probability sequences, like /gim/. Furthermore, children in Storkel's study were better able to match the high probability wordforms to semantic referents and to later recall that semantic knowledge. Thus, high frequency sound patterns appear to support the acquisition of both word form and meaning.

Frequency effects have also been observed in other languages, and Edwards and Beckman (2008) document these effects in two- and three-year-old children acquiring Cantonese, English, Greek, and Japanese. Recently, Másdóttir and Stokes (2016) observed robust correlations between frequency and production accuracy in two- and three-year-olds acquiring Icelandic. In sum, numerous studies have documented a dependency between expressive speech sound development on the one hand, and the frequency of sounds in the ambient language on the other hand. The dependency between frequency and production accuracy must be more narrowly defined, however, as the literature above could be true because children hear frequent sequences more often, because they produce them more often, or due to a combination of perceptual experience and production practice.

The relevance of production practice notwithstanding, the study of statistical learning has, to date, been primarily concerned with perceptual learning, and we will maintain our focus there. As such, when considering the literature on frequency effects in child speech development, the statistical learning framework suggests a basic hypothesis. Namely, if we control the frequency that children are exposed to a sound sequence *within an experiment*, we could see production accuracy change, depending on those experimental frequencies. That is, experimentally-implemented perceptual frequencies could simulate the relationship between speech accuracy and broader language statistics; the same relationship that has been documented by Edwards and Beckman (2008), Munson (2001), Zamuner et al. (2004), and others. So, within an experimental setting, child speech accuracy could be subject to statistical learning. Perhaps surprisingly, however, statistical learning has rarely been used to study expressive speech sound patterns.

Relative to the infant literature, the literature examining statistical learning in older children and adults is smaller. Statistical learning has been applied to pattern learning in typical adults (Eidsvåg, Austad, Plante, & Asbjørnsen, 2015; Finley, 2013; Onishi, Chambers, & Fisher, 2002; Richtsmeier, 2011; 2016). Statistical learning has also been studied in both typical and impaired populations, including for both receptive learning (Evans, Saffran, & Robe-Torres, 2009) and expressive learning (Plante, Bahl, Vance, & Gerken, 2011; Richtsmeier, Gerken, Goffman, & Ohala, 2009; Richtsmeier, Gerken, & Ohala, 2011). Focusing on expressive speech development, Richtsmeier et al. (2011) examined statistical learning in typically developing four-year-olds. Children were familiarized with novel words that were described as the names of make-believe animals. The words contained target word-medial consonant sequences, for example, the /fp/ in /neɪfpən/. Across three experiments, children's speech accuracy improved significantly following exposure to a combination of multiple talkers producing multiple related words (e.g., /mæfpəm/, /baɪfpəm/, and /gɪfpæk/ for /fp/), but not from exposure to multiple talkers producing a single word (e.g., just /mæfpəm/), or from exposure to single talkers producing multiple words. The authors concluded that statistical learning—incorporating multiple types of variability—can facilitate accurate speech production.

In a study of statistical learning that incorporated talker variability, Plante et al. (2011) applied a similar paradigm to children with language impairments. Four- and five-year-old children with and without language impairments were familiarized with novel words like /mæfpəm/ and /bɒktəm/. Children heard those words produced either once by a single talker or 10 times by multiple talkers. When they heard many examples of the words produced by many different talkers, both typically developing children and children with language impairments increased their accuracy and shortened their production latencies. Plante and colleagues therefore provide compelling evidence that talker variability specifically, and statistical learning more generally, may play a role in addressing language learning deficits (cf. also Evans et al., 2009 and Plante et al., 2014).

To summarize, statistical learning has been observed to facilitate language learning at multiple developmental stages, first allowing infants to learn ambient language structures and subsequently allowing children to increase production accuracy for within-experiment high frequency patterns. Statistical learning might facilitate speech production in a variety of ways, for example, by means of the multi-word input employed by Richtsmeier et al. (2011), or by means of the multi-talker input employed by Plante et al. (2011, 2014). But can it account for some of the gains children make across a more protracted developmental span, for example, Munson's (2001) data from 3- and 8-year-olds? Two recent analyses of child speech corpora suggest that it might.

Ota and Green (2013) looked at production accuracy for word-initial consonant clusters in speech corpora for three children, beginning just before their first birthday and ending when the clusters were produced with 80% accuracy. The authors were able to predict each child's level of accuracy using the rate that those children heard their mothers producing those clusters, that is, using a measure of perceptual input frequency. In another corpus-based analysis, Van Severen et al. (2013) also reported that input frequency had predictive power when looking at the development of word-initial consonants. In both studies, the effect of input frequency indicates that statistical learning is at work in expressive speech sound development.

The studies by Ota and Green (2013) and Van Severen et al. (2013) provide correlational evidence for the role of statistical learning in speech development. This correlation is also implied by the wider literature on frequency effects reviewed above (Edwards & Beckman, 2008; Munson, 2001; Zamuner et al., 2004, etc.). However, these lines of research do not directly assess a particular mechanism for learning from the statistics of the ambient language. In fact, both Ota and Green and Van Severen et al. emphasize the complexity of identifying the unique contribution of any one mechanism among highly correlated factors. These factors include functional load, articulatory complexity, perceptual salience, neuromotor skill, and representational complexity, among others. Given these concerns, an experimental approach may be optimal for delineating causal factors such as particular learning mechanisms. The studies by Richtsmeier et al. (2011) and Plante et al. (2011) more directly test whether statistical learning facilitates speech production, and those authors report positive learning effects. However, learning in those studies was examined within a single experimental session. As such, it is unclear whether the benefits persist over time. If

we are to invoke statistical learning to explain the broader time course of expressive speech sound development, learning must be examined in more than a single session.

The present study addresses two central questions related to statistical learning over time. First, can we replicate previous findings showing that statistical learning leads to greater production accuracy? Second, does the effect of statistical learning persist over time? The results will thereby advance our understanding of the relative importance of perceptual experience in the larger context of expressive speech sound development.

