THE RELATIONSHIP BETWEEN PHONEME PERCEPTION AND PRODUCTION

By Aliza Lichtenstein

A Thesis Submitted to the

Graduate School – New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Karin Stromswold, M.D., Ph.D.

And approved by

____

____

____

New Brunswick, New Jersey

October, 2010

ABSTRACT OF THE THESIS

The Relationship Between Phoneme Perception and Production

By ALIZA LICHTENSTEIN

Thesis Director:
Karin Stromswold

Twenty-eight typically developing preschool children were tested in 2 experiments. In a perception experiment, the children heard phonologically minimal quartets of words and chose which of 4 pictures matched the target word (e.g., selecting a picture of a *snail* from among pictures of *snail*, *sail*, *nail*, and *mail*). In a production experiment, they repeated the target word (e.g., *snail*). These experiments revealed evidence of a link between perception and production at age 4 and 5, but not at age 3, suggesting that the link between speech perception and production develops during the preschool years.

Acknowledgement

Most importantly, I thank Dr. Karin Stromswold for her broad research and clinical expertise driving this project. Her patience, resilience, and careful attention to detail are remarkable. I am grateful to Dr. Shigeto Kawahara for the generous time he dedicated to guiding this work. The stimuli would be considerably different without his experience in phonetics. I also benefited from insightful theoretical discussions with Dr. Paul De Lacey who kindly shared specialized equipment at Rutgers Center for Cognitive Science. I appreciate the support from my committee members Dr. Julien Musolino and Dr. Judith Hudson and thank Diane Molnar for helping with the recordings. This project would not have been possible without the kind cooperation of the parents and children who participated in this study and without support from the National Science Foundation BCS-0042561.

Table of Contents

List of tables

List of Illustrations

The Relationship Between Phoneme Perception and Production

One of the central questions in experimental phonology is how people are able to perceive speech as speech despite the lack of acoustical invariance in the speech stream. The articulatory gestures used to produce phonemes also vary tremendously across speakers, situations (normal talking, whispering, shouting, etc.), speaking rate, and phonological environment. Thus, a second core question in experimental phonology is how people automatically and effortlessly produce the articulatory gestures needed in different situations (see Miller, 1990 for a review). Logically speaking, phoneme perception could be independent of phoneme production, phoneme perception could be parasitic on phoneme production (roughly speaking, the motor theory of speech perception, Liberman & Mattingly, 1985), phoneme production could be parasitic on phoneme perception, some third factor could influence both, or the 2 systems could be completely independent.

Previous Research

Consistent with phoneme perception and production being linked, some studies have found that children with phonological language delays or impairments (henceforth, "misarticulating children") have poorer phonemic perception and awareness for the phonemes they misarticulate. For example, Marquart & Saxman (1972) reported that the number of words misarticulating children misperceive and misproduce are correlated. Furthermore, Shuster (1998) found that misarticulating children who pronounce /r/ incorrectly have difficulty detecting their own mispronunciations. More recently, Rvachew et al. (2003) found that misarticulating children have poorer phonemic perception for both words they say correctly and incorrectly.

Several studies report that misarticulating children who mispronounce /r/ as /w/, have a less sharp categorical boundary for /r/ and /w/ (Hoffman et al., 1985; Monnin and Huntington 1974; Ohde & Sharfe, 1988). Similar results have been found for fricative-affricate contrasts (e.g., Raaymakers & Crul, 1988), fricative place contrasts (e.g., Rvachew & Jamieson, 1989) and coronal-velar contrasts (e.g., Whitehill et. al., 2003).

In contrast to misarticulating children who have discrete difficulty producing a handful of phonemes, dyspraxic children have profound difficulty producing all or most phonemes. Groenen et al. (1996) found that the frequency with which Dutch-speaking dyspraxic children made place of articulation mistakes was related to the frequency with which they misperceived place of articulation (Groenen et al., 1996).

Contrary to the findings reported above, there is evidence that suggests that phoneme perception and production are independent of one another. Anecdotally, misarticulating children sometimes deem unacceptable their own mispronunciations (e.g., the child whom Berko & Brown (1960) reported said /fIs/ for *fish* yet objected when an adult said *fish* as /fIs/). In an experimental study, Rvachew & Grawberg (2006) found that the frequency with which preschool-aged children incorrectly say a phoneme is not related to their phonological awareness for the same phoneme. In another study, Bird & Bishop (1992) found that preschool-aged children who mispronounced particular phonemes, nonetheless had intact perception and awareness for these same phonemes.

Some studies have suggested that dyspraxic children have intact phoneme perception despite profound speech impairments. For instance, Hoit-Dalgaard et. al (1983) found no significant relationship between phoneme perception and production of voice onset time (VOT) in dyspraxic children. Moreover, in a study of a mute child with severe developmental verbal dyspraxia, Stromswold (2009) found the child's performance on a phoneme minimal pairs test to be near perfect.

The lack of consensus about the phoneme perception-production link across these studies could have been a result of differences in participants. In general, the studies that found *no* link between the systems typically assessed younger pre-school children (e.g., Bird & Bishop, 1992; Rvachew & Grawberg, 2006; Stromswold, 2009) than studies that did find a link and included older children (e.g., Groenen et al., 1996; Shuster, 1998; Hoffman et al., 1985). Furthermore, studies that collapsed data across a wider age range (e.g., Groenen et al., 1996; Shuster, 1998; Hoffman et al., 1985) were more likely to find a link than studies that collapsed across a tight age range or tested individual participants (e.g., Bird & Bishop, 1992; Rvachew & Grawberg, 2006; Stromswold, 2009).

Regarding the nature of the tasks, most of these studies demonstrating a link used categorical speech perception (e.g., Groenen et al., 1996, Hoffman et al., 1985; Monnin and Huntington 1974, Ohde & Sharfe, 1988), whereas most studies that found no link used discrimination tasks (e.g., Bird & Bishop, 1992; Rvachew et al 2003; Rvachew & Grawburg, 2006; Stromswold, 2009). Moreover, whereas the majority of studies that found a link specifically targeted /r/,which children frequently mispronounce (e.g., Hoffman et al., 1985; Monnin and Huntington 1974; Ohde & Sharfe, 1988; Shuster, 1998), studies that found no link used different and/or multiple phonemes in their assessments (e.g., Bird & Bishop, 1992; Hoit-Dalgaard et.al, 1983; Rvachew & Grawberg, 2006; Stromswold, 2009). Finally, most of the aforementioned studies did not examine cluster reduction, even though this is a common type of speech error in both onset and coda position.

The Current Study

The current study sought to clarify the relationship between phoneme perception and phoneme production in children. In the first experiment, children listened to minimal quartets of words and chose which of 4 pictures matched the target word (e.g., choosing the picture of a snail from among the pictures *snail, sail, nail,* and *mail*). In the second experiment, the same group of children said the target words used in experiment 1 (e.g., in the example above, the word *snail*).

Experiment I: Perception Task

Method I

Participants

Twenty-eight (16 males and 12 females) monolingual, English-speaking children (16 boys, 12 girls) between the ages of 3;0-5;6 (mean age = 4;2) participated. All children were typically-developing, with no history of speech impairment, hearing impairment, cognitive impairment, or physical impairment that might influence language development or interfere with their ability to perform the experiments. Children were recruited from New York City neighborhoods and private schools.

Visual Stimuli

For each trial, 4 pictures were presented simultaneously on a 17" computer screen in a 2x2 grid with each picture occupying an equal amount of the computer screen. As shown in Figure 1 below, the pictures in each of the 4 quadrants were clear and vivid, and were uniform in size, style, and type (e.g., black-and-white versus color, photographs versus drawing). Note that the words in the figure did not appear on the display in the experiment.

Figure 1. Sample trial in the perception task. Target *snail* with distractors *nail, mail sail.*

Acoustic Stimuli

Stimuli were recorded by a native monolingual English-speaking woman. The speaker was blind to the nature of the experiments, and received no guidance regarding the pronunciation of the words. Phoneticists deemed her to have a typical New Jersey accent with no evidence of articulatory problems. She spoke clearly (no over-nasalization, no dropping of final consonants, etc.), but did not hyper-articulate.

Because the intent of the experiment is to investigate the perception-production link of phonemes, great care was taken in producing the high-fidelity recordings. Stimuli were recorded using a head-mounted *Shure Microphone* attached to a *Roland Edirol R09 Solid State Recorder* and were digitized with a 44.1 kHz audio sample rate and 16 bit audio sample size.

Words were recorded in a sound attenuated booth. The 45 target words were all recorded on the same day, with each target word being recorded 3 times in 3 different sessions for a total of 9 recordings for each word. To avoid list intonation, each target word was inserted in the carrier sentence *say the word_, twice.* The carrier sentence ended with the word *twice* in order to avoid phrase-final lengthening of the target word and to avoid the phonetic property of "creakiness" in the target word. To ensure the speaker recited all stimuli recordings at an even rate, the experimenter monitored the meter on the *Edirol* recorder so that no "peaking out" occurred during the recordings.

The software program *Praat* (version 5.0.40) was used to extract the target word from the carrier sentence (Boersma & Weenink, 1992, 2008). Only nonlinguistic noise was removed from the recordings and the best example of each extracted word was chosen using the criteria of naturalness, clarity, least background noise, and least aspiration. The amplitude of each word was then adjusted to a mean of 70 dB. Five monolingual English speakers with no background in linguistics and no knowledge of the experiment judged that these recordings of the target words were natural sounding, clear, and similar to one another.

Linguistic Stimuli

All of the words in the experiment were high frequency, monosyllabic words that are acquired at a young age and are easily depictable (see appendix A). To form phonologically minimal quartets, each of 45 target words was grouped with three distractor words that differed minimally from the target word. Minimal quartets were designed instead of traditional minimal pairs to examine the relationship between several phonemes simultaneously and to reduce chance-level performance from 50% to 25%.

Taken as a group, the 45 target words and their distractors assessed the following 3 key aspects of children's perception and production: the beginnings versus ends of syllables (i.e., onsets versus codas); simple consonants versus consonant clusters; articulatory features such as voicing, manner of articulation, and place of articulation. In addition, stimuli were chosen to correspond to the types of speech errors children frequently make.

Onset vs. Coda. We assessed phoneme perception in both onset and coda position for 3 reasons. First, in modern theories of phonology, words are composed of syllables, which can be further broken down into onsets, rimes, and codas (see Figure 2). Second, the acoustic features that distinguish between minimally contrasting phonemes in onset position are not always the same as the acoustic features that distinguish between the same minimally contrasting phonemes in coda position. For example, the key acoustic features that distinguishes between oral stops that differ only in voicing (e.g., the unvoiced /t/ and its voiced counterpart /d/) in onset position is voice onset time (VOT) and the key feature in coda position is the duration of the preceding vowel. Third, children sometimes mispronounce the same phoneme differently in onset vs. coda position. For example, when children make voicing errors, they tend to voice unvoiced consonants in onset position (e.g., mispronouncing ***p****ig* as ***b****ig*) and de-voice consonants in coda position (e.g. mispronouncing *ca**b*** as *ca**p***). Thus, from linguistic, acoustic, and developmental psycholinguistic perspectives, it is critical to assess children's performance on phonemes in both onset position (as is typically done) and in coda position. (We did not assess perception of vowels because the pronunciation of vowels differs from dialect to dialect).

Simple Consonant vs. Consonant Clusters. Stimuli further divided into those with simple consonants *(**l**ake; fi**n**),* CC clusters (***sn**ail; fi**st***) and CCC clusters *(**str**ip)*. Some clusters are only acceptable in onset position and some are only acceptable in coda position (e.g., *st, sl, sw* are acceptable in onset position in English, whereas *sb, sg, zb* are not). Thus, we included different simple C and complex CC and CCC combinations in both onset and coda positions. See figure 2.

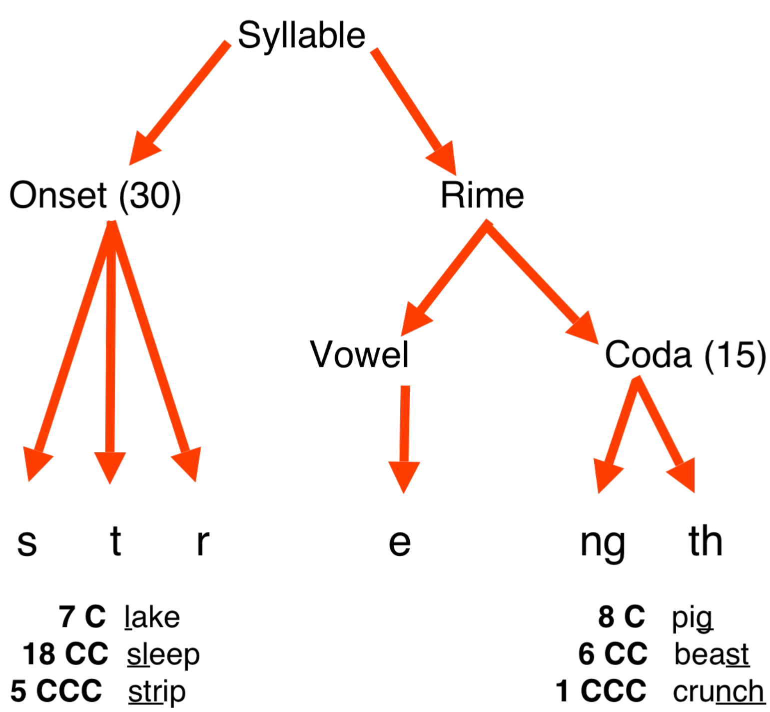

Figure 2. Structure of a syllable.

Articulatory Features. Each target word was grouped with 3 corresponding distractor words according to the articulatory features of English consonants. Consonants were classified according to whether or not the vocal folds vibrated (voicing feature), where airflow was obstructed (place of articulation or POA), and how airflow was obstructed (manner of articulation or MOA). See figure 3.

		Place of articulation					
		Bilabial	**Labiodental**	**Interdental**	**Alveolar**	**Palatal**	**Velar**
Manner of production	**Oral stop** voiceless voiced	p (pin) b (bin)			t (tin) d (din)		k (kin) g (get)
	Nasal stop voiced	m (map)			n (nap)		ŋ (sing)
	Fricative voiceless voiced		f (fin) v (van)	θ (thin) ð (than)	s (sin) z (zone)	ʃ (shin) ʒ (leisure)	
	Affricate voiceless voiced					tʃ (chin) dʒ (gin)	
	Liquid voiced				l (law) r (raw)		
	Glides voiced					j (yes)	w (we)

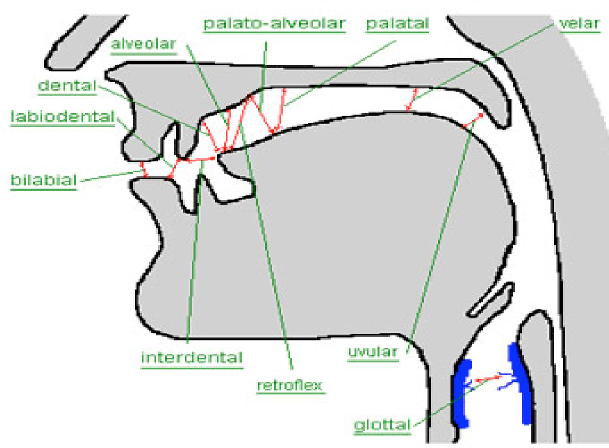

Figure 3. Consonant Classification in English

In Figure 3, POA for consonants are given in the columns with the point of articulation being increasingly far back in the mouth as the columns go to the right. Rows indicate MOA, with amount of obstruction decreasing the farther down in the table moves. Finally, when two phonemes occupy the same cell of the graph, the phoneme on the top is the voiceless version and the phoneme on the bottom is its voiced counterpart. For example, the top left cell of the chart contains the phonemes *p* and *b*. POA for both is bilabial, MOA for both is oral stop, with the *p* being voiceless and the *b* being voiced. Thus the phonemes *p* and *b* differ only in the single feature voicing.