3. An Experimental Study of Perceptual Statistical Learning over Time

Method

3.1 Participants

Twenty-two children between the ages of 4;4 (years; months) and 5;7 were recruited for the study ($M = 4;9.1$). All participants were native speakers of English who were raised in the American Midwest and who met criteria for typical development. These criteria included no diagnosis of speech disorders, language disorders, hearing loss, or intellectual disability. The children were also old enough to attend to a procedure that required five total sessions and 15–60 minutes per session. Children were recruited as part of a larger study incorporating acoustic, perceptual, and articulatory measures of statistical learning. Parents were contacted through local childcare centers and via a regional newsletter. Four children began the study but did not complete the experiment; they were removed from subsequent analyses. Data from two other children were not included in the analyses; one because they did not pass the hearing screening, the other due to a history of speech impairment. Results from the remaining 16 participants are analyzed below.

As shown in Table I, participants completed a battery of five standardized tests in the areas of expressive syntax and morphology (Structured Photographic Expressive Language Test, 3rd Edition; Dawson, Stout, & Eyer, 2003), expressive vocabulary (Expressive Vocabulary Test, 1st Edition; Williams, 1997), receptive vocabulary (Peabody Picture Vocabulary Test, 4th Edition; Dunn & Dunn, 2007), articulation (Bankson-Bernthal Test of Phonology; Bankson & Bernthal, 1990), and nonverbal cognitive skill (Columbia Mental Maturity Scale; Burgemeister, Blum, & Lorge, 1972). Participants also completed a nonword repetition task with nonwords from Dollaghan and Campbell (1998). All participants scored within normal limits on the five standardized tests, although the means in Table 1 suggest that participants were generally above average in terms of expressive language and vocabulary, receptive vocabulary, and nonverbal skill.

3.2 Materials

The learning targets in the experiment were word-medial consonant sequences inside phonologically unique novel words (Munson, 2001; Plante et al., 2011; Richtsmeier, 2011, 2016; Richtsmeier et al., 2011). The words were considered phonologically unique because they differed from each other by at least three phonemes. Four consonant sequences were selected as learning targets—/pt/, /zm/, /mk/, and /fp/. The familiarization and test words in which those sequences were embedded appear in Table II.

Expressive speech sound development is often concerned with individual speech sounds, but for several reasons, medial consonant sequences were chosen here. First, consonant sequences are relatively difficult, and children typically make errors when producing them (McLeod, 2013; Smit, 1993). Errors are necessary for accuracy to improve and for learning to be observed; they are therefore a necessary component of the present methodology. Second, medial consonant sequences are sensitive both to native-language phonotactic frequencies (e.g., Edwards & Beckman, 2008; Munson, 2001) and to experimentally-implemented phonotactic frequencies (Plante et al., 2011; Richtsmeier et al., 2009, 2011). Furthermore, the development of consonant sequences resembles the development of individual sounds in that they both take years to reach adult levels of mastery (Ota & Green, 2013; Smit, 1993). Most importantly for the study of statistical learning, previous work indicates that word-medial consonant sequences are ideal for the study of implicit, subconscious, statistical learning—which is the learning mechanism under consideration (Gómez & Gerken, 2000). Consonant sequences are therefore an ideal test case for investigating the influence of statistical learning on expressive speech sound development.

Experimental words were designed to be different from real English words. They had no phonological neighbors based on a search of the Washington University Speech and Hearing Laboratory's Neighborhood Database (<http://128.252.27.56/Neighborhood/NeighborHome.asp>).¹ They all had a CVCCVC shape and bore stress on the first syllable. Using phone and biphone probabilities obtained from an online phonotactic probability calculator for children (Storkel & Hoover, 2010; http://129.237.148.107/cml/post_ccc.vi), words were matched for position-specific phones and biphones, as well as whole-word phone and biphone sums. As can be seen in Table II, word-medial consonant sequences were selected to cover a range of biphone probabilities. This range allowed for some variation in accuracy based on native language phonotactics (Edwards & Beckman, 2008; Munson, 2001), with the expectation that children would produce relatively frequent sequences like /pt/ more accurately than infrequent sequences like /mk/ and /fp/. English frequency was not considered a variable in the statistical analyses, however.

Within the experiment, the frequency of word-medial sequences varied, referred to as the experimental frequency factor. For each participant, two sequences were low experimental frequency, meaning participants heard those sequences 5 times in a single word produced by a single talker (e.g., they heard /pt/ in five identical tokens of /sɛptəf/). The two other sequences were high experimental frequency, meaning participants heard those sequences 15 times, or in 3 phonologically different words that were each produced by 5 different talkers (e.g., they heard /mk/ in /gumkəs/, /tæmkən/ and /dɪmkəs/, and they heard five different talkers produce those words). Experimental frequency was therefore a within-subjects variable. Regardless of the experimental frequency of the familiarization words, children only heard the test words during the test block, and they produced all test words the same number of times. This design has at least two benefits. First, it controls production practice,

¹After data collection had begun, we realized that the test word /fomkəp/ is phonetically similar to the two-word phrase “foam cup”. No participant reported noticing the similarity, however, and accuracy for /fomkəp/ did not appear to be inflated relative to the other test words. It is also worth noting that neighborhood density is generally described as a measure of similarity among words, rather than the similarity of a word to a phrase, as is the case with /fomkəp/ to “foam cup”.

allowing for the results to be attributable to the perceptual manipulation of experimental frequency. Second, the use of novel test words requires children to generalize knowledge of the target consonant sequences to production in a word that was not familiarized. Generalization—in the form of higher accuracy—was expected for high experimental frequency sequences compared to low experimental frequency sequences.

The inclusion of different talkers was based on previous findings that talker variability can facilitate statistical learning (Houston & Jusczyk, 2003; Plante et al, 2011; Richtsmeier et al., 2009; 2011). Recordings of familiarization and test words were obtained from seven adult female speakers of a Midwestern dialect of American English. The recordings were made in a sound booth. To ensure some consistency, especially with respect to the stress pattern, the speakers produced three tokens of each word immediately after the first author provided a model. A single token of each word from each speaker was then chosen for the experiment and was scaled for intensity using Praat software (Boersma & Weenink, 2016; www.praat.org). Productions from five of the talkers were used for the familiarization words; productions from the other two talkers were used for the test words.