The phonetic detail by which all target words differed from its distractors corresponded to these 3 dimensions of English consonant phonetic classification. Specifically, place of articulation (POA) distractors differed in POA only (e.g., ***s**eat*> ***sh**eet)*, with the same voicing and manner of articulation (MOA), while MOA distractors differed in MOA only, with the same voicing and POA *(**t**oe* > ***s**ew)*. Voicing distractors differed in voicing only, with the same POA and MOA (e.g., ***g**rab* >***c**rab*). See table 1.

Table 1

Distractor Types

Distractor Type	Example (onset)	Example (coda)
Place of Articulation Distractor	***s**eat* > ***sh**eet*	*co**ke*** > *coa**t***
Manner of Articulation Distractor	***t**oe* > ***s**ew*	*wro**te*** > *roa**d***
Voicing Distractor	***g**rab* > ***c**rab*	*pi**g*** > *pi**ck***
Cluster Reduction Distractor	***sn**ail* > ***n**ail*	*bea**st*** > *bea**t***

Common Speech Errors. Items were chosen to be minimally distinct from the target word from an articulation standpoint, *and* to correspond to errors that preschool age children frequently make (Sander, 1972). For example, we included 7 cases testing gliding errors in which /r/ and /l/ are pronounced as /w/. See table 2.

Table 2

Processes Assessed in Experiment 1

Error Type	# Onset Cases	# Coda Cases	# Total Cases
Simple C Onset Deletion	2 (e.g., *seat > eat)*	0	2
Onset Cluster Reduction	51 (e.g., *snail* > *nail*)	0	51
Coda Cluster Reduction	0	14 (e.g., *beast* > *beat*)	14
Fronting	17 (e.g., *crash* > *trash*)	11 (e.g., *coke* > *coat*)	28
Stopping	5 (e.g., *fry*> *tie*)	11 (e.g., *rice* > *write*)	16
Gliding	7 (e.g., *lake* > *wake*)	0	7
Voicing	12 (e.g., *grab* > *crab*)	13 (e.g., *rice* > *rise*)	25

Summary. Each quartet (e.g., ***snail***, ***nail***, ***sail***, ***mail***) incorporated the above key aspects of perception. The phoneme types and target stimuli used in the study are listed in table 3 below. All stimuli quartets (target and distractors) are listed in appendix A.

Table 3

The 45 Stimuli Targets used in Experiment 1

Target Phoneme Type	**Phoneme Targets**	
C	**Onset (30)**	**Coda (15)**
fricative /s, f, z /	*seat sing fat (3)*	*rice buzz (2)*
stop /t, g, k, d, b /	*toe gas (2)*	*pig coke bad kid wrote robe (6)*
Liquid /l, r/	*lake rip (2)*	
CC		
fricative + stop /sp, sk, st/	*sp-ark sk-is (2)*	*bea-st cru-st fi-st (3)*
fricative + nasal /sn, sm/	*sn-ail sm-ell (2)*	
fricative + liquid /sl, fr, fl/	*sl-eep fr-y fl-ight (3)*	
stop + liquid /bl, gr, br, kr, tr, kl/	*bl-ed gr-ab br-eak cr-ash tr-ail tr-ee cl-ock (7)*	
stop + fricative /tsh, tsh, dž/	*chase chew chick chip (4)*	*catch badge (2)*
stop /nt/ + nasal		*pa-nt (1)*
CCC		
fricative + stop + liquid /skr, str, spr/	*scr-eam str-ip spr-ing (3)*	
fricative + stop + glide /skw/	*squ-eeze squ-irt (2)*	
Nasal+ stop + fricative /ntsh/		*cru-nch (1)*

Procedure

Children were tested individually in a quiet room in one session. They sat 18 inches at eye-level in front of a laptop computer wearing *Sennheiser HD 202* headphones.

As the children saw the visual stimulus on the screen, they heard the target word recording (audio stimulus) playing simultaneously at 70 dB via the headphones [e.g., "*snail*"]. The Children were asked to point to one picture in each quartet that they thought matched the word they heard. The picture remained on the screen until the experimenter marked the child's selection on the external keypad attached to the computer, which initiated a cartoon character to appear at the center of the screen. This served as a reward and a fixation point for the next trial. To initiate all subsequent trials, the experimenter pressed a designated key on the keypad. *Psyscope* was used to control the presentation of the 45 quartets and to record the data.

Quartet items that were not used in the main task and that were phonologically very different from each other (e.g., *star, bird, fork,* and *cheese*) were presented for practice. The same 4 practice trials were presented in random order until both of the following criteria were met: the children's selection accuracy on all quartets was 100% correct and children were acclimated to pointing procedure. All children reached both criteria within 1-2 minutes.

During testing, half of the participants were presented with the original stimuli presentation; the other half received the reverse order. Stimuli were presented in pseudo-random order and the target picture did not appear in same quadrant more than twice in a row. The entire session was video-recorded and recordings were used to check accuracy of coding.

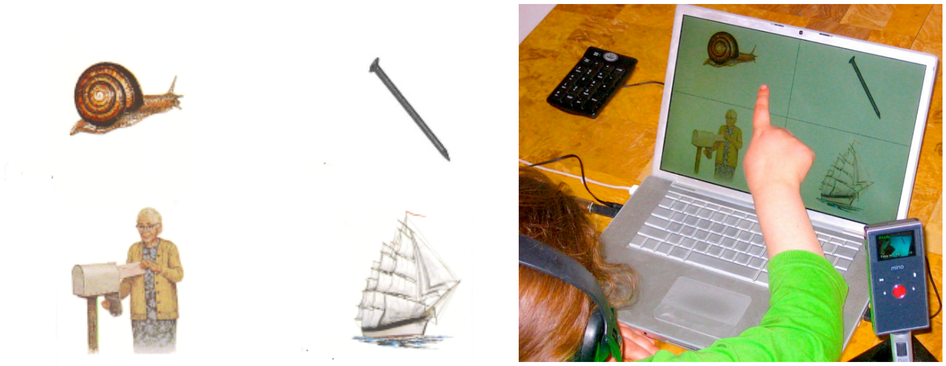

Figure 4. Perception task set-up. Target *snail* with distractors *nail, mail sail.*

Results I

For each phonological type, accuracy, and RT were analyzed using a 3 (age) x 2 (stimulus) x 2 (sex) ANOVA. There was no main effect of sex nor did sex interact with any phonological type, RT, or RT for correct trials only (all F's < 1). Thus, data from boys and girls were collapsed in subsequent analyses. See table 4 for age groups.

Table 4

Participant Age groups

Groups	Group 1	Group 2	Group 3
Age Range	3;0 – 3;6	3;8 – 4;6	4;9 – 5;3
Mean age	3;3	4;0	5;0

There was a significant main effect of age on accuracy ($F(2, 25) = 4.33$, p =.024), with the youngest children perceiving significantly more phonemes correctly than the older children. Pair-wise comparisons revealed that the 3 year olds were significantly less accurate on the perception task than the 4 year olds (F (1, 17) = 5.49, p =.032) and the 5 year olds (F (1, 17) = 7.13, p =.016). The 4 year olds, however, did not significantly differ from the 5 year olds in performance (F (1, 16) = 0.33, p =.573). See figure 5. Age did not have a significant effect on RT both when all trials were included and when only correct trials were included (both F's < 1).

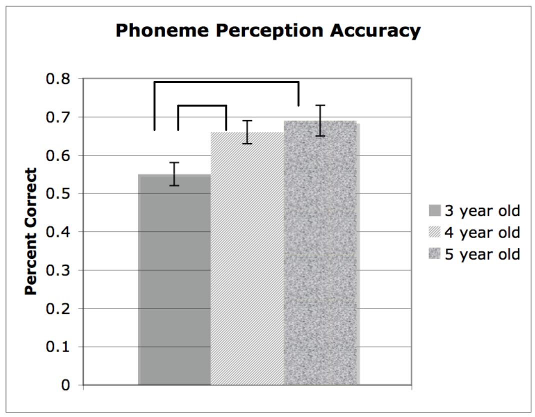

Figure 5. Phoneme perception developmental results

Perception Task Accuracy and RT

/S/ vs. Non-/S/. A 2 (*/s/-non-/s/*) x 3 (age) ANOVA revealed a significant main effect of age on phonological type, with the older children performing more accurately than the younger children (F (2, 25) = 3.84, p =.035). Post-hoc pairwise comparisons revealed this was a result of the 3 year olds (mean= 0.56; SE= 0.03) performing significantly less accurately than the 4 year olds (mean = 0.67; SE= 0.03) (F(1, 17) = 4.70, p = 0.045) and 5 year olds (mean = 0.69; SE= 0.03) (F(1, 17) = 6.50, p = 0.021). There was also a main effect of phonological type, with children performing more accurately on targets containing /s/ relative to non-/s/ targets (F(1, 25) = 6.77, p = 0.015). Age and phonological type did not interact. Similar analyses performed on RT data revealed no main effect or interaction when all trials and only correct trials were included

(both p's > .10). There was no significant effect of age on RT.

Liquids vs. Non-Liquids. A 2 (liquid/non-liquid) x 3 (age) ANOVA revealed a significant main effect of age on phonological type, with the older children performing more accurately than the younger children ($F(2, 25) = 4.79$, $p = 0.017$). Pairwise comparisons revealed this was a result of the 3 year olds (mean = 0.58; SE= 0.03) performing significantly less accurately than the ($F(1, 17) = 8.39$, $p = 0.010$) 4 year olds (mean =0.7333; SE= 0.0337) and 5 year olds (mean =0.72; SE= 0.04) ($F(1, 17) = 6.19$, $p = 0.024$). There was also a main effect of phonological type, with children performing significantly more accurately on targets that were not liquids than those that contained liquids ($F(1, 25) = 14.83$, $p = 0.001$). There was no interaction of age and stimulus type on accuracy. Similar analyses performed on RT data revealed that the children were significantly faster on non-liquids than liquids ($F(1, 25) = 7.39$, $p = 0.012$). The same was true when only correct trials were included: children were significantly faster on non-liquids than liquids ($F(1, 25) = 11.46$, $p = 0.002$). There was no interaction when all trials and only correct trials were included (both p's > .10). There was no main effect of age on RT.

Simple vs. Cluster. A 2 (simple/cluster) x 3 (age) ANOVA revealed a significant main effect of age on phonological type, with the older children performing more accurately than the younger children ($F(2, 25) = 3.71$, $p = 0.039$). Pairwise comparisons revealed this was a result of the 3 year olds (mean = 0.57; SE= 0.03) performing significantly less accurately than the 4 year olds (mean = 0.67; SE= 0.03) ($F(1, 17) = 5.19$, $p = 0.036$) and 5 year olds (mean = 0.69 ; SE= 0.03) ($F(1, 17) = 5.54$, $p = 0.031$).There was also a main effect of phonological type, with children performing more accurately on simple targets than clusters ($F(1, 25) = 8.50$, $p = 0.007$). Age and simple/complex did not interact. Similar analyses performed on RT data revealed that the children were significantly faster on simple than cluster target phonemes $F(1, 25)= 24.11$, $p = 0.000$). The same was true when only correct trials were included: children were significantly faster perceiving simple targets than cluster targets ($F(1, 25)= 18.64$, $p = 0.003$). There was no interaction when all trials and only correct trials were included (both p's > .10). There was no significant effect of age on RT.

Onsets vs. Codas. A 2 (onset/coda) x 3 (age) ANOVA revealed a significant main effect of age on phonological type, with the older children performing more accurately than the younger children ($F(2, 25)= 3.97$, $p = 0.032$). Pairwise comparisons revealed this was a result of the 3 year olds (mean = 0.56; SE= 0.03) performing significantly less accurately than the 5 year olds (mean =0.71 ; SE= 0.04) ($F(1, 17) = 7.82$, $p = 0.012$). There was also a main effect of phonological type, with children performing more accurately on onsets than codas ($F(1, 25) = 30.62$, $p = 0.000$). Age and onset/coda did not interact. Similar analyses performed on RT data revealed that the children were significantly faster at perceiving onsets than codas ($F(1, 25)= 8.37$, $p = 0.008$). When only correct trials were included, children were not significantly faster on onsets than codas ($F(1, 25)= 2.93$, $p = 0.100$).There was no interaction when all trials and only correct trials were included (both p's > .10). There was no significant effect of age on RT. (See table 5 and 6 for overall and developmental perception accuracy and RT results across all phonological types).

Table 5

Phoneme Perception Accuracy Results

Phoneme Type	All children	3 year olds	4 year olds	5 year olds
All Types	.63 (.02)	.55 (.03)	.66 (.03)	.69 (.04)*
S	.67 (.03)	.60 (.04)	.70 (.05)	.72 (.05) *
No S	.60 (.02) *	.52 (.03)	.63 (.03)	.67 (.05)
Liquid	.61 (.03) **	.53 (.03)	.68 (.04)	.63 (.05) *
No Liquid	.73 (.03)	.62 (.06)	.78 (.05)	.81 (.06)
Simple	.67 (.02) **	.61 (.05)	.72 (.03)	.70 (.04) *
Cluster	.61 (.02)	.52 (.03)	.63 (.04)	.68 (.04)
Onset	.73 (.03) **	.62 (. 06)	.78 (.05)	.81 (.06) *
Coda	.54 (.03)	.50 (.04)	.50 (.04)	.61 (.05)

**p* < .05; ***p* < .01. Mean and Standard Errors.

Table 6

Phoneme Perception RT Results

RT	Phoneme Type	All Children
All Trials	All Types	2874.62
Correct Trials	All Types	2643.07
All Trials	S	2965.04
	No S	2824.74
Correct Trials	S	2714.66
	No S	2597.78
All Trials	Liquid	2952.63*
	No Liquid	2680.20
Correct Trials	Liquid	2804.28**
	No Liquid	2488.91
All Trials	Simple	2654.88**
	Cluster	2984.50
Correct Trials	Simple	2409.11**
	Cluster	2765.56
All Trials	Onset	2680.20**
	Coda	2991.04
Correct Trials	Onset	2488.91
	Coda	2633.69

p < .05; **p* < .01

Experiment II: Production Task

Method II

Participants

Same as experiment I.

Visual Stimuli

The visual stimuli used in experiment II consisted of a series of the same target pictures from Experiment I, without the distractor pictures. For instance, only the picture of the target word *snail* [from the quartet *snail, mail, nail, sail* from experiment I] was presented. See figure 6. The picture was larger than its presentation in one of the 4 quadrants used in experiment I and it appeared at the center of the screen that was otherwise blank. Each of the 45 pictures was the same size.

Figure 6. Sample trial in production task. Target *snail.*

Acoustic Stimuli

The same acoustic stimuli and corresponding recording procedure that was used in experiment I was applied to experiment II. With respect to editing, the word *twice* was excluded from the phrase *say the word_ , twice* in which each token was originally recorded, which produced the audio instruction phrase for each token. For example *say* *the word snail, twice* was reduced to *say the word snail*, which became the audio component for the trial with the token *snail.* See table 7.

Table 7

Editing Recordings for Production Task

original recording >	**production task recording**
say the word snail, twice >	*say the word snail*

Linguistic Stimuli

Same stimuli as experiment I, including only target words (and not distractors).

Procedure

The standardized *Denver Articulation Screening Examination (DASE)* was administered as a familiarization to experiment II (Drumwright, 1971). The set-up and procedure for the production task was the same as experiment I. Additionally, each child wore a head-mounted *Shure* microphone held within 16 inches of the child's mouth, which was attached to the *Edirol State Recorder*.

The target pictures were sequentially presented in the exact order in which they appeared during the perception task. The children listened to the instructions *say the word*_____ while looking at the matching picture on the screen (and repeated the target word). This way, their errors could be attributed to mispronunciation and not mistakes in hearing. Once the child said the word, which was recorded on the *Edirol,* the experimenter pressed a key on the external keypad to terminate the trial, mark the length of time for each response, and remove the picture from the screen. The screen appeared blank until the experimenter pressed the key to initiate the next trial (and all subsequent trials). See figure 7.