Regarding counterbalancing, eight of the participants heard /pt/ and /zm/ as high experimental frequency and /mk/ and /fp/ as low experimental frequency. The other eight participants heard the reverse frequencies: /mk/ and /fp/ as high experimental frequency and /pt/ and /zm/ as low experimental frequency. These two assignments were then implemented within two lists each, or four total lists. For each list, the association of words to make-believe animals varied. Additionally, the talker for the test items varied; that is, each of the two lists with the same experimental frequency assignment had a different test talker.

Participants' productions of test words were collected across two sessions, a factor referred to hereafter as session. In the first session, productions were collected immediately following the familiarization. Participants returned one week later to produce the test items in a second session. There was no familiarization during the second session.

3.3 Procedure

Children were recruited to participate in a study lasting five weeks, one visit per week, and they completed two experiments over the course of the study. The first session included only standardized testing; this experiment was completed during the second and third sessions.

The experimental design is akin to infant statistical learning studies, but with speech production accuracy as the dependent variable. Participants completed the experiment in a quiet room. They were seated in a Rifton chair with an attachable tabletop, approximately 10 feet from a monitor and speakers. Before the start of the experiment, the experimenter explained that the child would hear the names of “funny, make-believe animals”, and that the child's task was to watch the animals and listen to their names.

At the end of the familiarization, the experimenter explained that the child would see new make-believe animals and hear their names. The child's task was to repeat the names of those animals. Children produced each of the four test words 9 times in a predetermined, pseudo-random order. In the second session, children were told they would “say the animals’

names again”, and they produced the four test words another 9 times in a different pseudo-random order. Throughout the experiment, parents could stay with their children to provide support or encouragement, but they did not produce the test words.

Both familiarization and test words were paired with pictures of make-believe animals (Ohala, 1999) to enhance learning (e.g., Heisler, Goffman, & Younger, 2010). However, cognitive demands related to word learning were minimized, and the data likely reflect phonological—as opposed to semantic—learning.

Children had 18 total opportunities to produce the test words. As described in the introduction, it is an open question whether native language frequency effects result from perceptual experience, production practice, or some combination of the two. The number of productions provided an opportunity to evaluate the influence of statistical learning over many productions, as well as over a week-long delay. The number of productions also offered a window into the influence of production practice, although production practice was not manipulated but was consistent for all test words and for all participants.

Productions were collected as imitations (i.e., participants produced a test word after hearing the test talker's production). Children were allowed as much time as they needed to make the productions. Some children were reticent at the beginning of the first test session and did not make the first few productions, but by the 10th overall production, all children produced the words immediately after hearing the test talker's prompt.

Children's productions were recorded and digitized directly to CD at 44.1 KHz. Video and audio were also recorded to DVD. Concurrent to collection of the phonetic transcriptions described here, kinematic data on lip and jaw movements were collected using an Optotrak camera system (Northern Digital Inc., Waterloo, Ontario, Canada). Those data will be reported elsewhere.

Both sessions of the experiment were controlled by Paradigm software (<http://www.paradigmexperiments.com>). Paradigm randomized the order of familiarization words and included 500 ms of silence between each word. The order of presentation of the test words in both sessions was randomized in advance of presentation and was thus consistent from participant to participant.

3.4 Analysis

Each child produced a maximum of nine tokens of each of the four test words in each test session. One child produced only five tokens of /fomkəp/ in the first session; otherwise at least eight tokens were available for each word for each session. Transcriptions were made by the first author using the high-quality recording on the CD. A second pass was made using the DVD recording to ensure that place of articulation was accurately transcribed. Statistical analyses were limited to the word-medial consonant sequences, although all consonants were transcribed and scored. Scoring of production accuracy followed a procedure described by Edwards, Beckman, and Munson (2004). Correct productions of each consonant received three points, for a total possible score of six points per consonant sequence. If a child's production of a consonant differed by one articulatory feature—place

of articulation, manner of articulation, or voicing—then the consonant received two points. For example, if a child produced /bɪptəm/ as [bɪktəm], then the intended /p/ would receive two points. Any other consonant or consonant sequence received one point. For example, the intended /p/ in either [bɪmtəm] or [bɪsp̥təm] would receive one point. Missing consonants received no points (e.g., no points for the /p/ absent in [bɪtəm]). A second transcriber independently transcribed 25% of the data using the same procedure described above. Overall agreement between the two transcribers was 88.0% and considered acceptable.

4. Results

Table III presents the production accuracy means and standard deviations for the word-medial consonant sequences, broken down by experimental manipulations: the high and low experimental frequency conditions across the two sessions. Although accuracy was numerically higher for words in the high experimental frequency condition, the difference was small. To determine whether the effect was significant, a 2×2 repeated measures ANOVA was conducted (Experimental Frequency [high, low] $\times$ Session [1, 2]). When these overall means were analyzed, no effects of experimental frequency or session were observed, nor was there a significant interaction (all $F_s(1, 15) < 2.5$, all $p_s > .15$). To confirm this finding, the data were recoded using two alternatives: correct/incorrect consonant scores and arcsine transformations. The results were similarly nonsignificant.

The results of the main analysis are puzzling. Given previous findings of statistical learning effects in child speech production (e.g., Plante et al., 2011; Richtsmeier et al., 2009, 2011), why would no effect be observable here? Two significant changes in the present study are worth considering: the addition of a second session one week after the learning phase, and the addition of more productions of the test items (9 productions per session here versus 4 total productions in Plante et al., 2011 and in Richtsmeier et al., 2009, 2011). It is possible that the present results are consistent with previous results, but the learning effect may have been limited to the first session or to just a few productions within a session.

To better understand learning at the level of individual productions, an experimental frequency effect was obtained for each participant for each production number (here *production number* refers to the accuracy for the productions of all four test words in order, with nine productions per session). For each production number, we calculated a difference score comparing the number of phonemes correct in the high experimental frequency condition to the low experimental frequency condition. For example, a child whose first production of the high experimental frequency words /bɪptəm/ and /pɛzməf/ were [bɪftəm] and [pɛzməf], respectively, would receive production accuracy scores of 5 and 6. The resulting high experimental frequency score would be 11 for production number 1. If the low experimental frequency words /fɒmkəp/ and /mæfpəm/ were produced as [fɒ ŋkəp] and [mæspəm], respectively, the child would receive production accuracy scores of 5 and 5, resulting in a low experimental frequency score of 10 for production number 1. The difference score for production number 1 would be $11 - 10 = 1$. The same calculation was made for production numbers 2-8 of session one and production numbers 1-9 of session 2. A graph of the average difference scores appears in Figure 1. The advantage of this analysis is

that it allows for statistical learning to be observed from production to production as well as across days.

Our first analysis considered whether the average experimental frequency effect was zero across all production numbers. A Wilcoxon signed-rank test was conducted for production numbers 1-9 for sessions 1 and 2. The analysis was significant ($p = .045$), revealing a small positive effect of experimental frequency (6/9 production numbers for the first session, 7/9 production numbers for the second session, cf. the number of shaded bars in Figure 1).

Based on a visual scan of the experimental frequency effect across production numbers, however, a robust experimental frequency effect probably did not persist through all 9 productions of session 1. To test this possibility, two additional ANOVAs were conducted on a limited number of productions from session 1. The first looked at production number 1, and the second looked at productions 1-3.

Considering production number 1 of session 1, children were significantly more accurate to produce high experimental frequency sequences $F(1, 15) = 6.22, p = .025$. Production numbers 1-3 of session 1 were analyzed next, similar to the four productions analyzed in previous studies (Plante et al., 2011; Richtsmeier et al., 2009, 2011). We also chose production numbers 1-3 because visual inspection of Figure 1 indicated that the first three productions of session 1 trended in the positive direction, whereas production number 4 did not. For productions 1-3, children were only marginally more accurate to produce high experimental frequency sequences, $F(1, 15) = 2.96, p = .106$. In sum, statistical learning did facilitate production accuracy for the first one-to-three productions of the first session. However, when we consider that the effect for the first three productions was marginal, and that the overall effect of experimental frequency was nonsignificant, a trajectory emerges in which an early effect of experimental frequency quickly tapers off.

5. General Discussion

In the introduction, we reviewed the state of statistical learning research and asked how it might inform the study of expressive speech sound development. The available literature suggests that statistical learning is relevant, but previous research has rarely considered statistical learning in speech production or over time. To address these empirical gaps, we completed a statistical learning experiment over one week. In this experiment we attempted to 1) replicate previous studies of statistical learning in child speech production and to 2) examine learning over time. The results revealed an initial advantage for frequently heard sequences; however, the effect tapered off. It was significant for the first production on the first day, marginal for the first three productions, but not significant for all productions across both days.

Regarding replication of previous studies, the statistical learning effect for the first few productions is similar to the results presented by Richtsmeier et al. (2011), where the combination of multiple words and multiple talkers improved production of low English frequency consonant sequences. The effect is also in line with similar statistical learning effects in child speech production reported by Richtsmeier et al. (2009) and Plante and

colleagues (2011, 2014). When statistical learning is examined in just a few productions—as has been the case in all previous statistical learning studies of child speech production—the present results are commensurate.

Regarding learning over time, a surprising result emerged. When we considered productions across the first and second sessions, the statistical learning effect tapered off. A Wilcoxon signed-rank test indicated a latent effect of statistical learning across both sessions, but robust benefits of the high experimental frequency condition clearly did not persist beyond the first few productions of session one. It is not clear how the statistical learning literature might predict such a finding. Nevertheless, if we look to other areas of research, and if we consider research over an expanded time scale, we begin to see some possible explanations for tapered statistical learning.

Research with adults suggests that perceptual learning may be limited or disrupted in the context of a speech production task. For example, Warker, Xu, Dell, and Fisher (2009) had participants produce lists of nonwords with co-occurrence constraints on phoneme sequences, for example, restricting /f/ to onsets and /s/ to codas (e.g., /fæs/ but not /sæf/). As a participant produced a list of restricted words, another participant produced lists with the opposite restrictions (e.g., /sæf/ but not /fæs/). When the restrictions in production and perception conflicted, participants' errors only corresponded to what they produced, suggesting that the act of production was more relevant to learning the co-occurrence constraints. Baese-Berk and Samuel (2016) asked Spanish-speaking participants to learn the /*ʎ*/ - /*ɲ*/ contrast from Basque. Some participants completed an initial perceptual learning phase, while other participants produced the contrast concurrently to the perceptual training. A clear decrement in perceptual learning was observed in the participants whose training included simultaneous perception and production of the contrast. These two studies suggest that production-based learning may sometimes outweigh perceptual learning, at least within an experimental setting. Given the high number of productions collected here, the statistical learning effect may have tapered off because it was overwhelmed by a normalizing production effect.

In fact, several studies indicate that repeated productions can have a normalizing effect. In articulation studies of nonword repetition, at least three studies (Richtsmeier & Goffman, 2015; Walsh, Smith, & Weber-Fox, 2006; Sasisekaran, Smith, Sadagopan, & Weber-Fox, 2010) have revealed greater production stability following repeated productions. Developmental research has also shown that children tend to become faster and more stable articulators over time, presumably as the result of increased speech production practice (A. Smith & Goffman, 1998; A. Smith & Zelaznik, 2004; B. L. Smith & Kenney, 1999; Sosa & Stoel-Gammon, 2012). Sosa and Stoel-Gammon (2012) examined intraword variability in transcriptions of the speech of two-year-olds. They looked at factors related to the number of phonetically different productions that children made of the same word. They found that high lexical frequency predicted low levels of variability, which they attribute to a motor practice effect. That is, as children produced a word often, they may have developed a “well-practiced motor routine” (p. 604), even when that motor routine was incorrect. In the present experiment, the nine productions that children made in each session could have resulted in a

well-practiced motor routine that was sometimes incorrect, but increasingly less susceptible to perceptual statistical learning for improving accuracy.