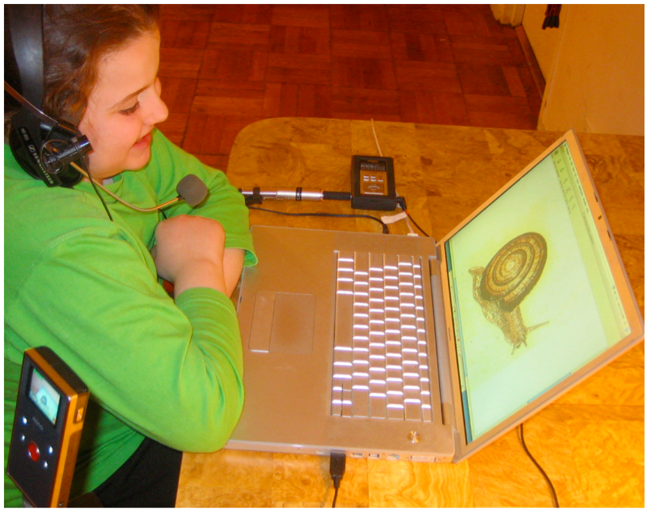

Figure 7. Production task set-up. Target *snail.*

Results II

For each phonological type, accuracy and RT were analyzed using a 3 (age) x 2 (stimulus type) x 2 (sex) ANOVA. There was no main effect of sex nor did sex interact with accuracy on any phonological type, RT, or RT for correct trials only (all F's < 1). Thus, data from boys and girls were collapsed in all analyses such that subsequent analyses were all 3 (age) x 2 (stimulus type).

In contrast with the perception data, where the younger children performed less accurately than the older children for all phonological types, there was no significant effect of age on production scores ($F(2, 25) = 1.98$, $p = 0.16$). However, there was a trend revealing the older children performing better than the younger children. See table 8 and figure 8.

Table 8

Phoneme Production Developmental Results

	3 year olds	4 year olds	5 year olds	Overall
% Correct	Mean = .78 SE = .06	Mean = .86 SE = .06	Mean = .93 SE = .03	Mean = .85 SE = .03

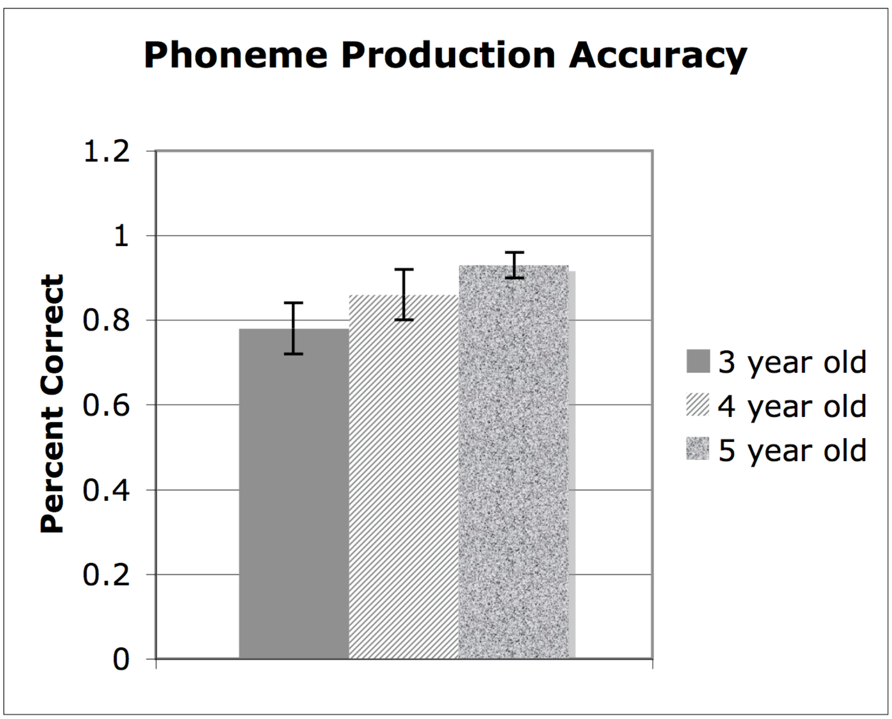

Figure 8. Phoneme production developmental results

Production Task Accuracy

/S/ vs. Non-/S/. A 2 (*/s/-non-/s/*) x 3 (age) ANOVA revealed no significant effect of age on phonological type, (F (2, 25) = 2.11, p = 0.143). There was no main effect of phonological type for target phoneme /s/ (mean= 0.8594, SE= 0.0409) vs. non-/s/ target phonemes (mean= 0.85, SE= 0.03) (F(1, 25) = 0.21, p = 0.648).

Liquid vs. Non-Liquid. A 2 (liquid/non-liquid) x 3 (age) ANOVA revealed no significant effect of age on phonological type, (F (2, 25) = 1.60, p = 0.221). There was no main effect of phonological type for liquids (mean= 0.81, SE= 0.05) vs. non-liquids (mean= 0.87, SE= 0.03) (F(1, 25) =1.82, p = 0.190).

Simple vs. Clusters. A 2 (simple/cluster) x 3 (age) ANOVA revealed no significant effect of age on phonological type, (F (2, 25) =2.23, p = 0.129). There was a significant main effect of phonological type for simple (mean= 0.91, SE= 0.02) vs. complex (mean= 0.82, SE= 0.04) (F(1, 25) = 6.98, p = 0.014) targets.

Onsets vs. Codas. A 2 (onset/coda) x 3 (age) ANOVA revealed no effect of age on phonological type, (F (2, 25) =2.57, p = 0.096). There was no main effect of phonological type for onset (mean= 0.87, SE= 0.03) vs. coda (mean= 0.87, SE= 0.02) (F(1, 25) = 0.11, p = 0.744) targets.

Results III: Perception-Production

The previous sections (results I and II) described the outcomes of the two experiments separately. To investigate the *link* between phoneme perception and production, we performed analyses across both tasks.

Overall and Phoneme Types. Regression analyses performed on all children's data for both tasks showed no evidence of a link between phoneme perception and production. When overall accuracy (all children included) on the production and perception tasks was analyzed, there was no relationship between children's performance on the two tasks (*r =.21, p =.27*). See figure 9. This was also true for fine-grain analyses of children's accuracy on specific types of phonological targets, including non- liquids (*r* =.34*, p*=.07) and liquids (*r, l)* (*r* =.16*, p* =.40) (figure 10a), codas (*r* =.17*, p* =.39) (figure 10b), clusters (*r* =.28, *p* =.15) (figure 10c), target words with /s/ (*r* =.27*, p* =.16) and target words without /s/ (*r* =.31*, p* =.12) (figure 10d).