If we expand our frame of reference further still, to include the full time course of expressive speech sound development, we find additional reasons to expect statistical learning to taper off. Recall that our goal was to examine perceptual statistical learning in the context of expressive speech sound development. Many studies of expressive speech sound development have been carried out in the form of normative sampling across childhood (e.g., Sander, 1972; Smit, Hand, Freilinger, Bernthal, and Bird, 1990). These studies use single-word elicitation across large numbers of children to estimate production accuracy for consonants and consonant clusters in various age groups. In our experiment, perceptual statistical learning was examined over one week, a much smaller time frame. Nevertheless, estimates of weekly gains are obtainable from those normative studies. See the Appendix for the details of this analysis. Using the ages of customary production and production mastery for consonant singletons from Sander (1972), we estimate that children increased their accuracy for consonant singletons by about 0.33% per week, 95% CI [0.08%, 0.59%]. Smit's (1993) normative study focused on consonant clusters, and from her data we developed estimates for children ages 4;6 to 5;6—a range that is similar to the ages of the participants in our study. Within that range, children increased their accuracy by about 0.17% per week, 95% CI [-0.02% to 0.36%].

To compare the data from Sander (1972) and Smit (1993) to our study, we first calculated an average experimental frequency difference for each of our participants across all productions of all words. We used the experimental frequency difference, rather than mean accuracies for the two sessions, so that the weekly gain could be attributable to perceptual statistical learning. On average, participants increased their accuracy for high experimental frequency sequences 1.49% more than for low experimental frequency sequences, 95% CI [-24.9%, 27.8%]. Note that the variance of our experimental data is quite high, and based on Levene tests, it is significantly higher than the Sander and Smit norms (for Sander, 1972, $F = 30.22$, $p < .001$; for Smit, 1993, $F = 32.82$, $p < .001$). It is nevertheless notable that the 1.49% figure from this study is similar to the estimates of 0.33% and 0.17% from the normative studies. In sum, normative studies of expressive speech sound development reveal a time course that is slow and likely relatively insensitive to short-term changes in statistics, indicating that a tapered statistical learning effect is quite sensible.

Returning to the statistical learning literature, it is important to consider why something that infants do so well might have a relatively small and tapered effect in the context of expressive speech sound development. One possibility is that statistical learning is robust early in development but less important as children reach preschool age. The significant effect of input frequency in the Ota and Green (2013) study, which focused on children who were under 4;0, is consistent with this possibility. It is also reflective of data from the Iowa-Nebraska Articulation Norms Project (Smit et al., 1990). Smit et al. present acquisition curves for consonant singletons and consonant clusters, broken down by gender and word position. For many sounds and sound sequences, the figures indicate an initially steep rise in accuracy up to about 65-70%, followed by a slower rate of improvement towards mastery. Although speculative, we suggest that the plateau in accuracy may correspond to a reduced

role for perceptual statistical learning, and at the same time, a greater effect of the normalizing forces of speech production (Sosa & Stoel-Gammon, 2012). As discussed in the introduction, well documented effects of English frequency on expressive speech sound development could be due to either perceptual learning or production practice (cf. discussion of Edwards & Beckman, 2008; Munson, 2001). Perhaps both factors are important, but at different developmental stages. An important direction for future research will be to identify the unique contributions of production practice and perceptual learning across different age groups.

Another explanation for the tapered effect is that statistical learning may be limited by other cognitive processes, such as memory. The tapered quality of the statistical learning effect—present only in the first few productions following familiarization—suggests that the perceptual statistical learning is subject to forgetting, or to interference, or perhaps to both. One approach for studying the relationship between statistical learning and memory would be to have child participants complete multiple statistical learning studies with different stimuli. If less evidence for learning is present in the second study that they complete (and assuming that the order of presentation of the different stimuli is counterbalanced), then statistical learning may be subject to the kinds of memory interference effects often observed in list learning studies (for a general review cf. Howes, 2006).

A third and final explanation for the tapered effect comes from the corpus studies by Van Severen et al. (2013) and Ota and Green (2013). Those authors remind us that statistical learning occurs within a complex phonological milieu, with multiple factors contributing to a sound's developmental course. Van Severen et al. (2013) controlled for the correlation between input frequency and functional load in a statistical model of word-initial consonant acquisition, and they concluded that functional load was driving the acquisition process beyond the effect of input frequency. However, they also warned of additional uncontrolled-but-correlated factors that might explain the course of expressive speech sound development, including articulatory complexity and perceptual salience. Still more correlated factors include the age of acquisition of a sound, the frequency that a child produces that sound, neighborhood density, and word length (Stoel-Gammon, 1998). Intriguingly, the work by Sosa and Stoel-Gammon (2012) suggests that production practice may sometimes work to reduce production accuracy, thereby counteracting potentially positive influences from mechanisms like perceptual statistical learning. In sum, numerous factors likely contribute to the trajectory of expressive speech sound development. These factors may overwhelm perceptual statistical learning in real-world learning, but they may also explain why typical speech sound learning is slower than what statistical learning experiments with infants might suggest.

The present experiment provides the statistical learning enterprise with a new level of ecological validity in that it expands the statistical learning framework to include underutilized dependent variables like speech accuracy, and it measures the effects of statistical learning over time. Although infant studies have demonstrated that perceptual statistical learning is a powerful mechanism during the first year of life, there is little data concerning how statistical learning is used by young children who are immersed in expressive speech and language development. Furthermore, speech accuracy is an inherently

linguistic behavior, making it ideal for drawing conclusions about real-world learning. Finally, only a handful of studies have looked at statistical learning in production and over time. By addressing each of the concerns above, this study represents an important step for bridging the gap between the statistical learning framework and the existing literature on expressive speech sound development.

5.1 Clinical Implications

The present results may place some limitations on the clinical value of the auditory input, but they do not preclude a role for perceptual statistical learning in a clinical context. For example, Plante et al. (2011) observed that children with specific language impairment were more accurate to produce novel words that they first heard produced by 10 different talkers. Those authors used a familiarization procedure like the one used here. Furthermore, their approach resonates with the “focused auditory input” component of the Cycles Phonological Remediation Approach (Hodson & Paden, 1991), also commonly referred to as auditory bombardment. In the Cycles approach, children listen to careful-speech productions of approximately 20 words containing their target sound or sound pattern.

The present results may also relate to the use of recasting, a clinical technique in which an adult provides a grammatically correct and pragmatically appropriate response to a child's errorful production. For example, if a child leaves off the copula *be* from their speech and says, “The dog running,” a clinician might respond with the correct form by saying, “Yes, the dog is running very fast.” A recent meta-analysis by Cleave, Becker, Curran, Van Horne, and Fey (2015) supports the use of recasting in intervention. The present study supports a basic assumption inherent to recasting, namely, that children can learn accurate forms from what they hear. Furthermore, when a recast is contingent upon an error, the recast highlights both the error and the correct form. Ongoing research by Plante and her colleagues (Plante et al., 2014) shows how statistical learning can be effectively combined with recasting.