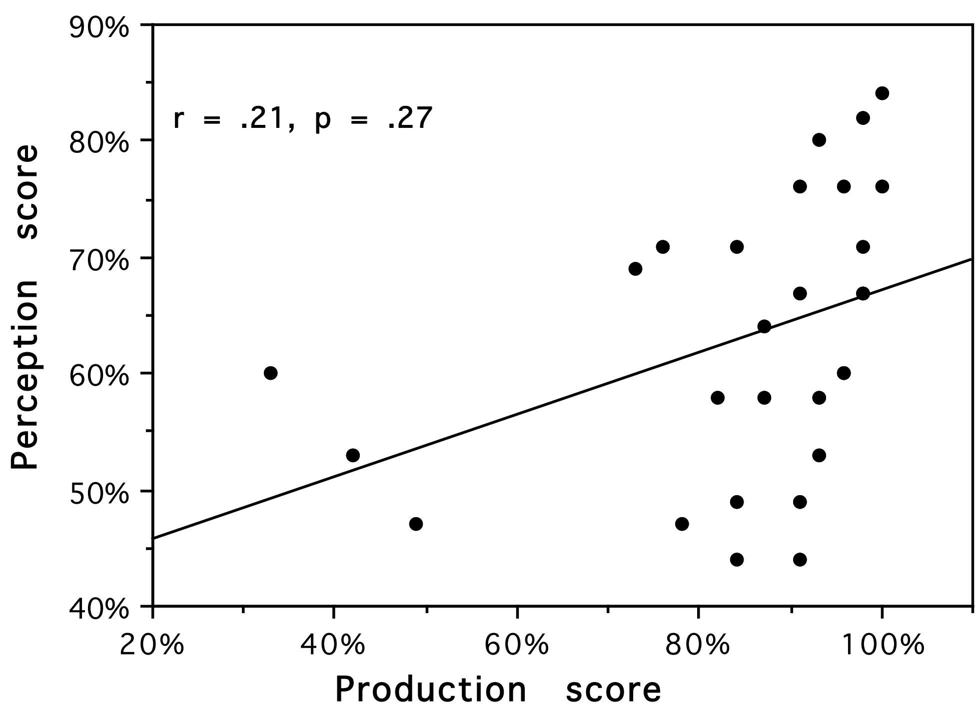

Figure 9. Cverall perception-production correlation

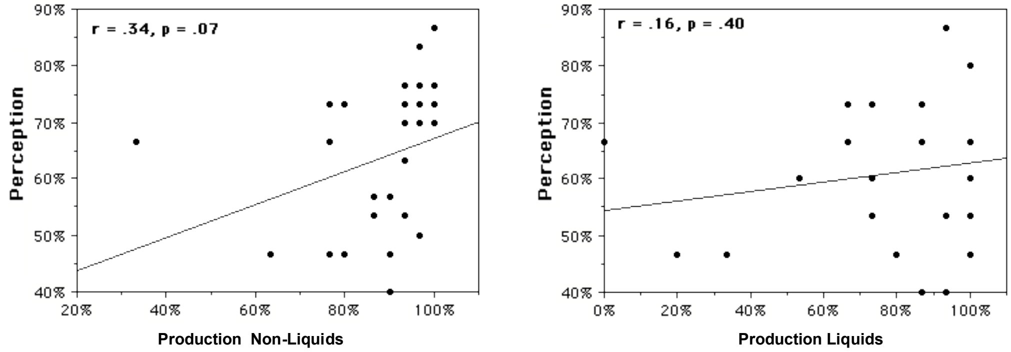

Figure 10a. Perception-production correlation for non-liquids and liquids

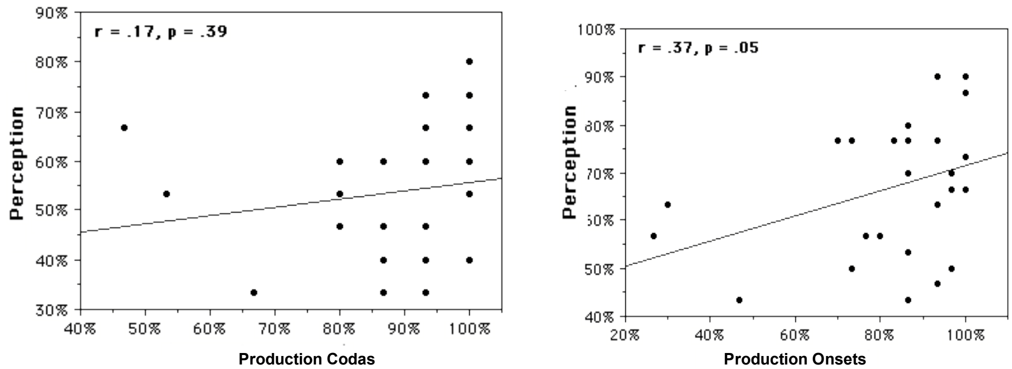

Figure 10b. Perception-production correlation for codas and onsets

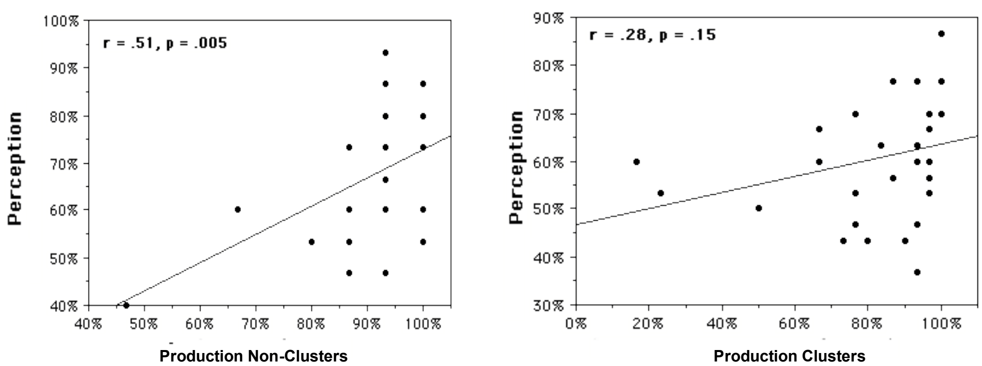

Figure 10c. Perception-production correlation for non-clusters and clusters

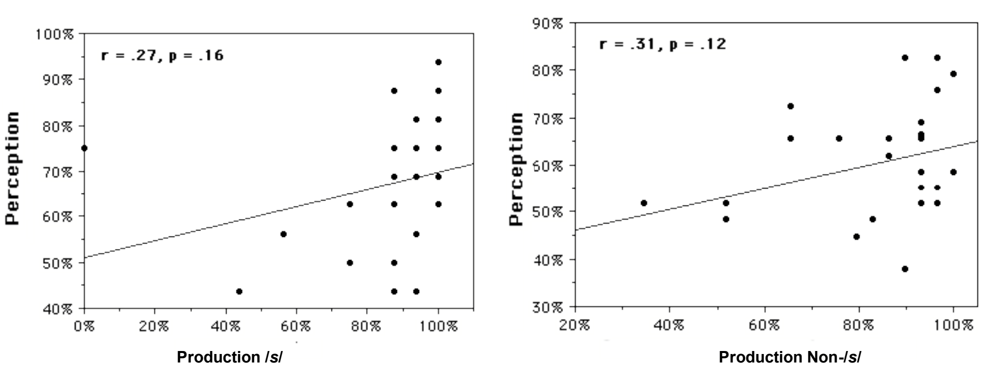

Figure 10d. Perception-production correlation for words with and without */s/*

Within Child Analyses

Perhaps there was a significant perception-production link for some children, but not for others. However, Spearman's *r* of each child's accuracy on the 45 perception and production trials revealed within child correlation coefficients of between -.16 and .31 (mean $r = .04$). These analyses simply addressed whether, for particular items, children misperceived the same items they mispronounced and vice versa.

Linguistically, a more interesting question is whether children similarly misperceive the same items they misproduce. For example, determining whether a child who misperceives *snail* as *sail* (i.e. chooses the *sail* picture) also mispronounces *snail* as *sail*. We investigated this question by examining the data for the 3 children who made the most errors because the fact that they made these errors increases the number of relevant trials and hence our ability to determine whether children misperceive and misproduce words in the same way. In figure 11, the 3 children who misperceived the most trials are circled in black (corresponding to table 9) and the children who misproduced the most trials are circled in gray (corresponding to table 10). The first thing to notice is that the 2 groups are completely disjunct. The lowest-scoring perceivers (mean=.44, .44, .77), ranked among the top half in production (mean= .91, 84, .78); (r = *.12, .01, -.04*) and the lowest-scoring producers (mean=*.33, .42, .49)*, were among the top half in perception (mean= .60, .53, .47); *(r* = .00, -.01, .07). Furthermore, these children made different types of errors in perceiving and producing phonemes. For example, if we examine the types of errors made by the child who scored lowest on the perceptual task (table 9), we find are only 3 cases in which this child made errors on the same target word across tasks, only 1 of which involved the same type of error (voicing).