6. Conclusion

In conclusion, we observed tapered perceptual statistical learning in typically developing children. The children learned word-medial consonant sequences and, after hearing them in multiple words produced by multiple talkers, were initially more accurate to produce those sequences. The learning effect was only observable for the first three productions of the first session. Yet, the rate of learning was comparable to learning rates calculated from previous normative studies of expressive speech sound development. Thus, the results set limitations on our expectations for perception-based statistical learning.

If we return to Romberg and Saffran's (2010) broad definition of statistical learning as the “discovery of patterns in the input”, a new path opens up for the future study of this important cognitive skill. Rather than favoring perceptual learning and existing methodologies, the present results suggest that statistical learning research should expand to include new questions and methodologies. With regards to expressive speech sound development, a focus on the relative contributions of different learning modalities, that is, both perception and production, may be especially fruitful.

Acknowledgments

This research was supported by grants from the National Institute on Deafness and Other Communication Disorders; R03DC011898 to Peter T. Richtsmeier and R01DC04826 to Lisa Goffman. The authors thank Antigone Fleck, Erica Berlin, Janna Berlin, Meredith Saletta, and Shie Kantor for their assistance with various aspects of data collection and analysis; the authors thank LouAnn Gerken for helpful commentary.

Appendix: Calculating Weekly Gains in Accuracy from Normative Studies of Expressive Speech Sound Development

We used normative data from Sander (1972) and Smit (1993) to estimate weekly gains in expressive speech sound accuracy across development. Sander formulates a range for thinking about child production accuracy. He graphed 24 consonants of American English starting at an age of “customary production”, or the age at which at least 50% of children in an age group correctly produce the sound in two out of three environments (initial, medial, and final positions). He ended each range at the age of “production mastery”, when 90% of children in an age group correctly produced the sound in all environments. In other words, each range represents the number of years and months it took for most children to move from producing the sound to producing the sound correctly. These age ranges can be transformed into weeks (Y years multiplied by approximately 52 weeks), and given the fixed growth in accuracy from 50% to 90%, or a 40% change, we divide 40% by the total number of weeks to determine the percent gained per week. Note that this analysis does not take the ages of the children into account, but Sander's earliest data is for children who are 1;6 (years; months) up to children who are 8;6. Based on Sander's data, we estimate that children increased their accuracy for consonant singletons by about 0.33% per week, 95% CI [0.08%, 0.59%].

Smit (1993) presents norms for gains in accuracy for word-initial consonant clusters. Rather than reporting ages based on fixed accuracy values, Smit reported accuracy values for each cluster at regular age intervals from 2;0 to 9;0. Smit's data allowed us to match our estimates for weekly gains in accuracy to the ages of our participants. Therefore, we measured the change in accuracy for consonant clusters from 4;6 to 5;6. First, the fixed age range was transformed into 52.14 weeks. Change in accuracy was calculated for each consonant cluster, and then each percent change was divided by 52.14. Children in Smit's study increased their accuracy by about 0.17% per week, 95% CI [-0.02% to 0.36%].

The raw data from the Sander (1972) and Smit (1993) studies are presented in Table A1 below. We note several caveats for interpreting these estimates of weekly gains in accuracy. First, the learning targets, age ranges, and procedures differ between the Sander and Smit studies compared to the present study. Furthermore, both datasets are from normative samples of speech development, which themselves entail a number of weaknesses related to the assessment of expressive speech sound development. These weaknesses include the single-word elicitation procedure, biases in the selection of children when developing the norms (Peña, Spaulding, & Plante, 2006), and the high degree of individual variability that they collapse. Thus, we encourage readers to use caution when interpreting the relatively narrow range of values obtained.

Table A1

Duration of development for consonant singletons (Sander, 1972) and percent change over one year for word-initial consonant clusters (Smit, 1993). Means and standard deviations for the percent weekly change for each study are presented in the bottom row.

Sander (1972)			Smit (1993)		
Target	Years from customary production (50%) to production mastery (90%)	Percent of weekly change	Target	Percent change from 4;6 to 5;6	Percent of weekly change
p	1.5	0.51	tw-	2.5	0.05
m	1.5	0.51	kw-	7	0.13
h	1.5	0.51	sp-	3.5	0.07
n	1.5	0.51	st-	4.5	0.09
w	1.5	0.51	sk-	0	0.00
b	2.5	0.31	sm-	10	0.19
k	2	0.38	sn-	3	0.06
g	2	0.38	sw-	5	0.10
d	2	0.38	sl-	8.5	0.16
t	4	0.19	pl-	11	0.21
ŋ	4	0.19	bl-	8.5	0.16
f	1.5	0.51	kl-	9	0.17
j	1.5	0.51	gl-	10.5	0.20
ɹ	3	0.26	fl-	15	0.29
l	3	0.26	pr-	14	0.27
s	5	0.15	br-	12	0.23
tʃ	3.5	0.22	tr-	15.5	0.30
ʃ	3.5	0.22	dr-	3.5	0.07
z	4.5	0.17	kr-	6.5	0.12
dʒ	3	0.26	gr-	15.5	0.30
v	4	0.19	fr-	15.5	0.30
θ	2.5	0.31	θr-	16.5	0.32
ð	3	0.26			
ʒ	2.5	0.31			
<i>M</i> (<i>SD</i>)		0.33 (0.13)	0.17 (0.096)		

References

- Baese-Berk MM, Samuel AG. Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*. 2016; 89:23–36. <http://dx.doi.org/10.1016/j.jml.2015.10.008>.
- Bankson, NW., Bernthal, JE. Bankson-Bernthal Test of Phonology [Measurement instrument]. Austin, TX: Pro-Ed; 1990.
- Boersma, P., Weenink, D. Praat: Doing phonetics by computer (Version 5.4.18) [Computer program]. 2015. Retrieved from <http://www.praat.org/>