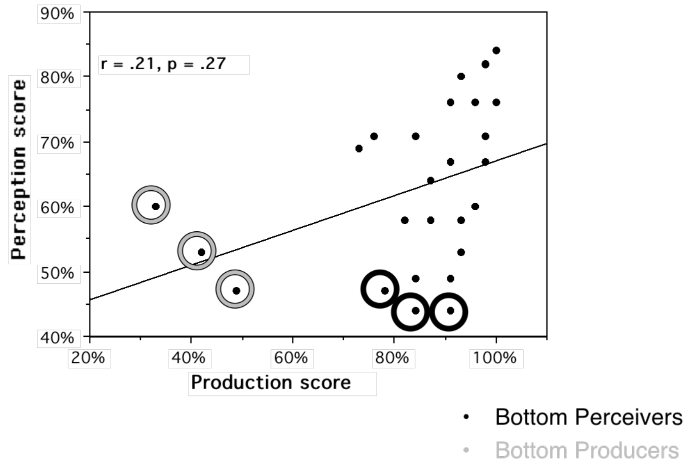

Figure 11. Lowest-scoring children

Table 9

Bottom Perceivers

Lowest Scorers	Common Error Type	# Same Target Errors Across Tasks	# Error Types in Common	Target Error Type in Common
1	<u>Cluster Reduction:</u> - Reduces cluster but retains liquid *r* or *l* >*r* - Reduces *s* - Reduces *tsh* to *sh*	3	1	- voicing
2	<u>Cluster Reduction:</u> - Reduces cluster but retains liquid *r* or *l* - Reduces *tsh* to *sh* - Gliding	4	1	- adds liquid
3	<u>Cluster Reduction:</u> - Reduces *s*	5	0	

Table 10

Bottom Producers

Lowest Scorers	Common Error Type	# Same Target Errors Across Tasks	# Error Types in Common	Target Error Type in Common
1	<u>Cluster Reduction:</u> - Reduces *s* - Reduces *tsh* to *sh*	12	4	- Reduces *tsh to sh* - Reduces *s* - Reduces *t* - Reduces coda *dge*
2	<u>Cluster Reduction:</u> - Reduces liquid *r* - Gliding	12	2	- Reduces *skw to w* - Voicing
3	- Gliding - Voicing	12	2	- Reduces *s* - Gliding

Developmental Link

Perhaps no apparent perception-production link across the tasks was a result examining so large an age range. Participants were thus divided into 3 age groups to investigate developmental differences in performance based on age. In these post-hoc analyses, we compared the extent to which children's performance on the perception and production of individual target words matched. Perception and production for a target word was considered a match if the child got both right (e.g., correctly pointing to the *snail* in the perception task and correctly saying the word *snail*) or got both incorrect (e.g., pointing to the *sail* in the perception task and saying the word *snail* as *sail*). Otherwise, production and perception on a trial were treated as a mismatch (table 11). We used absolute difference because a child need not get an item correct across the experiments for a link to be designated. We calculated the percent of trials in which the child scored the same on both the tasks. For example, a score of .50 means that for 50% of the trials (23 trials) the particular child got a target item both correct or both incorrect. Regression analyses revealed that the mean absolute difference for the 45 perception-production items decreased with age (r = -.064; p =.0003). Thus, the link appeared to emerge as children developed, with youngest children having significantly more mismatches than the older children. See figure 12.

Table 11

Perception-Production Difference Scores

	% Perception-Production Matching Trials		*% Perception-Production Mismatching Trials*	
Mean Age [Range] *N*	Perception-Production **Correct-Correct 1-1**	Perception-Production **Incorrect-Incorrect 0-0**	Perception-Production **Correct-Incorrect 1-0**	Perception-Production **Incorrect-Correct 0-1**
3;33 [3;0 – 3;5] *n =10*	193 **42.89%**	44 **9.78%**	56 **12.44%**	157 **34.89%**
	52.67%		**47.33%**	
4;05 [3;75 – 4;5] *n =9*	236 **58.27%**	26 **6.42%**	30 **7.41%**	113 **27.90%**
	64.69%		**35.31%**	
5;0 [4;75 – 5;25] *n =9*	262 **64.69%**	14 **3.46%**	16 **3.95%**	113 **27.90%**
	68.15%		**31.85%**	

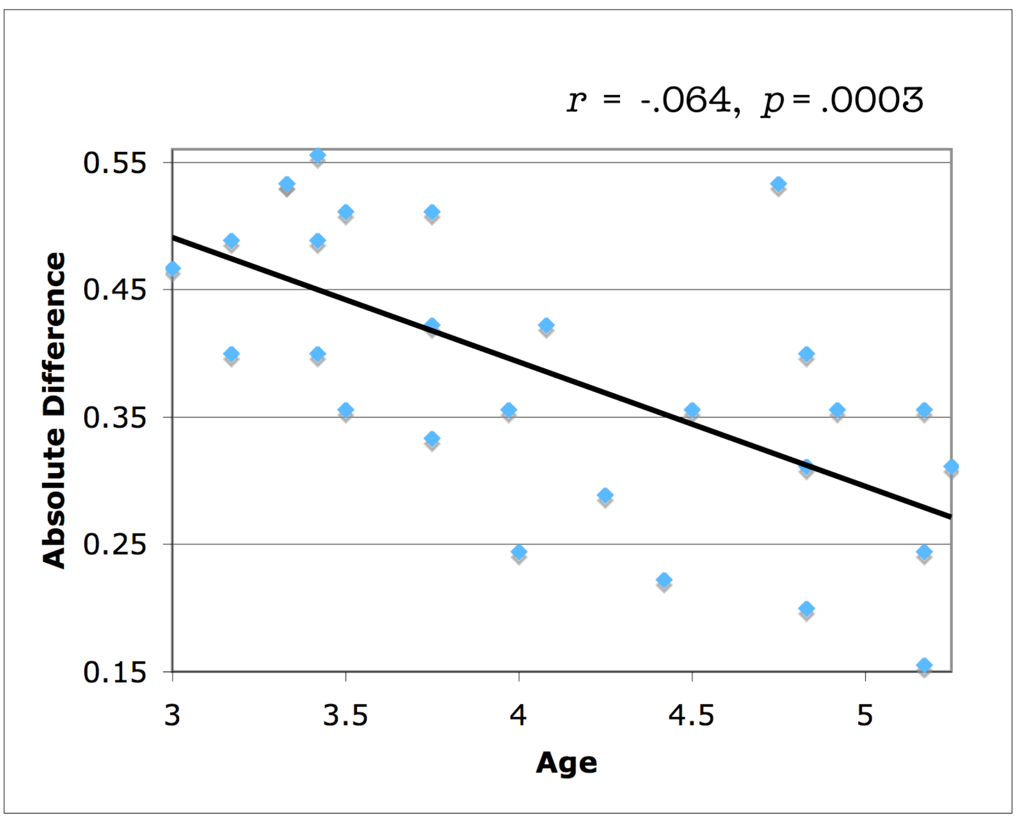

Figure 12. Difference Scores by Age

Discussion

The results of some experiments (see introduction section) suggest speech perception and production are linked, whereas others suggest the converse. Rather than examine the causal direction between the systems, our study sought to examine whether or not there was a link at the *phoneme* level. Analyses of overall performance, specific phonological subtypes of stimuli, and individual children's data initially revealed little evidence of a phoneme perception-production link. Consistent with Kuijpers (1996), however, when children were divided into 3 groups by age, we found evidence of a link in the older children but not in the younger children,

According to motor theory, speech perception occurs via knowledge of the articulatory commands used to produce speech and this system is considered to be innate (Liberman & Mattingly, 1985). Accordingly, perceptual errors should be tightly and specifically correlated with errors in production. Contrary to motor theory, the results of our study suggest that the production-perception link develops over time. In our study, we found that children's perception of phonemes depends on the phonological subtypes examined, whereas children's performance on the production task does not. For the phoneme perception task, children were more accurate in their perception of onsets vs. codas, simple vs. complex cases, non-liquids vs. liquids, and target phonemes containing /*s*/ vs. those without /*s*/. There was also a linear effect of age on perception accuracy, with the youngest children misperceiving more phonemes than the older children. There was no age effect for speed on the perception task. For the phoneme production task, there were no significant differences in the phonological types examined (except for simple vs. complex cases) and no age effect for accuracy.

First, our findings are consistent with theoretical accounts of normal phonological development suggesting children's awareness of the units of speech develops. According to this explanation, young children only have access to the largest units (prosodic structure of clauses and words). Over time, children develop awareness of syllables, followed by onsets and codas, and finally phonemes themselves (e.g.Waterson, 1987). For example, studies of children's speech production exhibit this shift in processes affecting the structure of the word as a whole entity (e.g. final consonant deletion) to those affecting specific segments or segment classes (e.g., stopping of fricatives) (Vihman, 1996).

For several reasons, researchers have assumed early storage of words to be holistic or underspecified, during which children may have a preference for certain fixed word templates (e.g., Vihman et al. 1994). Accordingly, another possible explanation is that as children's vocabulary increases, it becomes necessary to break words down into smaller units. In other words, children cannot just use simple syllables- they need distinguish words phonemically (e.g. *bat* and *rat*). The process of phonemic representation develops and may not be complete until much later in childhood (e.g., Vihman, 1996). Thus, the youngest children likely had the ability to think about larger units but were not yet able to distinguish single segments at the phoneme level.

A third reason for the absence of a link in the younger children could be the possible difference between synthesis (blending sounds together) and segmentation (taking apart units) skills. In most children, the former ability occurs earlier than the latter (e.g., Caravolas & Bruck, 1993). Thus, it is easier for the child to respond with the word *fat* when presented with the sounds *f - at* or *f-a-t*, than it is to supply *f-a-t* when asked what sounds are heard in *fat*. It is possible that children are required to break down the target word to succeed in perception, and simply blend sounds of the target word together to succeed in production. Conceivably, whereas the youngest children in our experiment could have succeeded on the production task using synthesis skills, they perhaps misperceived more phonemes because the perception task required segmentation ability that they had not yet acquired. This account is also consistent with higher overall accuracy on the perception vs. production tasks and with the fact that the highest proportion of perception-production difference scores for all children was more markedly driven by a mismatch in incorrect perception scores and correct production scores (see table 11).

The last potential source of the developmental link that we consider in this paper is phonemic awareness and/or reading. In educational settings, skill-building phoneme awareness exercises have been included in curricula for preschool children, particularly between the ages of 4 and 5 years-old in preparation for reading (Lundberg 1998). If the older children in our study are developing phonemic awareness skills, then these children can conceivably more readily generate an articulatory programme to produce phonemes on the basis of knowledge of the articulatory configuration corresponding to each segment, plus application of co-articulatory adjustments (Bird & Bishop, 1995). With heightened phonemic awareness that the word *fat* is comprised of the phoneme constituents *f-a-t*, for example, the older children may more readily generate the articulatory program to produce *fat* as *f-a-t*. Alternatively, or in addition, articulating or 'sounding out' *fat* as *f-a-t* (either out loud or by subvocalization) may encourage the children to more readily hear their own articulations (auditory feedback) of *fat* as *f-a-t* and/or thereby hear *fat* as *f-a-t* when articulated by others.

Given that there is some evidence of a perception-production link emerging over time, future work should include younger children (children who are still at the earliest stages of speech perception), school-age children who have the ability to read single words, and adults. Simpler stimuli that are easier to detect (such as minimal pairs) should be used to test the younger children (i.e. encourage them to do the perception task). By using eye tracking, one might also be able to reduce the task demands. As discussed in the introduction, another way to look at the development of the link is to consider children who have impaired production. An obvious extension of this work would be to study children with minor speech impairments (misarticulating children) and children with more severe speech impairments (dyspraxic and dysarthric children).

Thus far, work comparing speech perception and production has treated children across the ages in one group and compared performance on general perception-production measures. Our study approaches the long-standing controversy about the relationship between phoneme perception and production using meticulous measures across the tasks at high-level scrutiny to address the link at various stages of development. Our findings and directions for future work suggest that the extent to which fine-grained observations are conducted to scrutinize the relationship between the systems will directly impact the precision of their characterization as distinct or linked.

References

Berko, J. & Brown, R. (1960). Psycholinguistic Research Methods. *Handbook of Research Methods in Child Development*. Ed. Paul Henry Mussen. New York: Wiley and Sons, *517-557.*

Boersma, P., Weenink, D. (1992-2008). Praat: doing phonetics by computer (version 5.0.40). Computer program. Retrieved October, 2008, from http://www.praat.org

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. Journal of Speech and Hearing Research, 38, 446–462.

Bird, J., & Bishop, D. (1992). Perception and awareness of phonemes in phonologically impaired children. *Journal of Language & Communication Disorders, 27,* 289-311.

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55,* 1–30.

Drumwright, A. (1971). Denver Articulation Screening Examination. Denver, CO: University of Colorado Medical Center.

Groenen, P., Maassen, B., & Crul, T. (1996). The specific relation between perception and production errors for place of articulation in developmental apraxia of speech. *Journal of Speech and Hearing Research, 39, 468-482.*

Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. H. (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech & Hearing Disorders*, *50*, 46-53.

Hoit-Dalgaard, J., Murry, T., & Kopp, H. G. (1983). Voice onset time production and perception in apraxic subjects. *Brain and Language, 20,* 329-339.

Kuijpers, C.T.L. (1996). Perception of the voicing contrast by Dutch children and adults. *Journal of Phonetics, 24 (3),* 367-382.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Lundberg, I., Frost, J., & Peterson, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.

Miller, J.L. (1990). Speech perception. In D. Osherson & H. Lasnick (eds.), An invitation to Cognitive Science, 1, 69-93.

Monnin, L. M., & Huntington, D. A. (1974). Relationship of articulatory defects to speech-sound identification. *Journal of Speech and Hearing Research*, 17, 352-366.

Ohde, R. N., & Sharf, D. J. (1998). Perceptual categorization and consistency of synthesized /r-w/ continua by adults, normal children and /r/- misarticulating children. *Journal of Speech, and Hearing Research*, *31,* 556-568.

Raaymakers, E. M. J. A., & Crul, Th. A.M. (1988). Perception and production of the final /s-ts/ contrast in Dutch by misarticulating children. *Journal of Speech and Hearing Disorders*, 53, 262-270.

Rvachew, S., & Grawberg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research, 49,* 74-87.

Rvachew, S., & Jamieson, D. G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders*, 54, 193–208.

Rvachew, S., Ohberg, A., Grawberg, M., & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. American *Journal of Speech-Language Pathology*, *12,* 463-471.

Sander, E.K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37,* 55-63.

Shuster, L.I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research, 41,* 941-950.

Stromswold, K. (2009). Lessons from a mute child. Paper presented at *Rich Languages from Poor Inputs: A Workshop in Honor of Carol Chomsky*. MIT, Cambridge, MA.

Vihman, M. M. (1996). Phonological development: The origins of language in the child. Oxford, U.K.: Blackwell.

Vihman , M . M . & Boysson-Bardies, B. de (1994). The nature and origins of ambient language influence on infant vocal production and early words, *Phonetica, 51,*159 – 169.

Waterson, N. (1987). Prosodic Phonology: The Theory and Application to Language Acquisition and Speech Processing. Newcastle upon Tyne: Grevatt & Grevatt.

Whitehill, T.L., Francis, A.L., & Ching, C.K.-Y. (2003). Perception of stop placement by children with cleft palate and posterior placement. *Journal of Speech, Language, and Hearing Research, 46*, 451-461.

Appendix A: The 45 Stimuli Quartets

7 Simple Onset (C):	**s-eat** sheet street eat
	s-ing string sting wing
	f-at flat rat sat
	t-oe sew row throw
	g-as grass class glass
	l-ake rake wake ache
	r-ip lip whip zip
18 Cluster Onset (CC):	**sp-ark** bark park shark
	sk-is keys seas squeeze
	sn-ail sail mail nail
	sm-ell shell sell bell
	sl-eep leap weep sweep
	fr-y lie tie fly
	fl-ight light white write
	bl-ed red bread bed
	gr-ab crab cab stab
	br-eak bake rake wake
	cr-ash trash cash rash
	tr-ail tail rail whale
	tr-ee tea dee three
	cl-ock block lock rock
	ch-ase face lace vase
	ch-ew shoe zoo two
	ch-ick sick tick trick
	ch-ip sip tip ship
5 Cluster Onset (CCC)	**scr-eam** cream steam stream
	str-ip trip rip tip
	spr-ing swing sing wing
	squ-eeze skis sees keys
	squ-irt shirt skirt hurt
8 Simple Coda (C)	**ri-ce** write ride rise
	bu-zz bug bus bud
	pi-g pick pin pit
	co-ke cone coat comb
	ba-d bat bag badge
	k-id kick kit king
	wro-te road rope robe
	ro-be wrote road rope
6 Coda Cluster (CC)	**bea-st** bees beat beach
	cru-st crushed crunch crutch
	fi-st fish fin fizz
	pa-nt pats pans pants
	ca-tch cash cat cap
	ba-dge bad bat bag
1 Cluster Coda (CCC)	**cru-nch** crust crutch crush