- Burgemeister, BB., Blum, LH., Lorge, I. Columbia Assessment of Nonverbal Abilities [Measurement instrument]. New York, NY: The Psychological Corporation; 1972.
- Cleave PL, Becker SD, Curran MK, Van Horne AJO, Fey ME. The Efficacy of Recasts in Language Intervention: A Systematic Review and Meta-Analysis. *American Journal of Speech-Language Pathology*. 2015; 24(2):237–255. DOI: 10.1044/2015_AJSLP-14-0105 [PubMed: 25654306]
- Dawson, JL., Stout, CE., Eyer, JA. The Structured Photographic Expressive Language Test—Third Edition [Measurement instrument]. DeKalb, IL: Janelle Publications; 2003.
- Dollaghan C, Campbell TF. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1136–1146. DOI: 10.1044/jslhr.4105.1136
- Dunn, LM., Dunn, DM. The Peabody Picture Vocabulary Test—4th Edition [Measurement instrument]. Minneapolis, MN: Pearson; 2007.
- 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. DOI: 10.1044/1092-4388(2004/034)
- Edwards J, Beckman ME. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language learning and development*. 2008; 4(2):122–156. <http://dx.doi.org/10.1080/15475440801922115>. [PubMed: 19890438]
- Eidsvåg SS, Austad M, Plante E, Asbjørnsen AE. Input variability facilitates unguided subcategory learning in adults. *Journal of Speech, Language, and Hearing Research*. 2015; 58(3):826–839. DOI: 10.1044/2015_JSLHR-L-14-0172
- Evans JL, Saffran JR, Robe-Torres K. Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2009; 52(2):321–335. DOI: 10.1044/1092-4388(2009/07-0189)
- Finley S. Generalization to unfamiliar talkers in artificial language learning. *Psychonomic bulletin & review*. 2013; 20(4):780–789. DOI: 10.3758/s13423-013-0402-7 [PubMed: 23456327]
- Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA*. 2002; 99:15822–15826. DOI: 10.1073/pnas.262659399
- Gómez, RL. Memory, sleep and generalization in language acquisition. In: Arnon, I., Clark, EV., editors. *Experience, Variation and Generalization: Learning a first language*. Amsterdam: John Benjamins Publishing; 2011. p. 261–276.
- Gómez RL, Gerken LA. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*. 2000; 4(5):178–186. [http://dx.doi.org/10.1016/S1364-6613\(00\)01467-4](http://dx.doi.org/10.1016/S1364-6613(00)01467-4). [PubMed: 10782103]
- Heisler L, Goffman L, Younger B. Lexical and articulatory interactions in children's language production. *Developmental Science*. 2010; 13(5):722–730. DOI: 10.1111/j.1467-7687.2009.00930.x [PubMed: 20712738]
- Hodson, BW., Paden, EP. A phonological approach to remediation: Targeting intelligible speech. Austin, TX: Pro Ed; 1991.
- Howes, MB. Human memory: Structures and images. Thousand Oaks, CA: Sage Publications; 2006.
- Houston DM, Jusczyk PW. Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*. 2003; 29(6):1143–1154. <http://psycnet.apa.org/doi/10.1037/0096-1523.29.6.1143>. [PubMed: 14640835]
- Másdóttir T, Stokes SF. Influence of consonant frequency on Icelandic-speaking children's speech acquisition. *International journal of speech-language pathology*. 2016; 18(2):111–121. <http://dx.doi.org/10.3109/17549507.2015.1060525>. [PubMed: 26220368]
- McLeod, S. Speech sound acquisition. In: Bernthal, JE, Bankson, NW., Flipsen, P., editors. *Articulation and phonological disorders: Speech sound disorders in children*. 7th. Pearson: Upper Saddle River, NJ; 2013. p. 58–113.
- Munson B. Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*. 2001; 44:778–792. DOI: 10.1044/1092-4388(2001/061)
- Ohala DK. The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*. 1999; 32(6):397–422. [http://dx.doi.org/10.1016/S0021-9924\(99\)00018-0](http://dx.doi.org/10.1016/S0021-9924(99)00018-0). [PubMed: 10560714]

- Onishi KH, Chambers KE, Fisher C. Learning phonotactic constraints from brief auditory experience. *Cognition*. 2002; 83(1):B13–B23. [http://dx.doi.org/10.1016/S0010-0277\(01\)00165-2](http://dx.doi.org/10.1016/S0010-0277(01)00165-2). [PubMed: 11814489]
- Optotrak. Waterloo, Ontario, Canada: Northern Digital Inc; 2011. [Apparatus and software]
- Ota M, Green S. Input frequency and lexical variability in phonological development: a survival analysis of word-initial cluster production. *Journal of Child Language*. 2013; 40(3):539–566. DOI: 10.1017/S0305000912000074 [PubMed: 22452825]
- Lawrence, KS: Perception Research Systems; Paradigm (Version 2.4). [Computer software]Retrieved from <http://www.paradigmexperiments.com/>
- Peña ED, Spaulding TJ, Plante E. The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology*. 2006; 15(3):247–254. DOI: 10.1044/1058-0360(2006/023) [PubMed: 16896174]
- Plante E, Bahl M, Vance R, Gerken L. Beyond phonotactic frequency: Presentation frequency effects word productions in specific language impairment. *Journal of Communication Disorders*. 2011; 44(1):91–102. DOI: 10.1016/j.jcomdis.2010.07.005 [PubMed: 20727533]
- Plante E, Ogilvie T, Vance R, Aguilar JM, Dailey NS, Meyers C, Lieser AM, Burton R. Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*. 2014; 23:530–545. DOI: 10.1044/2014_AJSLP-13-0038 [PubMed: 24700145]
- Richtsmeier PT. Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology*. 2011; 2(1):157–183. DOI: 10.1515/labphon.2011.005
- Richtsmeier PT. Phonological and Semantic Cues to Learning from Word-Types. *Laboratory Phonology*. 2016; 7(1):1–23. <http://dx.doi.org/10.5334/labphon.39>.
- Richtsmeier PT, Gerken LA, Goffman L, Hogan T. Statistical frequency in perception affects children's lexical production. *Cognition*. 2009; 111:372–377. <http://dx.doi.org/10.1016/j.cognition.2009.02.009>. [PubMed: 19338981]
- Richtsmeier PT, Gerken LA, Ohala DK. Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language*. 2011; 38:951–78. <http://dx.doi.org/10.1017/S0305000910000371>. [PubMed: 21126387]
- Romberg AR, Saffran JR. Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2010; 1(6):906–914. DOI: 10.1002/wcs.78 [PubMed: 21666883]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1928. DOI: 10.1126/science.274.5294.1926 [PubMed: 8943209]
- Sander EK. When are speech sounds learned? *Journal of Speech and Hearing Disorders*. 1972; 37(1): 55–63. DOI: 10.1044/jshd.3701.55 [PubMed: 5053945]
- Sasisekaran J, Smith A, Sadagopan N, Weber-Fox C. Nonword repetition in children and adults: Effects on movement coordination. *Developmental Science*. 2010; 13(3):521–532. DOI: 10.1111/j.1467-7687.2009.00911.x [PubMed: 20443972]
- Smit AB. Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research*. 1993; 36(5):931–947. DOI: 10.1044/jshr.3605.931
- 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. DOI: 10.1044/jshd.5504.779 [PubMed: 2232757]
- Smith A, Goffman L. Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*. 1998; 41:18–30. DOI: 10.1044/jslhr.4101.18
- Smith A, Zelaznik H. Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*. 2004; 45:22–33. DOI: 10.1002/dev.20009 [PubMed: 15229873]
- Smith BL, Kenney MK. A Longitudinal Study of the Development of Temporal Properties of Speech Production: Data from 4 Children. *Phonetica*. 1999; 56:73–102. DOI: 10.1159/000028442 [PubMed: 10450077]

- Sosa AV, Stoel-Gammon C. Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*. 2012; 55(2):596–608. DOI: 10.1044/1092-4388(2011/10-0113)
- Stoel-Gammon, C. Sounds and words in early language acquisition: The relationship between lexical and phonological development. In: Paul, R., editor. *Communication and Language Intervention Series: Volume 8 Exploring the speech-language connection*. Baltimore, MD: Brookes; 1998. p. 25-52.
- Storkel HL. Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*. 2001; 44:1321–1337. DOI: 10.1044/1092-4388(2001/103)
- Storkel HL, Hoover JR. An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods*. 2010; 42(2):497–506. DOI: 10.3758/BRM.42.2.497 [PubMed: 20479181]
- Van Severen L, Gillis JJM, Molemans I, van den Berg R, de Maeyer S, Gillis S. The relation between order of acquisition, segmental frequency and function: The case of word-initial consonants in Dutch. *Journal of Child Language*. 2013; 40(4):703–740. DOI: 10.1017/S0305000912000219 [PubMed: 22853849]
- Walsh B, Smith A, Weber-Fox C. Short-term plasticity in children's speech motor systems. *Developmental Psychobiology*. 2006; 48(8):660–674. DOI: 10.1002/dev.20185 [PubMed: 17111401]
- Warker JA, Xu Y, Dell GS, Fisher C. Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*. 2009; 112:81–96. <http://dx.doi.org/10.1016/j.cognition.2009.03.009>. [PubMed: 19398099]
- Williams, KT. *Expressive Vocabulary Test – 2nd Edition* [Measurement instrument]. Minneapolis, MN: Pearson; 2007.
- Zamuner TS, Gerken LA, Hammond M. Phonotactic probabilities in young children's speech production. *Journal of Child Language*. 2004; 31:515–536. DOI: 10.1017/S0305000904006233 [PubMed: 15612388]

Highlights

- Statistical learning may explain how children increase speech accuracy over time
- In a statistical learning study, children improved accuracy by 1.5% over one week
- The effect was similar to estimates of weekly gains from previous normative studies

Figure 1.

Graph of the experimental frequency effect (high experimental frequency - low experimental frequency) for each production number. The nine productions from the first session are on the top row; the nine productions from the second session are on the bottom row. Shaded bars indicate positive effects of high experimental frequency; empty bars indicate negative effects. Error bars reflect mean standard errors.

Table I
Standardized Test Means and Confidence Intervals

	<i>n</i>	<i>M</i> (SD)	95% Confidence Interval
Expressive language (SPELT-III)	16	114.56 (8.49)	110.04 - 119.08
Expressive vocabulary (EVT-1)	16	110.88 (9.43)	105.85 - 115.90
Receptive vocabulary (PPVT-4)	16	114.31 (9.96)	109.00 - 119.62
Articulation (BBTOP)	16	102.88 (9.42)	97.86 - 107.89
Nonverbal skill (CMMS)	16	111.31 (9.71)	106.14 - 116.49
Nonword repetition (% correct)	15	74.65 (12.91)	67.506 - 81.80

Note. SPELT-III = Structured Photographic Expressive Language Test, Third Edition; EVT-1 = Expressive Vocabulary Test; PPVT-4 = Peabody Picture Vocabulary Test, Fourth Edition; BBTOP = Bankson Bernthal Test of Phonology; CMMS = Columbia Mental Maturity Scale. The nonword repetition task was adopted from Dollaghan and Campbell (1998) and was not completed by one participant.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table II
Familiarization and Test Words. Biphone probabilities were obtained from Storkel and Hoover's (2010) online calculator, http://129.237.148.107/cml/post_ccc.vi. Familiarization words in italics appeared in the low experimental frequency condition

Target Sequence	Sequence Biphone Probability	Familiarization Words	Test Words
/pt/	.0015	daptən zɛptəs <i>septəs</i>	bɪptəm
/zm/	.0003	kozɪmən lɪzməs <i>taɪzmək</i>	pɛzməf
/mk/	.0000	gumkəs tæmkən <i>dɪmkəs</i>	fomkəp
/fp/	.0000	nɪfpən ʃɛfpəs <i>kɔɪpət</i>	mæfpəm

Table III
Means and standard deviations for consonant sequence accuracy scores across all productions for the four within-subjects conditions

	Session 1	Session 2	Experimental Frequency Totals
High Experimental Frequency	5.43 (.55)	5.52 (.53)	5.48 (.52)
Low Experimental Frequency	5.34 (.71)	5.42 (.66)	5.38 (.68)
Session Totals	5.39 (.49)	5.47 (.44)	5.43 (.45)